2015

Patterns, causes, and consequences of connectivity within a coral reef fish metapopulation

BOSTON UNIVERSITY

GRADUATE SCHOOL OF ARTS AND SCIENCES

Dissertation

PATTERNS, CAUSES, AND CONSEQUENCES OF CONNECTIVITY WITHIN A CORAL REEF FISH METAPOPULATION

by

CASSIDY D'ALOIA

B.A., Middlebury College, 2010

Submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

2015

Approved by

First Reader

Peter Buston, PhD
Assistant Professor of Biology

Second Reader

Sean Mullen, PhD
Assistant Professor of Biology

Acknowledgments

I would first like to thank my advisor, Pete Buston, for his mentorship throughout my graduate studies. Pete has provided invaluable guidance in all aspects of my development as a scientist. I am also grateful for the guidance and support of my Ph.D. committee members: Rick Harrison, Phil Lobel, Sean Mullen, and Randi Rotjan.

Thanks to the past and present members of the Buston Lab at Boston University for being excellent colleagues and friends: Tina Barbasch, Rebecca Branconi, John Majoris, Alissa Rickborn, Jeremiah Seymour, and Marian Wong.

Many thanks to the Harrison Lab for hosting me at Cornell University. Rick Harrison has been a tremendous mentor to me and I am grateful for his support and insightful comments. Likewise, this dissertation would not have been completed in five years without Steve Bogdanowicz. Steve's expertise and enthusiasm have had a huge impact on the success of this project.

Thanks to my wonderful undergraduate assistants. Diana Acosta and Derek Scolaro were troopers in the field. Robin Francis provided an invaluable contribution to this dissertation through her otolith work. Ellie Bondra played a critical role in genotyping.

I am grateful to my extended Belizean family for their generosity and support during five summers of field work. Alben David, Kevin David, and Udell Foreman were incredible boat captains and friends. Celia Zelaya, Rosie Tillet, and Earl (Junior) David always took excellent care of us. I am so grateful to all the staff at Wee Wee Caye, Carrie Bow Caye, International Zoological Expeditions, Calabash Caye, and all the other places we visited. Many thanks to Belize Fisheries for allowing us to work on the Belizean reefs.

I would also like to thank our collaborators on the NSF Population Connectivity project for their keen insight: David Lindo, Claire Paris, Robert Warner, and Colleen Webb.

Finally, I thank my partner Noel, who makes me a better scientist and a happy person.

PATTERNS, CAUSES, AND CONSEQUENCES OF CONNECTIVITY WITHIN A CORAL REEF FISH METAPOPULATION

(Order No.)

CASSIDY D'ALOIA

Boston University, Graduate School of Arts and Sciences, 2015

Major Professor: Peter Buston, Assistant Professor of Biology

ABSTRACT

Population connectivity influences virtually all ecological and evolutionary processes within metapopulations including population dynamics, persistence, and divergence. A comprehensive analysis of connectivity must consider the exchange of both individuals and alleles among populations, representing demographic and genetic connectivity, respectively. For many marine species, connectivity is driven by larval dispersal. However, despite the widespread recognition that dispersal is key to predicting metapopulation dynamics and effectively managing networks of marine reserves, empirical data are scarce due to the methodological challenges of tracking larvae. This dissertation is an integrative study of the patterns, causes, and consequences of marine connectivity using the sponge-dwelling reef fish *Elacatinus lori* as a study system. I begin by describing the distribution and abundance patterns of *E. lori* and its host sponge on the Belize barrier reef. Next, I study demographic connectivity by using genetic parentage analysis to quantify dispersal. I conduct an intra-population study to identify self-recruiting dispersal trajectories and develop a method to approximate a dispersal kernel based on the distribution of habitat patches. I then complete a large-scale parentage analysis to produce the first statistically-robust marine dispersal kernel. I find that dispersal declines exponentially with respect to distance in *E. lori*, with no dispersal events exceeding 16.2 km. Notably, dispersal probabilities are unrelated to the number of days an individual spends in the larval phase and other biological variables. Finally, to elucidate the long-term microevolutionary consequences of

genetic connectivity, I investigate spatial genetic structure in the Belizean metapopulation. In a preliminary study based on mitochondrial and microsatellite data, I find high levels of pairwise genetic differentiation between sites separated by only 20 km. In a follow-up study, I use a high-throughput multiplex approach to resolve fine-scale patterns of genetic structure throughout the species' range. Seascape genetic analyses reveal that genetic connectivity is consistent with the shape of the dispersal kernel. Collectively, this dissertation generates novel insights regarding the spatial scale at which marine fish populations are connected. Given the alarming rate of population declines on coral reefs globally, these results have important and time-sensitive conservation implications.

Contents

List of Tables

List of Figures

List of Abbreviations & Symbols

λ	decay rate of exponential probability density function
Φ_{ST} ..	Excoffier's measure of genetic differentiation
π	# of pairwise differences between haplotypes
AIC ..	Akaike information criterion
AMOVA	analysis of molecular variance
ANCOVA	analysis of covariance
ANOVA	analysis of variance
AUC ..	area under the curve
BBR ..	Belize barrier reef
BIC ..	Bayesian information criterion
bp ...	base pair
CI ...	confidence interval
CM ...	connectivity matrix
COI ..	cytochrome oxidase I
cytb ..	cytochrome b
DAPC .	discriminant analysis of principal components
ddRAD .	double digest restriction associated DNA
DNA ..	dioxyribonucleic acid
ESU ..	evolutionarily significant unit
EtOH ..	ethanol
F_{ST} ...	Wright's fixation index
GIS ...	geographic information system

GLM .. generalized linear model

GLMM . generalized linear mixed model

G'_{ST} .. Hedrick's measure of genetic differentiaion

H_E ... expected heterozygosity

H_O ... observed heterozygosity

HWE .. Hardy-Weinberg equilibrium

IBB .. isolation by barrier

IBD .. isolation by distance

IBDOD . isolation by derived oceanographic distance

IBDP .. isolation by dispersal probability

IBE .. isolation by environment

IBOD . isolation by oceanographic distance

IMaRS . Institute for Marine Remote Sensing

k .... # of alleles per locus, or # of genetic clusters

LD ... linkage disequilibrium

LDD .. long-distance dispersal

LMM .. linear mixed models

LOD .. log-odds ratio

LOOCV leave-one-out cross-validation

MBR .. Mesoamerican barrier reef

MCMC . Markov Chain Monte Carlo

ML ... maximum likelihood

MMRR . multiple matrix regression with randomization

MPA .. marine protected area

MS-222 . tricaine mesylate

mtDNA . mitochondrial dioxyribonucleic acid

NLS .. nonlinear least squares

p.d.f. ..	probability density function
PCA ..	principal component analysis
PCR ..	polymerase chain reaction
PLD ..	pelagic larval duration
s	# of segregating sites
s.d. ...	standard deviation
SGS ..	spatial genetic structure
SL ...	standard length
SNP ..	single nucleotide polymorphism
SPAG .	spawning aggregation site
SR ...	self-recruitment

Chapter 1

Introduction

For many taxa, dispersal is the primary life history trait that drives the movement of alleles and individuals across space (Nathan & Muller-Landau 2000, Clobert *et al.* 2001, Hanski & Gaggiotti 2004, Clobert *et al.* 2012). In this dissertation, I focus on *natal* dispersal, which is defined as the movement of an individual from its natal site to a destination site, where it typically stays until breeding. I distinguish *natal* dispersal from *breeding* dispersal, in which an individual disperses from one breeding location to another (Clobert *et al.* 2001).

Dispersal has profound impacts on multiple levels of biological organization. For individuals, dispersal influences access to mates and resources (Matthysen 2005), proximity to kin (Szulkin & Sheldon 2008), and the capacity to find suitable habitat (Roderick & Caldwell 1992). For populations and metapopulations, dispersal drives population dynamics and persistence (Hastings & Botsford 2006) as well as the ability of local populations to 'rescue' each other after stochastic extinctions and/or drastic population declines (Hanski 1999). For species, dispersal affects range expansions (Saura *et al.* 2014) and evolutionary potential (Berthouly-Salazar *et al.* 2013) via long-distance dispersal.

Given the myriad ways in which dispersal influences biological processes, a large body of empirical and theoretical work has been devoted to the study of dispersal (reviewed in Clobert *et al.* 2001, 2012). Indeed, dispersal has been central to the development of entire fields of biology, including spatial ecology (Tilman & Kareiva 1997) and metapopulation ecology and evolution (Hanski 1999; Hanski & Gaggioti 2004).

More recently, much research has focused on dispersal as a driver of population connectivity in the marine environment (Cowen *et al.* 2007; Cowen & Sponaugle 2009). Here,

connectivity is defined as the rate of exchange of individuals (demographic connectivity) and alleles (genetic connectivity) among populations. This focus on marine connectivity is partly due to the relatively recent acknowledgment that groups of marine populations function as metapopulations (Kritzer & Sale 2006). While marine metapopulations generally do not conform to Levins' classic model of a group of populations undergoing stochastic extinctions and colonizations (Levins 1969), it is now recognized that *rates* of exchange among populations play a critical role in the ecological and evolutionary dynamics of marine metapopulations

There is now a widespread research effort focused on marine population connectivity that is driven by both pure and applied scientific motivations. The *pure* motivation is the scientific knowledge gap: we have a limited understanding of the patterns and drivers of connectivity. This knowledge gap is attributable to the fact that many marine organisms have a bi-partite life cycle, whereby individuals undergo a dispersive larval phase, followed by a relatively sedentary phase on the benthic habitat. Thus, connectivity is driven by larval dispersal, and quantifying dispersal patterns of tiny larvae traveling in vast oceans is logistically challenging. However, there is methodological precedent, as this type of propagule dispersal, in which connectivity is restricted to one or more discrete phases in the life cycle, is analogous to seed and pollen dispersal in plants (Nathan & Muller-Landau 2000), and spore dispersal in fungi (Rieux *et al.* 2014). Equally important, the *applied* motivation is the potential to use connectivity data to inform marine management plans (Sale *et al.* 2005; Botsford *et al.* 2009; White *et al.* 2014). Around the globe, marine populations are suffering from acute population declines (Hoegh-Guldberg *et al.* 2007; IPCC 2014). To combat these population declines, connectivity data can be used to help optimize effective networks of marine reserves; yet, progress has been limited due to a lack of empirical data.

However, despite the logistical challenges associated with the empirical study of larval dispersal, the above-mentioned motivations have fueled steady progress over the past 15 years. After the first groundbreaking studies showed that some fish returned to their natal

population after the larval phase (Jones *et al.* 1999; Swearer *et al.* 1999), scientists began to apply genetic parentage analysis to marine dispersal research, leading to a proliferation of empirical work (Jones *et al.* 2009). Research progressed from documenting recruitment back to a focal site to describing connectivity networks that depict the magnitude and direction of dispersal trajectories between a set of sites. Over time, the spatial scale of these networks has increased, from local networks (∼1 km) that show self-recruiting dispersal trajectories within a single population (Jones *et al.* 2005; Planes *et al.* 2009), to regional networks (∼30 km) that identify the exchange of individuals across populations (Planes *et al.* 2009; Saenz-Agudelo *et al.* 2011) and between marine reserves (Harrison *et al.* 2012). More recently, researchers have begun to adopt a more quantitative approach to estimating dispersal curves, drawing on methodology from the terrestrial plant dispersal literature (Nathan & Muller-Landau 2000). For example, logistic models have been used to show that the probability of dispersal declines rapidly as a function of distance from source at a local scale (∼1 km) (Buston *et al.* 2012) and alternative probability density functions were fit to observed dispersal data at a regional scale (∼30 km) (Almany *et al.* 2013). These studies have laid the foundation for the quantitative analysis of marine dispersal data.

However, more rigorous analyses of the patterns, predictors, and consequences of dispersal are needed in order to gain a complete understanding of marine population connectivity. Several major challenges remain. First, the description of a complete dispersal kernel is still lacking for any marine species (Botsford *et al.* 2009, Jones *et al.* 2009). An empirical dispersal kernel is defined as a probability density function that can be integrated to yield the probability of dispersing over a given distance. Dispersal kernels are also referred to as phenomenological models, and are considered complete descriptive models of dispersal patterns. To date, true kernels have been lacking; published work has either restricted the kernel shape to a single functional form, or has lacked sufficient sample size for a rigorous statistical analysis. A second major goal is to move beyond these phenomenological models towards an understanding of the causes of variation in dispersal and connectivity. To date,

mechanistic approaches to modeling marine dispersal have focused on bio-physical oceanographic models (Cowen *et al.* 2006; Paris *et al.* 2013). However, a potentially useful, albeit simpler approach, is the application of statistical models to identify predictive covariates. Finally, a third major goal is to explore the demographic and genetic consequences of population connectivity, including the short-term implications for marine reserve design, as well as the long-term implications of genetic exchange.

In this dissertation, I use the neon goby *Elacatinus lori* as a tractable study organism to investigate the patterns, causes, and consequences of population connectivity within a marine metapopulation. The species was described in 2002 as a sponge-dwelling goby of the genus *Elacatinus*, distributed throughout the Mesoamerican reef (Colin 2002). Prior to this dissertation research, no further work had been conducted on *E. lori*. Thus, in Chapter 2, I begin by describing the basic distribution and abundance pattern of *E. lori* and its primary host sponge, *Aplysina fistularis*. While distribution and abundance studies are commonplace for microhabitat specialists such as *E. lori*, few studies simultaneously describe patterns for both the focal species and the invertebrate host. Focusing on a 100 m wide x 300 m long transect at Carrie Bow Caye, Belize, I use underwater survey data and multivariate statistical analyses to identify the patterns and predictors of *E. lori* and *A. fistularis* abundance. This chapter lays the foundation for using *E. lori* as a study species.

Having described these basic distribution and abundance patterns, in Chapter 3 I conduct a small-scale dispersal study. In a 125 m wide x 500 m long transect on Curlew Caye, Belize, I map the *E. lori* and *A. fistularis* populations using underwater GPS technology. I use genetic parentage analysis (>600 individuals) to identify self-recruiting dispersal trajectories and then develop a method to approximate a dispersal kernel using a measure of self-recruitment and the distribution of nearby habitat patches.

Informed by the hypothesis of a dispersal kernel generated in Chapter 3, I conduct a large-scale empirical dispersal study in Chapter 4. To study dispersal up to 30-km from source, I systematically collect tissue samples (>7,100) along a 41 km transect on the Belize barrier reef. Using genetic parentage analysis to trace offspring to parents, I quantify a

complete dispersal kernel. Next, I explore variation in the pattern of dispersal and use predictive statistical modeling approaches to disentangle the predictors of dispersal. I then use graph theory to build connectivity matrices in order to explore the demographic and genetic consequences of dispersal, as well as the conservation implications.

In Chapter 5, I shift from direct dispersal studies to analyses of spatial genetic structure in order to explore the linkage between demographic and genetic connectivity. I conduct a preliminary seascape genetic analysis within a portion of the Belizean *E. lori* metapopulation. After characterizing spatial genetic structure with mitochondrial sequence data and 14 microsatellites, I test three alternative hypotheses for the correlates of genetic structure. I find a strong correlation between the continuity of reef habitat and genetic connectivity.

With a preliminary understanding of the correlates of genetic structure and a measurement of the dispersal kernel complete, in Chapter 6 I proceed to conduct a large-scale seascape genetic study of the entire Belizean metapopulation. Sampling from 39 sites along the entire barrier reef and all three atolls, I use sequence data from >1,000 individuals at 74 loci to describe genetic structure in the metapopulation. With more extensive spatial sampling, expanded genomic coverage, and sophisticated statistical analyses, I test the full suite of alternative seascape genetic hypotheses. With this alternative hypothesis-testing framework, I explore how dispersal, the seascape, oceanographic barriers, and environmental variables shape the spatial distribution of genetic variation.

Taken together, these chapters represent a comprehensive study of the patterns, causes, and consequences of larval dispersal and population connectivity in a single marine metapopulation.

Chapter 2

Predictors of the distribution and abundance of a tube sponge and its resident goby

2.1 Abstract

Microhabitat specialists offer tractable systems for studying the role of habitat in determining species' distribution and abundance patterns. While factors underlying the distribution patterns of these specialists have been studied for decades, few papers have considered factors influencing both the microhabitat and the inhabitant. On the Belizean barrier reef, the obligate sponge-dwelling goby *Elacatinus lori* inhabits the yellow tube sponge *Aplysina fistularis*. We used field data and multivariate analyses to simultaneously consider factors influencing sponge and goby distributions. Sponges were non-randomly distributed across the reef with density peaking at a depth of 10–20 m. Sponge morphology also varied with depth: sponges tended to be larger and have fewer tubes with increasing depth. Knowing these patterns of sponge distribution and morphology, we considered how they influenced the distribution of two categories of gobies: residents ($\geq$18 mm S.L.) and settlers ($<$18 mm S.L.). Maximum tube length, number of sponge tubes and depth were significant predictors of resident distribution. Residents were most abundant in large sponges with multiple tubes, and were virtually absent from sponges shallower than 10 m. Similarly, maximum tube length and number of sponge tubes were significant predictors of settler distribution, with settlers most abundant in large sponges with multiple tubes. The presence or absence of residents in a sponge was not a significant predictor of settler distribution. These results

provide us with a clear understanding of where sponges and gobies are found on the reef and support the hypothesis that microhabitat characteristics are good predictors of fish abundance for species that are tightly linked to microhabitat.

2.2 Introduction

Microhabitat-attached species are especially tractable systems for population studies of distribution and abundance. The phenomenon of strong attachment to sessile invertebrates is widespread throughout marine taxa. Some crustaceans exhibit microhabitat association, with shrimp inhabiting the mantle of oysters (Baeza 2008) and crabs living within the tentacles of sea anemones (Baeza *et al.* 2001). Several fish groups also exhibit microhabitat specialization. Pearlfishes live in the body cavities of sea cucumbers and sea stars (Lavett Smith *et al.* 1981; Colleye *et al.* 2008), anemonefishes inhabit anemones (Fautin 1992; Elliott & Mariscal 2001; Buston 2002), and damselfishes and gobies often associate with corals (Sweatman 1983; Booth 1992; Kuwamura *et al.* 1994; Munday *et al.* 1997; Wilson *et al.* 2008) or sponges (Colin 2002, 2010). Studies that predict microhabitat and inhabitant distribution may provide initial insights into interesting ecological interactions such as microhabitat preference, the effect of microhabitat on the fitness of inhabitants, and the relationship between population density and the environment. Distribution and abundance patterns of these systems can be influenced by three broad categories of variables: environmental gradients, microhabitat characteristics, and conspecific interactions.

Environmental gradients have been shown to influence the distribution of both microhabitats and the fish that associate with them. Distribution and abundance patterns of invertebrate microhabitats, such as corals and sponges, may be influenced by depth (Wilkinson & Chesire 1989; Wilkinson & Evans 1989; Duckworth & Wolff 2007; Roth & Knowlton 2009), substrate (Roth & Knowlton 2009), light availability (Wilkinson & Evans 1989), turbulence (Palumbi 1984; Wilkinson & Evans 1989; Mercado-Molina & Yoshioka 2009) and food availability (Lesser 2006). Fish settlement, distribution and abundance pat-

terns are also influenced by environmental gradients including depth (Roberts & Ormond 1987; Srinivasan 2003; González-Sansón *et al.* 2009), reef zone (Tolimieri 1998; Belmaker *et al.* 2007), reef continuity (Levin *et al.* 2000; Belmaker *et al.* 2009), and availability of sheltered areas (Roberts & Ormond 1987). For fish living in microhabitats, the interesting question becomes, what is the effect of environmental variables on the fish over-and-above the effect of these variables on the microhabitat?

Site-attached fish may be most profoundly influenced by microhabitat distribution (Munday *et al.* 1997). In some systems of obligate association, there is a triangle of correlates between microhabitat size, fish size, and group size. Several hypotheses that are not mutually exclusive could explain these positive correlations. First, in cases where microhabitat size is positively related to the number and the size of fish inhabitants (Fautin 1992; Kuwamura *et al.* 1994; Elliott & Mariscal 2001), microhabitat size may be the driver if larger microhabitats offer more resources. Second, fish biomass may be a driver of microhabitat growth and survival (Holbrook & Schmitt 2005). For example, anemonefishes have a symbiotic relationship with their host anemones, and an experimental removal of fish resulted in anemone shrinkage, possibly because the fish protect the anemones from predation and provide them with nutrients (Porat & Chadwick 2004, 2005). Third, there could be an external driver of these correlations, such as a 'site effect' where large invertebrate microhabitats and fish are simply located at favorable sites on the reef (van Noordwijk & de Jong 1986; Buston & Elith 2011).

Intraspecific interactions may also influence fish distribution and abundance patterns. Many studies have shown that conspecific interactions affect spatial patterns of settlement and recruitment. Conspecific presence may have a positive (Sweatman 1983, 1985, 1988; Booth 1992, 1995; Öhman *et al.* 1998), negative (Elliott *et al.* 1995; Öhman *et al.* 1998; Buston 2003) or insignificant correlation (Öhman *et al.* 1998) with settlement and recruitment depending on the ecology of the species. Adult conspecific attraction (Gardiner & Jones 2010) and density-dependent survival of juveniles and adults due to intraspecific competition have also been shown to influence distribution and abundance (Holbrook &

Schmitt 2002; Hixon & Jones 2005; Wilson 2005).

The yellow tube sponge *Aplysina fistularis* (Class: Demospongiae) is a common species of the Caribbean reef community. It is a keratose tube sponge with a distinct yellow-green coloration. This sponge exhibits a high level of morphological variation that may be strongly influenced by its environment (Neigel & Schmahl 1984). Sponges exist as single tubes or clusters of tubes and show wide variation in tube length. Sponges also tend to be bigger and have a faster growth rate with increasing depth, up to 30 m, as a result of greater food availability (Lesser 2006). *A. fistularis* provides habitat to a number of marine organisms including the sponge-dwelling goby *Elacatinus lori*. The sponge may benefit from hosting *E. lori* if the goby consumes the sponge's polychaete parasite *Haplosyllis spongicola* (López *et al.* 2001), as has been documented for other sponge-associated *Elacatinus* species (Smith & Tyler 1972; Colin 1975; Whiteman & Cotê 2002).

Elacatinus lori (Family: Gobiidae) is one of at least 25 neon goby species in the tropical western North Atlantic (Colin 2010). Distributed throughout the Gulf of Honduras, *E. lori* is distinguished from sympatric *Elacatinus* species (*E. colini*, *E. lobeli*, and *E. louisae*) by having a thin, white stripe running laterally along the body length and a white stripe on the snout (Colin 2002). Like most coral reef fishes, *E. lori* has a bipartite life cycle with a dispersing larval phase and a relatively sedentary reef resident phase. At the end of the pelagic phase, larvae settle to the reef and attempt to recruit to the resident population (Williams & Sale 1981; Buston 2003). For this system, we distinguish between settlement and recruitment, with settlement defined as the point when a late-stage larva settles near or outside the microhabitat and recruitment defined as the transition when a settler joins the group inside of the microhabitat. As residents, *E. lori* are obligate sponge dwellers (Colin 1975, 2002, 2010) found almost exclusively in *A. fistularis*. Beyond this, little is known about the distribution, abundance and group structure of this goby in yellow tube sponges. *E. lori* is strongly attached to the sponge, making it a good system to test the relationship between microhabitat distribution and fish distribution.

The purpose of this study is to evaluate the influence of environmental variables (e.g.,

depth), microhabitat characteristics and intraspecific interactions on the distribution of two reef species at Carrie Bow Cay, Belize: the yellow tube sponge *A. fistularis* and its neon goby inhabitant *E. lori*. We simultaneously consider sponge and goby distribution and specifically address the following questions:

1. How are sponges distributed across the reef and does sponge morphology change with an environmental variable (depth)?

2. What variables (depth and microhabitat characteristics) influence resident distribution and abundance?

3. What variables (depth, microhabitat characteristics, and conspecific interactions) influence settler distribution and abundance?

2.3 Methods

2.3.1 Study site

The Belize barrier reef complex is the largest contiguous reef in the Northern hemisphere (Rützler & Feller 1999), extending nearly 250 km along the coast of Belize from the Yucatán Peninsula to the Guatemalan border. Carrie Bow Cay (16° 48' 10" N 88° 05' 45" W) sits on the main barrier reef and is bordered by two channels: South Water Cut to the north and Curlew Cut to the south. Here, the barrier reef has several distinct reef zones, defined by Rützler & Macintyre (1982): reef crest, inner fore reef (high spur and groove, low spur and groove), outer fore reef (inner reef slope, sand trough, outer ridge, fore reef slope). Our study area was haphazardly chosen among barrier reef sites off Carrie Bow Cay. We sampled across all reef zones and had no prior data on the distribution of sponges and gobies. We expect the sponge and goby distribution patterns found at this study site to be representative of other locations across the barrier reef.

2.3.2 Sponge population

Initial surveying of all tube sponges on the reef revealed that *E. lori* was found almost exclusively in *A. fistularis*: *A. fistularis* was the only sponge species that harbored more *E. lori* than expected by chance, assuming an equal abundance of fish in all tube sponges (Chi-Square: $\chi^2 = 54.52$; df = 4; $p < 0.001$). Given this, we exclusively mapped *A. fistularis* sponges. We surveyed a 100 m wide x 300 m long transect of reef around Carrie Bow Cay by SCUBA to map the distribution of the *A. fistularis* population. The transect extended due east from the reef crest and included each reef zone out to the outer ridge of the outer fore reef (maximum depth 30 m). Three divers swam along sub-transects, approximately 3 m apart, until the entire area was surveyed. Sponge waypoints (n = 509), defined as the x and y coordinates of each sponge, were recorded using the Garmin GPSMAP®76Cx in an underwater housing produced by Sound Ocean Systems. We mapped every *A. fistularis* regardless of whether *E. lori* were present. At each sponge, we recorded the following data: depth at base of sponge (m), number of tubes, length (nearest cm), and width (nearest cm) of each tube. We focused on the role of the environmental parameter depth which has been shown to be correlated with food availability and growth rates for *A. fistularis* (Lesser 2006). We measured depth not because we believe that depth alone offers a biological explanation of distribution patterns, but because depth is a proxy for several other environmental gradients, including light availability, wave exposure, and reef zone.

2.3.3 Fish population

To map the distribution and abundance of the *E. lori* population, we recorded the number of fish at each tube, and assigned them into two size-based categories: residents or settlers. Based on the logic of previous investigators (Williams & Sale 1981; Buston 2003), residents were defined as being 18 mm standard length (S.L.) or greater, while settlers were defined as being less than 18 mm (S.L.). To test this size cut-off for *E. lori*, we collected a sub-set of residents and settlers using hand-nets and slurp guns, and measured the S.L. of each

individual with calipers to confirm that all of the fish we classified as settlers visually were $<$ 18 mm S.L. and all of the fish we classified as residents visually were $\geq$ 18 mm S.L.. This cut-off between residents and settlers coincided with an apparent shift in life history: in general, large fish (residents) lived inside sponge tubes while small fish (settlers) lived on the outside of tubes.

2.3.4 Data analysis

We analyzed sponge distribution by looking at the relationship between the independent variable 'depth' and the dependent variable 'number of sponges'. The observed number of sponges at each depth was determined from raw data, with depth categorized into four zones: 1) 0 – 10 m; 2) 10.1 – 15 m; 3) 15.1 – 20 m; 4) 20.1 – 30 m. These categories with unequal bin sizes were chosen to balance the data set and did not affect the results. The expected number of sponges at each depth was estimated by assuming that, if sponges were randomly distributed with respect to depth, the density of sponges (number sponges m^{-2}) would be the same at each depth. To generate the expected number of sponges in each zone, we multiplied the proportion of total reef area in each depth zone by the total number of sponges. We calculated the reef area in each zone by plotting the depth of each sponge against its position on the transect (Fig. 2.1), and fitting a smoothing spline through the data (JMP 8.0.1). From the fitted curve we estimated the horizontal distance encompassed by each depth zone and then estimated the surface area in each depth zone using Pythagoras' theorem. We then compared the observed distribution of sponges at each depth zone to the expected distribution of sponges at each depth zone using a chi-square test for multiple classes with subsequent subdivision of the chi-square analysis (Zar 1984).

Next, we assessed the variables that influenced sponge size and morphology using analysis of variance (JMP 8.0.1). First, treating sponges as individual entities, we used 'maximum tube length' per sponge as a metric of sponge size and investigated the relationship between 'depth' and 'maximum tube length'. Second, using 'number of tubes' as a metric

of sponge morphology, we investigated the relationship between 'depth' and 'number of tubes'. Finally, treating tubes as individual entities, we tested the relationship between 'depth' and 'tube length' to know if, and how, tube size changed. In this latter analysis, 'Sponge ID' was entered as a random effect to account for the lack of independence among tubes from the same sponge. Tukey-Kramer tests were used for all post-hoc comparisons (JMP 8.0.1).

Having determined sponge distribution, we tested which variables could be used to explain resident distribution and abundance patterns using a two-step approach. First, we investigated the probability of a resident occupying a sponge using logistic regression (JMP 8.0.1). Considering whole sponges as entities, we investigated the effect of the independent variables 'depth', 'number of sponge tubes', 'maximum tube length' and their interactions on the dependent variable 'probability of occupancy by resident(s)'. Variables were removed in a backward stepwise fashion if they did not have a significant effect ($p > 0.05$). Second, considering only the sponges occupied by residents, we investigated the predictors of the abundance of residents using analysis of covariance (JMP 8.0.1). Again, we investigated the relationship between 'depth', 'number of sponge tubes', 'maximum tube length', and their interactions on the number of residents per sponge. We used this two-step approach for two reasons: 1) to deal with the large number of zeros in our dataset; and 2) because distinct biological processes might drive presence and abundance when present. We completed similar analyses to investigate settler distribution and abundance, with the independent variable 'presence or absence of residents' included as an additional predictor.

2.4 Results

2.4.1 Sponge distribution and morphology

The observed distribution of sponges at each depth was significantly different from the expected distribution (Chi-square test: $\chi^2 = 434$, df = 3, $p < 0.001$) (Fig. 2.1). The expected density across all zones was 0.02 sponges m^{-2}. Subsequent subdivision of the

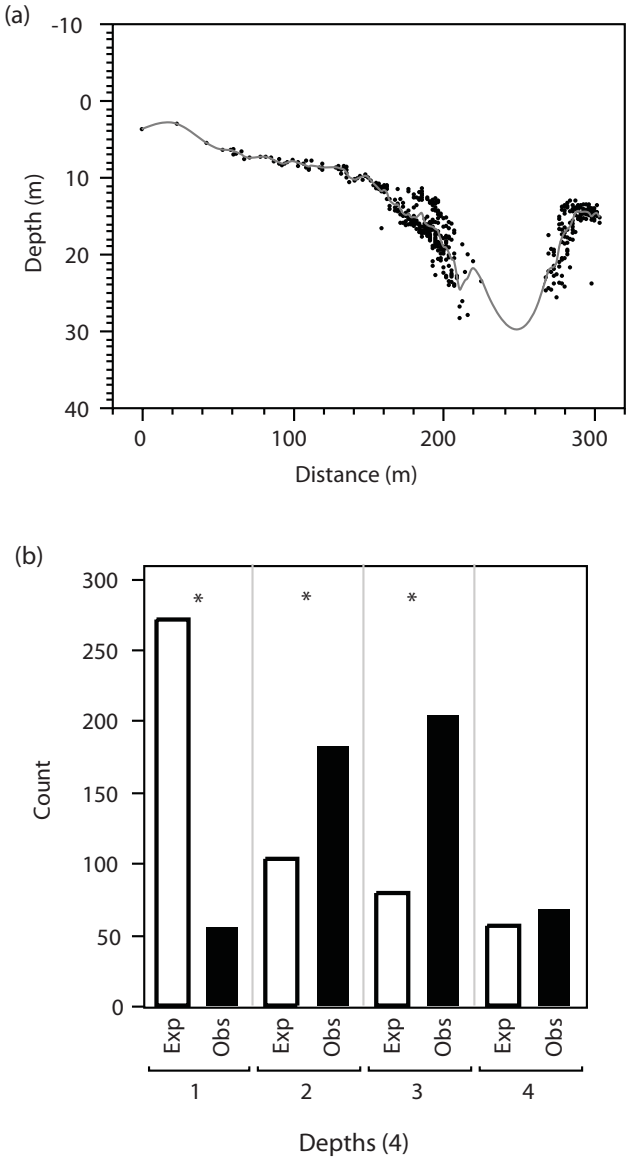

Figure 2.1: a) Distribution of sponges along the transect at Carrie Bow Cay, Belize. Dots represent individual sponges (n = 509), with a smoothing spline best fit line. Each sponge is shown at its location along the West-East transect relative to the first sponge of the transect (x = 0 m). b) Expected (white) and observed (black) sponge distribution at each depth zone, with * representing a significant difference within each zone. Depth zones are divided into the following depth categories: 1) 0 – 10 m; 2) 10.1 – 15 m; 3) 15.1 – 20 m; 4) 20.1 – 30 m.

chi-square revealed that sponges were observed less often than expected by chance from 0 – 10 m (0.004 sponges m^{-2}; $\chi^2 = 172$, df = 1, $p < 0.001$), more often than expected by chance from 10.1 – 15 m (0.03 sponges m^{-2}; $\chi^2 = 61$, df = 1, $p < 0.001$) and 15.1 – 20 m (0.05 sponges m^{-2}; $\chi^2 = 198$, df = 1, $p < 0.001$), and no different than expected by chance from 20.1 – 30 m (0.02 sponges m^{-2}; $\chi^2 = 2.57$, df =1, $p > 0.05$).

At the scale of the sponge, the mean 'maximum tube length' was 21.19 cm ± 14.38 (mean ± s.d.). 'Maximum tube length' of the sponge was also related to 'depth' ($r^2 = 0.08$; ANOVA: F = 16.11, df = 3, $p < 0.0001$), with sponges at greater depths generally having longer maximum tube lengths than sponges at shallower depths (Fig. 2.2a). Specifically, a post-hoc Tukey-Kramer test revealed that sponges at depths greater than 15 m had longer maximum tube lengths than sponges shallower than 15 m, and sponges at 10 – 15 m had longer maximum tube lengths than sponges at 0 – 10 m ($p < 0.05$).

The mean number of tubes per sponge was 1.66 ± 1.05 (mean ± s.d.). The variable 'number of tubes per sponge' was related to 'depth' ($r^2 = 0.06$; ANOVA: F = 11.51, df = 3, $p < 0.0001$), with sponges at greater depths having fewer tubes than sponges at shallower depths (Fig. 2.2b). A post-hoc Tukey-Kramer test showed that sponges at 0 – 10 m have significantly more tubes than sponges in all other depth categories ($p < 0.05$).

At the scale of the tube, the mean tube length was 18.43 cm ± 13.94 (mean ± s.d.). We found that 'tube length' was positively related to 'depth' (Fixed Effect Test: F = 22.04, df = 3, $p < 0.0001$). Consistent with the sponge-level analysis, a post-hoc Tukey HSD test revealed that individual tubes at depths greater than 15 m were significantly longer than tubes at 15 m or shallower, and tubes from 10.1 – 15 m were longer than tubes from 0 – 10 m. This model provided the strongest explanation for variation in sponge morphology ($r^2 = 0.67$). A reduced model without the random effect explained 8% of the variation in tube length, consistent with analyses at the scale of the sponge. This indicates that the majority of variation was accounted for by the random effect 'Sponge ID'.

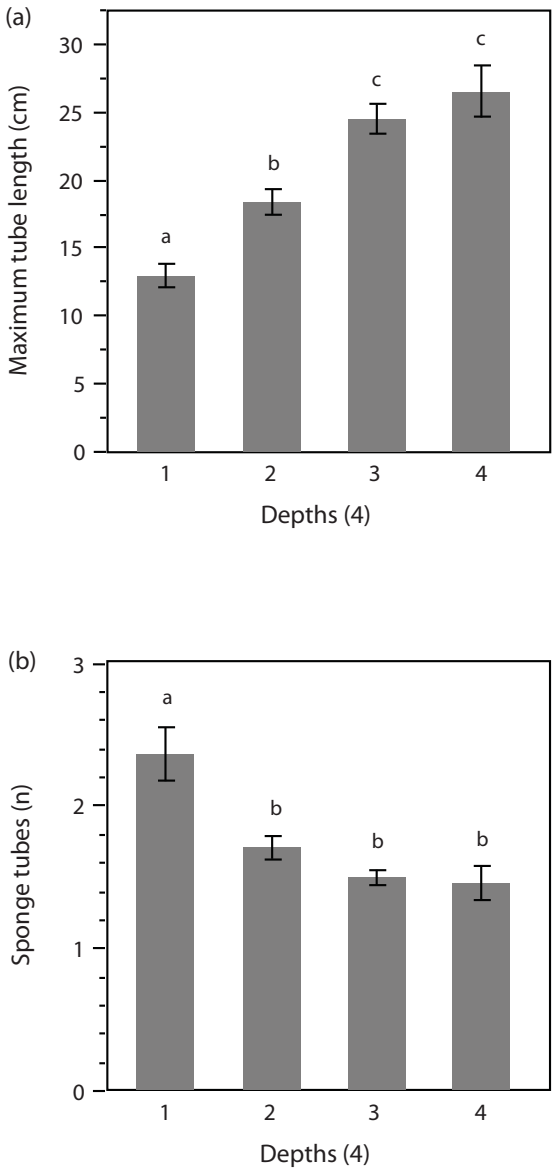

Figure 2.2: Sponge size and morphology versus depth. Depth zones are divided into the following depth categories: 1) 0 – 10 m; 2) 10.1 – 15 m; 3) 15.1 – 20 m; 4) 20.1 – 30 m. Zones not connected by letters are significantly different in both panels. Gray bars represent the least squares mean estimate in each zone and vertical bars represent the standard error (S.E.) of the mean. a) Maximum tube length per sponge at four depth categories; b) Number of tubes per sponge at four depth categories.

2.4.2 Resident distribution

The mean probability of one or more residents occupying a sponge was 0.32. The probability of occupancy was related to 'depth', 'number of sponge tubes', 'maximum tube length' and the interaction between 'depth' and 'maximum tube length' ($r^2 = 0.36$) (Fig. 2.3; Table 2.1). The interaction indicates that the relationship between the probability of occupancy and maximum tube length varies with depth (Fig. 2.3a). Specifically, the initial rate of increase in the probability of occupancy with maximum tube length was much lower at depths less than 10 m than at all other depths. In addition to the interacting effects of depth and maximum tube length, the probability of resident occupancy increased with the number of sponge tubes (Fig. 2.3b).

Table 2.1: Likelihood ratio tests for logistic regression model predicting the probability of resident occupancy on sponges. These tests demonstrate whether the whole model is a significantly better fit than a model without a given variable.

Variable	**df**	$\boldsymbol{\chi^2}$	**p**
Depth	3	16.75	0.0008
Number of sponge tubes	1	4.53	0.003
Maximum tube length	1	16.35	<0.0001
Maximum tube length * Depth	3	13.18	0.004

Considering only occupied sponges, the mean number of residents per sponge was 1.27 $\pm$ 0.60 (mean $\pm$ s.d). The number of residents per sponge was related to 'number of sponge tubes' ($df = 1$; $F = 59.26$; $p < 0.0001$), 'maximum tube length' ($df = 1$; $F = 8.47$; $p = 0.004$), and the interaction between 'number of sponge tubes' and 'maximum tube length' ($df = 1$; $F = 17.48$; $p < 0.0001$) (ANCOVA: $r^2 = 0.41$). The interaction between the number of tubes and maximum tube length indicates that the effect of one sponge metric on the number of residents varies with the other sponge metric. Specifically, the effect of maximum tube length on the number of residents is small when the number of tubes is small, and *vice versa*. In other words, the number of residents was highest in sponges with many, large tubes.

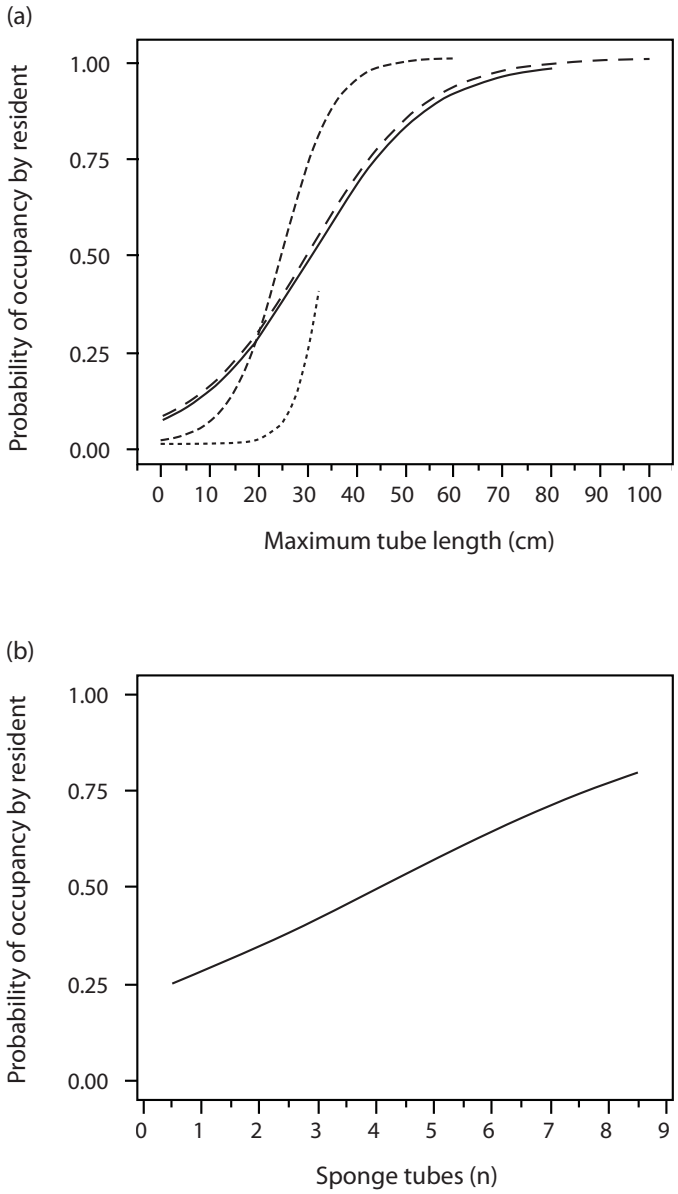

Figure 2.3: Probability of residents occupying a sponge as predicted by three variables. a) Probability of occupancy by resident is related to the interaction between maximum tube length and depth. There are four depth zones: 1) 0 – 10 m, short dashed line; 2) 10.1 – 15 m, medium dashed line; 3) 15.1 – 20 m, long dashed line; 4) 20.1 – 30 m, solid line; b) Probability of occupancy by resident is related to the number of sponge tubes. In both a and b, lines represent the relationship between the probability of occupancy and independent variables estimated from the parameters of the logistic model.

2.4.3 Settler distribution

The mean probability of one or more settlers occurring at a sponge was 0.06. Their occurrence was related to the 'number of sponge tubes' and 'maximum tube length' (r^2 = 0.45) (Fig. 2.4), but not the presence or absence of residents (Table 2.2). The main effects indicate that the probability of settlers occurring at a sponge increased with both the number of sponge tubes and maximum tube length, with maximum tube length being a stronger predictor. Considering only settler-occupied sponges, the mean number of settlers per sponge was 1.26 ± 0.44 (mean ± s.d.). The 'number of settlers' was related to 'maximum tube length', with the number of settlers increasing with maximum tube length (r^2 = 0.18; ANCOVA: df = 1; F = 6.32; p = 0.018).

Table 2.2: Likelihood ratio tests for logistic regression model predicting the probability of settler occurrence on sponges. These tests demonstrate whether the whole model is a significantly better fit than a model without a given variable.

Variable	**df**	$\boldsymbol{\chi^2}$	**p**
Number of sponge tubes	1	16.08	<0.0001
Maximum tube length	1	79.86	<0.0001

2.5 Discussion

In this study, we simultaneously consider factors influencing sponge and goby distribution on the reef around Carrie Bow Cay, Belize using a large empirical data set and multivariate analyses. Given that the goby, *E. lori*, is an obligate sponge-dweller, we also use sponge characteristics to predict fish distribution and abundance. We found the sponge *A. fistularis* to be distributed non-randomly across the reef, with sponge density peaking at intermediate depths (10 – 20 m). Sponge morphology also varied across the reef, and sponges tended to have fewer, but longer tubes with increasing depth. Considering only *A. fistularis* as fish habitat, fish presence also increased with depth (over and above the effect of depth on sponges), and fish were more common in bigger sponges with multiple tubes. There was no significant effect of conspecific fish interactions on settler distribution patterns.

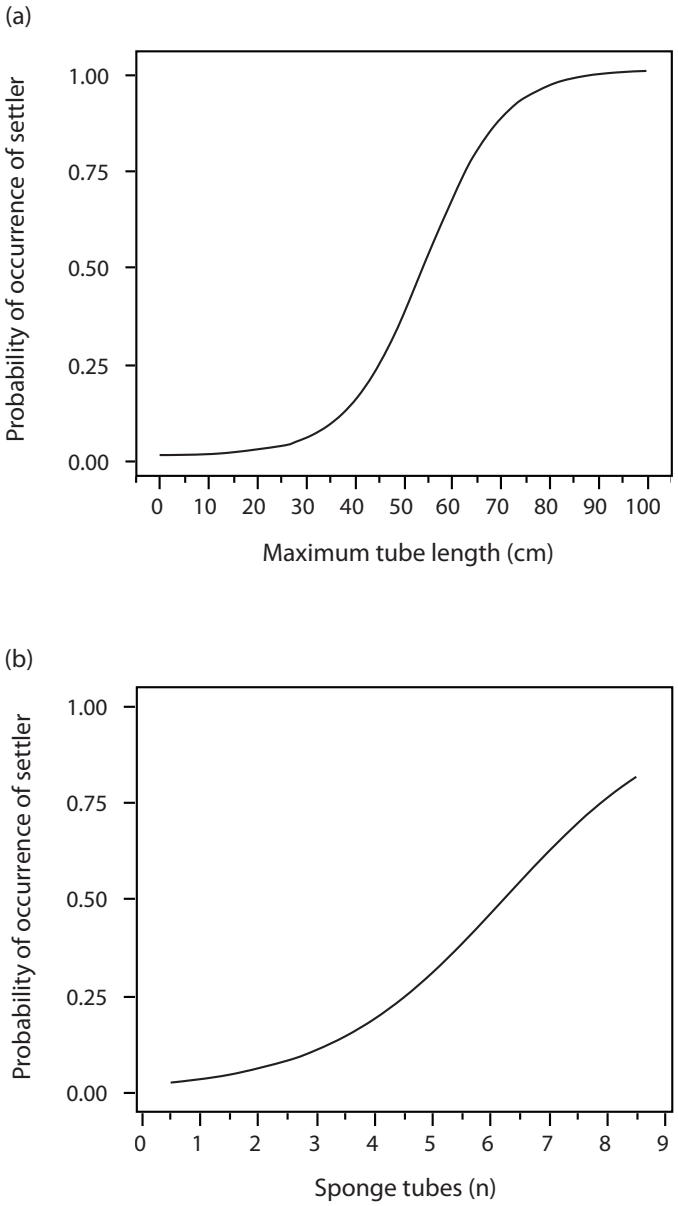

Figure 2.4: Probability of settlers occurring at a sponge as predicted by two variables. a) Probability of occurrence is related to maximum tube length; b) Probability of occurrence is related to the number of sponge tubes. In both a and b, lines represent the relationship between the probability of occurrence and independent variables estimated from the parameters of the logistic model.

Given our results, we can consider why sponge distribution and morphology might vary with depth. We found that sponges were sparsely distributed above 10 m, and sponges at these shallow depths were significantly shorter than sponges at greater depths. It is possible that this morphology, and the low density of sponges at shallow depths, can be attributed to close proximity to strong surface currents (Palumbi 1984; Wilkinson & Evans 1989; Mercado-Molina & Yoshioka 2009). Sponges were densely distributed along the reef at depths of 10 – 20 m. This pattern of sponge distribution, with fewer sponges shallower than 10 m and a density peak at intermediate depths, is consistent with previous research (Wilkinson & Chesire 1989; Wilkinson & Evans 1989). *A. fistularis* is usually not found below 40 m, though it has been shown to survive at up to 100 m in transplantation experiments (Maldonado & Young 1998). Sponge size also increased with depth, a pattern that has been observed in several species of tube sponges. Lesser (2006) found that *A. fistularis* and other tube sponges are bigger and grow faster with increasing depth up to 30 m because their primary food source, picoplankton, is more abundant. If food is more abundant at 10 – 20 m, and sponges are protected from near-surface turbulence, sponges may simply live longer and grow more at these depths, which could explain the patterns we observe.

The models describing the relationship between depth and sponge metrics explained low levels of variation in our data. This was particularly true for the whole sponge-level analyses. Depth explained only 6% of the variation in number of tubes per sponge, and 8% of the variation in maximum tube length. Interestingly, the tube-level analysis was a better model in terms of explaining variance: the model explained 67% of the variation in tube length. However, most of that variation was explained by the random effect 'Sponge ID'. This result is indicative of a site effect, where tubes from a single sponge were generally more similar to each other in comparison to tubes at other sponges. Tubes from a single sponge live in the same small-scale microhabitat on the reef and are likely to be close in age. They may therefore have a shared history of exposure to biotic and abiotic parameters such as predation and wave action (none of which we measured). Site effects may be particularly

influential for long-lived organisms such as sponges and sea anemones.

Given the resident distribution results, we can hypothesize why *E. lori* have high occupancy rates and are most abundant in large, multi-tubed sponges at depths greater than 10 m. Residents were less likely to occur in sponges at depths shallower than 10 m than in sponges at all other depths. However, a closer look at the statistical interaction between depth and maximum tube length indicates that depth per se may not be limiting resident distribution. Within this shallow depth zone, the slope of the relationship between maximum tube length and the probability of resident occupancy initially increases slowly, but sharply increases when maximum tube length reaches 30 cm (Fig. 2.3). This suggests that it may be the lack of large sponges above 10 m (Fig. 2.1 & 2.2) that explains why there are few residents found there.

Large sponges could offer fitness advantages to the fish by influencing survival and reproductive success. First, resident survival rates may be higher in large sponges if their deep oscula offer enhanced protection from predation. Second, larger tubes may offer more nutrients to *E. lori*, influencing both survival and reproductive success. Food can be a limiting resource to fish even within suitable habitat (Jones 1986; Wong *et al.* 2008). Larger tubes may offer more nutrients if they harbor more *H. spongicola* polychaetes, and if *E. lori* eats these polychaetes. Stomach content analyses will be needed to test this hypothesis, but analyses of closely-related *Elacatinus spp.* have shown *H. spongicola* to constitute a high percentage of the gut contents (Smith & Tyler 1972; Colin 1975; Whiteman & Cotê 2002). Another plausible hypothesis is that being associated with a large microhabitat enables the residents to grow large, an effect seen in other microhabitat-associated species (Fautin 1992; Elliot & Mariscal 2001; Buston 2002). In such species, female body size can be positively correlated with the number of eggs hatched by breeding pairs (Buston & Elith 2011). If these ecological processes lead to differential survival and reproductive success, then they will lead to natural selection for pre-settlement, post-settlement or post-recruitment preferences for large sponges.

For *E. lori*, our data suggest that settlers were most abundant on the same sponge

archetype as residents (long sponges with multiple tubes). However, there was no significant relationship between the presence of residents and the distribution of settlers. Even though there is no significant effect of conspecifics on settlement, preliminary behavioral observations indicate that there may be a negative effect on recruitment into the sponge. Residents evicted settlers that were pushed into tubes, and settlers only recruited into tubes when residents were absent. Resident fish of several fish species have been documented to push settlers and/or juveniles to the edges of their microhabitat, and this has been shown to have a negative effect on recruitment (Holbrook & Schmitt 2002; Almany 2003, 2004; Buston 2003; Wong *et al.* 2007; Ben-Tzvi *et al.* 2008). Taken together, these observations suggest that although conspecific interactions might influence recruitment, they are not a strong predictor of *E. lori* settler distribution and abundance patterns.

One interesting aspect of this study system, in contrast to other systems (e.g., anemonefishes), is that the sponge habitat is not totally saturated by fish. This is all the more remarkable given that settlers remain on the outside of resident-occupied sponges, presumably exposed to predators, while there are nearby sponges with no residents. Settlers may stay at occupied sponges because the benefits of staying outweigh the costs of leaving (Emlen 1982, 1991). In this case, the chance of inheriting the tube or becoming the residents mate outweighs the energy cost or risk of predation in searching for another suitable sponge (Buston 2004; Wong *et al.* 2007; Wong 2010). It has also been shown in other fishes that settlers remaining close to residents may gain protection from predators because of resident aggression towards predators (Sandin & Pacala 2005; White & Warner 2007), though the observation that residents rarely leave their tubes makes this hypothesis less plausible in *E. lori*.

Overall, these results provide a framework for understanding how different ecological factors may influence the distribution and abundance of two interacting reef species. All reef species are interconnected with members of their community, and this system provides an excellent opportunity to explore the relationship between two tightly linked reef species — a sponge goby and the tube sponge it associates with. Given the fish's strong site-

attachment, this could be a good system for marine ecological studies. Future studies will benefit from the natural history framework that this study provides, and will focus on the behavioral ecology and population ecology of these fish.

2.6 Acknowledgments

We thank the Belizean government and Fisheries Department for permission to dive on the reefs of Belize and the staff at the Wee Wee Cay Marine Station for use of their facilities. We are also grateful to Phil Lobel, Lisa Lobel, and Amparo Carillo Gavilán for assistance in the field. Jelle Atema, Marian Wong, and three anonymous reviewers provided helpful comments on this manuscript. Funding was provided by a start up award to Peter Buston from Boston University. All work was approved by Belize Fisheries and the Boston University IACUC.

2.7 Publication

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.

2.8 Author Contributions

CCD, JEM, and PMB conducted field work. PMB and CCD analyzed the data. CCD wrote the paper and all authors contributed to revisions.

Chapter 3

Self-recruitment in a Caribbean reef fish: a method for approximating dispersal kernels accounting for seascape

3.1 Abstract

Characterizing patterns of larval dispersal is essential to understanding the ecological and evolutionary dynamics of marine metapopulations. Recent research has measured local dispersal within populations, but the development of marine dispersal kernels from empirical data remains a challenge. We propose a framework to move beyond point estimates of dispersal towards the approximation of a simple dispersal kernel, based on the hypothesis that the structure of the seascape is a primary predictor of realized dispersal patterns. Using the coral reef fish *Elacatinus lori* as a study organism, we use genetic parentage analysis to estimate self-recruitment at a small spatial scale (less than 1 km). Next, we determine which simple kernel explains the observed self-recruitment, given the influx of larvae from reef habitat patches in the seascape at a large spatial scale (up to 35 km). Finally, we complete parentage analyses at 6 additional sites to test for export from the focal site and compare these observed dispersal data within the metapopulation to the predicted dispersal kernel. We find 4.6% self-recruitment ($CI_{95\%}$: $\pm$ 3.0%) in the focal population, which is explained by the exponential kernel $y=0.915^x$ ($CI_{95\%}$: $y=0.865^x$, $y=0.965^x$), given the seascape. Additional parentage analyses showed low levels of export to nearby sites, and the best-fit line through the observed dispersal proportions also revealed a declining

function y=0.77^x. This study lends direct support to the hypothesis that the probability of larval dispersal declines rapidly with distance in Atlantic gobies in continuously distributed habitat, just as it does in the Indo-Pacific anemonefishes in patchily distributed habitat.

3.2 Introduction

Connectivity influences virtually all ecological and evolutionary processes in metapopulations, including population growth rates, resilience to perturbations, gene flow, local adaptation, and divergence (Hanski 1999; Hanski & Gaggiotti 2004; Cowen *et al.* 2007; Cowen & Sponaugle 2009). For many organisms with bipartite life cycles, population connectivity is primarily driven by propagule dispersal. Therefore, quantifying spatial patterns of propagule dispersal is essential to predicting metapopulation dynamics and to designing effective management plans (Fahrig & Merriam 1994; Botsford *et al.* 2001, 2009; Roberts *et al.* 2003; Sale *et al.* 2005; Blowes & Connolly 2012).

While much research has focused on the predictors, patterns, and consequences of dispersal in terrestrial species (Nathan & Miller-Landau 2000; Clobert *et al.* 2001), marine dispersal research has lagged behind. This lag is attributable to the fact that most marine species disperse via larvae, and sampling marine larvae—extremely small propagules that travel in the pelagic environment for weeks to months—is logistically challenging. Recently, the application of genetic parentage analysis to marine systems has led to great progress in quantifying local patterns of dispersal for marine fishes (Jones *et al.* 2005; Planes *et al.* 2009; Christie *et al.* 2010; Saenz-Agudelo *et al.* 2011; Buston *et al.* 2012; Harrison *et al.* 2012). The advantage of genetic parentage analysis over other methods lies in the ability to directly uncover demographic connectivity, or the movement of individuals, which is essential for understanding metapopulation dynamics (Hanski 1999; Hanski & Gaggiotti 2004; Hedgecock *et al.* 2007; Broquet & Petit 2009; Leis *et al.* 2011).

A simple comparison of published marine dispersal estimates suggests that there is large variation in local dispersal patterns, measured by self-recruitment. Self-recruitment

is defined as the percentage of locally settling individuals that were spawned by local parents, and is a relatively easy dispersal metric to measure. Empirical data have revealed self-recruitment estimates ranging from 0% (*Amphiprion polymnus*, Saenz-Agudelo *et al.* 2011) to upwards of 60% (*Amphiprion percula* and *Chaetodon vagabundus*, Almany *et al.* 2007), and spanning the range in between (Jones *et al.* 2005; Planes *et al.* 2009; Saenz-Agudelo *et al.* 2011; Harrison *et al.* 2012). It has been suggested that major determinants of variation in self-recruitment may be species traits (e.g., pelagic larval duration) and seascape. Understanding the causes of the observed variation is, however, challenging because most of the studies have been conducted in one seascape (Kimbe Bay) using one genus (*Amphiprion*), though new studies are emerging (Christie *et al.* 2010; Harrison *et al.* 2012). An important next step in dispersal research is to collect empirical data from different systems, e.g., different taxa and seascapes, and test how potential predictors of dispersal affect variation in local dispersal estimates.

The idea that habitat patchiness, or fragmentation, can have profound impacts on patterns of demographic and genetic connectivity has strong foundations in the terrestrial literature (Hanski 1999; Hanski & Gaggiotti 2004; Manel *et al.* 2003; Storfer *et al.* 2007), but has only recently come to the fore in marine ecology (Jones *et al.* 2009; Saenz-Agudelo *et al.* 2011; Pinsky *et al.* 2012). A recent marine model explicitly illustrates that isolated populations in patchy seascapes are predicted to have high self-recruitment, whereas populations situated in continuous seascapes are predicted to have low self-recruitment (Pinsky *et al.* 2012). This follows from the fact that the denominator of the self-recruitment fraction increases with the influx of larvae from additional patches on the seascape, creating a dilution effect (Fig. 3.1[a1] versus Fig. 3.1[b1]). Further, if the probability of dispersal declines with distance, the presence or absence of *nearby* reef patches will have a disproportionate effect on self-recruitment estimates, relative to the presence or absence of *distant* reef patches. The same applies to measures of larval export (Fig. 3.1[a2] versus Fig. 3.1[b2]). Importantly, this dilution phenomenon will hold even if the underlying dispersal kernel is the same in these two scenarios. Following this logic, it may be possible to estimate marine

dispersal kernels from empirical self-recruitment estimates and the seascape structure.

We propose a framework for incorporating seascape analyses into quantitative marine dispersal estimates using the sponge-dwelling neon goby *Elacatinus lori* as a study organism. *E. lori* has several characteristics that make it a tractable study organism for direct dispersal studies. First, *E. lori* is an endemic to the Mesoamerican barrier reef (MBR) system, where it is a specialist on tube sponges (Colin 1975, 2002, 2010), particularly the yellow tube sponge *Aplysina fistularis* (D'Aloia *et al.* 2011). This microhabitat association enables efficient location and collection of individuals within a study area. Second, a life history transition that is correlated with body size facilitates genetic parentage analysis—in this system we define 'settlers' as individuals that have recently settled on the outside of sponges upon completing the larval phase (S.L. $<$ 18 mm), and 'residents' as individuals that have recruited inside the sponge (S.L. $\geq$ 18 mm) (D'Aloia *et al.* 2011). Parentage analyses can therefore focus on assigning settlers to potential resident parents. Third, its location on the MBR, a reef system that is relatively continuous over a thousand kilometers, offers an important spatial contrast to dispersal studies that have focused on patchy seascapes in the Pacific (Jones *et al.* 2005; Almany *et al.* 2007; Planes *et al.* 2009; Harrison *et al.* 2012). Capitalizing on these characteristics of *E. lori*, we develop a method for combining estimates of self-recruitment with knowledge of the seascape structure to approximate a simple dispersal kernel.

We characterize *E. lori* dispersal in three steps: 1) we apply genetic parentage analysis to a focal study population on the MBR to describe the pattern of local dispersal (i.e. self-recruitment) along a 0.5 km stretch of reef; 2) we assess which simple negative exponential function can explain the observed self-recruitment percentage, accounting for influx from all other sites; 3) we complete additional parentage analyses to test for export to 6 nearby sites, and compare the best-fit line through the observed dispersal data with the approximated dispersal function. This study provides further direct empirical support for the hypothesis that there is a rapid decline in the probability of marine larval dispersal with respect to distance.

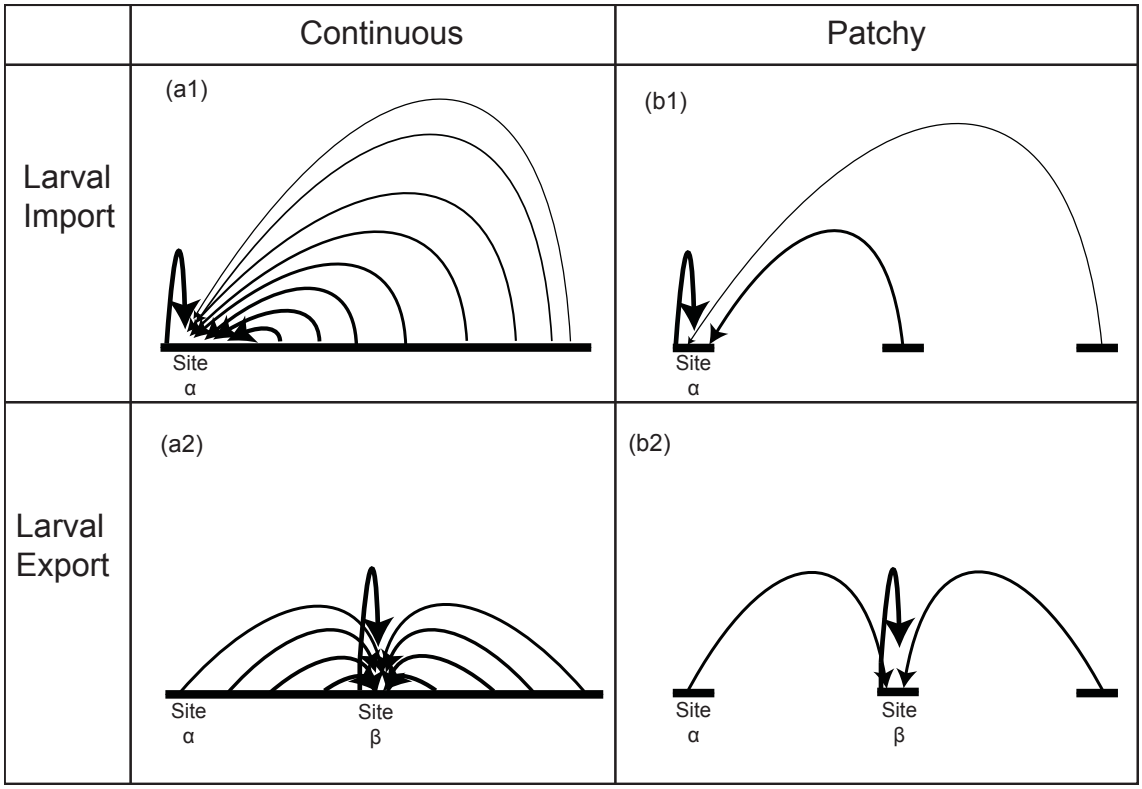

Figure 3.1: Hypothesized effect of seascape patchiness on self-recruitment and export. In all panels, thickness of the arrows is proportional to the number of dispersive larvae; (a1) Self-recruitment at site α will be low on a continuous reef when there is a dilution effect by the influx of larvae from many nearby sites; (b1) Self-recruitment at site α will be high on a patchy reef, when few nearby patches produce larvae that settle at site α; (a2) Site α will export a small proportion of the total larval settlement at any additional site (e.g. site β) on a continuous reef, due to the dilution effect; (b2) Site α will export a larger proportion of the total larval settlement at an additional site (e.g. site β) on a patchy reef because there are few nearby sites contributing larvae.

3.3 Methods

3.3.1 Study metapopulation

This study was focused on a population of *E. lori* inhabiting *A. fistularis* within a 500 m long by 125 m wide area, between depths of 10 m and 25 m, on the barrier reef at Curlew Caye, Belize (16° 47' 23" N 88° 04' 33" W) (Fig. 3.2). This depth range represents the depths at which the fish and their host sponge are most abundant (D'Aloia *et al.* 2011); the width of the study area simply represents the width of the barrier reef between these depths. The area between 10 and 25 m was sampled completely, i.e., all *E. lori* individuals in *A. fistularis* greater than 10 cm in length were sampled. We excluded sponges less than 10 cm in length from our study because they are rarely occupied by *E. lori* (D'Aloia *et al.* 2011).

To characterize self-recruiting dispersal trajectories in three dimensions (distance, direction and depth change) we measured depth at the base of each sponge (m) and marked the location of every sponge in the study area using a GPSMAP®76Cx (Garmin, USA) in an underwater housing (Sound Ocean Systems, Inc., Redmond, Washington, USA). From here, we calculated the net distance, depth displacement, and directional change (N-S and E-W) between the origin and destination sponges of the self-recruiting trajectories.

To uncover self-recruiting dispersal trajectories, we collected tissue from all residents and settlers in the study area. All residents (n = 425) were collected using slurp guns and fin clipped using scissors. Then, they were placed in a plastic bag and their standard length (S.L.) was measured to the nearest mm before they were returned to their sponge. Settlers (n = 194) were collected using slurp guns and transported to the surface in plastic bags. There, they were euthanized using MS-222 because they were too small for non-lethal fin clipping. All tissue samples were stored in 95% EtOH.

Additional tissue samples were collected from nearby sites along the barrier reef to test whether export from the Curlew population could be detected. We sampled 3 sites to the north and 3 sites to the south of Curlew (Fig. 3.2b). These additional sites were chosen

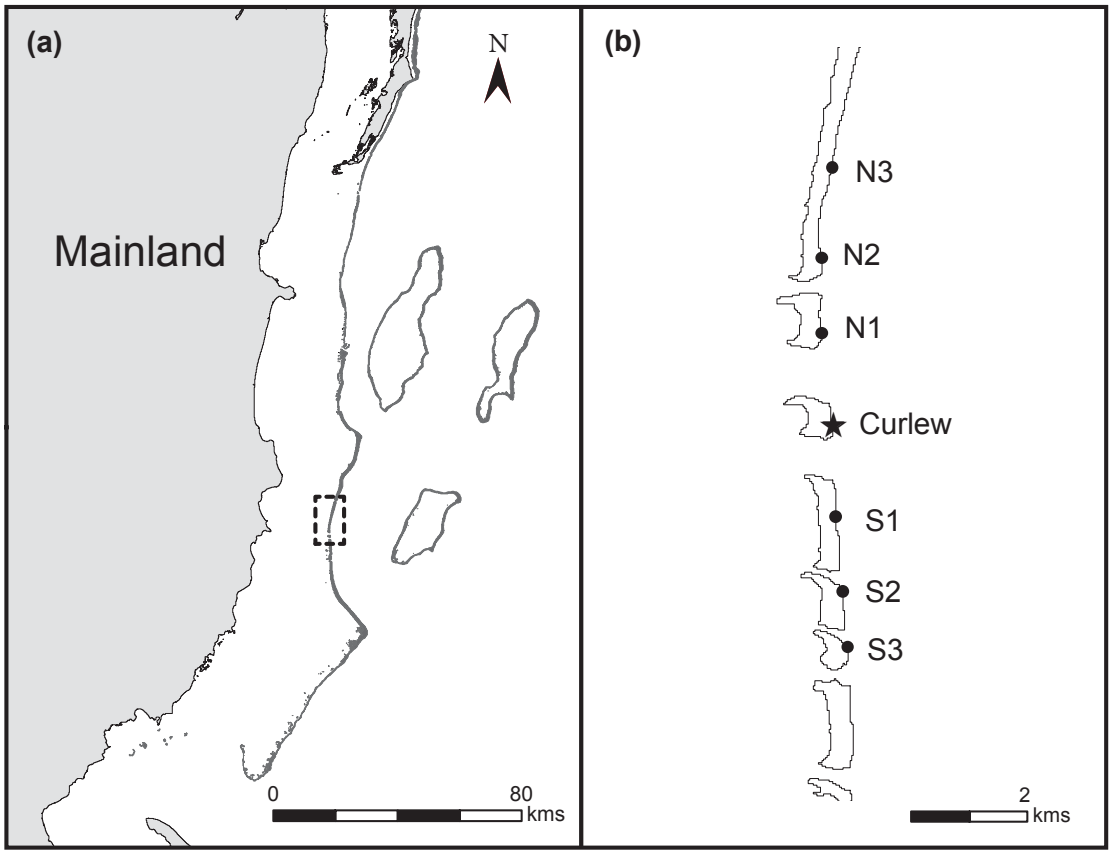

Figure 3.2: Study area on the Mesoamerican Barrier Reef (MBR). (a) Map of the Belizean section of the MBR, including the three oceanic atolls. The study area lies within the dashed box. (b) Close-up of the study area around Curlew Caye. The black star represents the focal population sampled at Curlew Caye and black dots represent additional sites where settlers were collected for export analysis. Site N1 represents Carrie Bow Caye, site of the Smithsonian's CCRE field station.

based on their proximity to the focal site and were spaced 1 – 1.5 km apart. Approximately 30 settlers were collected at each site.

3.3.2 Isolation and characterization of microsatellites

E. lori genomic DNA was digested with the restriction enzyme *Hinc II* and ligated to a double-stranded linker (Andrés & Bogdanowicz 2011). Digested, ligated DNA fragments were enriched for microsatellites by hybridization to 3'-biotinylated oligonucleotide repeat probes representing two dimers, five trimers, five tetramers and two pentamers. Enriched

fragments were captured by streptavidin-coated magnetic beads, amplified by PCR, and ligated to a Roche/454 rapid library adapter. The sample was submitted to the Sequencing and Genotyping Facility at the Cornell Life Sciences Core Laboratory Center (CLC) for Titanium 454 sequencing.

Raw data were imported to SeqMan Pro (Lasergene v 8.1.1, DNASTAR, Inc.) and assembled (average Q score= 16 for quality trimming, mer size= 120, minimum match percentage=94, repeat handling on). Post assembly, sequences smaller than 150 base-pairs were discarded; both multi-read and single-read contigs above this size were kept. We used the program msatcommander (version 1.0.3, for Mac OSX) to scan the data for all dimeric, trimeric, tetrameric, pentameric and hexameric microsatellites and design primers. Minimum perfect repeat lengths were six for dimers, and five for all other repeat types selected. Minimum, optimum, and maximum primer lengths were set to 22, 23, and 24 bases, respectively. All other settings in msatcommander were kept at defaults.

We used the three-primer PCR method (Schuelke 2000) to assay loci for PCR quality and levels of polymorphism, on 3100 and 3730xl DNA Analyzers (Life Technologies). Loci that appeared "clean" and polymorphic had one primer re-synthesized with a 5' fluorescent dye compatible with the G5 dye set (Life Technologies). All reverse primers were "pig-tailed" (Brownstein *et al.* 1996) to promote adenylation of the fluorescently-labeled strand. Allele sizing data at the three-primer PCR stage were used to inform PCR multiplexing of loci for high throughput genotyping.

3.3.3 Genotyping

Genomic DNA was extracted with the Agencourt DNAdvance kit (Beckman Coulter, Inc., USA) and individuals were genotyped at 14 microsatellite loci (Table 3.1) using two sets of multiplex PCR with the Type-It Microsatellite PCR Kit (Qiagen, USA). Each 10 μL PCR reaction contained: 1 μL DNA, 3 μL H_2O, 5 μL 2x Type-it Multiplex PCR Master Mix, and 1 μL 10x primer mix. Thermal cycling followed a touchdown protocol: 95 °C for 5 min; six cycles of 94 °C for 50 s, 59-54 °C for 90 s (annealing temperature reduced

1 °C each cycle), 72 °C for 30s; 23 cycles of 94 °C for 50 s, 53 °C for 90 s, 72 °C for 30 s; 60 °C for 30 min. Diluted amplicons (1 PCR product: 35 H_2O) were combined with formamide and genescan LIZ-500 size standard (Applied Biosystems, USA), and screened on an ABI 3730 automated sequencer. Allele sizes were determined with GENEMAPPER v. 4.0 (Applied Biosystems, USA).

Summary statistics (allele frequencies, observed and expected heterozygosity) were generated in CERVUS v. 3.0 (Kalinowski *et al.* 2007). Deviations from Hardy-Weinberg equilibrium (HWE) were tested with GENEPOP v. 4.1.4 (Raymond & Rousset 1995) using Markov chain methods (dememorization number = 10,000; number of batches = 10,000; iterations per batch = 1,000). The presence of null alleles, large allele dropout, and stuttering was checked on MICRO-CHECKER (van Oosterhout *et al.* 2004).

3.3.4 Parentage analysis and self-recruitment

To assign self-recruiting offspring from Curlew to their parents, a categorical likelihood-based parentage analysis was conducted in CERVUS v 3.0. Prior to running the parentage analysis, a simulation was run in order to determine the critical LOD value (natural log of the likelihood ratio) for assigning offspring to parents at a 95% confidence level. The simulation generated both true parents and unrelated candidate parents for 10,000 hypothetical offspring created from the observed allele frequencies in the population. The critical LOD value was determined by comparing the LOD distribution when the true parents were assigned with the LOD distribution when unrelated candidates were assigned. Because CERVUS requires an estimate of the proportion of candidate parents sampled, we tested a range of proportions (0.01 – 0.70) but found that the input value did not ultimately affect the results of the subsequent parentage analysis. However, parameter inputs do affect the number of expected assignments through the parentage simulation, so we use a conservative sampling proportion of 0.01 to report the simulation results.

Once the simulation was complete, a parentage analysis was run to test the pool of offspring (n = 194) against potential parents (n = 425) with a mistyping percentage of

Table 3.1: Polymorphic microsatellite markers developed for *Elacatinus lori*. Reported characteristics at each locus are based on the population sampled at Curlew Caye, Belize (n = 619 individuals): number of alleles (k), observed (H_O) and expected (H_E) heterozygosity, and deviations from Hardy-Weinberg equilibrium (HWE). An asterisk * indicates a significant deviation from HWE ($p<0.05$). Italicized sequences at the 5' end of most upper primers are not part of the *E. lori* genome, and serve to optimally separate allele sizes across loci. The sequence GTTTCTT for some lower primers promotes adenylation of the labeled strand (Brownstein *et al.* 1996).

Locus	**Repeat motif**	**Primer Sequence (5'-3')**	**k**	**H_O**	**H_O**	**HWE**
1419tri	AAC	*CGAGTTTTCCCAGTCACGAC*GCGACGGGGAGCCTCAAAT *GTTTCTTCGAGTTTTCC*ATGATTCGGCCGATACGATGGA	17	0.323	0.688	<0.001*
14528tet	AAAC	*CGAGTTTTCCCAGTCACGAC*GCGCATGAGCCCGTTTTT *GTTTCTTCGAGTTTTCC*GTGCCACCGGTTGCTCTTG	19	0.856	0.874	0.518
18144tri	AAC	*CGAGTTTTCCCAGTCACGAC*GACCCGGATTAGTCCTGGTTTG *GTTTCTTCGAGTTTTCCGAC*CGGAGTAAATGTTGGCTCAC	13	0.567	0.572	0.282
21378tri	AAC	*CGAGTTTTCCCAGTCACGAC*CCTCCCTCCCCAGCACA *GTTTCTT*TTTGTCCAAGTCTAGCAGGTATTC	35	0.811	0.935	<0.001*
23415tet	AACT	CACCGTTATTGACTAAAGTGTTCT *GTTTCTTT*ACAGATCCAGATCATCATCCA	25	0.889	0.893	0.364
23889tri	AAAC	GGTGACTCCCGTGTGAAGAGC *GTTTCTT*AACATGCTATGGCTAACACTGACG	21	0.901	0.885	0.643
25362tri	AAC	GCTAAAGGCGGTCCATCAAAA *GTTTCTT*CAGCGTCCAACAGTGTCTTCAGT	29	0.890	0.923	0.382
25632tet	AAAC	TCCAGTGCTATTTGTGGCATTGTT *GTTTCTT*CTTATTCCCGGTTCCTCCACTGAT	33	0.947	0.945	0.053
25745tet	AAAC	*CAGTCACGAC*GGGGGAGGCATGTGATGT *GTTTCTT*AAACTTCCTCTCGCCATAGTGA	10	0.629	0.772	<0.001*
29109tet	AAAC	*CGAGTTTTCCCAGTCACGAC*AGCCGGGGATTTAGGCAGGAATAA *GTTTCTTCGAGTTTTCC*GAGAGGCTAAATTGATTAAATGCTC	43	0.961	0.959	0.300
6231tet	AAAC	CCAGGGATAGTTTCACGATGTT *GTTTCTT*ATGGATTTCCCCTTCAGTGTTT	29	0.942	0.941	0.720
6266tri	AAC	*CGAGTTTTCCCAGTCACGAC*GTAGACGGACGTGTGAGGTTGTT *GTTTCTTTCGAGTTT*TCCCCCCCAGACTTGTGAATAATGTGA	22	0.816	0.842	0.787
6326tet	AAAC	GGTCCATGGTCCCAAAGAAAC *GTTTCTT*CAATGTTAGCGCCCAATGTTCG	10	0.711	0.700	0.799
985tet	AGAT	*CGAGTTTTCCCAGTCACGAC*TTTCCCTGCAGCTGTCAGACT *GTTTCTTCGAGTTTTCC*AAAGCTCCACATCCGATTTCA	76	0.917	0.982	<0.001*

1% to account for potential genotyping errors. This level of mistyping is conservative, but consistent with percentages used in previous studies of this type (Jones *et al.* 2005; García *et al.* 2007; Christie *et al.* 2010; Saenz-Agudelo *et al.* 2011). Putative parent-offspring assignments were only accepted if they had a LOD score exceeding the critical value. Once assignments were made, the observed self-recruitment (SR_O) was calculated as $SR_O = n_C / n_T$, where n_C is the number of settlers assigned to Curlew parents and n_T is the total number of settlers collected from Curlew.

3.3.5 Approximating a dispersal kernel

To generate a first approximation of a dispersal kernel for *E. lori*, we created a map of Curlew and all nearby reef patches, so that larval exchange between the focal population and these nearby populations could be simulated. A map of the MBR was obtained from the IMaRS Millennium Coral Reef Mapping Project (Andréfouët *et al.* 2005) and subsequent GIS analyses were performed in ArcMap 10 (ESRI). We focused on the Belizean portion of the MBR, and only used areas of the reef designated as 'forereef' by the IMaRS. We then divided the reef into 500 m by 500 m square grids to break the reef into patches approximating the size of our study area. Some areas of the 'forereef' on the IMaRS maps are wider than 500 m; however, we know that *E. lori* are only abundant on the outer reef slope, which is never greater than 500 m wide (D'Aloia *et al.* 2011). Therefore, to ensure that this gridded reef represented the true width of the MBR where *E. lori* are present (i.e. less than 500 m wide), we manually removed excess reef patches whenever the gridded forereef was greater than 500 m wide. Next, a distance matrix was generated that included the distance between Curlew and every other reef patch, restricted to within 35 km (inclusion of longer distances will not significantly alter the outcome of our analyses because the expected influx of larvae from such distant sites has a negligible effect on self-recruitment estimates).

Given the distribution of all reef patches, we tested which simple exponential decay function could capture the observed self-recruitment at Curlew. Specifically, we estimated

the relative probability of dispersal (y) from any patch (j) to the focal Curlew patch as:

$$y_j = y_0 b^{x_j},$$

where $y_0 = 1$, x_j = distance to Curlew, and b = the decay rate, ranging from 0.00 to 1.00 by intervals of 0.005. In this way, we iteratively simulated dispersal to Curlew from all reef patches for each potential value of the decay rate. This approach assumes that all reef patches are occupied, population density and reproductive output are equal among patches, and larval dispersal is represented by the same function from each patch. *E. lori* is well-suited to such a simple approach, as personal surveys throughout the MBR revealed that *E. lori* is densely distributed all along the barrier reef and atolls. Further, previous studies on simulated dispersal data suggest that spatial variation in reproductive output does not significantly alter dispersal kernel estimates, provided that the variation is randomly distributed (Robledo-Arnuncio & García 2007; Robledo-Arnuncio 2008).

Once the relative probability of dispersal from all patches to Curlew was calculated, the predicted self-recruitment (SR_P) from each exponential decay function was estimated as:

$$SR_P = \frac{y_0}{\sum_{j=0}^{n} y_j} ,$$

where $y_0 = 1$, and n = 224 reef patches. The dispersal function (y) that accurately predicted the observed self-recruitment was accepted as the putative dispersal kernel.

3.3.6 Export from focal site

To test for export from the focal site, we used the same genotyping and assignment methods described above to assign settlers from 6 additional sites on the MBR to the resident population at Curlew (Fig. 3.2b). For each export site, we calculated the proportion of settlers originating from Curlew (# settlers assigned to Curlew parents / total # settlers collected). Finally, we added an exponential best fit line through the observed self-recruitment and export proportions, using nonlinear least-squares regression, to determine whether a function similar to the approximated dispersal kernel captured the observed data.

3.4 Results

3.4.1 Summary statistics and tests of HWE

The 14 microsatellite markers were highly polymorphic: there was an average of 27 alleles per locus, ranging from 10 to 76 (Table 3.1). Exact tests for deviations from HWE revealed that 4 loci were potentially out of HWE, but MICRO-CHECKER analyses suggested that all 4 deviations were due to the presence of null alleles. This set of highly polymorphic markers provided a combined non-exclusion probability of P = 1.8 x 10^{-7} for the first parent, and P = 1.17 x 10^{-17} for a parent pair. Because the probability of incorrectly assigning an offspring to a parent is so low, the potential deviations from HWE due to null alleles should not significantly bias parentage assignments.

3.4.2 Self-recruitment

The local parentage analysis assigned nine settlers collected at Curlew to parents, revealing a self-recruitment percentage of 4.6% (9 assigned settlers / 194 total settlers) (95% CI: 1.7%, 7.6%). All assigned parent-offspring pairs mismatched at 0 or 1 loci and had LOD scores exceeding the critical value for 95% confidence ($LOD_{critical} = 4.00$). Further, the simulation revealed that offspring were assigned to true parents 96% of the time, and that there was a 0% chance of false exclusion (Type II error) when the true parent was sampled, and only a 4% chance of false assignment (Type I error) when the true parent was unsampled. As mentioned in the Methods section, however, these error estimates are dependent on parameter input values.

A local connectivity network of these nine self-recruiting trajectories reveals the pattern of dispersal in terms of distance, direction and depth (Table 3.2; Fig. 3.3). While there are too few data points to statistically test for the association between these variables and the probability of dispersal, the network shows that seven out of the nine trajectories have a northward directional change. Interestingly, one parent produced two of the self-recruiting offspring. There were also two self-recruiting offspring that settled on the same sponge.

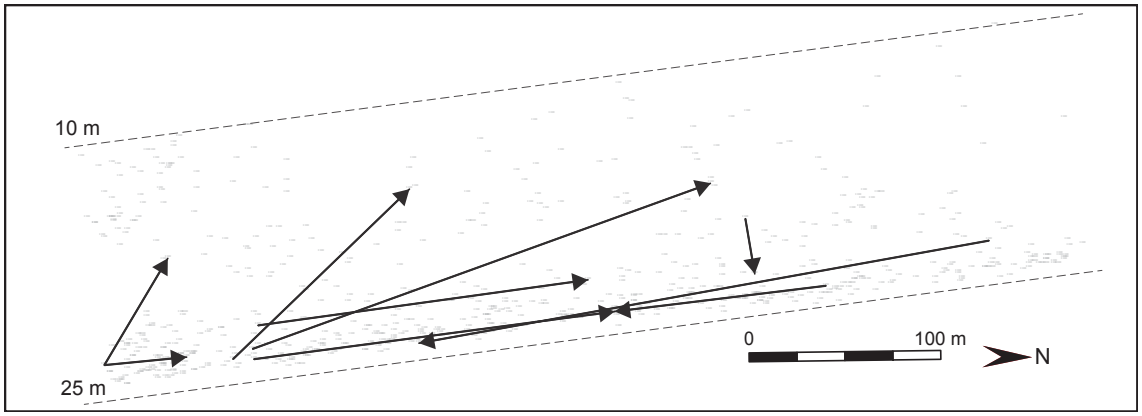

Figure 3.3: Local connectivity network of *Elacatinus lori* at Curlew Caye, Belize. Grey dots represent all sponges (*Aplysina fistularis*) in the study area and dashed lines represent the 10 m and 25 m depth contours. The net distance and direction change of nine self-recruiting dispersal trajectories are shown by the black arrows.

These data provide a preliminary description of the pattern of local dispersal and provide a foundation for future tests of the predictors of local dispersal.

3.4.3 Predicted dispersal kernel

Of the simple exponential functions tested, $y = 0.915^x$ was the best predictor of observed self-recruitment (Fig. 3.4). This dispersal kernel explains the observed self-recruitment percentage of 4.6% at Curlew, accounting for the influx of offspring from nearby reef patches. The 95% CI around this approximation of the dispersal kernel was generated by determining which exponential functions explained the 95% CI around the observed self-recruitment percentage (Fig. 4; Lower 95% CI: SR = 1.7%; $y = 0.965^x$; Upper 95% CI: SR = 7.6%; $y = 0.865^x$).

3.4.4 Export analyses

Additional parentage analyses revealed an overall export percentage of 1.1% to 6 nearby sites within the metapopulation (2 settlers assigned to Curlew parents / 184 total settlers collected; 95% CI: 0%, 2.6%). Of the two settlers assigned to Curlew parents, one was collected from the North (site N2) and one was collected from the South (site S1) (Fig.

Table 3.2: Description of nine self-recruiting dispersal trajectories. Characteristics of origin (parent size, origin sponge size) and destination (destination sponge size) are shown, as well as characteristics of the trajectory in three dimensions (distance, depth, direction). Note that a positive depth change indicates that the larva settled at a shallower destination relative to its origin, while a negative value indicates that it settled deeper.

	Offspring Size (mm)	**Parent(s) size (mm)**	**Origin Sponge Length (cm)**	**Destination Sponge Length (cm)**	**Distance (m)**	**Depth change (m)**	**N-S change**	**E-W change**
1	9	40	34	56	256.5	-2.5	N	W
2	9	48	46	47	19.5	1.8	N	E
3	10	51	46	18	174.3	-2.6	N	W
4	10	52	38	26	112.2	3.3	S	E
5	10	54; 42	42	22	127.2	9.1	N	W
6	10	29	17	46	303.1	-7.5	S	E
7	12	36	25	46	189.7	7.6	N	W
8	14	35	37	16	44.7	-1.1	N	W
9	17	35	37	40	63.8	6.8	N	W
$\overline{x}$	11.2	42.2	35.2	38.3	—	—	—	—

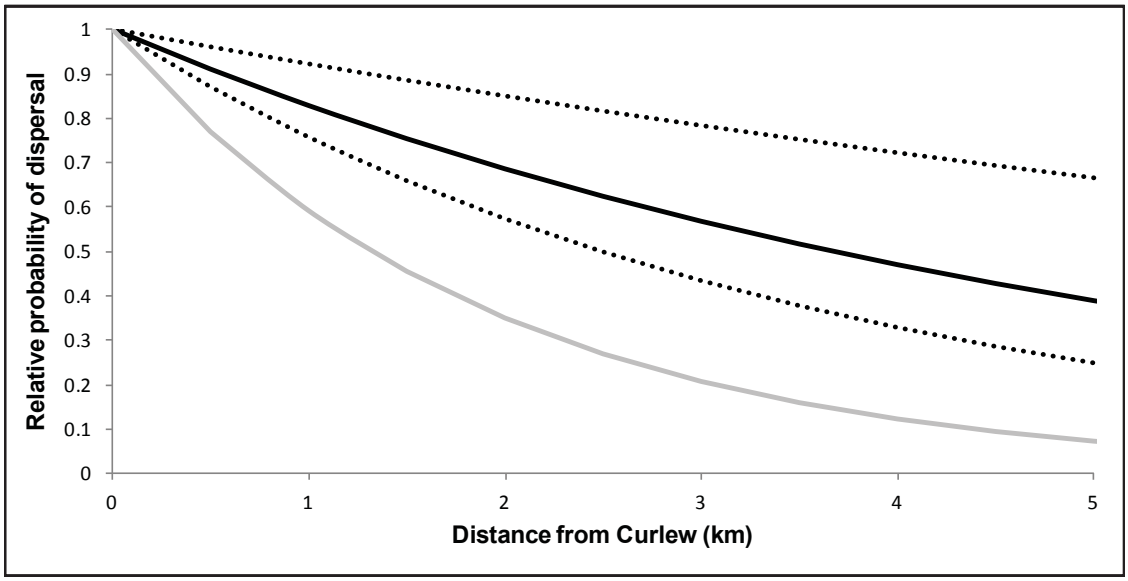

Figure 3.4: Simple exponential dispersal kernel for *E. lori*. The solid black line represents the function ($y = 0.915^x$) that explains the observed level of self-recruitment at Curlew Caye (SR=4.6%), given the seascape structure. The dashed black lines represent the dispersal kernels that predict the 95% CI around the observed self-recruitment proportion (Lower CI: SR = 1.7%; $y = 0.965^x$; Upper CI: SR = 7.6%; $y = 0.865^x$). The solid grey line represents the exponential best-fit line through the observed self-recruitment and export data ($y = 0.77^x$).

3.2b). The exponential best-fit curve through the observed self-recruitment and export proportions was $y = 0.77^x$ (Fig. 3.4; NLS regression: $t = 679$; $df = 5$; $p = 0.001$). While the low sample size limits our ability to estimate a kernel directly from the observed data points, this best-fit line also reveals a rapid decline in the probability of dispersal, suggesting that the dispersal kernel approximated in this study may coarsely capture dispersal patterns within the metapopulation.

3.5 Discussion

Quantifying patterns of larval dispersal remains a major goal of marine ecology because it is essential for a complete understanding of marine population dynamics, and because of the potential benefits of incorporating connectivity data into marine reserve design. To date, the logistical difficulties of collecting marine dispersal data have limited most empirical

studies to point estimates of self-recruitment or export. In order to advance our understanding of marine population connectivity, the focus must progress from making point estimates to fitting dispersal kernels to empirical data and testing the predictions derived from such kernels. In this study, we take a first step towards approximating a dispersal kernel for the coral reef fish *E. lori* along an 8 kilometer stretch of reef on the barrier reef in Belize. Through genetic parentage analyses, we uncovered a 4.6% self-recruitment percentage. Working under the hypothesis that seascape patchiness is a primary determinant of observed measures of self-recruitment, we used habitat maps to estimate a simple exponential function, $y = 0.915^x$, that would explain the observed self-recruitment and took that function to be a first approximation of the dispersal kernel. An analysis of export from the focal site suggests that an even more rapidly declining kernel may explain observed levels of dispersal within the metapopulation, though this analysis lacks statistical power due to the limited data. Taken together, the results of this study indicate a rapid decline in the probability of dispersal with respect to distance for *E. lori*.

These results add to the growing body of research that suggests the scale of marine larval dispersal is much shorter than previously assumed, as evidenced by other direct genetic estimates of dispersal (Jones *et al.* 2005; Planes *et al.* 2009; Buston *et al.* 2012), patterns of genetic differentiation (Taylor & Hellberg 2003, 2006), and bio-physical oceanographic models (Cowen *et al.* 2006). Our data are consistent with the only other empirically-derived marine dispersal kernel, which was constructed for another microhabitat specialistthe anemonefish *A. percula*. As in *E. lori*, the probability of dispersal with respect to distance declined rapidly—fivefold over the first kilometer from source (Buston *et al.* 2012). Moreover, indirect dispersal data specific to the genus *Elacatinus spp.*, suggests that neon gobies may have particularly short dispersal distances, as evidenced by their limited distributions (Colin 1975, 2010), and remarkably high levels of genetic differentiation between populations separated by as little as 23 km (Taylor & Hellberg 2003, 2006). With few obvious geographic barriers, short larval dispersal distances were initially proposed as the mechanism behind the unusually high genetic differentiation in *Elacatinus spp.* (Taylor &

Hellberg 2003, 2006), but direct dispersal data have been lacking until now. A rapidly declining dispersal kernel such as the one proposed in this study could help explain observed genetic structure patterns found in this genus.

We found that the rapidly declining exponential function $y = 0.915^x$ was the best approximation of an *E. lori* dispersal kernel. This approximation was based on the observed self-recruitment percentage, and accounted for the influx of larvae from all nearby reef patches on the MBR. We do not, however, claim that this simple function represents the true, underlying dispersal kernel. Instead, we believe that it serves two valuable purposes. First, it provides an initial approximation of the *E. lori* dispersal kernel. This approximation will guide the sampling design of future large-scale studies that will rigorously fit dispersal distributions to larger empirical data sets. Second, it lays out a quantitative framework for incorporating measurable empirical data and the seascape into kernel approximations. Shifting dispersal research towards more quantitative analyses will enable researchers to make predictions about dispersal. A predictive approach is essential to making progress in the field because it is the key to forecasting dispersal patterns under different scenarios, and generalizing across systems (Nathan & Muller-Landau 2000; Botsford *et al.* 2009; Jones *et al.* 2009).

Eventually, a rigorous approximation of a marine dispersal kernel must be based on empirical dispersal events at many distance classes and should evaluate alternative functional forms of the kernel. A framework for these analyses can already be found in the terrestrial seed dispersal literature (Nathan & Muller-Landau 2000; Austerlitz *et al.* 2004; Greene *et al.* 2004). We chose to solely evaluate the negative exponential distribution in this study because it is a potential dispersal kernel form that is a 1-parameter probability distribution. Other distributions that are frequently used in seed dispersal studies, including the lognormal (Greene & Johnson 1989), Weibull (Ribbens *et al.* 1994), and 2Dt (Clark *et al.* 1999), require multiple parameter estimates. As marine ecologists continue to make progress in collecting empirical dispersal data, we will eventually be able to use a maximum likelihood framework to evaluate alternative functional forms of the dispersal kernel. This study sets

a foundation for this future quantitative work by testing whether the simplest functional form of the potential dispersal kernels can explain an observed empirical pattern.

The simple exponential kernel proposed in this study, and the probability distributions described above, represent phenomenological models of dispersal, which differ importantly from mechanistic models of dispersal (Nathan & Muller-Landau 2000). Phenomenological models characterize the function that best describes the observed relationship between dispersal and distance. These models are an important first step in quantifying dispersal patterns, especially because the spatial pattern of dispersal is what ultimately influences ecological and evolutionary processes such as recruitment dynamics, competition, mating, gene flow and local adaptation (Nathan & Muller-Landau 2000; Clobert *et al.* 2001; Hanksi 1999; Hanski & Gaggiotti 2004). To begin generalizing across species and seascapes, the next step will be to assess how potential predictors shape these observed patterns (Nathan & Muller-Landau 2000). Because marine larval dispersal is a bio-physical process, realized dispersal patterns are certainly influenced by oceanographic processes and species-specific larval behaviors, in addition to seascape structure (Cowen & Sponaugle 2009; Jones *et al.* 2009).

In conjunction with other recent work (Buston *et al.* 2012) this study begins to build a framework for approximating marine dispersal kernels from empirical data. We show that simple kernels can be approximated from point estimates of self-recruitment and the structure of the seascape. Such simple kernels may serve an important role in conservation planning if a coarse approximation of larval exchange within an area is the goal, or if there are limited resources to quantify dispersal for many target species. Simple rules of thumb could also help identify key areas for stepping stone reserves that are well connected in a reef network. Quantitative estimates of dispersal kernels will also be useful in cases when the aim is to explicitly maximize connectivity in a reserve network because they can be used to assign connectivity values between potential reserve sites. Ultimately, understanding the patterns and predictors of dispersal across species and seascapes will enable us to build better predictive models of metapopulation dynamics and will strengthen our abil-

ity to predict the impacts of marine reserve design on connectivity and metapopulation persistence.

3.6 Acknowledgments

We thank the staff at Wee Wee Cay Marine Lab for use of their facilities, Marian Wong, Udell Foreman and Alissa Rickborn for assistance in the field, and Phil Lobel for logistical diving support. We are also grateful to Sean Mullen's lab for assistance with lab work, and two anonymous reviewers for helpful comments. Funding for this project was provided by a Boston University start-up award to PMB. CCD was supported by a NSF Graduate Research Fellowship and a Boston University Deans Fellowship. JEM was supported by a Warren McLeod summer research fellowship. All work was approved by Belize Fisheries and the Boston University IACUC.

3.7 Data Accessibility

Microsatellite data, dispersal simulations, sampling GIS waypoints, raw field data, and all relevant GIS shape files are available via Dryad (doi:10.5061/dryad.k8s1b).

3.8 Publication

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

3.9 Author Contributions

CCD and PMB planned the study. CCD, JEM, and PMB conducted the field research. CCD, SMB, and RGH contributed to genetic work. SMB developed the microsatellite markers. CCD analyzed the data and wrote the paper. All authors contributed to revisions.

Chapter 4

Patterns, causes, and consequences of marine larval dispersal

4.1 Abstract

Quantifying the probability of larval exchange among marine populations is key to predicting local population dynamics and optimizing networks of marine protected areas. The pattern of connectivity among populations can be described by the measurement of a dispersal kernel. However, a statistically-robust, empirical dispersal kernel has been lacking for any marine species. Here, we use genetic parentage analysis to quantify a dispersal kernel for the reef fish *Elacatinus lori*, demonstrating that dispersal declines exponentially with distance. The spatial scale of dispersal is an order of magnitude less than previous estimates—the median dispersal distance is just 1.7 km and no dispersal events exceed 16.2 km despite intensive sampling out to 30 km from source. Overlaid on this strong pattern is subtle spatial variation, but neither pelagic larval duration nor direction is associated with the probability of successful dispersal. Given the strong relationship between distance and dispersal, we show that distance-driven logistic models have strong power to predict dispersal probabilities. Moreover, connectivity matrices generated from these models are congruent with empirical estimates of spatial genetic structure, suggesting that the pattern of dispersal we uncovered reflects long-term patterns of gene flow. These results challenge assumptions regarding the spatial scale and presumed predictors of marine population connectivity. We conclude that if marine reserve networks aim to connect whole communities

of fishes, protected areas should be close in space (<10 km) to accommodate short distance dispersers.

4.2 Introduction

Quantifying patterns of marine larval dispersal is a major goal of ecology and conservation biology (Botsford *et al.* 2001; Sale *et al.* 2005; Cowen *et al.* 2007). Many marine species have a bipartite life cycle that is characterized by a dispersive larval phase and a relatively sedentary adult phase. Thus, larval dispersal drives the exchange of individuals and genes (i.e., connectivity) among populations within many marine metapopulations (Kritzer & Sale 2006). In turn, estimates of connectivity are critical to predicting population dynamics, understanding microevolutionary processes, and designing effective networks of marine reserves.

Ecologists have long-recognized that dispersal kernels offer a useful approach to quantifying patterns of dispersal (Clark *et al.* 1999; Nathan & Muller-Landau 2000). Here, an empirical dispersal kernel is defined as a probability density function that can be integrated to yield the probability of successful dispersal over a given distance. Estimating a dispersal kernel requires that sampling be spatially extensive to capture long-distance dispersal (LDD) events–the tail of the kernel. Capturing the tail is essential to understanding ecological and evolutionary processes that are driven by LDD (Nathan 2006). Sampling must also be intensive to tighten the confidence intervals associated with low-frequency LDD events. Despite a decades-long research effort (Jones *et al.* 1999, 2005; Planes *et al.* 2009; Buston *et al.* 2012; Almany *et al.* 2013), the description of a complete dispersal kernel, including the tail, remains a major challenge in the field of marine ecology (Botsford *et al.* 2009; Jones *et al.* 2009).

Given the paucity of empirical dispersal kernels, the causes of variation in patterns of dispersal also remain poorly studied. Dispersal distance data, combined with measurements of relevant biological and spatial variables, will enable the test of key hypotheses

related to the predictors of marine dispersal, analogous to work on terrestrial seed dispersal (Jordano *et al.* 2007; Nathan *et al.* 2008). For example, oceanographers have hypothesized that anisotropy in ocean currents will lead to asymmetry in dispersal and population connectivity (Cowen *et al.* 2006; Huret *et al.* 2010). Ecologists have also hypothesized that dispersal costs and/or post-settlement selection may lead to higher mortality rates for long-distance dispersers (Burgess *et al.* 2012). Finally, there is a long-standing hypothesis that the number of days an individual spends in the larval phase affects its dispersal capacity (Siegel *et al.* 2003; Miterai *et al.* 2008; Selkoe & Toonen 2011). To date, there are few rigorous empirical tests of these hypotheses, because dispersal data have been lacking.

In addition to testing these key predictors, there is growing interest in the demographic and genetic consequences of dispersal, give the precipitous decline in the health of coral reef ecosystems globally (Hoegh-Guldberg *et al.* 2007; IPCC 2014). Because dispersal has profound impacts on both ecological (i.e., population dynamics) and evolutionary (i.e., local adaptation) processes within metapopulations, dispersal kernels can potentially play a key role in developing effective management strategies. For example, recent research has advocated the use of simple spacing guidelines in marine reserve design (Moffitt *et al.* 2011, Anadón *et al.* 2013), and empirical dispersal data for tractable taxa can be used to test the accuracy of these guidelines. However, thus far, the lack of data has limited the incorporation of dispersal patterns into reserve design.

To tackle these remaining challenges, we used the neon goby *Elacatinus lori* as a tractable study organism. *E. lori* is endemic to the Mesoamerican barrier reef and representative of the most speciose marine fish family (Gobiidae) (Colin 2002). Adults live and breed within sponges (D'Aloia *et al.* 2011). The male tends to demersal eggs for ∼7 days until they hatch. Individuals then spend ∼26 days in a larval phase–a duration that approximates the median value in reef fishes (Brothers & Thresher 1985). Upon completing the larval phase, individuals settle onto sponges. The close association with sponges makes the fish easy to locate and capture. Capitalizing on these characteristics, we addressed three objectives: (i) quantify the pattern of dispersal with a complete dispersal kernel; (ii)

identify the predictors of dispersal; and (iii) explore the evolutionary consequences and conservation implications of dispersal.

4.3 Results and Discussion

4.3.1 Pattern of Dispersal

To quantify a complete dispersal kernel, we conducted a massive field study in Belize during 2013. We sampled intensively along a transect that was designed to capture dispersal trajectories up to 30 km from source (Fig. 4.1). We collected 3,033 tissue samples from potential parents and 4,112 samples from potential offspring (Appendix A). We then genotyped individuals at 20 microsatellite loci (Appendix A). Using genetic likelihood-based parentage analysis, we assigned 120 offspring to parents and calculated the net distance between all parent-offspring matches. We estimated the dispersal kernel by fitting alternative probability density functions to the observed distribution of dispersal distances (Appendix A), controlling for variation in sampling effort. The best-fit functional form of the dispersal kernel was exponential,

$$f(x) = \lambda e^{-\lambda x},$$

where f(x) = probability density, x = distance (km), and the decay parameter $\lambda = 0.36$ (95% CI=0.30, 0.43) (Fig. 4.2A). The decay parameter can be interpreted as one over the mean dispersal distance (mean=2.8 km). The modal dispersal distance class was 0-1 km and the median distance was just 1.7 km. We observed no dispersal events beyond 16.2 km, despite intensive sampling up to 30 km from source, indicating that we sampled deeply into the tail of the kernel. Thus, despite having an average 26 day larval phase, and therefore the potential to disperse far via ocean currents (Siegel *et al.* 2003), *E. lori* exhibits a spatially-restricted leptokurtic pattern of dispersal.

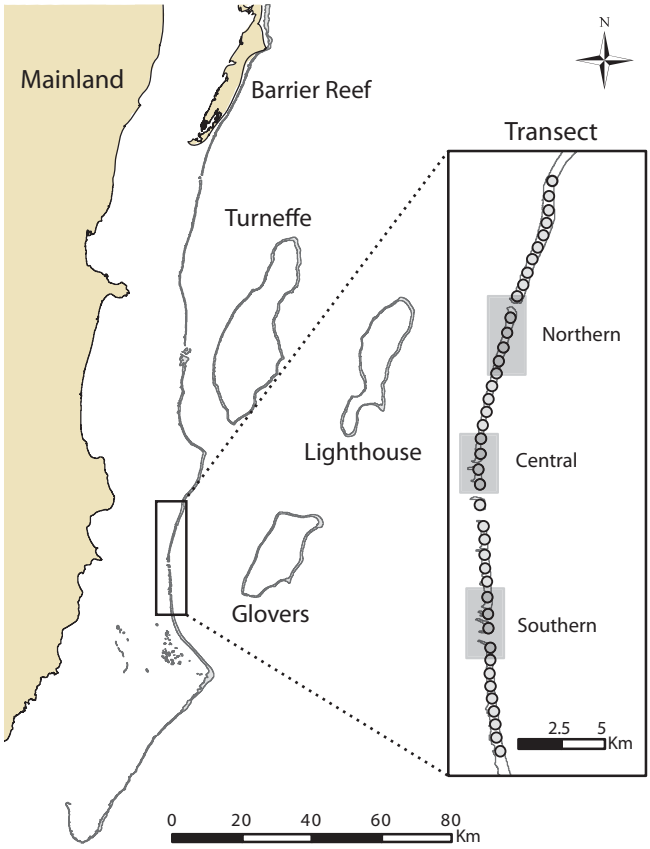

Figure 4.1: Transect on the Belize barrier reef. The inset map shows approximate sampling locations along the transect, which extended for 41 km, centered on Carrie Bow Cay (16°48'10"N 88°05'45"W). Parent tissue samples were collected from three regions, represented by the shaded boxes (n ≈ 1,000 per region). Offspring tissue samples were collected every km, represented by the grey circles (n ≈ 100 per site).

4.3.2 Causes of Variation in Dispersal

To begin to disentangle the predictors of dispersal, we first explored whether the shape of the dispersal kernel varied with spatial and biological variables (Jordano *et al.* 2007). We found subtle evidence of spatial variation in the magnitude of the decay rate (Fig. 4.2B). However, there was no evidence for any effect of direction, settler standard length (a proxy for age), or pelagic larval duration (PLD) on the kernel shape (Fig. 4). Likewise, no variables related to the microhabitat at origin influenced the kernel shape (Appendix A). This subdivision of the dispersal kernel was consistent with multivariate analyses that

showed these same variables were not significantly associated with the dispersal distances of individuals (Appendix A). Taken together, these results demonstrate that distance is the primary explanatory variable of the dispersal pattern.

These results lend new insight into major hypotheses related to causes of variation in dispersal. For example, we found that direction (North/South) does not influence the shape of the kernel (Fig. 4.2C), despite reported directionality in ocean currents around the Belize barrier reef (Ezer *et al.* 2005). This suggests that dispersal is either isotropic despite anisotropic currents, or that currents were isotropic over the timeframe of our study. We also found no evidence to support the hypothesis that long-distance dispersers experience elevated post-settlement mortality rates. If long-distance migrants suffer higher post-settlement mortality rates, the dispersal kernel for larger, older settlers should have a significantly larger decay rate (λ) than the kernel for smaller, younger settlers. Instead, we found that time since settlement (measured indirectly by settler standard length) does not influence the shape of the kernel (Fig. 4.2D). Similarly, we show that PLD (i.e., the number of days an individual spends in the larval phase) does not influence the shape of the kernel (Fig. 4.2E).

Given the widespread use of PLD as a proxy for dispersal distance, we conducted additional analyses to test the robustness of our findings. While we observed substantial intraspecific variation in PLD (mean = 26 days $\pm$ 3.6 s.d), we found that individuals' PLDs are not correlated with the net distance traveled (Fig. 4.3). These results call into question the pervasive use of PLD as a proxy for dispersal potential (e.g. Siegel *et al.* 2003; Miterai *et al.* 2008).

Next, to move beyond description towards a predictive dispersal model, we adopted a logistic regression approach and explored the predictors of the probability of successful dispersal (Buston *et al.* 2012). We identified potential dispersal trajectories by generating a distance matrix between every parent collection location (n=64) and every offspring collection location (n=69), and identified actual dispersal trajectories from the parentage analysis. We then built a set of logistic models to identify predictors of a dependent

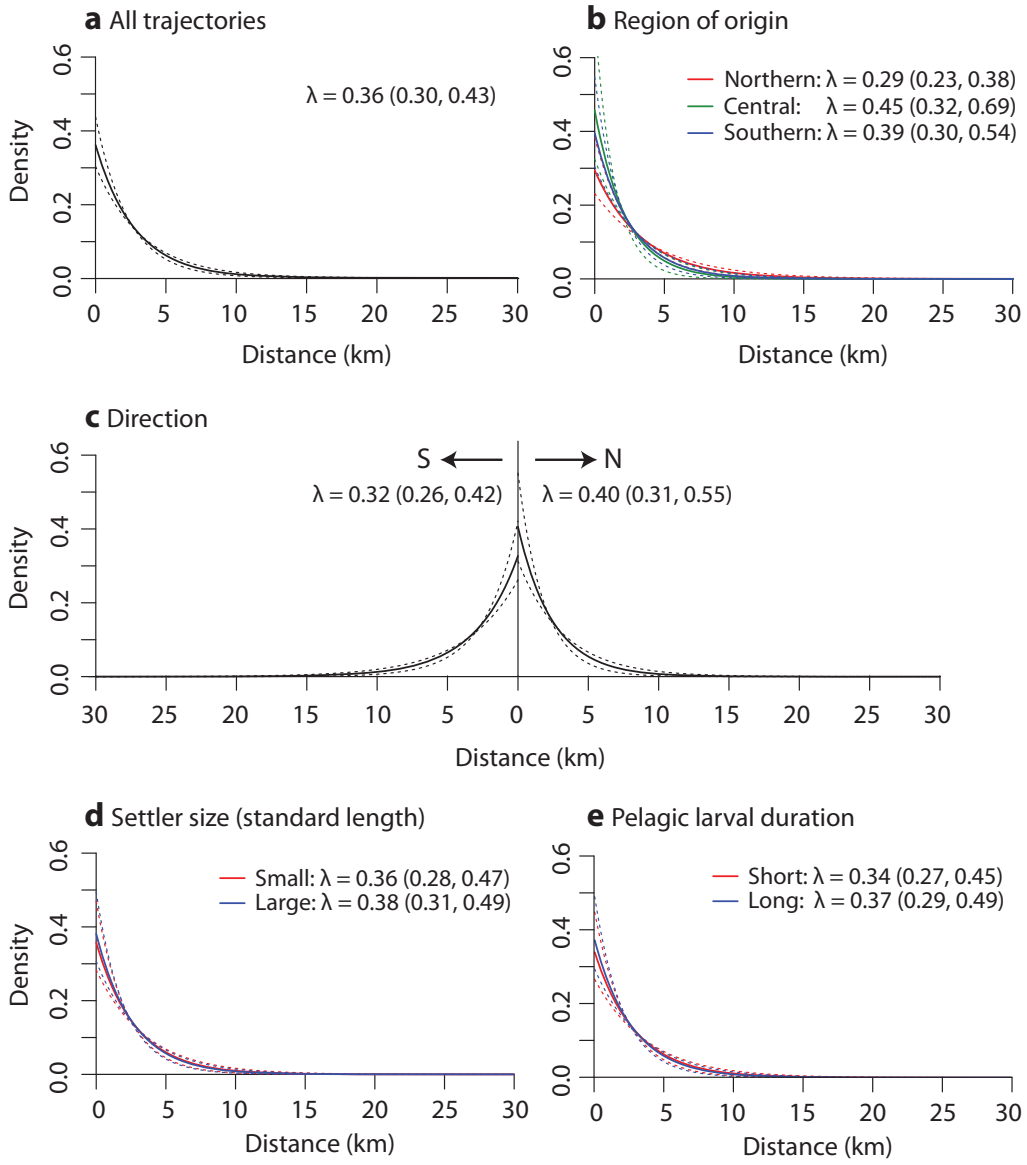

Figure 4.2: Empirical dispersal kernels for *Elacatinus lori*. Solid lines represent best estimates for the exponential dispersal kernel, $f(x) = \lambda e^{-\lambda x}$, and dashed lines represent 95% bootstrapped confidence intervals. Values for λ with 95% CI are provided. (A) Kernel for all dispersal trajectories; (B-E) Kernels fit to subdivided data reveal that the shape exhibits (B) subtle spatial variation, but consistency with respect to (C) direction, (D) settler size and (E) pelagic larval duration based on overlapping curves and confidence intervals.

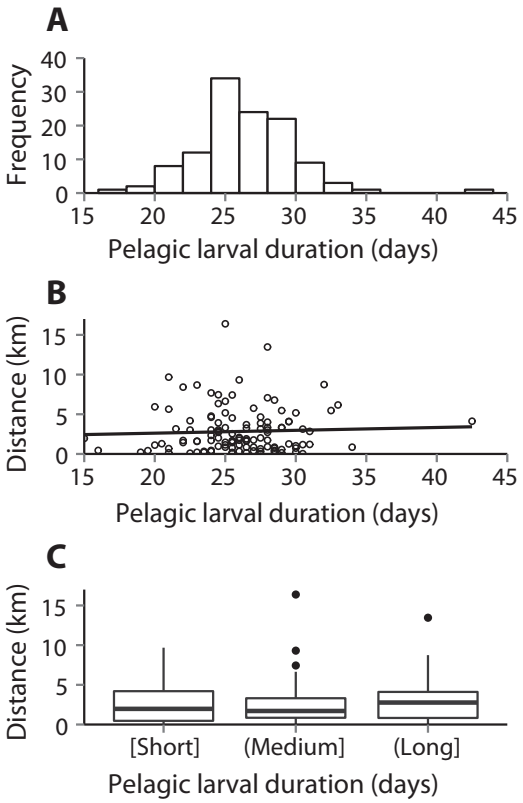

Figure 4.3: No relationship between PLD and dispersal distance. (A) Observed PLD for offspring assigned to parents (n=118). (B) No correlation between PLD and dispersal distance (Spearman's correlation: $\rho = 0.03$; df = 116; p = 0.74). (C) No relationship between PLD and dispersal distance when PLD is binned (Kruskal-Wallis test: $\chi^2 = 0.49$; df = 2; p = 0.78). These results are robust to the inclusion/exclusion of outliers and alternative binning strategies.

variable that noted whether a potential dispersal trajectory was actually used (1 or 0). Model selection revealed that distance, parent region, and sampling effort significantly predict the probability of successful dispersal (Appendix A).

The logistic model had strong predictive power. Cross validation—10-fold and leave-one-out—revealed a low average prediction error (0.025). The area under the receiver operating characteristic curve (AUC) was 0.89 (95% CI=0.86-0.91), indicating excellent predictive accuracy (Hosmer *et al.* 2013). Notably, when the regional predictor was removed from the model, there was no substantive reduction in predictive accuracy (Appendix A), revealing that distance is the primary predictor of dispersal between sites.

4.3.3 Consequences of Dispersal

Finally, we explored the consequences of dispersal by using the logistic model to generate a population connectivity matrix for the Belizean seascape. High probabilities of larval exchange are predicted to occur between nearby sites, but a stark lack of connectivity is predicted between most sites (Fig. 4.4A). The continuity of reef habitat provides a structural basis for multi-generational stepping-stone dispersal along the barrier reef and within each atoll. While some connectivity is predicted between the barrier reef and one proximate atoll (Turneffe), no connectivity is predicted between the two atolls that lie further offshore (Lighthouse and Glovers) and other regions.

The predicted levels of demographic connectivity suggest that there will be low levels of genetic connectivity between the two distant atolls (Lighthouse and Glovers) and other sites, and that these atolls will emerge as evolutionarily significant units (ESUs) (Fig. 4.4B). This prediction is qualitatively consistent with previous estimates of spatial genetic structure for *E. lori* in Belize (D'Aloia *et al.* 2014). This consistency suggests that the logistic model predicts both short-term patterns of dispersal and long-term patterns of gene flow. Moreover, it strongly supports the conclusion that we captured the tail of the dispersal kernel–if there were a non-trivial number of long-distance dispersal events beyond 16 km each generation, then there would be no genetic structure between the barrier reef and the two distant atolls at neutral genetic markers. Instead, the genetic structure data suggest that the observed dispersal kernel is real and temporally stable.

To explore connectivity within the existing network of marine protected areas (MPAs) in Belize, we highlighted nodes in the matrix that fell within the boundaries of Belize's current MPA network. Despite an extensive network, there are major gaps in connectivity (Fig. 4.4C). While individual MPAs are predicted to self-replenish due to short-distance dispersal, most reserves will not be connected to others. Only 9 out of 136 pairs of MPAs are predicted to have any level of connectivity. Simply stated, the current network of marine reserves is not predicted to be demographically connected for *E. lori* or other species with

similar dispersal patterns.

4.4 Conclusions

Here, we address a critical knowledge gap in the field of marine ecology—the measurement of a statistically-robust empirical larval dispersal kernel. Our findings demonstrate that the probability of successful dispersal declines exponentially and predictably as a function of distance from source. We conclude that some marine fish populations may be strongly demographically connected on the scale of only 0-10 km—an order of magnitude less than current estimates (Cowen *et al.* 2006). Thus, we argue that if networks of marine reserves aim to connect whole communities of reef fishes, individual reserves will need to be closer in space ($<$ 10 km apart) to accommodate short distance dispersers.

4.5 Methods

4.5.1 Transect

To investigate potential dispersal trajectories up to 30 km from source, tissue samples for genetic parentage analysis were collected along a 41 km stretch of the Belize barrier reef (BBR), centered on Carrie Bow Caye, Belize. The sampling design included three regions for adult tissue collection, spaced roughly 10 km apart, and forty-one sites for settler collection, spaced roughly 1 km apart (Fig. 4.1). All sampling was conducted using SCUBA at an average depth of 16.03 meters $\pm$ 2.19 s.d.. Sampling was focused at those depths where *E. lori* density is the highest (D'Aloia *et al.* 2011). A waypoint was recorded at the beginning and end of every collection dive, with the midpoint of each dive taken as the location for all individuals sampled on that dive.

4.5.2 Tissue Collection

Adults were found on the inside of yellow tube sponges, captured using slurp guns, and restrained in aquarium nets while a non-lethal fin clip was obtained from the caudal fin.

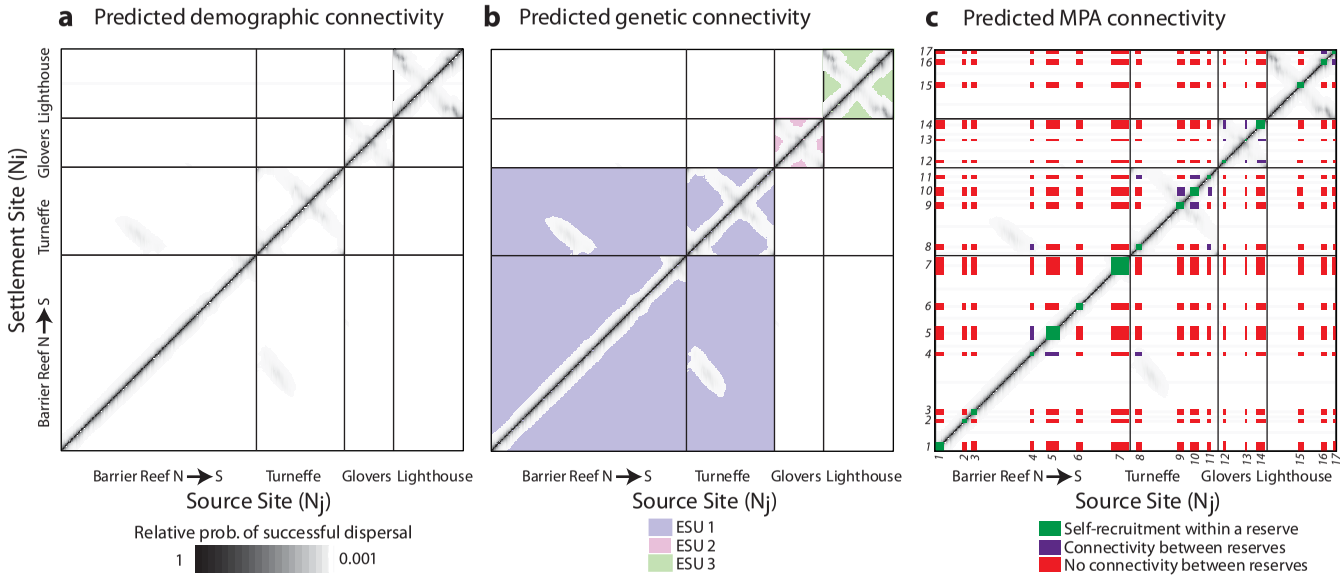

Figure 4.4: Population connectivity matrices on the Belize barrier reef. Matrices show predicted connectivity between all potential source (N_j) and settlement sites (N_i), where j=i=516 1-km^2 sites. The relative probability of successful dispersal between any two sites was visualized by grayscale intensity (relative probabilities < 0.001 were excluded). (A) Predicted demographic connectivity based on the logistic model. (B) Predicted genetic connectivity based on the logistic model and the assumption that continuous reef habitat provides a structural basis for multi-generational stepping stone dispersal along the barrier reef and within each atoll. (C) Predicted connectivity between MPAs in Belize, with individual MPAs numbered (Appendix A).

Fin clips were transferred to 95% EtOH at the surface. Approximately 1,000 adult fin clips were collected in each region (Appendix A). Settlers were found on the outside of yellow tube sponges and captured using slurp guns. Because settlers were too small for non-lethal fin-clipping, individuals were collected in plastic bags underwater and euthanized with MS-222 at the surface. At every site (n = 41 sites), approximately 100 settlers were obtained (Appendix A).

4.5.3 Genotyping

Genomic DNA was extracted with a rapid HotSHOT protocol following the methods described by Truett *et al.* (2000). Briefly, tissue samples were lysed through an incubation in hot sodium hydroxide and treated with a neutralizing agent (Tris-HCl). Individuals were then genotyped following a two-step protocol for parentage analysis. First, all individuals were genotyped at 14 microsatellite loci using the Type-It Microsatellite PCR Kit (Qiagen, USA), following the methods of D'Aloia *et al.* (2013). Second, following an initial parentage analysis that identified potential parent-offspring matches (see below), all individuals from these putative matches (n = 576) were genotyped at 6 additional loci in order to reduce the possibility of false positive assignments (Appendix A). We calculated summary statistics (number of alleles, observed heterozygosity, expected heterozygosity) in Cervus v.3.0 (Kalinowski *et al.* 2007), identified deviations from Hardy-Weinberg Equilibrium in Genepop v.4.2 (Raymond & Rousset 1995) and tested for the presence of null alleles in Microchecker v.2.2.3 (van Oosterhout *et al.* 2004) (Appendix A).

4.5.4 Parentage Analysis

To assign offspring from the entirety of the 41-km transect to potential parents from the Northern, Central, and Southern regions, a categorical likelihood-based parentage analysis was conducted in CERVUS. An initial parentage analysis was run using all potential offspring (n= 4,110) and parents (n=3,031) that amplified successfully. To determine the critical LOD value at 95% confidence, a simulation was run for 100,000 offspring and 100,000

candidate parents based on the observed allele frequencies. For the parentage analysis, conservative input values were used for the proportion of candidate parents sampled (0.03) and the mistyping proportion (0.01); however, previous results have shown that results are robust to a range of input values (D'Aloia *et al.* 2013). Parent-offspring assignments with an LOD score exceeding the critical value at 95% confidence were selected for additional genotyping and run through a second parentage analysis based on the expanded genotypes at 20 loci. Final parent-offspring assignments were selected if their LOD score exceeded the critical value and met additional filtering criteria (Appendix A).

4.5.5 Predictors of Dispersal

To calculate the distance between all potential dispersal trajectories, we generated a distance matrix between each parent collection location (n = 64) and each offspring collection location (n = 69). The location for each individual was taken as the midpoint of the SCUBA dive for every collection. The Euclidian distance between all pairwise possibilities was generated using ArcMap 10.1 (ESRI), for a total of n=4,416 potential trajectories. Next, to determine whether any microhabitat characteristics at origin were associated with the net distance travelled by a larva, depth, number of tubes, and maximum tube length were measured at every sponge where an adult was fin-clipped.

To test the relationship between larval duration and dispersal, we measured larval duration in all settlers assigned to parents (n = 120). Here, we refer to the larval duration as pelagic larval duration (PLD) to be consistent with the broader literature; however, we do not know definitively whether larvae stay within the pelagic zone. Otoliths were dissected under an Olympus SZX10 Dissecting Microscope. Each otolith was cleared of tissue in 95% EtOH, dried until the EtOH had evaporated, then mounted on a slide and immersed in oil for 2-7 days. Rings were counted under a 50X oil immersion lens using the Manual Tag measuring tool in Image-Pro Plus 6.3. For each settler, one randomly-selected otolith was read twice, and PLD was estimated as the average of these two reads. Counting out from the otolith core, clear rings began at approximately the sixth ring in

all individuals. This first clear ring was recorded as the hatching mark; counts began at this clear ring and continued until the settlement mark. If a settlement mark was not distinguished, the count was made until the last clearly defined, uniformly-circular line. To address potential biases from technique refinement over the course of measurement, additional reads were collected (Appendix A).

4.5.6 Statistical Analysis 1—Descriptive models

We estimated an empirical dispersal kernel by fitting alternative probability density functions (p.d.f.s) to the observed dispersal distances between parent-offspring assignments. A suite of alternative p.d.f.s were fit to the vector d of net dispersal distances using a maximum likelihood framework, including exponential, Gamma, Gaussian, logistic, lognormal, and Weibull distributions. We selected an exponential distribution as the best-fit p.d.f. based on AIC and BIC (Appendix A) (Burnham & Anderson 2002). To test whether the transect design, which led to unequal sampling at different distances, affected our estimate of the kernel, we applied a sampling correction (Appendix A). This correction did not affect the estimate of the kernel, so additional analyses were based on the raw data.

We then explored whether the shape of the dispersal kernel was consistent across a number of measured spatial and biological variables. We fit p.d.f.s to subdivided data based on the following variables: (i) Northerly / Southerly dispersal trajectory; (ii) larval region of origin (Northern, Central, and Southern); (iii) settler size, divided by the median value; and (iv) PLD, divided by the median value. Similar analyses were repeated for variables related to the sponge microhabitat (Appendix A). We also considered all the above-mentioned variables in a multivariate analysis to investigate the relationship between these predictors and the mean and variance of dispersal distance (Appendix A).

4.5.7 Statistical Analysis 1—Predictive models

For the logistic models, we considered all potential dispersal trajectories (n = 4,416) as described above. We constructed a set of generalized linear models (family = binomial; link

= logit) with a binary dependent variable that noted whether a dispersal trajectory was used (1) or not (0), based on the outcome of the parentage analysis. Potential predictors included distance (km), direction (North/South), parent region of origin (Northern, Central, Southern), and interaction terms. Sampling effort was also included as a covariate in all models to control for unequal sampling between potential trajectories, and was defined as the total number of offspring collected times the total number of parents collected for each trajectory, scaled from 0 to 1. We tested the predictive power of the best-fit model using 10-fold cross-validation and leave one out cross-validation (LOOCV), with a cost function for binomial data (Davison & Hinkley 1997). We also calculated the AUC with 95% confidence intervals (n = 10,000 bootstrap replicates) as a metric of the predictive accuracy of the model (Hosmer *et al.* 2013).

4.5.8 Connectivity Matrix Analyses

To investigate patterns of population connectivity across the Belize barrier reef system, we generated a connectivity matrix between all potential source sites (N_j) and all potential settlement sites (N_i). To define this set of sites, we obtained a map of the Belize barrier reef system from the IMaRS Millennium Coral Reef Mapping Project (Andréfouët *et al.* 2005). Focusing only on the 'forereef' zone where *E. lori* resides, we converted the forereef polygon into $1km^2$ grid cells using a fishnet tool in ArcMap 10.2 (ESRI). Next, we manually removed excess reef patches to ensure that the reef width never exceeded 1 km, given that *E. lori* are only abundant on the outer reef slope (D'Aloia *et al.* 2011). This led to a total of 516 sites. We generated a distance matrix between all sites by calculating the Euclidian distance between the centroids of all grids. We then used the logistic model to predict the probability of successful dispersal between all sites. In order to scale up from transect to the whole BBR, we assumed that dispersal did not vary regionally. Importantly, this assumption did not lead to a loss of predictive power in the logistic model (Appendix A). The output from the logistic model was scaled from 0 to 1 to represent relative probabilities of dispersal. We visualized connectivity using graph theory approaches to construct connectivity matrices

(Appendix A).

4.6 Acknowledgments

We thank Diana Acosta, Alben David, Kevin David, Alissa Rickborn, and Derek Scolaro for assistance with field work; Eliana Bondra for assistance with molecular work; and Peter Carlson for assistance with otolith work. We are grateful to Noel Anderson, David Lindo, Claire Paris, Robert Warner, and Colleen Webb for comments on this manuscript. This work was supported by NSF (OCE-1260424) and CCD was supported by a NSF GRF (DGE-1247312). All work was approved by Belize Fisheries and Boston University IACUC.

4.7 Author Contributions

CCD, PMB, RGH and SMB designed research; CCD, PMB, SMB, RKF and JEM performed research; CCD analyzed data; CCD wrote the paper; all authors contributed to revisions.

Chapter 5

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

5.1 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 $\Phi_{ST} = 0.12$, microsatellite $F_{ST} = 0.0056$. However, there was high variability among pairwise estimates, ranging from no differentiation between sites on contiguous reef ($\Phi_{ST} = 0$) to strong differentiation between sites separated by ocean expanses $\geq$ 20 km (maximum $\Phi_{ST} = 0.65$). 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.

5.2 Introduction

Current and historical patterns of gene flow influence the evolutionary dynamics of metapopulations and can result in the non-random distribution of alleles across space — spatial genetic structure (SGS) (Hanski & Gaggiotti 2004; Vekemans & Hardy 2004). Analyzing 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 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 kilometers (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 geographic 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 modelsthis 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. 5.1). 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.

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 patterns 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

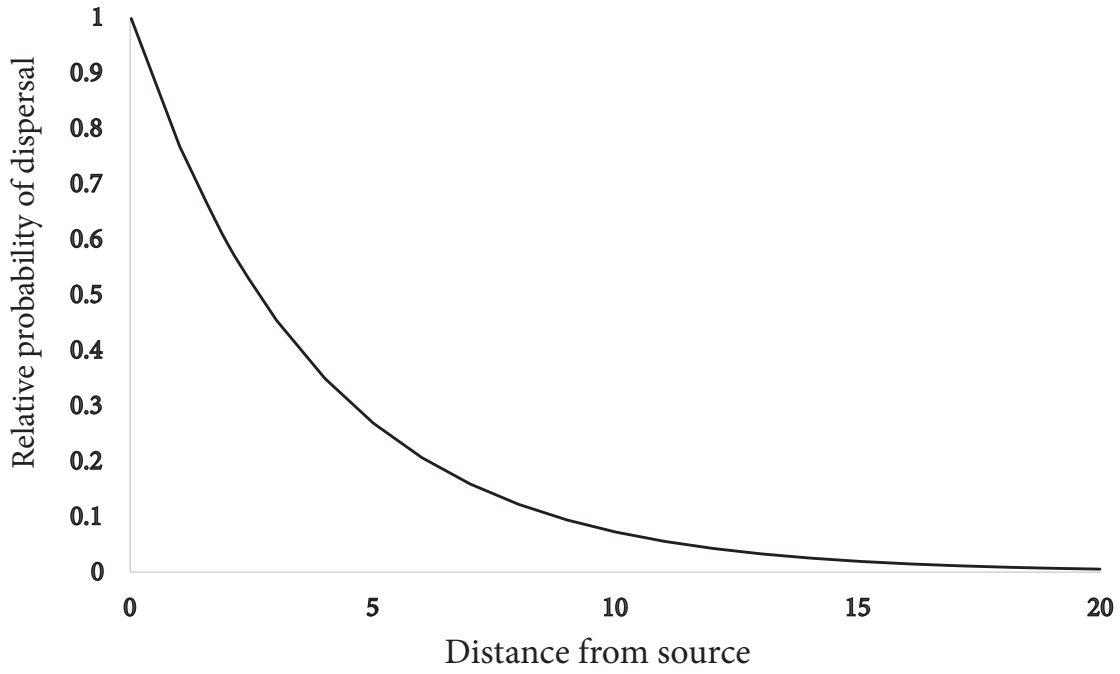

Figure 5.1: First Approximation of the *E. lori* dispersal kernel. The solid black line represents a first approximation of the kernel estimated from genetic parentage analysis. The kernel was measured up to 5 km from source; here, we extrapolate out to 20 km for comparison with SGS data. Figure modified from D'Aloia *et al.* (2013).

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 which 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 geographic barriers to genetic connectivity can occur at a much smaller scale than has been assumed in reef fishes.

5.3 Methods

5.3.1 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. 5.2). There were a total of 5 barrier reef sites and 5 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, non-lethal 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.

5.3.2 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 1,102-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' – TAGAGGGAAAAAGGC-CAAGAAAATAGAAA – 3'). Standard PCRs were run with the following reagents per reaction: 6.9 μl H_20, 1 μl 10X 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), 72°C for 1 min; 29 cycles of 95°C for 40 s, 55°C for 45 s, 72°C 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 10X 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.

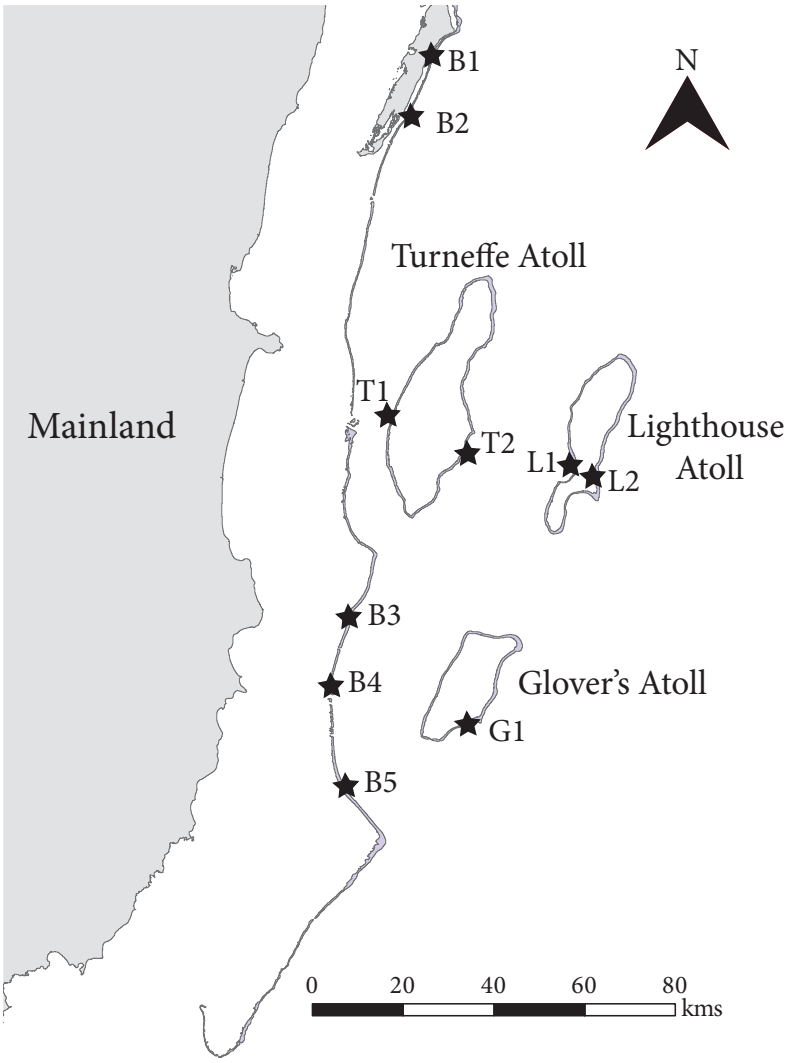

Figure 5.2: Map of the Belizean barrier reef complex. Reef is represented by offshore grey lines. Sampling locations (n=10) are represented by black stars. Barrier reef sites are labeled with a B, while atoll sampling sites are labeled with the first letter of their name.

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 5X 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.* 2004), 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.

5.3.3 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. 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 = 1,001,000; 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.

5.3.4 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).

5.3.5 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 = 1,000) (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). In order 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: 1) 'haplotype 1'; 2) 'haplotypes that were one mutation away from haplotype 1'; 3) 'haplotypes that were two or more mutations away from haplotype 1'; 4) 'haplotype 2'; 5) 'haplotypes that were one mutation away from haplotype 2'; and 6) 'haplotypes that were two or more mutations away from haplotype 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 B).

5.3.6 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 structure values (R: *lme4* and *AICcmodavg* packages). Here, we explicitly accounted for the non-independence

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 non-mutually 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 $\geq$ 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 Glovers 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 (S2). 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'.

5.4 Results

5.4.1 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 B). Similarly, observed heterozygosity was high, ranging from 0.78 – 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 B). 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 out of 91 pairwise comparisons was significant (Appendix B). 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 B). Within sampling locations, the number of private substitution sites ranged from zero to five.

5.4.2 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 5.1). 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 = 1,000; 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: Φ_{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 B). 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 5.2); 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.

5.4.3 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. 5.3a). 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. 5.2). In contrast, sites along the barrier reef are predominantly characterized by two different genetic clusters ('yellow' and 'blue') (Fig. 5.3a). There is some evidence for a cline in neutral markers

Table 5.1: 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 was assessed by a permutation test at $= 0.05$ (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 5.2: Analysis of molecular variance (AMOVA). AMOVA analyses 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 (1,023 permutations).

a)	***mtDNA***				***b)***	***nDNA***			
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%

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 curve (Fig. 5.1; 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 5 point mutations (Appendix B). 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. 5.3b). 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.

5.4.4 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

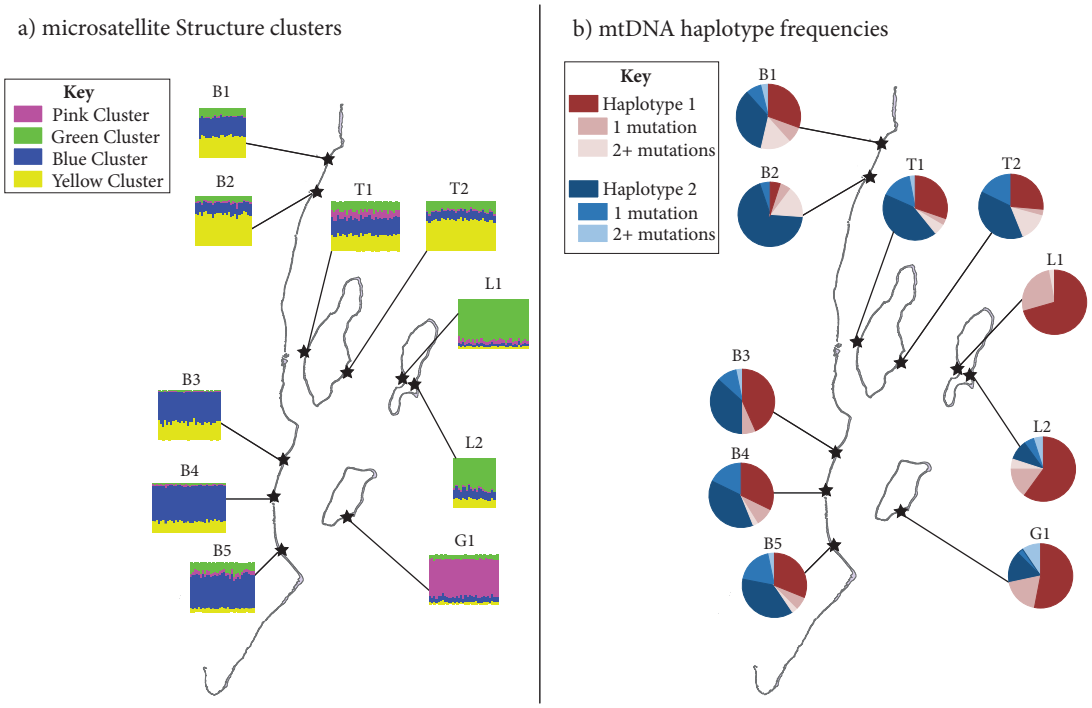

Figure 5.3: Qualitative analyses of spatial genetic structure. a) Distribution of k=4 genetic clusters across all ten sampling sites, based on Structure analysis of microsatellites. These data show distinct clusters on Glover's Atoll (predominantly pink) and Lighthouse Atoll (predominantly green). The 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 B). 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.

break $\geq$ 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); s.e. = 0.97; $p < 0.001$). There was no significant improvement over this 'oceanic break' model when 'Euclidian distance' was included as a predictor ($\chi^2 = 0.77$, df = 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 non-zero value of F_{ST} relative to sites that are not separated by a break

(coefficient = 2.92 (odds ratio = 18.52); s.e. = 1.53; p = 0.007) and that adding Euclidian distance as a predictor did not improve model fit (χ^2 = 1.12, df = 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 which posits that SGS is associated with oceanic breaks.

Building upon these findings, linear mixed models revealed that oceanic breaks $\geq$ 20 km were also significantly associated with the magnitude of non-zero structure estimates between sites. The presence of a break was associated with a 93% increase in pairwise Φ_{ST} (Table 5.3a) and a 56% increase in pairwise F_{ST} (Table 5.3b), relative to pairs of sites without a break. For Φ_{ST} models, the 'ocean break' model (H_1; ΔAIC_c = 0.00) was a better fit than a null model of only random effects (H_0; ΔAIC_c = 2.51), and an alternative model that also included Euclidian distance (H_2; ΔAIC_c = 2.69). A comparison of F_{ST} models revealed the same pattern: the 'ocean break' model (H_1; ΔAIC_c = 0.00) was a better fit than a null model of only random effects (H_0; ΔAIC_c = 6.00), and an alternative model that also included Euclidian distance (H_2; ΔAIC_c = 2.75). Taken together, these results show strong statistical support for H_1 across both genetic markers.

Table 5.3: Linear mixed model output. The dependent variable is non-zero 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	0.93	0.34	2.74	Oceanic break	0.56	0.10	5.45
Random effects				***Random Effects***			
Group	**S.D.**			**Group**	**S.D.**		
Site 1	0.22	—	—	Site 1	0.00	—	—
Site 2	0.48	—	—	Site 2	0.00	—	—

Importantly, these results were not robust to shortening the distance cut-off for the definition of an 'oceanic break' (Appendix B). This result is congruent with the qualitative

clustering analyses. Together, they demonstrate that only substantively 'large' oceanic breaks ($\geq$ 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*.

5.5 Discussion

Investigating the patterns and predictors of spatial genetic structure (SGS) concurrently is essential to 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 bio-physical 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 (DAloia *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. 5.1; DAloia *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 in order 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 strong structure. 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. 5.1). 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 geographic 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. 5.3a). 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 5.1). Despite this difference in magnitude, the results of the structure analyses, the clustering analyses and the statistical modeling 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.

5.6 Acknowledgments

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. 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.

5.7 Data Accessibility

Supplemental tables and figures references in the text are available in Appendix B. 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).

5.8 Publication

D'Aloia CC, Bogdanowicz SM, Harrison RG, Buston PM (2014). Seascape continuity plays an important role in determining patterns of spatial genetic structure in a coral reef fish. *Molecular Ecology* 23: 2902–2913.

5.9 Author Contributions

CCD and PMB planned the study. CCD conducted field work with assistance from PMB and others. CCD conducted laboratory work with supervision by SMB and RGH. CCD

analyzed the data and wrote the paper, with all authors contributing to revisions.

Chapter 6

A complete seascape genetic analysis of SGS in the *E. lori* Belizean metapopulation

6.1 Abstract

The exchange of alleles via larval dispersal is fundamental to microevolutionary processes in marine metapopulations. While conventional wisdom posits that stochastic long-distance recruitment pulses should lead to strongly-connected metapopulations across large spatial scales (100s of kilometers), recent research has documented genetic structure at relatively small spatial scales (10s of kilometers). The objectives of this study are to quantify fine scale patterns of spatial genetic structure (SGS) across the entire metapopulation of the coral reef fish *Elacatinus lori* and to identify the bio-physical correlates of SGS. This research builds upon a preliminary study (D'Aloia *et al.* 2014) by expanding the spatial scale of sampling and increasing the intensity of sampling (both individuals and loci). We used a ddRAD approach to identify polymorphic loci for a multiplex PCR experiment and then sequenced 1231 individuals at 59 anonymous loci, 14 microsatellite loci, and mitochondrial cytochrome b. First, to the describe macro-scale patterns of SGS, we reconstructed intraspecific phylogenies and haplotype networks. Unexpectedly, we discovered a unique clade of *E. lori* in the Southern region of Belize and Honduras that appears to be genetically isolated from the rest of Belize. Second, we used multiple matrix regression with randomization to test alternative predictors of SGS within the Belizean portion of the *E. lori* metapopulation. These analyses revealed that a dummy variable capturing oceanic

breaks between sites explained most of the variance in pairwise genetic structure, with an additional weak signal of isolation by distance. We suspect that oceanic breaks are associated with SGS because of the low probability of dispersal across these breaks in habitat. Future work will integrate these data with connectivity matrices derived from parentage analyses and bio-physical models to test the full suite of alternative seascape genetic hypotheses. Ultimately, this project will advance our understanding of the processes driving long-term genetic connectivity in a spatially-complex marine metapopulation.

6.2 Introduction

In recent years, the field of seascape genetics has proliferated (Selkoe *et al.* 2008; Riginos & Liggins 2013). At its core, this burgeoning field is focused on understanding how features of the seascape influence patterns of genetic connectivity and the emergence of spatial genetic structure (SGS). While the role that spatial structure can play in microevolution has long been recognized in population genetics (Wright 1943), the expanded conceptual framework of seascape genetics is derived from the field of landscape genetics, which seeks to evaluate the effect of landscapes on the spatial distribution of genetic variation (Manel *et al.* 2003; Storfer *et al.* 2007; Manel & Holderegger 2013).

While many of the established principles and analytical approaches from landscape genetics can be transferred to seascape genetics, there are notable differences that must be considered for organisms living in a fluid environment. First, many environmental features of the ocean, such as current speed and direction, salinity, and sea surface temperature, can fluctuate rapidly in comparison to features of the terrestrial landscape (Riginos & Liggins 2013). A second major difference is that genetic structure tends to be consistently weak in marine metapopulations; this can make it challenging to detect subtle levels of genetic structure. Additionally, when there *is* evidence of genetic structure, there may be drastic temporal changes and/or no clear geographic pattern, termed 'chaotic genetic patchiness' (Johnson & Black 1982; Selkoe *et al.* 2008). There is some evidence, however,

that these so-called chaotic patterns may simply reflect the fact that many marine species do not follow an isolation-by-distance (IBD) pattern, and the 'chaos' may disappear as alternative predictors are considered in seascape genetic studies (Selkoe *et al.* 2010). These characteristics of marine metapopulations represent important potential deviations from landscape studies and must guide *a priori* hypothesis development and sampling design.

In addition to these seascape-specific considerations, there has also been a general call for researchers to shift from pattern-oriented studies toward process-oriented studies in landscape and seascape genetics (Pflüger & Balkenhol 2014). Traditionally, these fields have focused on correlative studies that relate features of the intervening landscape/seascape matrix to patterns of pairwise genetic structure. This focus on the movement of individuals (and alleles) *through* the landscape matrix ignores other key aspects of dispersal—mainly, emigration from the source site and immigration into the settlement site. One way to move towards a process-oriented focus is to quantify local population densities and environmental variables at source and settlement sites. The former may be important for taxa in which dispersal is density-dependent (Matthysen 2005), while the latter may be important for calculating environmental distances between sites if there is local adaptation (Wang & Bradburd 2014).

An alternative approach that may be more appropriate for marine species is the incorporation of probabilistic connectivity matrices into the seascape genetic analytical framework. High-resolution oceanographic models can be coupled with larval behavior to predict dispersal kernels and transition probability matrices (e.g. Paris *et al.* 2013). For species in which demographic connectivity data have been collected, the empirical connectivity matrix can also be used to predict connectivity. Importantly, the empirical connectivity matrix implicitly includes all aspects of dispersal (i.e. emigration-transfer-immigration) as it represents *realized* connectivity patterns. Relating these data to SGS will likely be a major new focus in seascape and landscape genetics, since these data are just becoming available for a few species.

Building on these emerging topics in seascape and landscape genetics, researchers are

actively developing a set of best-practices for spatial statistical analyses (Wagner & Fortin 2005, 2013; Balkenhol *et al.* 2009). Combining seascape/landscape feature data with genetic data to address questions related to the drivers of genetic connectivity poses several challenges. Most studies aimed at understanding gene flow between sites utilize link-level analyses, where researchers are interested in the lines, or distances, between two sampling sites (Wagner & Fortin 2013). For example, researchers may be interested in relating pairwise genetic distance (i.e. pairwise F_{ST}) to features of the seascape. Here, [Y] would represent the response matrix, with Y_{ij} describing the genetic distance between sites i and j. A suite of predictor matrices [X_1 ... X_p] could then be used to test specific seascape genetic hypotheses. However, within this analytical framework two major problems arise. First, the observations are non-independent, as they are pairwise comparisons. Second, the predictor variables are often spatially-autocorrelated to some degree. Given these statistical issues, a promising approach is multiple matrix regression with randomization (MMRR) (Wang 2013), which is well-suited for data in matrix format. Through the multivariate framework, Euclidean distance can be included as a covariate to control for spatial autocorrelation in all models. However, there is still progress to be made. Appropriate model selection for MMRR models is currently unclear, and the method still needs to be modified to account for non-linear relationships (Wagner & Fortin 2013; Wang 2013).

The neon goby *Elacatinus lori* is a tractable study organism for a comprehensive seascape genetic analysis for several reasons. First, spatial variation in genetic structure has already been documented for *E. lori* within Belize (D'Aloia *et al.* 2014). Second, an empirical dispersal kernel has been measured for *E. lori* (Chapter 4), and a connectivity matrix can be generated from that kernel. Third, a high resolution ocean-atmosphere model has been developed for the Belizean Barrier Reef (BBR) region (Lindo-Atichati, unpublished), and physical and bio-physical connectivity matrices can be generated from that model. In this study, we build upon these data to address three specific objectives: (1) quantify fine-scale patterns of spatial genetic structure across the entire Belizean metapopulation, (2) test alternative correlates of genetic structure and (3) test alternative connectivity matrices

as potential mechanistic predictors of genetic structure.

6.3 Methods

6.3.1 Sample collection

To quantify spatial genetic structure within the entire Belizean metapopulation, we collected tissue samples from adult *E. lori* at 39 sites out of 43 planned sites across the BBR (Fig. 6.1). Individuals were caught using slurp-guns and non-lethal fin-clips were taken from the caudal fin using scissors. Each fish was returned to its sponge after handling. In total, we collected 1231 fin clips, with an average of 31.51 $\pm$ 5.10 s.d. per site (Table 6.1).

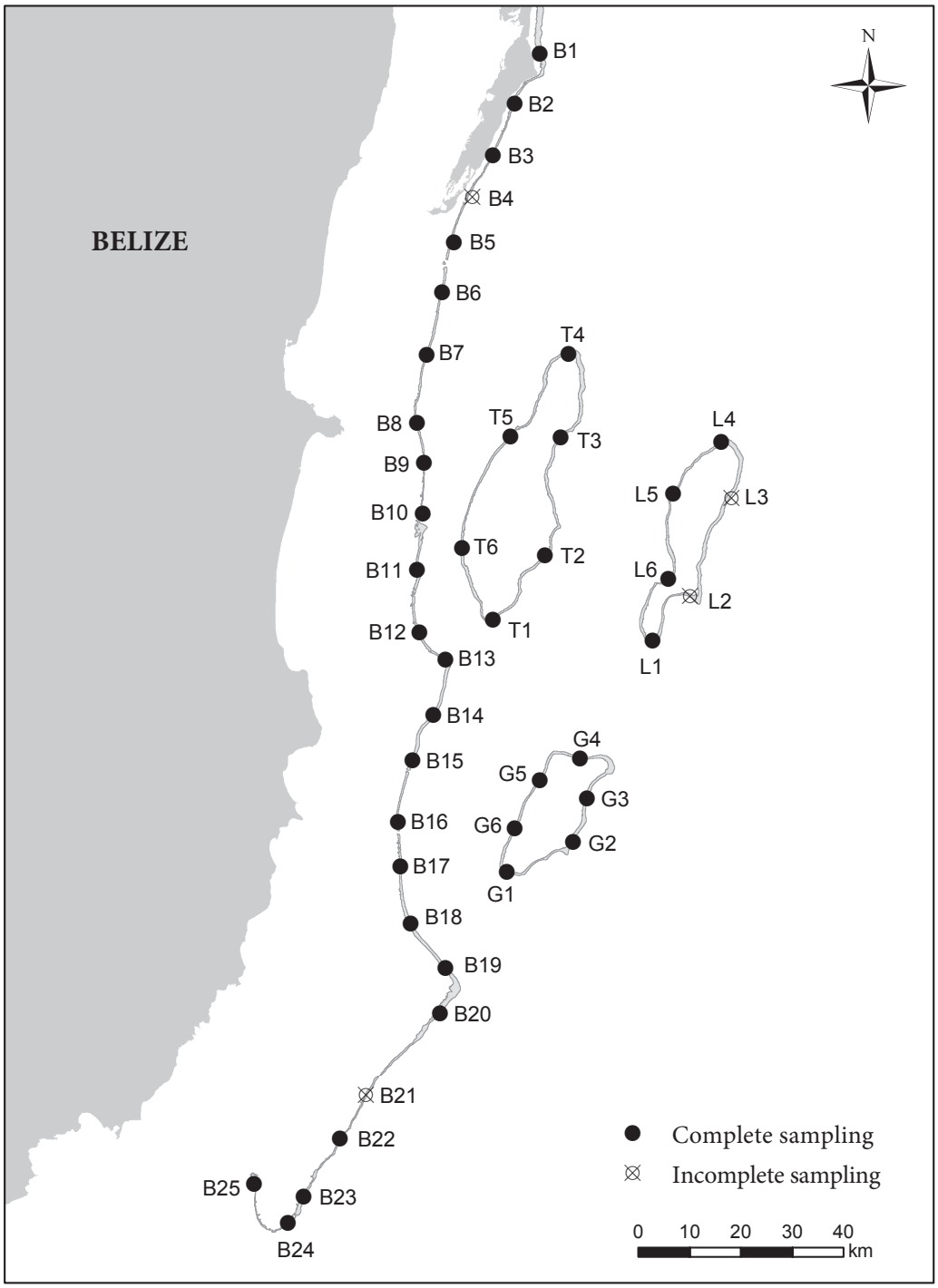

Figure 6.1: Sampling locations along the Belize barrier reef system in 2014. All samples were collected between May-June 2014. Locations where sampling was unsuccessful (4 out of 43 planned locations) were attributable to weather (B4, L3), lack of sponge microhabitat (B21), and an aggressive resident barracuda (L2).

Table 6.1: Number of individuals collected and sequenced per site. The four sites where only 0 or 1 individuals were collected are highlighted in gray as these were excluded from all analyses including the calculation of means and standard deviations. The number of individuals sequenced indicates the number of individuals with sequences retained after all quality filtering.

Site	**# individuals collected**	**# individuals sequenced**
B1	15	15
B2	34	30
B3	32	31
B4	0	0
B5	35	31
B6	35	31
B7	20	19
B8	35	31
B9	35	32
B10	21	21
B11	35	33
B12	35	30
B13	35	32
B14	31	31
B15	31	30
B16	34	29
B17	36	31
B18	31	31
B19	27	27
B20	33	31
B21	1	0
B22	22	17
B23	24	24
B24	23	16
B25	31	21
T1	35	30
T2	35	32
T3	34	32
T4	35	30
T5	35	32
T6	35	32
G1	32	32
G2	32	30
G3	32	32
G4	33	32
G5	35	32
G6	31	30
L1	34	32
L2	1	0
L3	0	0
L4	35	32
L5	35	32
L6	31	30
Mean	**31.51**	**28.87**
s.d.	**5.10**	**4.91**

6.3.2 Preliminary ddRAD sequencing of anonymous loci

We used a double-digest restriction-site associated DNA (ddRAD) sequencing approach to identify polymorphic loci for multiplex PCR. To reduce ascertainment bias, we selected eight individuals from a broad geographic range to generate an initial ddRAD library (IDs: B2-1, B6-3, B17-7, B20-7, B25-1, T2-1, L2-1, G3-1). Modifying the Peterson *et al.* (2012) protocol, we used Sbf1 and Msp1 restriction digests/adapter ligations and targeted relatively large fragments (size selected for 350-550 bp with Ampure XP). We then used the 3,800 resulting genomic contigs as a reference file to re-align all raw reads in order to assess coverage for each contig. We excluded the following contigs: (i) contigs with low coverage ($<$ 20X) to reduce sequencing error bias; (ii) contigs with extremely high coverage ($>$1,000X) to avoid repetitive regions in the genome; and (iii) monomorphic contigs to exclude uninformative regions. From the resulting pool of contigs, we randomly drew 136 contigs as anonymous loci for downstream multiplex PCR.

We developed primer pairs for each contig using BatchPrimer3 (You *et al.* 2008), and synthesized both forward and reverse primers with Nextera 5' tags (5'– TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG was appended to each forward primer, 5'–GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG to each reverse primer). Using individuals B10-1 and T1-11 as consistent test samples for all loci, we ran PCRs with touch-down protocols for every set of primers. We created a master mix of 0.5 μl DNA, 5.85 μl H_2O, 2 μl 5X buffer, 0.2 μl 10 mM dNTP, and 0.05 μl One Taq (New England Biolabs) per sample. We then combined 8.6 μl master mix with 1.4 μl primer mix (forward + reverse primers diluted in 0.5X AE buffer to 2.85 μM). PCRs were run with the following thermal cycler settings: heat to 90°C, then 6 touch-down cycles of 95°C for 40 s, 61°C - 56°C for 45 s (annealing temperature reduced 1°C each cycle), 68°C for 1 min; 29 cycles of 95°C for 40 s, 55°C for 45 s, 68°C for 1 min. We determined whether loci amplified by running PCRs out on a 1.3% agarose gel. We identified 109 candidate loci for further analyses.

6.3.3 Nuclear microsatellites and mitochondrial loci sequencing

To directly compare these data to SGS estimates from 2012, we also amplified the 14 microsatellite loci and mtDNA cytochrome b (*cytb*) locus used in the pilot study (D'Aloia *et al.* 2014). Because the original cytb sequence was 1102-bp long, we divided the sequence into 2 contigs for sequencing on the Miseq by developing two new nested primers (Appendix C).

6.3.4 Multiplex PCR and sequencing

Genomic DNA from all samples was extracted using the Agencourt DNAdvance DNA Isolation Kit at $\frac{1}{2}$ kit volumes. We conducted multiplex PCR reactions in 384-well plates using the QIAGEN Multiplex PCR Kit. The 125 loci were divided into five multiplex groups (Appendix C). For each multiplex group, we made diluted primer mixes by combining 2 μl of each primer (500 μM stocks) and enough Qiagen AE buffer (diluted 1:1 with water) for a total volume of 500 μl per multiplex mix. Primers for mitochondrial *cytb* were added at 0.5X to account for the increased copy number of the mitochondrial genome. For each multiplex run we ran 10 μl rxns; we created a master mix by combining 5 μl 2X MM, 2.5 μl RNase-free H_2O, and 1 μl diluted primer mix per sample. In each well of a 384-well plate we combined 8.5 μl of this master mix with 1.5 μl DNA.

We amplified each multiplex group with touchdown PCRs at the following settings: initial denaturation at 95°C for 15 min; six cycles of 94°C for 30 s, 61°C 56°C for 90 s, 72°C for 90 s; 22 cycles of 94°C for 30 s, 55°C for 90 s, 72°C for 90 s; a final extension at 72°C for 10 min.

We then pooled the 5 multiplex groups by combining 2 μl of amplified DNA from each multiplex into a single 384-well plate (2 μl x 5 multiplex groups = 10 μl), and then diluted the pooled DNA with 10 μl H_2O. Next, we used Illumina's S5 (n=16) and N7 (n=24) Nextera primers to barcode each individual. We made a master mix of 2666 μl H_2O, 860 μl 5x One Taq buffer, 86 μl 10mM dNTPS, and 22 μl One Taq HotStart. The master mix

was divided proportionately among 16 S5 and 24 N7 wells; we then added 1.5 μl each S5 primer (100 mM) and 1 μl each N7 primer and mixed. To each well in the 384-well plate, we combined 4.25 μl S5 master mix, 4.25 μl N7 master mix, and 1.5 μl DNA. Barcoding PCRs were run with the following settings: initial denaturation at 92°C for 2 min; six cycles of 94°C for 30 s, 62°C for 1 min, 68°C for 1 min; and a final extension at 68°C for 10 min.

Upon completion, we pooled all barcoded individuals by taking 3 μl from every barcoding rxn and combining all samples into a single 1.5 mL tube (384 x 3 μl = 1152 μl). We performed a clean-up using the AxyPrep Mag FragmentSelect Kit to remove small fragments (i.e. PCR artifacts) before sequencing. We cleaned 1 μg of DNA with a bead:DNA volume ratio of 0.65. Following the One-Step size selection protocol from the AxyPrep Kit, we eluted the size-selected DNA in 0.5X AE buffer.

Eluted DNA was diluted with Sigma H_2O and 0.1% Tween to achieve a 2nM concentration. All samples were then sequenced on an Illumina MiSeq with paired-end reads (2 x 300bp) at Cornell University's BioResource Center. We repeated this entire multiplex process 3 times to sequence 1152 individuals (384 individuals/run x 3 runs =1152 individuals).

6.3.5 Post-sequencing data processing

We ran a custom Perl script in order to extract reads from the Miseq run and assign them to the appropriate locus and individual. Specifically, the script trims adapters and low quality reads, creates contigs from overlapping reads (for paired-end sequencing), identifies reads corresponding to each locus, collapses identical reads for each individual, identifies the top two haplotypes for individuals at all loci (i.e. their diploid genotypes), and generates FASTA alignments of haplotypes for export to other programs.

We ran the script separately for the anonymous loci, microsatellites, and mtDNA to customize the input parameters. For all three runs, we applied a matching command (-x) that required 90% of the first 40 bp of a read to align to and match the reference contig,

thereby filtering most PCR artifacts and paralogs while retaining true SNPs and short indels. For the anonymous loci, we set a minimum read length of 225 bp to filter out PCR artifacts. For the microsatellites, we reduced the minimum read length to 100 bp to capture shorter alleles. For the mitochondrial loci, we kept the minimum read length at 225 bp, but relaxed the Q quality score from Q20 to Q12 in order to recapture missing data and boost read counts. We also ran a concatenate command (-w2) in order to join the 225 bp forward read and 225 bp reverse read, as these reads did not overlap.

Finally, we applied additional manual filters to the data prior to all downstream population structure analyses. Individuals that failed completely or had <20X coverage across all loci were excluded. Loci that failed in all individuals, had <20X coverage at most individuals, or appeared to be paralagous were also excluded. After these filters were applied, we retained sequence data for 74 loci (n = 59 anonymous loci; n = 14 microsatellites; n = 1 *cytb* fragment) for 1126 individuals.

6.3.6 Macro-scale patterns: investigating a new genetic clade

6.3.6.1 Additional Sanger sequencing to investigate the Southern region

Preliminary results revealed evidence of a divergent lineage of *E. lori* in the southern region of the Belize barrier reef (sites B22-B25; Fig. 6.1). To further explore this lineage through phylogenetic and haplotype network analyses, we completed additional Sanger sequencing of two mitochondrial genes: cytochrome b (*cytb*) and cytochrome oxidase 1 (*COI*). We also obtained four tissue samples of *E. lori* from Utila, Honduras, enabling us to place Southern haplotypes within the context of the greater Mesoamerican reef system. Honduran samples were provided by Benjamin C. Victor (collection date: 7/1/2008; location: 16.08 N -86.92 W).

For *cytb*, we used the sequences generated from the multiplex experiment for all Belizean individuals (n = 1126 individuals). We sequenced the four Utila samples with *E. lori*-specific primers, following the protocol detailed in D'Aloia *et al.* (2014). For *COI*,

we used previously-published sequence data for the four individuals from Utila (Genbank accession #s: KM894168.1, KM894165.1, KM894155.1, KM894154.1). To generate *COI* sequences from Belizean individuals, we randomly selected individuals from different regions of the BBR for sequencing, with a relatively large number of samples from the divergent Southern region (see Appendix C for list of individuals). We designed *E. lori*-specific primers (Elori-COIF 5'– ATGGTCGGCACAGCTCTTAGCCT and Elori-COIR 5'– CCGAAGAATCAGAATAGATGCTGATAAAGGATGG), and ran standard PCR and enzymatic clean-up reactions described in detail in D'Aloia *et al.* (2014). We ran forward sequencing reactions for each individual by combining 1.5 μl DNA, 2.1 μl Sigma H_2O, 0.5 μl BigDye terminator v. 3.1 Ready Reaction Mix (Life Technologies), 0.75 μl ABI 5X sequencing buffer and 0.15 μl forward primer. We cleaned sequencing reactions with Agencourt CleanSeq beads (Beckman Coulter), and sequenced on an ABI 3730 automated sequencer.

Sequences were aligned with the MUSCLE algorithm (Edgar 2004) in MEGA 6 (Tamura *et al.* 2013) and trimmed to the same length (n=379 bp for *cytb*; n=560 bp for *COI*). Some base pairs were cut from the middle of the *cytb* Sanger sequences, as they were not sequenced in the multiplex experiment (see Post-sequencing data processing details on the -w2 concatenate script command).

We visualized the relationship among haplotypes by generating a haplotype network for each locus in TCS v1.21 with 95% parsimony. Next, we estimated phylogenetic relationships in MEGA 6 (Tamura *et al.* 2013). For each locus, we chose the best-fit nucleotide substitution model based on BIC. We then estimated phylogenetic relationships using a maximum-likelihood framework to build trees, rooted with sequence data from the sister species *E. horsti*. We assessed support for nodes with bootstrapping (n=1,000 bootstrap replicates).

Because there was clear support for two clades of *E. lori*, we excluded the second clade (Southern Belize/Honduras) from all further analyses. The individuals from these regions represent a distinct metapopulation.

6.3.7 Micro-scale patterns: SGS within Belize

6.3.7.1 Genetic clustering analysis

To identify clusters of genetically-related individuals within the Belizean metapopulation, we used discriminant analysis of principal components (DAPC) (Jombart *et al.* 2010). DAPC maximizes between-cluster variance and minimizes within-cluster variance using synthetic variables (discriminant functions) that represent linear combinations of alleles. Briefly, PCA-transformed data are used in a sequential k-means clustering algorithm to choose the number of clusters (k) that maximizes inter-cluster variation. We considered k=1 to k=35 clusters and used BIC to choose the optimal value for k. Next, we ran an optimization test of the 'alpha-score' to determine the number of principal components (PCs) to retain for the discriminant analysis. The alpha-score represents the trade-off between the power of discrimination and over-fitting. Finally, a discriminant analysis, with n = 59 PCs retained, was run to assign individuals to clusters.

6.3.7.2 Pairwise genetic structure

Focusing exclusively on the Belizean metapopulation (n=35 sites), we calculated pairwise F_{ST} based on (1) the 59 anonymous markers, (2) the 14 microsatellite markers, and (3) all anonymous and microsatellite diploid markers combined. We estimated significance with permutation tests in Arlequin v 3.5 (n=10,000 permutations).

We built a response matrix of pairwise genetic distance [Y], using pairwise F_{ST} values based on the 59 anonymous loci. Future analyses will also use response variables based on alternative loci (i.e. microsatellites) and alternative differentiation metrics (i.e. D_{EST}).

6.3.7.3 Correlates of SGS

We constructed a set of predictor matrices [X_1, X_2, ... X_p] corresponding to three major seascape genetic hypotheses for the correlates of SGS:

- First, we created a dummy variable that described whether any two sites were sep-

arated by an ocean expanse $\geq$ 20 km (1) or not (0), representing the isolation by barrier (IBB) hypothesis, based on preliminary findings in *E. lori* (DAloia *et al.* 2014).

- Second, we generated a matrix of the Euclidean distance between all pairs of sites to test the classic isolation by distance hypothesis (IBD).

- Third, we created an environmental distance variable to explore the isolation by environment hypothesis (IBE). We focused on variables related to salinity and sea surface temperature from the MARSPEC data set, which is available at a $1km^2$ grid scale (Sbrocco & Barber 2013). These data represent contemporary environmental conditions; the salinity data are based on in-situ NOAA observations from 1955-2006 and the temperature data are based on NASA satellite data from 2002-2010. We extracted the annual means for sea surface temperature and salinity for each sampling location as a preliminary characterization of the local environment. To translate these data into an environmental distance measure between sites, we generated a dissimilarity matrix using Gower's dissimilarity coefficient (Gower 1971) in R.

6.3.7.4 Connectivity matrices as mechanistic predictors of SGS

We are constructing an additional set of predictor matrices [X_1, X_2, ... X_p] that represent alternative connectivity matrices (i.e. mechanistic predictors of SGS):

- Fourth, given that the relationship between dispersal and distance is non-linear for *E. lori* (Chapter 4), we tested whether the empirical connectivity matrix (CM) derived from parentage analyses predicted SGS (here, we call this hypothesis isolation by dispersal probability—IBDP). Briefly, to predict the CM, we divided the forereef of the BBR into $1km^2$ grids, and calculated the Euclidean distance between the centroids of all 516 grids. These distances were then used as input to the distance-driven logistic model that had been fit to empirical dispersal data. The resulting connectivity matrix (CM) from the logistic model represented the predicted probability of dispersal

between any two sites in one time step. We then divided each element of the matrix by the matrix sum, such that the entire matrix summed to 1. For these preliminary seascape analyses, we pulled out the 35 sites in the Belizean metapopulation in order to test whether a single temporal snapshot of the CM was related to SGS.

- Fifth, a physical oceanography connectivity matrix will be incorporated as a predictor once collaborators complete simulations of a high-resolution physical model in Belize (Lindo-Atichati, unpublished). This will enable the test of the isolation by oceanographic distance hypothesis (IBOD).

- Sixth, a biophysical oceanography connectivity matrix will be incorporated as a predictor once collaborators have determined the biophysical model that best predicts the empirical dispersal kernel derived from parentage analysis (Lindo-Atichati, unpublished).

6.3.7.5 Regression and correlation analyses

To test whether any of the predictor matrices was associated with genetic distance between sites in a multivariate framework, we used multiple matrix regression with randomization (MMRR) (Manly 1986, 1991; Legendre *et al.* 1994; Wang 2013). MMRR applies standard multiple regression analyses, but is appropriate when the data (both predictor and response variables) are in a pairwise matrix format. Given the non-independence of the pairwise observations, this technique implements randomized permutation to assess the significance of predictors. We performed MMRR on standardized matrices with 1,000 permutations using the function "MMRR" in R (Wang 2013). To meet model assumptions, we also tested for correlation among predictor matrices. After performing MMRR, we applied backward and forward stepwise selection to identify the best-fit model.

We also conducted traditional Mantel tests to investigate the correlation between each predictor matrix and the genetic distance matrix. Mantel tests were conducted in the *vegan* package in R based on the implementation in Legendre and Legendre (2012). We tested

both Pearson's product-moment correlation and Spearman's rank correlation with 1,000 permutations.

Given the linear and continuous distribution of the Belize barrier reef, we conducted an additional Mantel test to explicitly test the IBD hypothesis within a truly linear seascape. Theory predicts that IBD may be detectable on such a linear seascape (Rousset *et al.* 1997), even if the pattern is obscured in the broader, spatially-complex, two-dimensional seascape. Thus, we applied a reduced Mantel test for genetic distance and Euclidean distance, using only the n = 19 Northern and Central barrier reef sites.

6.4 Results

6.4.1 Macro-scale patterns: investigating a new genetic clade

Haplotype network analyses based on two mitochondrial genes revealed that there are two distinct clusters of haplotypes. Across the Northern/Central BBR (Sites B1-B20) and all three Belizean atolls (Fig. 6.1), individuals have haplotypes from one particular cluster, denoted by the color blue in Figs. 6.2 (*cytb*) and 6.3 (*COI*). There is a second cluster of haplotypes, denoted by the color red, that is separated from the blue cluster by a minimum of 5 (*cytb*) or 7 mutations (*COI*). Haplotypes from the second cluster are only found in individuals from the Southern tip of the BBR (Sites B22-B25) and Utila, Honduras, with two exceptions: two individuals from site B20 were found to have 'Southern' *cytb* haplotypes (Fig. 6.2). However, these two individuals were clearly assigned to the Northern-Central BBR/atoll group based on clustering analyses of nuclear markers (see below). These two individuals were not sequenced at *COI*. Overall, there are more haplotypes at *cytb* (n=40 haplotypes) than *COI* (n=17), even though the *cytb* fragment is substantially shorter than the *COI* fragment, indicating a faster mutation rate at *cytb*.

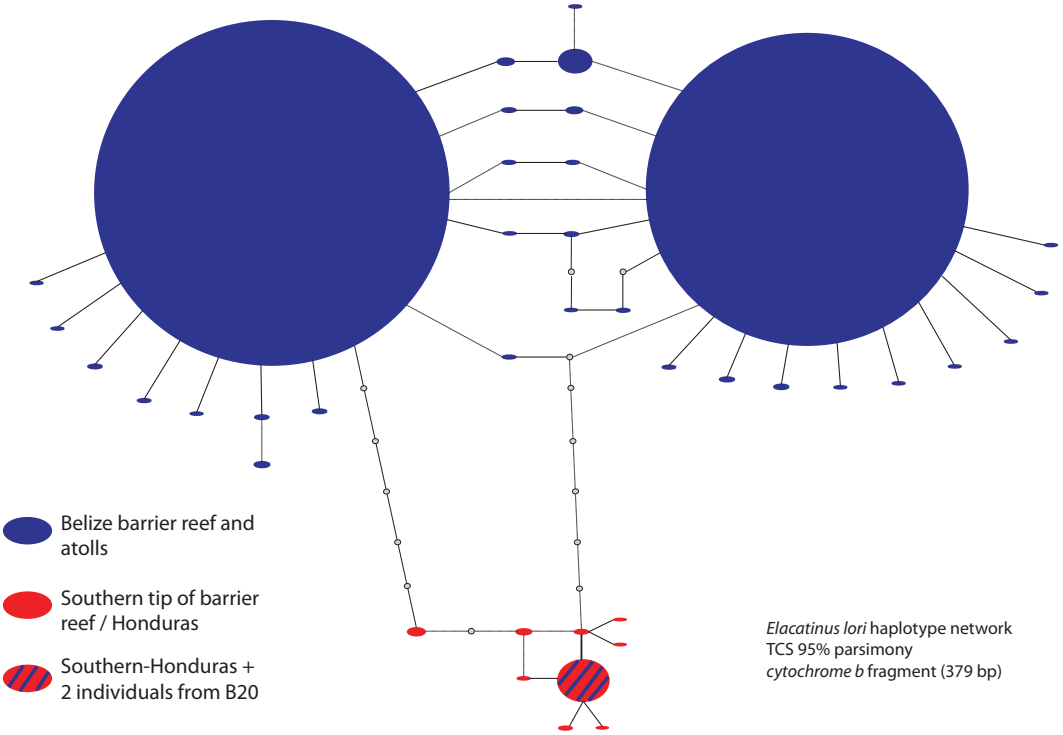

Figure 6.2: Haplotype network based on 379-bp of mitochondrial *cytb* gene. The network was generated in TCS v1.21 with 95% parsimony. The size of the circle is proportional to the frequency among individuals sequenced; each line segment represents a single point mutation; and small grey circles represent inferred haplotypes (n=1048 individuals from Northern-Central BBR & atolls; n=82 individuals from Southern BBR & Honduras).

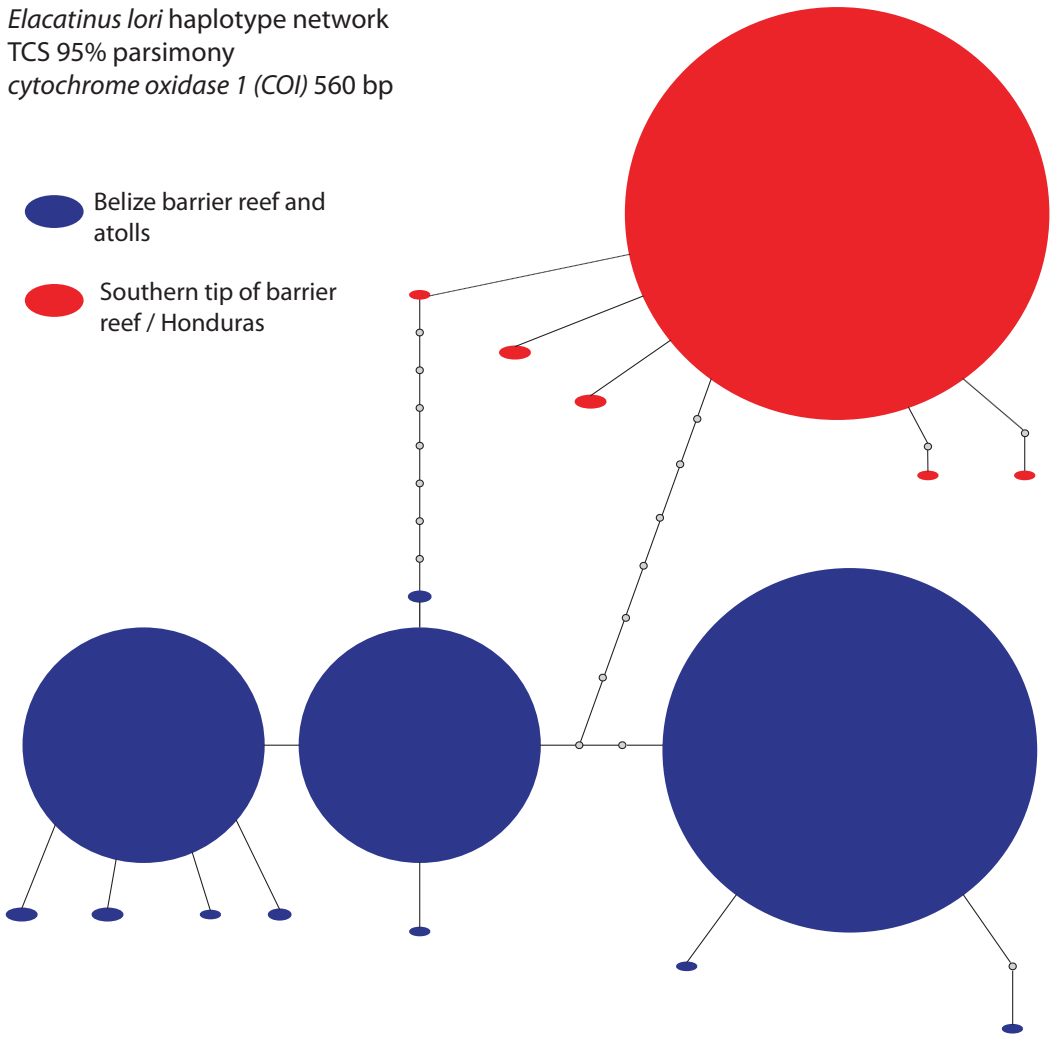

Figure 6.3: Haplotype network based on 560-bp of mitochondrial *COI* gene. The network was generated in TCS v1.21 with 95% parsimony. The size of the circle is proportional to the frequency among individuals sequenced; each line segment represents a single point mutation; and small grey circles represent inferred haplotypes (n=117 individuals from Northern-Central BBR & atolls; n=62 individuals from Southern BBR & Honduras).

Reconstructed phylogenetic trees were consistent with the haplotype networks. Maximum likelihood trees based on both mitochondrial markers showed evidence for two distinct clades of *E. lori*: one clade consists of the Northern/Central BBR and Belizean atolls and

another clade consists of the Southern BBR and Honduras (Figs. 6.4 and 6.5). There was bootstrap support for these two clades at a significance level of 70%. For both trees, we highlight the haplotypes observed in individuals from Utila in green, showing that they are either shared with Southern Belizean individuals, or very closely related.

Because there is clear evidence that the Southern Belizean region represents a distinct genetic clade, and the haplotype networks show that there are almost no shared haplotypes between the Southern region and the rest of the BBR, we exclude the Southern individuals from all further seascape analyses. Additionally, preliminary statistical analyses that grouped the Southern individuals with the rest of Belize resulted in spurious seascape genetic results (not shown). As there is likely little to no ongoing gene flow across these two regions, we consider the Southern Belize / Honduras clade as a distinct metapopulation.

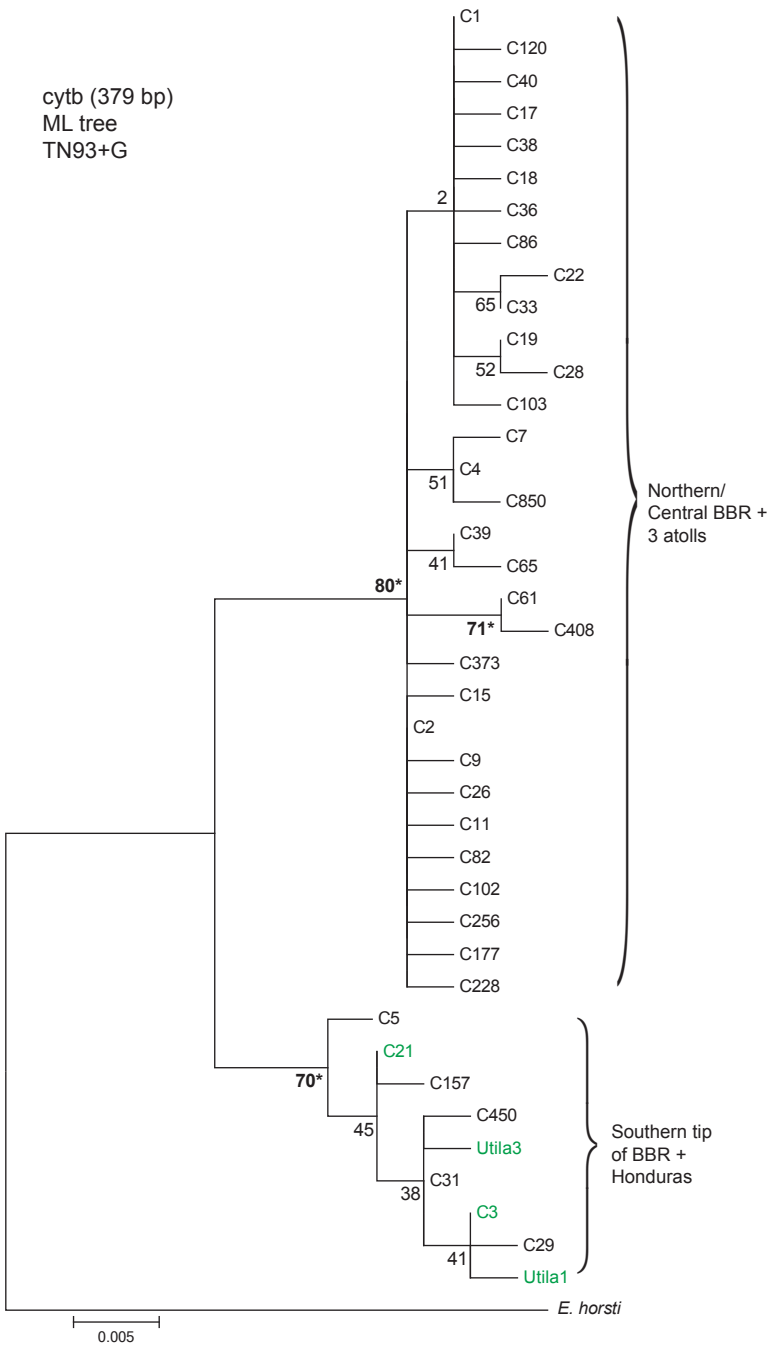

Figure 6.4: Maximum likelihood tree of *E. lori* haplotypes based on 379-bp of *cytb*. We used BIC to identify the Tamura-Nei model with a gamma distribution as the most appropriate substitution (TN93 + G). We used a discrete Gamma distribution (parameter=0.05). Bootstrap percentages are shown at nodes (n=1,000 bootstrap replicates). We represent 'significant' bootstrap values >70% with an * and bolded text. Haplotypes that are present in at least one individual from Utila are shown in green.

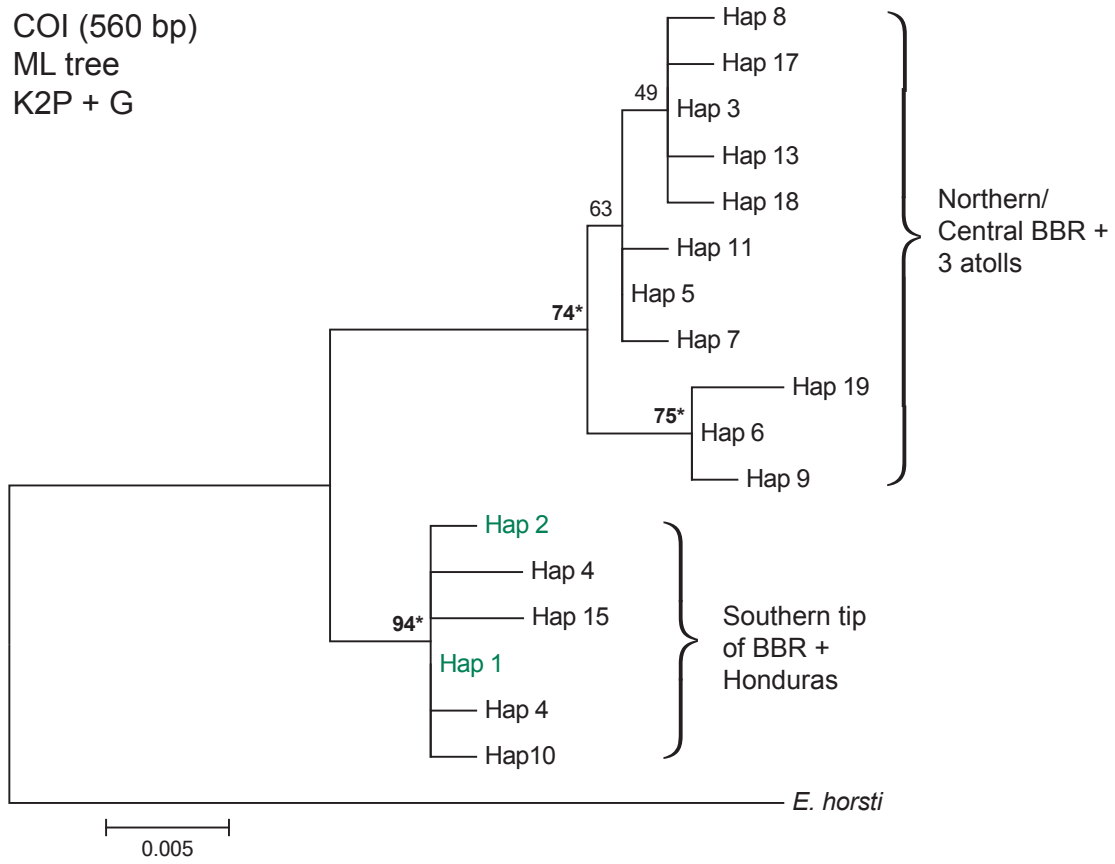

Figure 6.5: Maximum likelihood tree of *E. lori* haplotypes based on 560-bp of *COI*. We used BIC to identify the Kimura 2-parameter model with a gamma distribution as the most appropriate nucleotide substitution model (K2P + G). We used a discrete Gamma distribution (parameter=0.05). Bootstrap percentages are shown at nodes (n=1,000 bootstrap replicates). We represent 'significant' bootstrap values >70% with an * and bolded text. Haplotypes that are present in at least one individual from Utila are shown in green.

6.4.2 Micro-scale patterns: SGS within Belize

6.4.2.1 Genetic clustering analysis

K-means clustering based on 200 retained PCs indicated that there were k=3 genetic clusters in the Belizean metapopulation, based on anonymous and microsatellite allele frequencies (Fig. 6.6). The alpha-optimization test revealed that retaining 59 PCs for the

discriminant analysis optimized the trade-off between the power to discriminate (i.e. the power to actually distinguish between groups) and overfitting (i.e. retaining too many PCs, such that the model can discriminate any possible combination of clusters, but with low power to predict new observations) (Fig. 6.7).

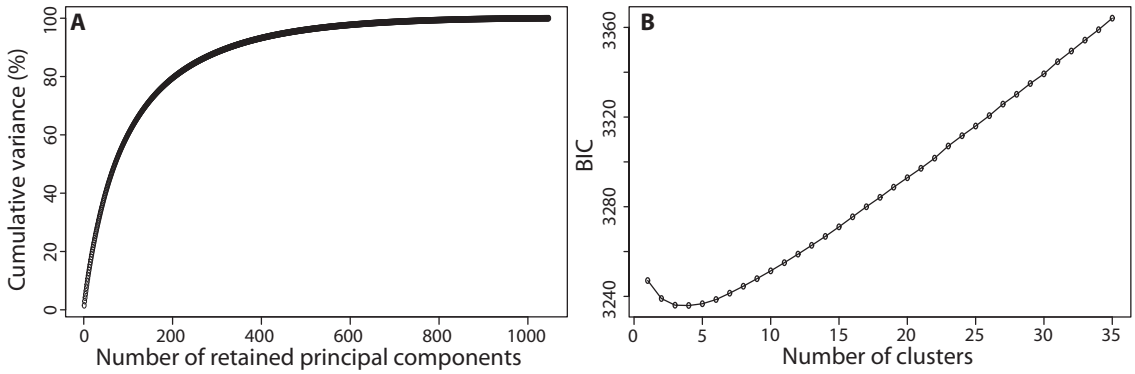

Figure 6.6: Preliminary clustering analysis for DAPC: (A) Variance explained by principal components for anonymous and microsatellite markers. The curve begins to asymptote around 200 PCs. (B) BIC versus number of clusters. Here, BIC is minimized at k=3 clusters.

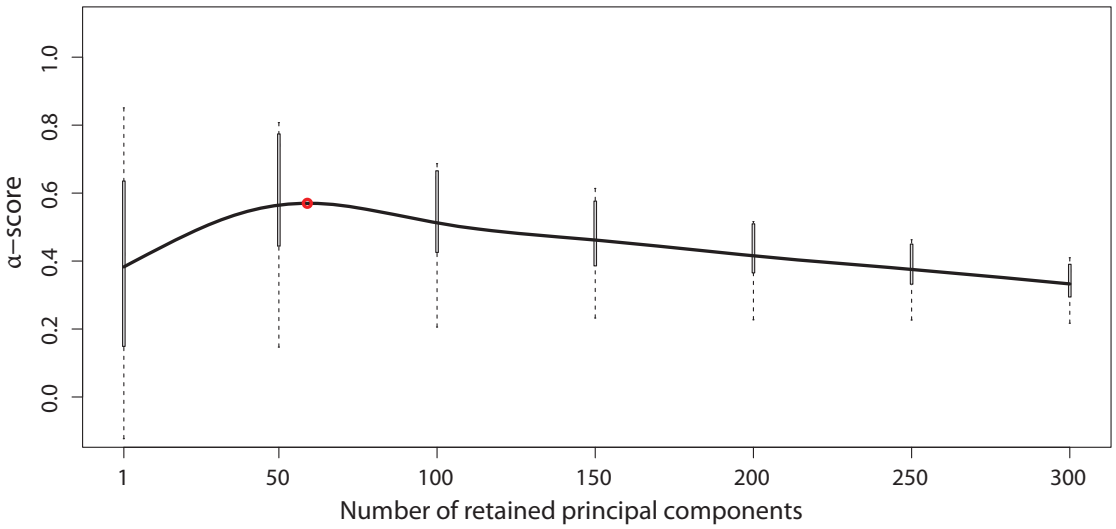

Figure 6.7: Alpha optimization to determine the optimal number of principal components to retain for the DAPC analysis. Using spline interpolation, we select a maximum α at n = 59 PCs, represented by the red circle; thus, 59 PCs are retained for the DAPC analysis to optimize the trade-off between the power of discrimination and overfitting.

The resulting DAPC analysis reveals that the three clusters are close in two dimensional space on a Cartesian plane (Fig. 6.8). This indicates that although the model is assigning individuals to clusters, the clusters themselves are not very different, consistent with estimates of low pairwise differentiation (see next section). To determine whether there is a spatial pattern associated with these three clusters, we plotted the frequency of individual assignments per cluster at each sampling location (Fig. 6.9). We see that the number of individuals assigned to cluster 3 (green) is relatively stable across sites, but sites on the barrier reef and Turneffe atoll have relatively more individuals assigned to cluster 2 (red), while sites on Glovers and Lighthouse atolls have relatively more individuals assigned to cluster 1 (blue). However, contrary to *structure* analyses from 2012 data (D'Aloia *et al.* 2014), Glovers and Lighthouse are not qualitatively distinguishable from each other.

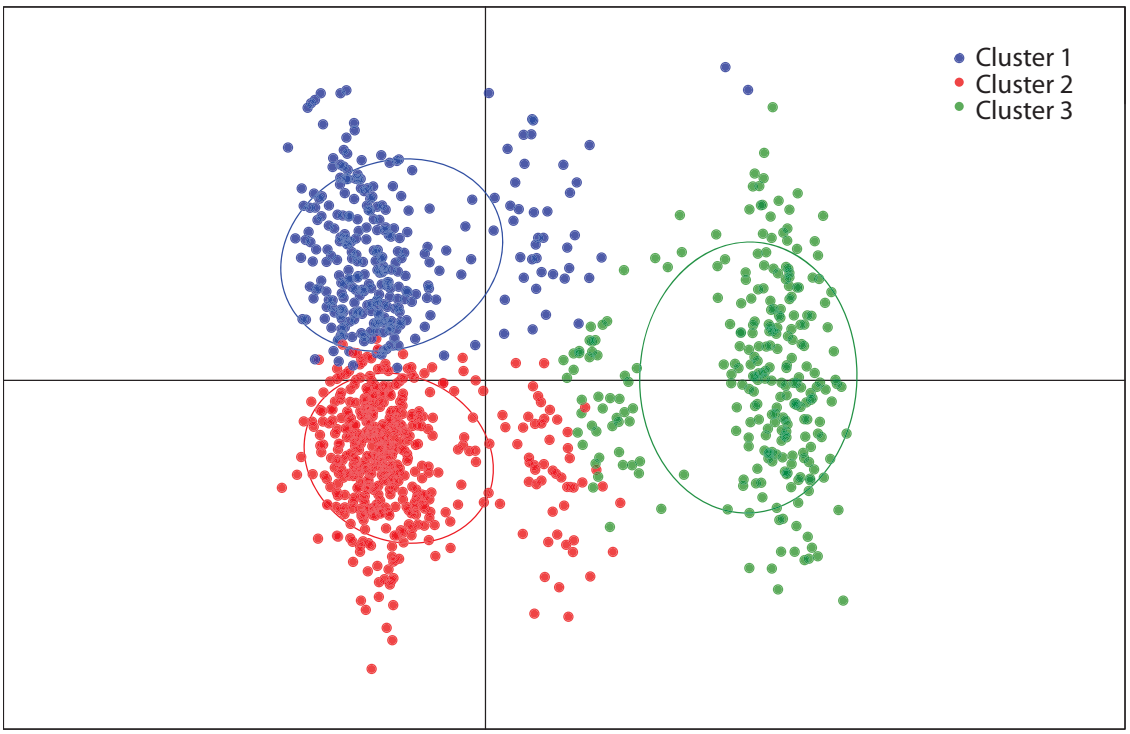

Figure 6.8: DAPC on k=3 clusters based on anonymous and microsatellite loci. Dots represent individuals and ellipses are drawn to encompass two thirds of the individuals in each cluster. We see that all three clusters are close in 2-dimensional space.

6.4.2.2 Pairwise genetic differentiation

We observed a low average pairwise F_{ST} (mean = 0.01 $\pm$ 0.009). Moreover, the range indicates that no sites exhibited high pairwise values (minimum = -0.004; maximum = 0.038). Here, F_{ST} is based on the 59 anonymous loci; however, the pattern was consistent across the microsatellite markers and cytb (data not shown). We visualized pairwise F_{ST} values with a heatmap to facilitate interpretation of the 35 x 35 matrix (Fig. 6.10). Consistent with D'Aloia *et al.* (2014), pairwise differentiation is higher for comparisons between Glovers Atoll, Lighthouse Atoll, and all other reef sites, indicated by the orange and red tiles on the heatmap.

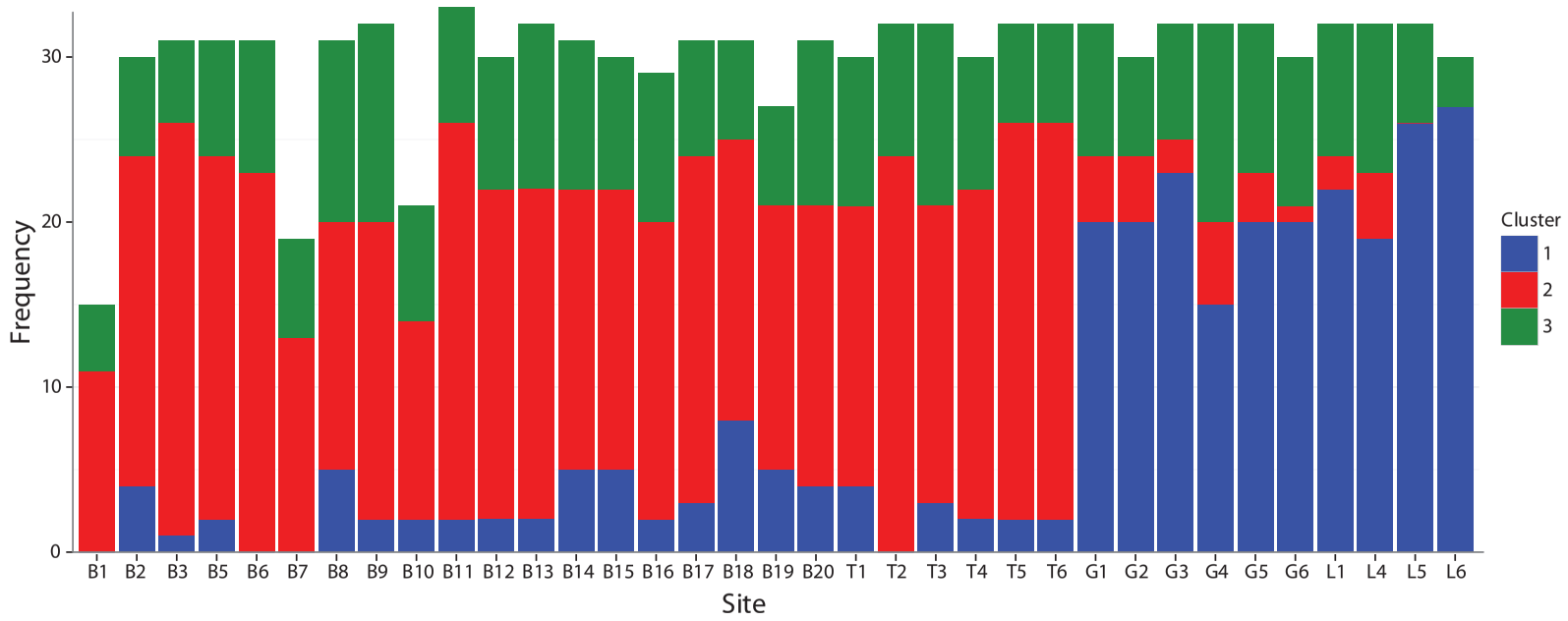

Figure 6.9: Stacked bar graph representing individual assignments to k=3 genetic clusters based on DAPC. Individuals are grouped by sampling location to investigate spatial patterns. The x-axis represents each site and the-y axis represents the frequency of individuals at each site, assigned to one of three clusters.

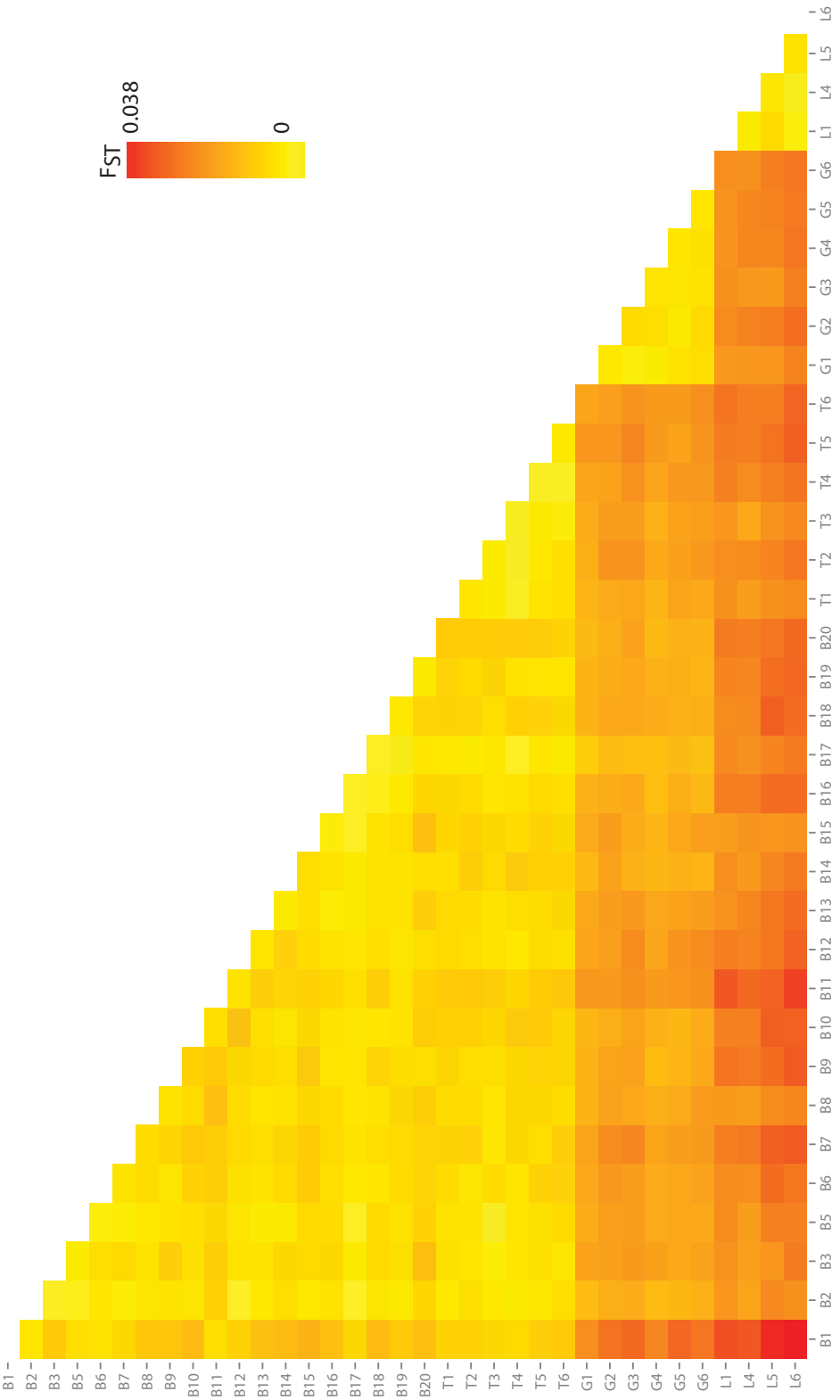

Figure 6.10: Heatmap of pairwise F_{ST} values based on 59 anonymous loci

6.4.2.3 Seascape genetic analysis

Backward and forward selection revealed that a MMRR model with Euclidean distance and the presence/absence of an oceanographic barrier $\geq$ 20 km explains a large amount of variance in pairwise genetic distance (R^2 = 0.782; F-statistic = 1067.08; p < 0.001). This model was primarily driven by the oceanographic barrier dummy variable (β_{BARRIER} = 0.868), with an additional weak association between Euclidean distance and genetic distance (β_{DISTANCE} = 0.083) (Table 6.2; Fig. 6.11). Adding the connectivity matrix and environmental distance to the model had no substantive effect on the explanatory power (R^2 = 0.790), and neither predictor variable was significantly associated with genetic distance between sites (β_{CM} = -0.05, p = 0.17; $\beta_{\text{ENVIRONMENT}}$ = -0.08, p = 0.12).

Table 6.2: Output from best-fit MMRR model. The model with standardized Euclidean distance and the presence/absence of an oceanographic barrier as predictors was largely driven by the dummy variable 'barrier'.

Parameter	**Coefficient**	**t-statistic**	**p-value**
Intercept	0.006	0.337	1.00
Euclidean distance	0.083	4.20	0.04
Barrier (1/0)	0.868	45.469	<0.001

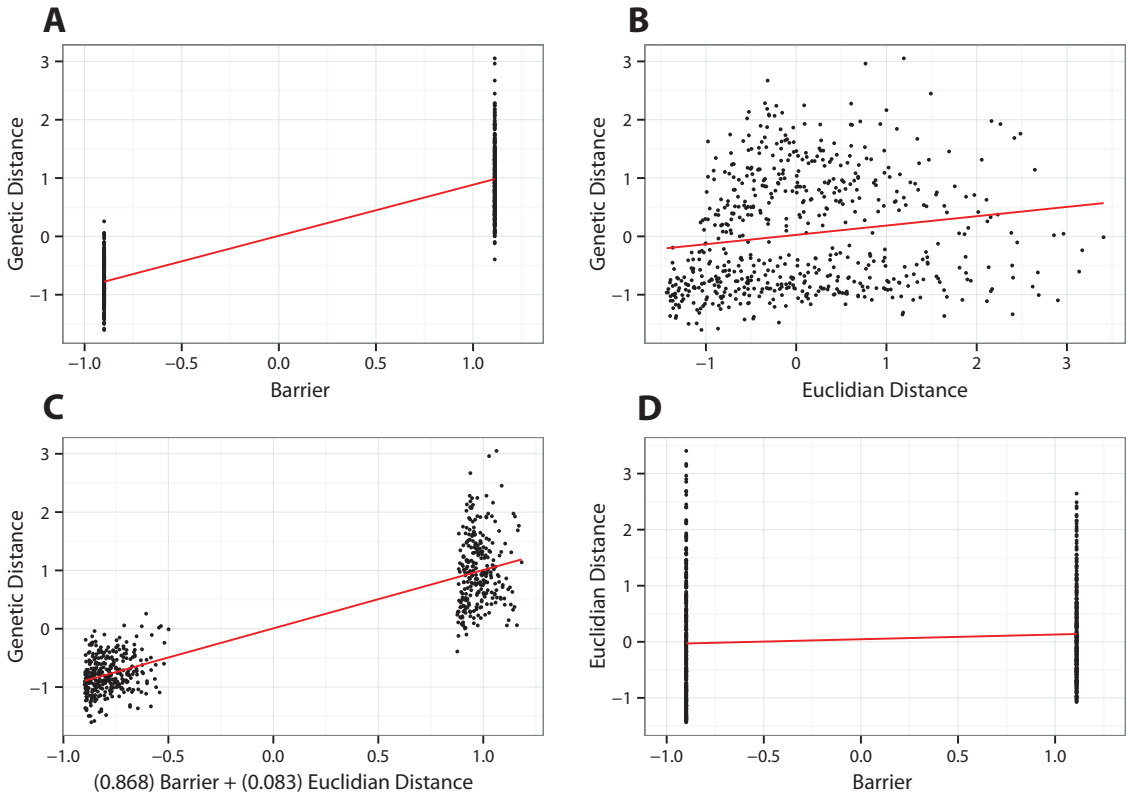

Figure 6.11: MMRR scatterplots using standardized F_{ST} as genetic distance. (A) Correlation between 20-km ocean barrier and genetic distance; (B) weak correlation between Euclidean distance and genetic distance; (C) combined effects of barrier and Euclidean distance on genetic distance; and (D) no correlation between the two predictor variables. All matrices were standardized for MMRR.

These MMRR results were congruent with individual Mantel tests that investigated the correlation between each predictor matrix and the genetic distance response matrix in isolation. The oceanographic barrier was strongly correlated with genetic distance, Euclidean distance was weakly correlated with genetic distance, and neither the connectivity matrix nor environmental distance was correlated with genetic distance (Table 6.3).

As noted by D'Aloia *et al.* (2014), the relatively linear shape of the BBR renders it an ideal study area for the test of the classic IBD hypothesis. Therefore, we conducted reduced Mantel tests, focusing exclusively on the Euclidean distance between the 19 sites

Table 6.3: Mantel tests for correlation between each individual predictor matrix (Xp) and the genetic distance matrix (Y). Here, Y is based on F_{ST} from n=59 anonymous loci. We used both Pearson's product-moment correlation and Spearman's rank correlation with 1,000 permutations to assess significance at an alpha-value of 0.05.

Xp	**Pearson's ρ**	**p-value**	**Spearman's ρ**	**p-value**
Barrier ($\geq$ 20 km)	0.881	0.001	0.862	0.001
Euclidean distance	0.157	0.045	0.223	0.003
Connectivity matrix	-0.150	0.999	-0.343	0.998
Environment	-0.1498	0.939	-0.1072	0.919

on the Northern and Central BBR (i.e., excluding the atoll reefs and Southern BBR). We found that when we focused on this 1-dimensional section of the habitat, there was a slightly stronger correlation between Euclidean distance and genetic distance (Pearson's ρ = 0.2789, p = 0.004; Spearman's ρ = 0.2479, p = 0.005). Overall, however, these results suggest that Euclidean distance is not a strong predictor of genetic connectivity between sites.

6.5 Discussion

Seascape genetic studies can provide insight into how dispersal patterns, habitat configuration, oceanographic features, and local environmental conditions interact to influence long-term patterns of genetic connectivity. In the context of this dissertation, this final chapter provides an opportunity to compare detailed demographic connectivity data to broad-scale, population-level genetic connectivity data. This study builds upon the pilot SGS study in *E. lori* (D'Aloia *et al.* 2014) with extensive geographic sampling and nearly an order of magnitude increase in the number of loci. We found consistent statistical results, where the presence of large oceanographic barriers between habitat patches is strongly associated with pairwise differentiation. However, this analysis remains preliminary, and the most exciting aspects of this analysis have yet to be integrated—mainly, oceanographic predictions and the projection of the empirical connectivity matrix over hundreds of gener-

ations. Surprisingly, we also found that there are two distinct clades of *E. lori* within the MAR. Here, we explore the main findings to date and highlight future research directions that will ultimately round-out two manuscripts: (1) a complete seascape genetic analysis of the Belizean metapopulation and (2) an exploration of the two distinct genetic lineages of *E. lori*.

Seascape analysis

Consistent with the pilot study, we found that a dummy variable representing oceanographic breaks $\geq$20 km between reef habitat patches captures most of the variation in pairwise genetic differentiation. A relevant question becomes, why is the dummy variable so strong, when this study also includes the empirical connectivity matrix (CM)? Intuitively, we would expect that the CM, which is based on systematic sampling of actual dispersal events along the BBR (Chapter 4), should be a more accurate predictor of dispersal and gene flow between sites. However, the CM included in this study is based on a single temporal snapshot of predicted connectivity, i.e., dispersal in a single generation, and therefore does not account for the inter-generational stepping-stone movement of alleles (Kimura & Weiss 1964). Moreover, we now know that the dummy variable, which was created prior to the large-scale dispersal study, is a rough proxy for the tail of the dispersal kernel, which begins at 15 km from source (Chapter 4; Fig. 4.2). In order to improve upon this coarse estimate of the kernel's tail, future work must use the CM to simulate dispersal and gene flow over hundreds of generations (e.g. Kool *et al.* 2010). Over time, the symmetry of the CM (Chapter 4; Fig. 4.4) should lead to differentiation between sites separated by at least 15 km of unsuitable habitat (i.e., the tail of the dispersal kernel), but a lack of structure between sites on continuous reef. Theoretically, this process should capture the variation currently accounted for by the dummy variable. However, we have extrapolated the model used to generate the CM over space and time; thus, the CM may not be a stronger predictor of pairwise differentiation if there is substantial spatial and/or temporal variation in dispersal. Regardless, we again show strong support for the role that

habitat continuity plays in driving patterns of SGS in the Belizean *E. lori* metapopulation.

We also found weak but significant support for an isolation by distance (IBD) hypothesis after controlling for habitat discontinuity barriers (Fig. 6.11), suggesting that within continuous regions of the reef, gene flow may decline slightly as geographic distance between sites increases. Results from the pilot study of SGS in *E. lori* demonstrated no statistical support for IBD; however, due to a potential cline in microsatellite allele frequencies along the barrier reef, it was suggested that more extensive spatial sampling may reveal an IBD pattern (D'Aloia *et al.* 2014). Indeed, the expanded sampling scheme in the present study revealed a subtle IBD pattern. Contrary to expectations, however, Mantel tests that focused exclusively on sites along the linear barrier reef did not show a substantially stronger IBD pattern than the entire region. Overall, this weak effect of geographic distance, relative to habitat continuity, aligns with evidence in other marine and terrestrial taxa. While IBD was long the prevailing spatial hypothesis in empirical population genetics, newer multivariate analyses often suggest that genetic data may be more strongly correlated with alternative predictor variables, which are themselves spatially auto-correlated (Selkoe *et al.* 2010; Wang & Bradburd 2014). In other words, the IBD signal may reflect spatial autocorrelation of alternative oceanographic and/or environmental variables.

Overall these results offer consistent findings regarding the role of the habitat continuity in driving emergent patterns of SGS, but the seascape analyses presented in this chapter must be expanded in several ways. First, regarding statistical analysis of the data, the robustness of the findings needs to be evaluated. One important test is to determine whether the results hold when alternative genetic distance metrics are used. We will also conduct redundancy analyses—an alternative multivariate approach for link-level data—as a secondary robustness check (Legendre & Legendre 2012). Finally, we will explore whether the MMRR framework can be extended to account for response variables with non-Gaussian distributions. Pairwise genetic distance variables tend to have positively-skewed distributions in marine species, with many observations tending to zero; thus, an important next-step for the application of MMRR to seascape genetic data should be to

account for non-linear models (Wang 2013). Taken together, these analyses will provide a more rigorous and comprehensive approach to data analysis.

Another future research direction is the incorporation of more data into the GIS analyses to test the isolation by environment (IBE) hypothesis. Our preliminary test included data on contemporary measures of sea surface temperature and salinity; however, a potential issue in seascape genetics is temporal discordance between *contemporary* predictor variables and measures of genetic structure, the latter of which may be the product of thousands of years of evolutionary forces (Riginos & Liggins 2013). To this end, we will also test estimated historical measures of temperature and salinity from the last glacial maximum (∼21 kya) and mid-Holocene (∼6 kya) (Sbrocco 2014). Genetic structure may also correlate with alternative environmental variables. For example, a recent study found that genetic structure in the anemonefish *Amphiprion bicinctus* followed a strong IBE pattern along a chlorophyll-a gradient within the Red Sea (Nanninga *et al.* 2014). While high-resolution environmental data layers are scarce for Belize, there are some 'expert-mapped' layers available that document certain features, such as elevated pollution from river run-off in Southern Belize (World Resources Institute 2005). Adding these additional variables will strengthen the test of IBE in the *E. lori* metapopulation.

While the abovementioned issues all represent important future research directions, the biggest gap in the present study is an integration of oceanographic predictions with genetic data—this integration is central to any comprehensive seascape genetic analysis. Indeed, the first studies that coined the term 'seascape genetics' explicitly focused on comparing oceanographic and genetic data (Galindo *et al.* 2006; Selkoe *et al.* 2006); the magnitude and direction of ocean currents have traditionally been viewed as the driving force of marine larval dispersal and gene flow. Since these initial studies, oceanography-oriented seascape genetic studies have become increasingly common, with most studies employing a coarse grid size for the oceanographic model. For example, White *et al.* (2010) derived oceanographic distances between sites based on an ocean flow-field model at a 5 km grid. Foster *et al.* (2012) predicted connectivity across the Caribbean by coupling a genetic

projection model to the global HYCOM ocean model, which operates at a 7 km grid. In our study, we will use a high-resolution model (Belize-HYCOM) at a 1 km grid, validated with in-situ drifters from summer 2014 (Lindo-Atichati, unpublished). This model can predict (1) pure physical connectivity via the movement of water parcels between sites and (2) simulated demographic connectivity by integrating the Belize-HYCOM with particle release simulations that include larval vertical migration, pelagic larval duration, and other biological attributes (Paris *et al.* 2013). These high-resolution oceanographic predictions will enable us to test whether oceanographic distance is a better predictor of fine-scale genetic structure in Belize, compared to the strong observed role of habitat continuity.

Genetic divergence in *E. lori*

The results from all analyses of the genetic data (haplotype networks, phylogenetic trees, DAPC, and pairwise structure) provided clear evidence for a second clade of *E. lori* in Southern Belize. Genetic isolation at this spatial scale is somewhat surprising, although new research is revealing cryptic genetic diversity in coral reef fishes (Victor 2015). Still, this finding was unexpected, because the shortest geographic distance between these two clades is only 30 km (site B20 to site B22). All available GIS layers indicate that there is continuous barrier reef habitat running between these sites, from Gladden Spit down to the Sapodilla Cayes (Fig. 6.1; Andréfouët *et al.* 2005; World Resources Institute 2005); however, we know that the reef habitat along this stretch is not suitable for *E. lori*. The reef gradually slopes off, without a clear drop-off on the outer fore reef; the benthic habitat is noticeably sandy and rocky, with a lack of live coral and sponge cover (D'Aloia & Buston, personal observation). Interestingly, the genetic break in *E. lori* does not occur at Gladden Spit itself. Preliminary oceanographic modeling suggests that there may be an East-West barrier to dispersal at Gladden (Lindo-Atichati, personal communication). Instead, the break occurs just South of Gladden (i.e., after site B20), where suitable *E. lori* habitat disappears. This underscores the importance of habitat continuity for this short-distance disperser within the Belizean metapopulation.

However, these results also indicate that the simple spacing patterns that are associated with genetic structure in the primary clade within Belize do not hold across the species' range. Although 20 km breaks in habitat are associated with weak but significant pairwise differentiation within most of Belize, there is evidence for long-term genetic isolation between the Northern/Central/Atoll clade and the Southern clade, which are separated by only 30 km. Moreover, individuals from Southern Belize share haplotypes with individuals from Utila, Honduras, which is >140 km away.

What, then, could be driving these patterns of genetic connectivity? One potential explanation for the observed Honduras-Southern Belize connectivity is dispersal via the Honduran Gyre. In a Caribbean-wide study based on the global HYCOM oceanographic model, Foster *et al.* (2012) predicted high connectivity for the coral *Montastraea annularis* between Honduras and Belize due to this gyre. Regarding the lack of connectivity between Southern Belize and the rest of Belize, one hypothesis is that the frequency of gene flow to this region is comparable to long-distance gene flow between the barrier reef and distant atolls, but that individuals are locally adapted to the Southern reef environment. Therefore, long-distance migrants may have reduced fitness in the South. We noticed qualitative habitat differences in the Southern region, compared to the rest of Belize (D'Aloia & Buston, personal observation). The water was noticeably murkier, in agreement with mapped pollution from river run-off (World Resources Institute 2005), and the sponge assemblage was markedly different. There were almost no *A. fistularis* present, and most *E. lori* had subsequently shifted hosts to the brown tube sponge *Agelas conifera*. However, these hypotheses remain speculative at this time, and future research is needed to explore the mechanism(s) behind these patterns of genetic connectivity.

Conclusions and future directions

This chapter lays the foundation for two forthcoming manuscripts. The first will be an integrative seascape genetic study that adds oceanographic predictions, the simulated empirical connectivity matrix, and additional local environmental variables to the anal-

yses. Together, these data will allow us to test the full suite of alternative seascape genetic hypotheses. Thus far, this chapter presents strong evidence for the IBB hypothesis, strengthening the original finding that habitat discontinuity (here, defined as oceanic barriers between reef patches $\geq$20 km) predicts pairwise genetic differentiation within the Belizean metapopulation (D'Aloia *et al.* 2014). This result underscores the important role that habitat configuration can play in driving patterns of dispersal and gene flow, even for a fish with a dispersive larval phase approximating one month. The second manuscript will investigate the unexpected but clear finding that there are two distinct genetic lineages of *E. lori* within the Mesoamerican reef. This manuscript will test the timing of divergence and the potential for species delimitation. Taken together, these manuscripts will explore the intrinsic and extrinsic barriers to gene flow in *E. lori*. They will generate novel insights into how the pattern of dispersal, documented in this dissertation, interacts with the seascape and the dynamic ocean environment to create genetic structure within metapopulations, and genetic divergence between metapopulations, at small spatial scales.

6.6 Acknowledgments

We thank Diana Acosta, Alben David, and Derek Scolaro for assistance with field work. We are also grateful to members of the Harrison Lab at Cornell University for feedback on lab work. This study was supported by NSF (OCE-1260424) and CCD was supported by a NSF GRF (Grant No. DGE-1247312). Research was approved by Belize Fisheries and the Boston University IACUC.

6.7 Author Contributions

CCD planned the study. CCD and PMB conducted field work with assistance from others. SMB and RGH planned laboratory work, and CCD and SMB conducted laboratory work. CCD analyzed the data and wrote the chapter.

Chapter 7

Conclusions and future research directions

Contributions of the dissertation

This dissertation is the first research project that integrates direct dispersal data with population-level genetic data for a marine species. These two components of the dissertation represent demographic and genetic connectivity, respectively. After describing the basic population ecology of the focal species, and thereby laying the foundation for using *E. lori* as a model study organism for marine population connectivity (Chapter 2), I gradually scale-up the demographic and genetic connectivity analyses. After conducting a detailed intra-population study of dispersal (Chapter 3), I expand the scope of the dispersal project by systematically sampling dispersal at a regional scale (Chapter 4). Likewise, after completing a pilot study of spatial genetic structure within a *portion* of the Belizean metapopulation (Chapter 5), I conduct a complete seascape genetic analysis across the *entire* metapopulation (Chapter 6). Taken together, these chapters describe patterns, predictors, and consequences of demographic and genetic connectivity in a coral reef fish.

This research project has made novel contributions to the field of marine ecology, as discussed within each chapter. Here, I briefly note a few salient contributions. First, I have shown that marine larval dispersal can be surprisingly spatially-restricted, even when a species is distributed along continuous habitat and has a larval phase lasting several weeks. This finding, primarily discussed in Chapter 4, challenges long-standing assumptions related to the predictors of marine dispersal. Second, data collected in Chapters 3, 4, 5, and 6 suggest that the pattern of connectivity is relatively stable over space and time,

challenging conventional wisdom that marine connectivity is stochastic (Siegel *et al.* 2008). Specifically, I found that the functional form of the estimated dispersal kernel was consistent across years and study regions (Chapters 3 and 4), and that the complete dispersal kernel was congruent with patterns of spatial genetic structure in Belize (Chapters 4, 5, and 6). Third, I discovered that the seascape itself is an incredibly powerful predictor of demographic and genetic connectivity for *E. lori*. Given the leptokurtic pattern of dispersal, habitat continuity plays an important role in driving demographic and genetic exchange, as migrants are unlikely to cross open habitat 'barriers' greater than 15 km (i.e., where the tail of the dispersal kernel begins). While it is difficult to assess the generalizability of this finding without empirical data in other species, I hypothesize that the seascape could play a similar role for other spatially-restricted benthic species with shared dispersal traits. Additionally, there is some evidence that seascape continuity can drive recruitment patterns across taxa (Pinsky *et al.* 2012) via the dilution effect described in Chapter 3. Thus, the distance between stretches of continuous habitat may be a simple but effective metric for spacing marine reserves. Collectively, these contributions generate new ideas that expand our conceptual understanding of marine connectivity, and provide useful data for marine conservation work.

Future directions

Here, I explore three subject areas with opportunity for interesting future research projects. First, I highlight remaining topics related to our understanding of population connectivity in *E. lori*. Second, I outline important next steps for studying population connectivity in other marine taxa. Third, I briefly describe interesting side-projects that have arisen from the large amount of genetic data collected in this dissertation.

Population connectivity in E. lori

Several lines of future research will enhance our understanding of population connectivity in *E. lori*. The first priority is to integrate the empirical dispersal and genetic structure

data with predictions from biophysical models. This integrative approach will allow us to link phenomenological and mechanistic approaches to studying connectivity. While both approaches are time-intensive, biophysical modeling has the potential to be more widely implemented as species-specific parameters can be adjusted once the oceanographic model is developed for a region. In Belize, we will compare predictions from a high-resolution bio-physical model to the observed dispersal kernel and SGS data. This work will enable us to identify key biological inputs to the bio-physical model that have the biggest effect on linking these two approaches. In turn, knowing these key inputs may be useful to the management community; it may guide future research for target species. An essential component to this integration will be experimental work on larval orientation and swimming behavior that will inform model parameterization.

Ultimately, these data must be shared with government agencies, research organizations, and conservation non-profits operating in Belize in order to incorporate connectivity data into actual MPA design. Relevant organizations include Belize Fisheries, Belize Coastal Zone Management Authority and Institute, University of Belize's Environmental Research Institute, Glover's Reef Research Station (Wildlife Conservation Society), Carrie Bow Cay Field Station (Smithsonian Institute), Healthy Reefs for Healthy People Initiative, and more.

Population connectivity in other marine systems

While the data generated in this dissertation have addressed long-standing questions in marine ecology, the techniques employed must be carefully and thoughtfully applied to other species. The generalizability of any single-species study is always questioned, and therefore there is much interest in exploring whether similar patterns emerge across taxa. For example, in connectivity research, there is particular interest in applying these methods to species with commercial value (i.e. Almany *et al.* 2013). In Belize, grouper represent potentially-tractable study species due to their site-specific spawning aggregations. However, much caution should be used in planning future dispersal studies; a careful analysis

would be required to assess whether the additional information gained would outweigh the tissue sampling, financial cost, and time required. Given the state of the world's coral reef ecosystems, I do not advocate the widespread implementation of direct dispersal studies.

Instead, I suggest that moving forward, dispersal studies should be targeted to focal species that represent a range of life history traits, distributed across distinct seascapes and ocean regions. Understanding the patterns and drivers of connectivity for a broader spectrum of species will strengthen our understanding of how the seascape, the environment, and species-specific traits interact to affect realized patterns. In turn, any emerging trends could enable us to develop simple rules of thumb for conservation purposes. For example, if the tail of most dispersal kernels begins at $\leq$ 20 km from source, a simple rule would be to space marine reserves such that they are $\leq$ 20 km apart.

There has also been a call from theoreticians for empiricists to obtain empirical measures of population density and individual reproductive success in tandem with dispersal data (Burgess *et al.* 2014). While empiricists tend to focus on self-recruitment (Chapter 3), modelers are interested in a different metric—local retention (# individuals retained locally / total # of individuals produced locally). The abundance and fecundity data associated with measuring retention would be very challenging to collect, but perhaps a few target species, which have already been the subject of intense dispersal studies, could be suitable candidates (i.e. *E. lori* and *A. percula*). Collecting these data would enable modelers to parameterize population persistence models. Persistence models are key to predicting metapopulation dynamics and MPA network dynamics under alternative climate and fishing scenarios.

Other extensions of the dissertation

The large population genetic data sets collected for this dissertation have the potential to answer interesting research questions beyond the framework of marine population connectivity. Most immediately, I will be exploring the strong evidence for a unique genetic lineage of *E. lori* in Southern Belize and Honduras, uncovered in Chapter 6. Despite

relatively close geographic proximity, I identified a distinct clade of *E. lori*, suggestive of long-term genetic isolation from the rest of the Belizean metapopulation. This research project will investigate the timing of divergence, evaluate evidence for species delimitation, and explore patterns of introgression in the boundary region.

A second research project will be an analysis of genetic relatedness within a single population of *E. lori*. Recent research has focused on the degree to which larval cohorts are composed of siblings, and, ultimately, how that affects the spatial distribution of related individuals. An early assumption of marine ecology was that dynamic oceanographic conditions and cohort mixing would disrupt kin associations in coral reef fish with a pelagic larval stage (Victor 1984; Leis 1991). However, the empirical evidence has been mixed; several studies have documented associations between close relatives at the larval (Planes *et al.* 2002), recruitment (Selkoe *et al.* 2006; Ben-Tzvi *et al.* 2012), and post-recruitment (Buston *et al.* 2009; Berry *et al.* 2012) stages, while others have found no genetic relatedness among recruits (Avise & Shapiro 1986) or post-recruitment group members (Buston *et al.* 2007). *E. lori* is an interesting study organism for analyses of post-settlement relatedness, given its incredibly restricted spatial pattern of dispersal. However, preliminary analyses show that (i) there is no evidence for relatedness at the sponge 'group' level, and (ii) there is no evidence for spatial autocorrelation of relatedness within the population (D'Aloia, unpublished).

Finally, in a third project I will explore my interest in indirect genetic estimates of dispersal. Given the substantial amount of time, money, and sampling required for direct studies of dispersal, population genetic data have been used for decades to infer dispersal patterns, despite theoretical criticisms (Whitlock & McCauley 1999; Neigel 2002). Briefly, under the isolation by distance (IBD) model, the slope b of the linear regression of linearized F_{ST} over geographic distance is assumed to be inversely proportional to Wright's neighborhood size (Wright 1943; Rousset 1997):

$$b = \frac{1}{4N_e\pi\sigma^2},$$

where σ^2 is the second moment of the dispersal kernel, or the variance of dispersal distances. From this we can also derive σ, the first moment of the dispersal kernel, also interpreted as the spread or standard deviation of the dispersal distribution. The data collected in this dissertation provide the first opportunity to test the validity of this assumption in the marine environment. I will use the population-level genetic data collected on the BBR (Chapter 6) to generate estimates of the moments of the dispersal kernel from the IBD approach, which in turn can be compared to moments of the empirical dispersal kernel for *E. lori* (Chapter 4). Additionally, we can test alternative metrics of 'distance' within the IBD framework by considering both geographic distance and oceanographic distance, estimated by collaborators at the University of Miami. Thus, these data may be used to conduct the first rigorous test of the utility of indirectly estimating dispersal in a marine species.

Appendix A

Supplement for Chapter 4

Additional Methods

Parentage analysis criteria. In addition to having an LOD score exceeding the critical value, all potential parent-offspring assignments were inspected manually and were only accepted if the following three additional criteria were met: 1) the parent-offspring pair was compared at a minimum of 15 loci, 2) the pair mismatched at only 0, 1, or 2 loci, and 3) at each mismatched locus, both individuals were homozygotes. This last rule allows for the possibility of true null alleles, but excludes putative nulls that amplified well in other individuals.

Sampling correction of dispersal kernel. The design of the 41-km linear transect led to unequal sampling at different distances. Specifically, there were more potential short distance dispersal trajectories than long-distance dispersal trajectories. To assess whether this unequal sampling strategy affected our estimate of the empirical dispersal kernel, we applied a sampling correction. First, we quantified unequal sampling. We began by considering all potential dispersal trajectories by comparing the midpoint of every parent collection dive (n=64) to the midpoint of every offspring collection dive (n=69) and generating a distance matrix using ArcMap 10.1 (ESRI), for a total of n=4,416 potential trajectories. Second, we pooled these 4,416 potential dispersal trajectories into 1-km bins, and used the number of potential trajectories per 1-km bin as a metric of sampling effort at that distance class. Third, considering 1-km bins out to 17 km (i.e., the point beyond

which we did not identify a dispersal event), we identified the bin with the smallest number of potential trajectories, and took that bin (bin = 15-16 km) as the distance bin with the minimum sampling effort (effort = 128 potential trajectories). Fourth, based on this minimum sampling effort, we wrote an R script to estimate iteratively the dispersal kernel while controlling for unequal sampling in the following steps: i) to sample evenly across bins, 128 trajectories (i.e., the minimum sampling effort) were randomly drawn without replacement from each distance bin; ii) within this randomly-drawn subset of the data, the observed dispersal trajectories based on parentage-offspring assignments were pulled out of all potential dispersal trajectories and their associated distances were used to build a vector *di* of net dispersal distances; iii) the decay rate, λ, for the exponential probability density function, was estimated for *di* using maximum likelihood. This process was repeated for 10,000 iterations (from i = 1 to i = 10,000), with the average parameter estimate taken as the best estimate for the sampling-corrected dispersal kernel ($\lambda = 0.37$; $CI_{95\%} = 0.29$, 0.45). We found that this sampling correction did not affect the kernel estimate.

Additional otolith reads. Because the otolith data collection process spanned several months, otoliths read at the beginning of the study were re-read to control for potential technique refinement. For all settlers, if the first two reads were inconsistent, then the otolith was read a third and fourth time. If the difference between the first two reads was less than or equal to three rings, the same otolith was read; otherwise, the second extracted otolith was read. If the difference between the averages of reads one and two and reads three and four exceeded five rings, then individuals were read a fifth and sixth time. The final pair of readings was averaged, taken as the best-estimate for PLD, and used in subsequent analyses.

Multivariate analysis of predictors of dispersal distance. To consider all the measured biological and spatial variables in a multivariate analysis, we constructed a set of generalized linear models (distribution = Gamma; dispersion = 1; link = log) with dis-

persal distance (km) as the dependent variable. These analyses investigate the relationship between the covariates and the mean and variance of dispersal distance, rather than the functional form of the dispersal kernel. We investigated all bivariate analyses, as well as biologically-meaningful groups of covariates: biological predictors consisting of fish and microhabitat characteristics (coded B) and spatial predictors (coded S).

Model selection. Model selection was based on ΔAIC_c values. To exclude uninformative variables, more complex models were rejected if 1) models with an additional variable had a ΔAIC_c score within 2 units of a simpler model without the variable (Burnham & Anderson 2002); and 2) adding a variable did not result in a substantial increase in the log-likelihood. When the top candidate models were nested, likelihood-ratio tests were also conducted.

Building connectivity matrices. We visualized demographic connectivity through the raw connectivity matrix. Nodes were defined as all potential source sites (N_j) versus all potential settlement sites (N_i), where j = i = 516 sites. Edge values were the relative probabilities of dispersal, visualized by color intensity in grayscale. We considered edge values of less than 0.001 as a negligible level of dispersal, and color-coded all of these edges as white. To explore the evolutionary consequences of dispersal, we then used the predicted levels of demographic exchange to identify regions of the reef that may experience reduced levels of gene flow at neutral markers. Here, we assume that if nodes from different geographic regions of the reef (e.g. Barrier reef, Turneffe Atoll, Glovers Atoll, Lighthouse Atoll) are predicted to be demographically connected, then subsequent stepping-stone dispersal within regions will lead to genetic connectivity at neutral markers over many generations. We drew colored boxes around the putative evolutionarily significant units (ESUs) and overlaid them onto the connectivity matrix. Finally, to gain insight into the conservation implications, we focused on connectivity between nodes lying within the boundaries of real marine protected areas (MPAs) in Belize. We color-coded connections between MPA nodes

to represent self-recruitment within a reserve (green), some connectivity predicted between reserves (purple), or no connectivity predicted between reserves (red), and overlaid these colors onto the connectivity matrix.

Tables and Figures

Table A.1: Number of tissue samples collected along the transect. (A) Parent samples collected at each region on the transect, spaced approximately 10 km apart (n=3 regions). (B) Offspring samples collected at each collection site along the length of the transect, spaced approximately 1 km apart (n=41 sites). Offspring collection sites are labeled from 1 (Northernmost site) to 41 (Southernmost site) based on latitude.

A) Parent tissue		***B) Offspring tissue***							
Region	**Samples**	**Site**	**Samples**	**Site**	**Samples**	**Site**	**Samples**	**Site**	**Samples**
Northern	1024	1	109	11	77	21	106	31	89
Central	1003	2	121	12	105	22	116	32	106
Southern	1006	3	103	13	80	23	104	33	89
—	—	4	88	14	92	24	106	34	112
—	—	5	93	15	103	25	104	35	110
—	—	6	95	16	97	26	80	36	110
—	—	7	84	17	104	27	82	37	115
—	—	8	111	18	105	28	110	38	83
—	—	9	108	19	106	29	120	39	110
—	—	10	104	20	100	30	91	40	91
—	—	—	—	—	—	—	—	41	67

Table A.2: Microsatellite markers developed for *E. lori* parentage analysis. (A) Primer sequences for the six new microsatellites developed for this study. Primers for the initial 14 markers used and details on primer design are available in D'Aloia *et al.* (2013). Each forward primer has a 5' dye group; the sequence GTTTCTT for each lower primer promotes adenylation of the labelled strand. (B) Summary statistics for all 20 polymorphic microsatellite markers for *E. lori*: number of alleles (k), observed (H_O) and expected (H_E) heterozygosity, and deviations from Hardy-Weinberg equilibrium (HWE). An asterisk * indicates a significant deviation from HWE (dememorization steps = 10,000; # batches = 10,000; # iterations/batch = 1,000; $p < 0.05$). The potential presence of null alleles was inferred from excess homozygosity across allele size classes.

A) Primer sequences for 6 new microsatellite loci

Locus	**Repeat motif**	**Primer Sequence (5'-3')**
24561tet	AATC	6FAM-CCGGTGAGAAAGCCAATAGTG
		GTTTCTTTCTGAATTATTTCCGTGCTGCAC
5796tet	AATC	VIC-TTTGTTGCTGCTTCTTACTCCAC
		GTTTCTTAAGTGGACGGATAGGTAAGTAGG
1184tet	ATCC	6FAM-GGACCATGGCATAAATGATCTGG
		GTTTCTTGACATGGTCTCTCGGCTGG
24777tet	AGAT	NED-ACAGACAGTAAGGGTTATCTGCC
		GTTTCTTAACATAACAACCAAGATGCGAGG
26721tet	AATC	VIC-GTTTGAGAACCACTGCCCTATTC
		GTTTCTTAAGTTCTGTAAATGTAAGGGCGG
25176tet	ACGC/AC	6FAM-CCATCTGCCTTATGAGCTTACAC
		GTTTCTTAAGGAGCTTAAATGTGTCTGTCG

B) Summary statistics for all 20 microsatellite loci

Locus	***k***	***H_O***	***H_E***	***HWE***	***Null Alleles***
25745tet	10	0.658	0.764	<.001*	Yes
29109tet	50	0.948	0.959	0.01*	No
1419tri	19	0.316	0.694	<.001*	Yes
18144tri	12	0.575	0.555	0.87	No
985tet	85	0.912	0.981	<.001*	Yes
6326tet	10	0.758	0.705	0.13	No
25632tet	31	0.937	0.941	0.26	No
6231tet	27	0.949	0.938	0.03*	No
23889tri	22	0.888	0.884	0.78	No
25362tri	30	0.927	0.917	0.74	No
21378tri	48	0.916	0.931	0.54	No
14528tet	18	0.863	0.859	0.82	No
6266tri	20	0.717	0.770	0.002*	Yes
23415tet	23	0.903	0.890	0.32	No
24561tet	13	0.784	0.817	0.14	Yes
5796tet	31	0.913	0.940	0.35	Yes
1184tet	97	0.598	0.962	<.001*	Yes
24777tet	11	0.724	0.753	0.23	No
26721tet	18	0.564	0.881	<.001*	Yes
25176tet	32	0.886	0.914	<.001*	Yes

Table A.3: Goodness of fit criteria for *E. lori* dispersal kernel. Alternative probability density functions are considered, with AIC and BIC criteria used.

p.d.f.	**AIC**	**BIC**
Exponential	489.45	492.23
Weibull	491.08	496.65
Gamma	491.08	496.66
Lognormal	504.84	510.41
Logistic	579.94	585.51
Gaussian	596.17	601.75

Table A.4: Results from GLM models of predictors of net dispersal distance. (A) Goodness-of-fit criteria to determine the underlying distribution of the dependent variable dispersal (km). (B) Model selection criteria for the set of candidate GLMs (distribution = gamma; dispersion = 1). Biological predictors related to the fish and/or microhabitat are denoted by a B label and spatial predictors are denoted by a S label. We did not conduct additional likelihood ratio tests because predictors in the best-fit model were all non-significant at an α-level of 0.05. (C) Model output for the best-fit GLM based on selection criteria described in the supplementary methods text.

A) Distribution of dependent variable dispersal (km)

Distribution	**AIC**	**BIC**
Exponential	**489.45**	**492.23**
Weibull	491.08	496.65
Gamma	491.08	496.66
Lognormal	504.84	510.41
Gaussian	596.17	601.75

B) Information theoretic approach

Model	**AICc**	**ΔAICc**	**AICc Weight**	**Log-likelihood**
1. Full model (All predictors) *B+S*	462.08	0.00	0.62	-219.94
2. All biological predictors B	**463.06**	**0.98**	**0.38**	**-223.99**
3. Settler size (S.L.) *B*	476.20	14.12	0.00	-234.99
4. Sponge depth at origin *B*	476.82	14.74	0.00	-235.30
5. Max sponge length at origin *B*	481.62	19.54	0.00	-237.70
6. # Sponge tubes at origin *B*	482.02	19.94	0.00	-237.90
7. PLD B	489.16	27.07	0.00	-241.47
8. Parent Region (N/C/S) *S*	492.55	30.47	0.00	-242.10
9. Null	492.91	30.83	0.00	-244.41
10. Direction (N/S) *S*	493.63	31.55	0.00	-243.71
11. All spatial predictors *S*	494.00	31.92	0.00	-241.74

C) Model output for best-fit model

Parameter	**Estimate**	**S.E.**	**p-value**
(Intercept)	0.060	0.989	0.95
PLD	0.008	0.029	0.789
S.L.	-0.033	0.053	0.534
Max sponge length (origin)	0.001	0.008	0.835
# Sponge tubes (origin)	-0.023	0.119	0.844
Sponge depth (origin)	0.021	0.013	0.103

Table A.5: Logistic model analyses. (A) Information theoretic approach for logistic model selection: we present second-order AICc values, ΔAICc, AICc weights and log-likelihoods. The best-fit model is bolded. (B) Pairwise likelihood ratio tests. Given that the top four models include distance, effort, and region as predictors, we also conducted likelihood ratio (LR) tests for nested models with those predictors. Models with alternative predictors cannot be compared with LR tests. The best-fit model is bolded. (C) Output from the best-fit logistic regression model. The dependent variable is binary (dispersal event or no dispersal event). We report coefficients (with exponentiated odds ratios), standard errors, and p-values for the parameters in the best-fit model from the group of candidate models. The * indicates that the regional factors are compared to the Northern region; when the baseline factor is changed in the logit, the Central and Southern regions are shown to not be significantly different from each other. (D) Predictive power of logistic models for GIS analyses. The predictive power of the best-fit logistic model and the model without spatial variation are shown. Both models correct for sampling effort.

A) Information theoretic approach

Model	**AICc**	**ΔAICc**	**AICc Weight**	**Log Likelihood**
1. Distance + Effort + Region	**791.23**	**0.00**	**0.47**	**-390.61**
2. Distance + Effort + Region + Direction	792.63	1.40	0.23	-390.31
3. Distance + Effort + Region + Dist*Reg	792.78	1.55	0.22	-389.38
4. Distance + Effort	795.90	4.66	0.05	-394.95
5. Distance + Effort + Direction	797.41	6.18	0.02	-394.70
6. Distance + Effort + Direction + Dist*Dir	798.99	7.75	0.01	-394.49
7. Null	1031.59	240.36	0.00	-514.79

B) Likelihood ratio tests

Models compared	χ^2	**d.f.**	**p-value**
7 vs 4	239.70	2	<0.001
4 vs **1**	8.67	2	0.01
1 vs 2	0.61	1	0.44
1 vs 3	2.46	2	0.29

C) Model output for best-fit model

Parameter	**Estimate (OR)**	**S.E.**	**p-value**
(Intercept)	-1.852(0.16)	0.313	<0.001
Distance (km)	-0.356(0.70)	0.036	<0.001
Sampling effort scaled	1.687(5.40)	0.564	0.003
Central Region*	-0.330(0.72)	0.231	0.154
Southern Region*	-0.763(0.47)	0.266	0.004

D) Predictive power of logistic models

Model	**LOOCV (10-fold)**	**AUC**
1. Distance + Parent Region + Effort	0.025	0.89 (95% CI= 0.86, 0.91)
2. Distance + Effort	0.025	0.88 (95% CI = 0.86, 0.91)

Table A.6: Marine protected areas listed in Figure 4.4. We only consider zones within marine reserves that are non-extractive, including conservation and preservation zones, as well as spawning aggregation sites (SPAGs); these latter areas typically have seasonal limits on fishing.

MPA #	**MPA Name**
1	Bacalar Chico
2	Hol Chan
3	Caye Caulker
4	Caye Glory
5	Southwater
6	Gladden Spit
7	Sapodilla Cayes
8	Caye Bokel
9	Maugre Caye SPAG
10	Cockroach & Dog Flea Cayes
11	Blackbird Caye
12	Glover's Reserve (Western side)
13	Glovers SPAG (North)
14	Glover's Reserve (Eastern side)
15	Lighthouse SPAG (North)
16	Halfmoon Caye
17	Lighthouse SPAG (South)

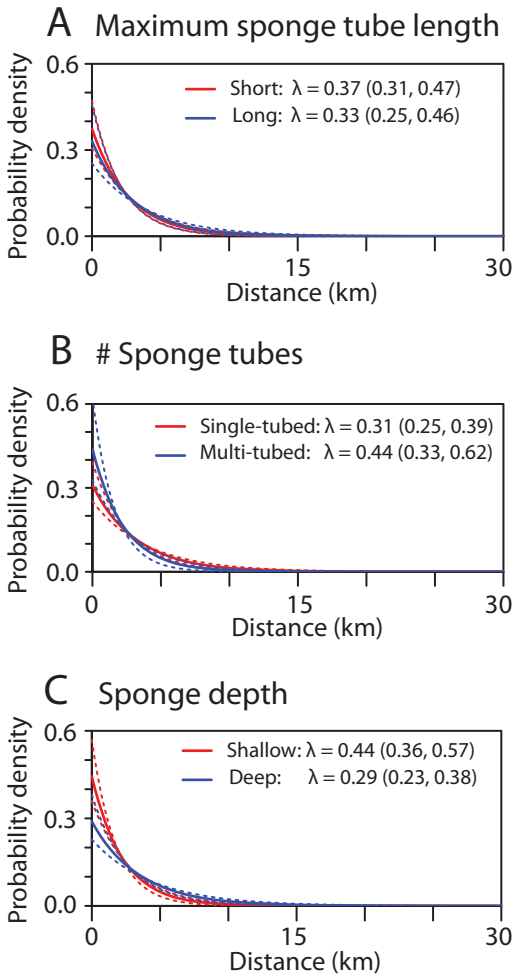

Figure A.1: Exponential dispersal kernels subdivided by sponge variables. Solid lines represent best estimates for the kernel and dashed lines represent 95% bootstrapped confidence intervals. Values for λ with 95% CI are provided. Subdivided kernels reveal a consistent shape with respect to (A) maximum sponge tube length at origin, (B) number of sponge tubes at origin, and (C) sponge depth at origin.

Appendix B

Supplement for Chapter 5

Overall summary statistics

Table B.1: Summary statistics across 10 sampling sites for 14 microsatellite loci. We present the mean (± s.d.) for the following metrics at each sampling location: number of alleles per locus per site (k), observed heterozygosity (H_O) and expected heterozygosity (H_E).

Site	**Mean k**	**Mean H_O**	**Mean H_E**
B1	14.50 (8.28)	0.79 (0.23)	0.83 (0.17)
B2	14.07 (6.91)	0.78 (0.24)	0.84 (0.16)
B3	16.29 (8.32)	0.80 (0.20)	0.85 (0.12)
B4	17.07 (10.22)	0.81 (0.19)	0.84 (0.14)
B5	16.14 (8.58)	0.78 (0.20)	0.86 (0.10)
T1	16.29 (9.92)	0.79 (0.21)	0.85 (0.13)
T2	15.64 (8.01)	0.79 (0.19)	0.86 (0.13)
G1	15.21 (8.33)	0.79 (0.20)	0.83 (0.13)
L1	14.50 (8.86)	0.79 (0.20)	0.83 (0.14)
L2	13.36 (6.13)	0.83 (0.18)	0.85 (0.10)

Table B.2: Molecular diversity indices in each population based on 960-bp region of *cytb* mtDNA sequences. At each site, we present mutation characteristics as well as theta estimates based on: expected number of alleles (k), expected homozygosity (H), the number of segregating sites (s), and the number of pairwise differences between haplotypes (π) (see Arlequin 2.3.4 manual for details).

Index	**B1**	**B2**	**B3**	**B4**	**B5**	**T1**	**T2**	**G1**	**L1**	**L2**	**Mean ± s.d**
# transitions	13	10	10	11	14	13	15	17	6	12	12.1 ± 3.1
# transversions	3	1	0	2	1	0	1	0	1	0	0.9 ± 1
# substitutions	16	11	10	13	15	13	16	17	7	12	13 ± 3.1
# indels	0	0	0	0	0	0	0	0	0	0	0 ± 0
# private sites	1	0	1	2	4	3	4	5	3	3	2.6 ± 1.6
Theta(k)	5.5	2.6	3.2	4.4	5.5	4.5	5.2	4.6	2.4	3.4	4.1 (1.9, 8.8)
Theta(H)	3.2	0.9	1.7	2.4	2.7	2.2	3.0	1.8	0.7	1.4	2 ± 0.8
Theta (s)	4.2	3.1	2.5	3.2	3.7	3.2	3.9	4.2	1.7	3.4	3.3 ± 0.8
Theta (π)	3.6	2.9	3.0	3.2	3.2	3.2	3.6	3.0	0.6	2.6	2.9 ± 0.8

Within-site summary statistics

Table B.3: Summary statistics of microsatellite loci (n=14) at site B1. Significant deviations from Hardy-Weinberg equilibrium (HWE) were determined with an exact test at α = 0.05 using a markov chain (chain length = 1,001,000; dememorization steps = 10,000).

Locus #	**k**	**H_O**	**H_E**	**HWE P-Value**
1419tri	6	0.20000	0.54694	**0.00010**
14528tet	12	0.85185	0.90287	0.08730
18144tri	5	0.40741	0.41649	0.50737
21378tri	15	0.77778	0.91405	**0.03045**
23415tet	12	0.92593	0.87701	**0.00106**
23889tri	11	0.88889	0.88330	0.93455
25362tri	18	0.92593	0.91474	0.86989
25632tet	19	1.00000	0.94759	1.00000
25745tet	6	0.76923	0.81976	0.39454
29109tet	23	0.96296	0.95877	0.69638
6231tet	19	0.92593	0.94200	0.50735
6266tri	13	0.81481	0.81901	0.51367
6326tet	8	0.62963	0.67156	0.78527
985tet	36	0.92593	0.98532	0.11427
Mean	14.50	0.78616	0.82853	—
s.d.	8.28	0.22916	0.16827	—

Table B.4: Summary statistics of microsatellite loci (n=14) at site B2. Significant deviations from Hardy-Weinberg equilibrium (HWE) were determined with an exact test at α = 0.05 using a markov chain (chain length = 1,001,000; dememorization steps = 10,000).

Locus #	**k**	$\mathbf{H_O}$	$\mathbf{H_E}$	**HWE P-Value**
1419tri	6	0.2	0.54483	**0.00010**
14528tet	11	0.86364	0.86469	0.57660
18144tri	6	0.40909	0.45243	0.16728
21378tri	19	0.95455	0.94503	0.76714
23415tet	13	0.86364	0.88795	0.06598
23889tri	12	0.90909	0.88266	0.24419
25362tri	18	0.95455	0.93552	0.37687
25632tet	17	0.95455	0.94503	0.94732
25745tet	7	0.63636	0.81184	**0.04233**
29109tet	23	1	0.963	1.00000
6231tet	17	1	0.9334	0.88389
6266tri	13	0.68182	0.89112	0.06545
6326tet	6	0.63636	0.7537	0.51349
985tet	29	0.81818	0.97569	**0.00000**
Mean	14.071	0.77727	0.84192	—
s.d.	6.911	0.2394	0.15843	—

Table B.5: Summary statistics of microsatellite loci (n=14) at site B3. Significant deviations from Hardy-Weinberg equilibrium (HWE) were determined with an exact test at α = 0.05 using a markov chain (chain length = 1,001,000; dememorization steps = 10,000).

Locus #	**k**	$\mathbf{H_O}$	$\mathbf{H_E}$	**HWE P-Value**
1419tri	7	0.34783	0.69662	**0.00068**
14528tet	15	0.96667	0.86271	0.89104
18144tri	9	0.56667	0.60226	0.17976
21378tri	16	0.8	0.9226	0.07615
23415tet	16	0.96667	0.90734	0.88772
23889tri	12	0.86667	0.8887	0.53370
25362tri	20	0.86667	0.92429	0.37788
25632tet	21	0.96667	0.94802	0.92040
25745tet	7	0.53333	0.75254	**0.02054**
29109tet	29	1	0.96836	0.51720
6231tet	20	0.9	0.93842	0.76674
6266tri	14	0.76667	0.82599	0.35249
6326tet	7	0.66667	0.71356	0.42376
985tet	35	0.96667	0.97684	0.27903
Mean	16.286	0.79865	0.85202	—
s.d.	8.315	0.19978	0.11662	—

Table B.6: Summary statistics of microsatellite loci (n=14) at site B4. Significant deviations from Hardy-Weinberg equilibrium (HWE) were determined with an exact test at α = 0.05 using a markov chain (chain length = 1,001,000; dememorization steps = 10,000).

Locus #	**k**	$\mathbf{H_O}$	$\mathbf{H_E}$	**HWE P-Value**
1419tri	8	0.32258	0.6367	**0.00000**
14528tet	12	0.82857	0.85507	0.18041
18144tri	7	0.54286	0.5118	0.91972
21378tri	17	0.82857	0.92795	**0.04130**
23415tet	17	0.97143	0.93126	0.97395
23889tri	13	0.91429	0.8795	0.45160
25362tri	16	0.8	0.90518	0.54050
25632tet	26	1	0.96025	0.86083
25745tet	7	0.74286	0.8207	0.80368
29109tet	30	0.97143	0.96232	0.95297
6231tet	20	0.88571	0.93582	0.17268
6266tri	16	0.82857	0.80538	0.29366
6326tet	7	0.8	0.69193	0.60587
985tet	43	0.97143	0.98509	0.654908
Mean	17.071	0.81488	0.8435	—
s.d.	10.224	0.18507	0.13981	—

Table B.7: Summary statistics of microsatellite loci (n=14) at site B5. Significant deviations from Hardy-Weinberg equilibrium (HWE) were determined with an exact test at α = 0.05 using a markov chain (chain length = 1,001,000; dememorization steps = 10,000).

Locus #	**k**	$\mathbf{H_O}$	$\mathbf{H_E}$	**HWE P-Value**
1419tri	9	0.26667	0.73503	**0.00000**
14528tet	12	0.93548	0.86356	0.12509
18144tri	7	0.67742	0.67795	0.23826
21378tri	19	0.90323	0.93919	0.08733
23415tet	15	0.83871	0.86991	0.37912
23889tri	11	0.77419	0.88789	0.05025
25362tri	19	0.87097	0.93125	0.29988
25632tet	22	0.93548	0.94976	0.23855
25745tet	8	0.58065	0.7964	0.17672
29109tet	27	0.96774	0.95505	0.59077
6231tet	20	0.96774	0.94183	0.88301
6266tri	13	0.58065	0.7578	**0.00213**
6326tet	7	0.70968	0.72819	0.60340
985tet	37	0.93548	0.98149	0.19989
Mean	16.143	0.78172	0.85824	—
s.d.	8.583	0.20161	0.10052	—

Table B.8: Summary statistics of microsatellite loci (n=14) at site T1. Significant deviations from Hardy-Weinberg equilibrium (HWE) were determined with an exact test at α = 0.05 using a markov chain (chain length = 1,001,000; dememorization steps = 10,000).

Locus #	**k**	**HO**	**HE**	**HWE P-Value**
1419tri	6	0.2	0.64653	**0.00000**
14528tet	13	0.90909	0.87179	0.68210
18144tri	5	0.54545	0.55431	0.35199
21378tri	19	0.90909	0.93193	0.36297
23415tet	13	0.90909	0.88298	0.05395
23889tri	11	0.90909	0.88811	0.82351
25362tri	18	0.93939	0.94312	0.98252
25632tet	20	0.9697	0.94079	0.93896
25745tet	8	0.66667	0.81632	0.18023
29109tet	29	0.90909	0.95618	0.43147
6231tet	21	0.93939	0.94452	0.76072
6266tri	18	0.75758	0.89231	**0.02438**
6326tet	6	0.66667	0.70256	0.61710
985tet	41	0.84848	0.98228	**0.00000**
Mean	16.286	0.79134	0.85384	—
s.d.	9.918	0.21315	0.12939	—

Table B.9: Summary statistics of microsatellite loci (n=14) at site T2. Significant deviations from Hardy-Weinberg equilibrium (HWE) were determined with an exact test at α = 0.05 using a markov chain (chain length = 1,001,000; dememorization steps = 10,000).

Locus #	**k**	**HO**	**HE**	**HWE P-Value**
1419tri	10	0.34615	0.75189	**0.00002**
14528tet	14	0.84848	0.91702	0.18970
18144tri	6	0.45455	0.49277	0.36133
21378tri	16	0.87879	0.92401	0.14636
23415tet	11	0.84848	0.85734	0.12826
23889tri	13	0.81818	0.89604	0.40291
25362tri	20	0.78788	0.92308	**0.02751**
25632tet	19	1	0.94825	0.99550
25745tet	7	0.72727	0.78974	0.42361
29109tet	25	0.96875	0.95734	0.97401
6231tet	19	0.93939	0.94685	0.70235
6266tri	15	0.84848	0.88019	0.71695
6326tet	8	0.63636	0.77296	0.07981
985tet	36	0.9697	0.97622	0.37951
Mean	15.643	0.79089	0.85955	—
s.d.	8.006	0.19348	0.12722	—

Table B.10: Summary statistics of microsatellite loci (n=14) at site G1. Significant deviations from Hardy-Weinberg equilibrium (HWE) were determined with an exact test at α = 0.05 using a markov chain (chain length = 1,001,000; dememorization steps = 10,000).

Locus #	**k**	**HO**	**HE**	**HWE P-Value**
1419tri	5	0.26667	0.58701	**0.00002**
14528tet	13	0.93939	0.87459	0.05117
18144tri	8	0.54545	0.5366	0.26212
21378tri	17	0.90909	0.92727	0.75264
23415tet	15	0.90909	0.85221	0.94855
23889tri	10	0.75758	0.8331	0.28297
25362tri	16	0.75758	0.89697	**0.01875**
25632tet	20	0.9697	0.93893	0.66129
25745tet	8	0.72727	0.79534	0.80546
29109tet	25	1.00000	0.938	**0.03759**
6231tet	18	0.93939	0.92821	0.74419
6266tri	13	0.75758	0.77156	0.41541
6326tet	8	0.72727	0.7683	**0.03903**
985tet	37	0.87879	0.97622	**0.00323**
Mean	15.214	0.79177	0.83031	—
s.d.	8.331	0.1965	0.13143	—

Table B.11: Summary statistics of microsatellite loci (n=14) at site L1. Significant deviations from Hardy-Weinberg equilibrium (HWE) were determined with an exact test at α = 0.05 using a markov chain (chain length = 1,001,000; dememorization steps = 10,000).

Locus #	**k**	**HO**	**HE**	**HWE P-Value**
1419tri	5	0.36364	0.69044	**0.00004**
14528tet	14	0.91176	0.8863	**0.00968**
18144tri	7	0.41176	0.4899	0.60276
21378tri	17	0.88235	0.92801	0.45718
23415tet	11	0.85294	0.8626	0.63039
23889tri	8	0.73529	0.76207	0.47947
25362tri	19	0.76471	0.92142	**0.00029**
25632tet	22	1	0.95347	0.98823
25745tet	7	0.76471	0.79807	0.46943
29109tet	23	0.97059	0.94864	0.87762
6231tet	20	0.91176	0.93459	0.29631
6266tri	8	0.85294	0.77173	0.71593
6326tet	6	0.64706	0.69579	0.44155
985tet	36	0.94118	0.9741	0.43397
Mean	14.5	0.78648	0.8298	—
s.d.	8.856	0.19514	0.13739	—

Table B.12: Summary statistics of microsatellite loci (n=14) at site L2. Significant deviations from Hardy-Weinberg equilibrium (HWE) were determined with an exact test at $\alpha = 0.05$ using a markov chain (chain length = 1,001,000; dememorization steps = 10,000).

Locus #	**k**	**HO**	**HE**	**HWE P-Value**
1419tri	9	0.33333	0.76825	**0.00000**
14528tet	12	1.00000	0.91538	0.96900
18144tri	9	0.80000	0.78333	0.35099
21378tri	16	0.95000	0.92949	0.83391
23415tet	10	1.00000	0.87436	0.88893
23889tri	10	0.95000	0.88590	0.53302
25362tri	16	0.75000	0.88462	0.08488
25632tet	20	1.00000	0.93333	1.00000
25745tet	6	0.75000	0.77051	0.06366
29109tet	17	0.85000	0.94359	0.34071
6231tet	18	0.95000	0.95513	0.52716
6266tri	10	0.65000	0.67692	0.84023
6326tet	6	0.75000	0.65897	0.84471
985tet	28	0.95000	0.98077	**0.04561**
Mean	13.357	0.83452	0.85433	—
s.d.	6.134	0.18472	0.10420	—

Table B.13: Linkage disequilibrium tests among loci at site B1, with significance assessed by permutations using the EM algorithm in Arlequin (10,000 permutations). We applied a sequential Bonferroni test to account for multiple pairwise comparisons. Significant values are shown in bold.

Locus	1419tri	14528tet	18144tri	21378tri	23415tet	23889tri	25362tri	25632tet	25745tet	29109tet	6231tet	6266tri	6326tet	985tet
14528tet	0.19	—												
18144tri	0.09	0.48	—											
21378tri	0.02	0.62	0.43	—										
23415tet	0.03	0.87	0.60	0.58	—									
23889tri	0.81	0.40	0.62	0.74	0.36	—								
25362tri	0.006	0.36	0.009	0.05	0.52	0.53	—							
25632tet	0.65	0.98	0.13	0.37	0.78	0.62	0.37	—						
25745tet	0.62	0.06	0.96	0.11	0.10	0.33	0.59	0.35	—					
29109tet	0.15	0.99	0.75	0.04	0.83	0.54	0.87	0.02	0.99	—				
6231tet	0.30	0.71	0.14	0.76	0.31	0.72	0.30	0.36	0.92	0.67	—			
6266tri	0.16	0.17	0.39	0.27	0.84	0.59	0.84	0.36	0.33	0.06	0.37	—		
6326tet	0.07	0.73	0.20	0.53	0.75	0.26	0.50	0.62	0.51	0.96	0.31	0.83	—	
985tet	0.09	1.00	0.08	0.19	0.78	0.20	0.09	0.42	0.52	0.70	0.55	0.59	0.79	—

Table B.14: Linkage disequilibrium tests among loci at site B2, with significance assessed by permutations using the EM algorithm in Arlequin (10,000 permutations). We applied a sequential Bonferroni test to account for multiple pairwise comparisons. Significant values are shown in bold.

Locus	1419tri	14528tet	18144tri	21378tri	23415tet	23889tri	25362tri	25632tet	25745tet	29109tet	6231tet	6266tri	6326tet	985tet
1419tri	—													
14528tet	0.68	—												
18144tri	0.66	0.29	—											
21378tri	0.007	0.55	0.57	—										
23415tet	0.47	0.49	0.31	0.64	—									
23889tri	0.82	0.39	0.05	0.64	0.85	—								
25362tri	0.83	0.06	0.12	0.23	0.13	0.78	—							
25632tet	0.81	0.54	0.54	0.20	0.28	1.0	0.40	—						
25745tet	0.09	0.11	0.39	0.99	0.56	0.90	0.40	0.86	—					
29109tet	0.11	0.78	0.85	0.93	0.22	0.82	0.46	0.02	0.24	—				
6231tet	0.56	0.86	0.39	0.55	0.97	0.41	0.69	0.96	0.20	0.35	—			
6266tri	0.44	0.47	0.29	0.43	0.65	0.60	0.12	0.81	0.51	0.45	0.73	—		
6326tet	0.21	0.69	0.53	0.92	0.05	0.96	0.48	0.48	0.35	0.68	0.03	0.21	—	
985tet	0.08	0.12	0.50	0.68	0.45	0.36	0.83	0.42	0.11	0.55	0.66	0.23	0.84	—

Table B.15: Linkage disequilibrium tests among loci at site B3, with significance assessed by permutations using the EM algorithm in Arlequin (10,000 permutations). We applied a sequential Bonferroni test to account for multiple pairwise comparisons. Significant values are shown in bold.

Locus	1419tri	14528tet	18144tri	21378tri	23415tet	23889tri	25362tri	25632tet	25745tet	29109tet	6231tet	6266tri	6326tet	985tet
1419tri	—													
14528tet	0.33	—												
18144tri	0.37	0.10	—											
21378tri	0.03	0.63	0.46	—										
23415tet	0.61	0.60	0.90	0.37	—									
23889tri	0.89	0.52	0.64	0.21	0.41	—								
25362tri	0.03	0.92	0.80	0.68	0.70	0.98	—							
25632tet	0.002	0.42	0.46	0.01	0.78	0.52	0.59	—						
25745tet	0.12	0.31	0.71	0.37	0.38	0.78	0.16	0.66	—					
29109tet	0.45	0.25	0.40	0.93	0.51	0.10	0.29	0.13	0.83	—				
6231tet	0.45	0.26	0.42	0.07	0.69	0.93	0.22	0.006	0.03	0.95	—			
6266tri	0.78	0.88	0.85	0.21	0.74	0.56	0.40	0.01	0.40	0.32	0.67	—		
6326tet	0.73	0.67	0.23	0.11	0.29	0.50	0.30	0.26	0.08	0.70	0.32	0.03	—	
985tet	0.07	0.24	0.05	0.09	0.11	0.80	0.26	0.40	0.16	0.52	0.15	0.11	0.69	—

Table B.16: Linkage disequilibrium tests among loci at site B4, with significance assessed by permutations using the EM algorithm in Arlequin (10,000 permutations). We applied a sequential Bonferroni test to account for multiple pairwise comparisons. Significant values are shown in bold.

Locus	1419tri	14528tet	18144tri	21378tri	23415tet	23889tri	25362tri	25632tet	25745tet	29109tet	6231tet	6266tri	6326tet	985tet
1419tri	—													
14528tet	0.24	—												
18144tri	0.74	0.62	—											
21378tri	0.10	0.35	0.01	—										
23415tet	0.81	0.98	0.65	0.07	—									
23889tri	0.70	0.90	0.57	0.53	0.08	—								
25362tri	0.47	0.61	0.63	0.66	0.37	0.59	—							
25632tet	0.003	0.18	0.90	0.55	0.94	0.75	0.87	—						
25745tet	0.19	0.20	0.54	0.67	0.92	0.27	0.49	0.14	—					
29109tet	0.95	0.57	0.06	0.36	0.53	0.82	0.66	0.67	0.56	—				
6231tet	0.18	0.19	0.15	0.32	0.07	0.76	0.40	0.27	0.93	0.88	—			
6266tri	0.73	0.52	0.32	0.38	0.58	0.76	0.79	0.87	0.71	0.43	0.06	—		
6326tet	0.80	0.77	0.37	0.24	0.88	0.96	0.72	0.77	0.09	0.58	0.51	0.34	—	
985tet	0.48	0.50	0.85	0.07	0.85	0.31	0.60	0.96	0.51	0.29	0.06	0.41	1.0	—

Table B.17: Linkage disequilibrium tests among loci at site B5, with significance assessed by permutations using the EM algorithm in Arlequin (10,000 permutations). We applied a sequential Bonferroni test to account for multiple pairwise comparisons. Significant values are shown in bold.

Locus	**1419tri**	**14528tet**	**18144tri**	**21378tri**	**23415tet**	**23889tri**	**25362tri**	**25632tet**	**25745tet**	**29109tet**	**6231tet**	**6266tri**	**6326tet**	**985tet**
1419tri	—													
14528tet	0.88	—												
18144tri	0.92	0.39	—											
21378tri	0.31	0.43	0.59	—										
23415tet	0.65	0.85	0.51	0.16	—									
23889tri	0.39	0.33	0.58	0.69	0.13	—								
25362tri	0.45	0.57	0.81	0.79	0.10	0.20	—							
25632tet	0.20	0.79	0.59	0.18	0.09	0.06	0.56	—						
25745tet	0.73	0.95	0.03	0.79	0.87	0.53	0.26	0.86	—					
29109tet	0.01	0.16	0.34	0.68	0.03	0.77	0.22	0.73	0.89	—				
6231tet	0.38	0.05	0.14	0.66	0.55	0.81	0.42	0.15	0.97	0.51	—			
6266tri	0.14	0.03	0.25	0.71	0.31	0.97	0.34	0.64	0.50	0.43	0.77	—		
6326tet	0.34	0.35	0.38	0.72	0.62	0.60	0.45	0.02	0.99	0.82	0.76	0.99	—	
985tet	0.0008	0.08	0.30	0.03	0.92	0.19	0.18	0.19	0.20	0.57	0.07	0.22	0.90	—

Table B.18: Linkage disequilibrium tests among loci at site T1, with significance assessed by permutations using the EM algorithm in Arlequin (10,000 permutations). We applied a sequential Bonferroni test to account for multiple pairwise comparisons. Significant values are shown in bold.

Locus	**1419tri**	**14528tet**	**18144tri**	**21378tri**	**23415tet**	**23889tri**	**25362tri**	**25632tet**	**25745tet**	**29109tet**	**6231tet**	**6266tri**	**6326tet**	**985tet**
1419tri	—													
14528tet	0.62	—												
18144tri	0.24	0.67	—											
21378tri	0.04	0.46	0.23	—										
23415tet	0.71	0.07	0.22	0.69	—									
23889tri	0.07	0.46	0.85	0.64	0.58	—								
25362tri	0.68	0.68	0.97	0.26	0.46	0.13	—							
25632tet	0.28	0.80	0.61	0.18	0.21	0.83	0.35	—						
25745tet	0.08	0.48	0.50	0.40	0.70	0.89	0.96	0.68	—					
29109tet	0.38	0.54	0.28	0.88	0.93	0.20	0.17	0.86	0.01	—				
6231tet	0.47	0.69	0.68	0.67	0.66	0.32	0.14	0.85	0.27	0.08	—			
6266tri	0.45	0.53	0.96	0.47	0.69	0.63	0.13	0.66	0.99	0.62	0.61	—		
6326tet	0.84	0.28	0.40	0.66	0.42	0.96	0.36	0.25	0.57	0.38	0.71	0.92	—	
985tet	0.002	0.64	0.85	0.27	0.32	0.51	0.43	0.63	0.53	0.72	0.76	0.41	0.78	—

Table B.19: Linkage disequilibrium tests among loci at site T2, with significance assessed by permutations using the EM algorithm in Arlequin (10,000 permutations). We applied a sequential Bonferroni test to account for multiple pairwise comparisons. Significant values are shown in bold.

Locus	1419tri	14528tet	18144tri	21378tri	23415tet	23889tri	25362tri	25632tet	25745tet	29109tet	6231tet	6266tri	6326tet	985tet
1419tri	—													
14528tet	0.51	—												
18144tri	0.58	0.94	—											
21378tri	0.77	0.29	0.65	—										
23415tet	0.80	0.51	0.40	0.35	—									
23889tri	0.99	0.44	0.89	0.66	0.77	—								
25362tri	0.0007	**0.0003**	0.54	0.31	0.95	0.34	—							
25632tet	0.54	0.61	0.29	0.58	0.52	0.22	0.03	—						
25745tet	0.04	0.32	0.86	0.92	0.82	0.44	0.02	0.55	—					
29109tet	0.009	0.81	0.74	0.94	0.91	0.62	0.10	0.33	0.91	—				
6231tet	0.18	0.15	0.03	0.56	0.80	0.09	0.79	0.24	0.29	0.68	—			
6266tri	0.10	0.55	0.78	0.25	0.49	0.07	0.57	0.91	0.21	0.14	0.22	—		
6326tet	0.43	0.23	0.08	0.82	0.89	0.09	0.20	0.77	0.25	0.31	0.20	0.70	—	
985tet	0.02	0.76	0.49	0.05	0.33	0.48	0.01	0.57	0.29	0.73	0.47	0.08	0.12	—

Table B.20: Linkage disequilibrium tests among loci at site G1, with significance assessed by permutations using the EM algorithm in Arlequin (10,000 permutations). We applied a sequential Bonferroni test to account for multiple pairwise comparisons. Significant values are shown in bold.

Locus	1419tri	14528tet	18144tri	21378tri	23415tet	23889tri	25362tri	25632tet	25745tet	29109tet	6231tet	6266tri	6326tet	985tet
1419tri	—													
14528tet	0.81	—												
18144tri	0.13	0.15	—											
21378tri	0.10	0.49	0.73	—										
23415tet	0.39	0.47	0.39	0.47	—									
23889tri	0.96	0.75	0.29	0.99	0.26	—								
25362tri	0.32	0.98	0.89	0.55	0.31	0.24	—							
25632tet	0.33	0.38	0.65	0.96	0.85	0.71	0.76	—						
25745tet	0.68	0.49	0.05	0.35	0.73	0.34	0.21	0.20	—					
29109tet	0.14	0.80	0.22	0.96	0.27	0.51	0.14	0.29	0.46	—				
6231tet	0.49	0.70	0.82	0.06	0.74	0.34	0.34	0.93	0.67	0.11	—			
6266tri	0.78	0.62	0.95	0.95	0.29	0.80	0.05	0.71	0.93	1.00	0.90	—		
6326tet	0.25	0.44	0.29	0.39	0.42	0.42	0.70	0.63	0.74	0.43	0.19	0.60	—	
985tet	0.09	0.27	0.78	0.75	0.43	0.22	0.30	0.70	0.60	0.85	0.64	0.09	0.23	—

Table B.21: Linkage disequilibrium tests among loci at site L1, with significance assessed by permutations using the EM algorithm in Arlequin (10,000 permutations). We applied a sequential Bonferroni test to account for multiple pairwise comparisons. Significant values are shown in bold.

Locus	1419tri	14528tet	18144tri	21378tri	23415tet	23889tri	25362tri	25632tet	25745tet	29109tet	6231tet	6266tri	6326tet	985tet
1419tri	—													
14528tet	0.34	—												
18144tri	0.69	0.07	—											
21378tri	0.85	0.06	0.30	—										
23415tet	0.78	0.92	0.64	0.30	—									
23889tri	0.65	0.78	0.33	0.10	0.98	—								
25362tri	0.42	0.31	0.64	0.03	0.92	0.53	—							
25632tet	0.43	0.79	1.0	0.55	0.20	0.27	0.29	—						
25745tet	0.23	0.91	0.93	0.18	0.30	0.12	0.21	0.36	—					
29109tet	0.74	0.53	0.77	0.53	0.003	0.71	0.43	0.24	0.16	—				
6231tet	0.12	0.14	0.001	0.13	0.26	0.59	0.92	0.20	0.25	0.21	—			
6266tri	0.16	0.61	0.42	0.32	0.95	0.43	0.90	0.69	0.97	0.45	0.50	—		
6326tet	0.03	0.77	0.83	0.97	0.68	0.34	0.47	0.61	0.19	0.38	0.91	0.20	—	
985tet	0.20	0.62	0.81	0.49	0.32	0.22	0.23	0.89	0.51	0.18	0.40	0.86	0.33	—

Table B.22: Linkage disequilibrium tests among loci at site L2, with significance assessed by permutations using the EM algorithm in Arlequin (10,000 permutations). We applied a sequential Bonferroni test to account for multiple pairwise comparisons. Significant values are shown in bold.

Locus	1419tri	14528tet	18144tri	21378tri	23415tet	23889tri	25362tri	25632tet	25745tet	29109tet	6231tet	6266tri	6326tet	985tet
1419tri	—													
14528tet	0.95	—												
18144tri	0.01	0.47	—											
21378tri	0.66	0.78	0.91	—										
23415tet	0.92	0.68	0.33	0.66	—									
23889tri	0.26	0.79	0.58	0.34	0.38	—								
25362tri	0.15	0.31	1.0	0.52	0.15	0.83	—							
25632tet	0.45	0.57	0.26	0.16	0.008	0.68	0.83	—						
25745tet	0.07	0.67	0.73	0.42	0.61	0.50	0.24	0.92	—					
29109tet	0.51	0.47	0.60	0.28	0.28	0.75	0.29	0.06	0.29	—				
6231tet	0.25	0.99	0.10	0.88	0.19	0.41	0.08	0.60	0.88	0.002	—			
6266tri	0.27	0.01	0.93	0.50	0.21	0.02	0.92	0.50	0.46	0.04	0.20	—		
6326tet	0.11	0.49	0.35	0.05	0.66	0.11	0.93	0.75	0.72	0.80	0.46	0.75	—	
985tet	0.006	0.72	0.90	0.44	0.95	0.38	0.26	0.98	0.04	0.52	0.74	0.58	0.88	—

Microchecker analyses

Table B.23: Microchecker analyses of null allele presence within sites at all microsatellite loci. Significance is reported for $\alpha = 0.05$.

	B1	**B2**	**B3**	**B4**	**B5**	**T1**	**T2**	**G1**	**L1**	**L2**
1419tri	**YES**	**YES**	**YES**	**YES**	**YES**	**YES**	**YES**	**YES**	**YES**	**YES**
14528tet	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO
18144tri	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO
21378tri	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO
23415tet	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO
23889tri	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO
25362tri	NO	NO	NO	NO	NO	NO	**YES**	**YES**	**YES**	NO
25632tet	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO
25745tet	NO	**YES**	**YES**	NO	**YES**	**YES**	NO	NO	NO	NO
29109tet	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO
6231tet	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO
6266tri	NO	**YES**	NO	NO	**YES**	**YES**	NO	NO	NO	NO
6326tet	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO
985tet	NO	**YES**	NO	NO	NO	**YES**	NO	**YES**	NO	NO

Hedrick's G'_{ST}

Table B.24: Hedricks G'_{ST} as an alternative metric for pairwise differentiation (based on microsatellites).

	B1	**B2**	**B3**	**B4**	**B5**	**T1**	**T2**	**G1**	**L1**	**L2**
B1	–	–	–	–	–	–	–	–	–	–
B2	0.002	–	–	–	–	–	–	–	–	–
B3	0.000	0.027	–	–	–	–	–	–	–	–
B4	0.000	0.011	0.000	–	–	–	–	–	–	–
B5	0.040	0.040	0.019	0.012	–	–	–	–	–	–
T1	0.006	0.033	0.000	0.004	0.029	–	–	–	–	–
T2	0.002	0.032	0.016	0.017	0.026	0.000	–	–	–	–
G1	0.046	0.077	0.076	0.056	0.055	0.030	0.067	–	–	–
L1	0.058	0.049	0.082	0.059	0.029	0.069	0.069	0.082	–	–
L2	0.053	0.080	0.036	0.047	0.025	0.040	0.063	0.094	0.040	–

Structure analyses

Table B.25: Mean LnP and standard deviation (s.d.) for all values of k (k=1 to k=10) based on Structure microsatellite clustering analyses. Model parameter details can be found in the main text

K	LnP	s.d.
1	-20058.37	1.39
2	-20043.70	14.89
3	-20042.90	43.84
4	-20110.97	72.85
5	-20317.64	263.85
6	-20378.27	252.80
7	-20328.04	213.23
8	-20420.30	255.29
9	-20423.51	382.06
10	-20382.75	245.66

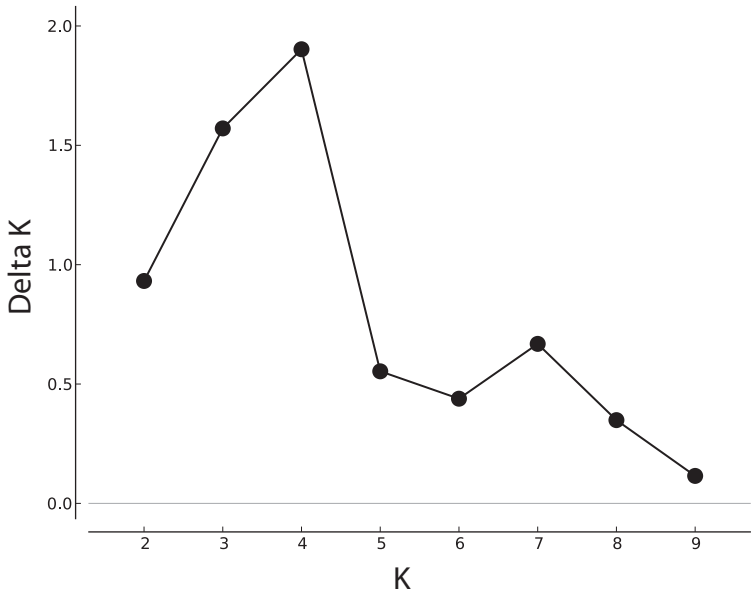

Figure B.1: Evanno plot for Structure analyses. Plot of delta k versus k from Structure Harvester based on the Evanno method (Evanno *et al.* 2005) with a mode at k=4.

Haplotype network

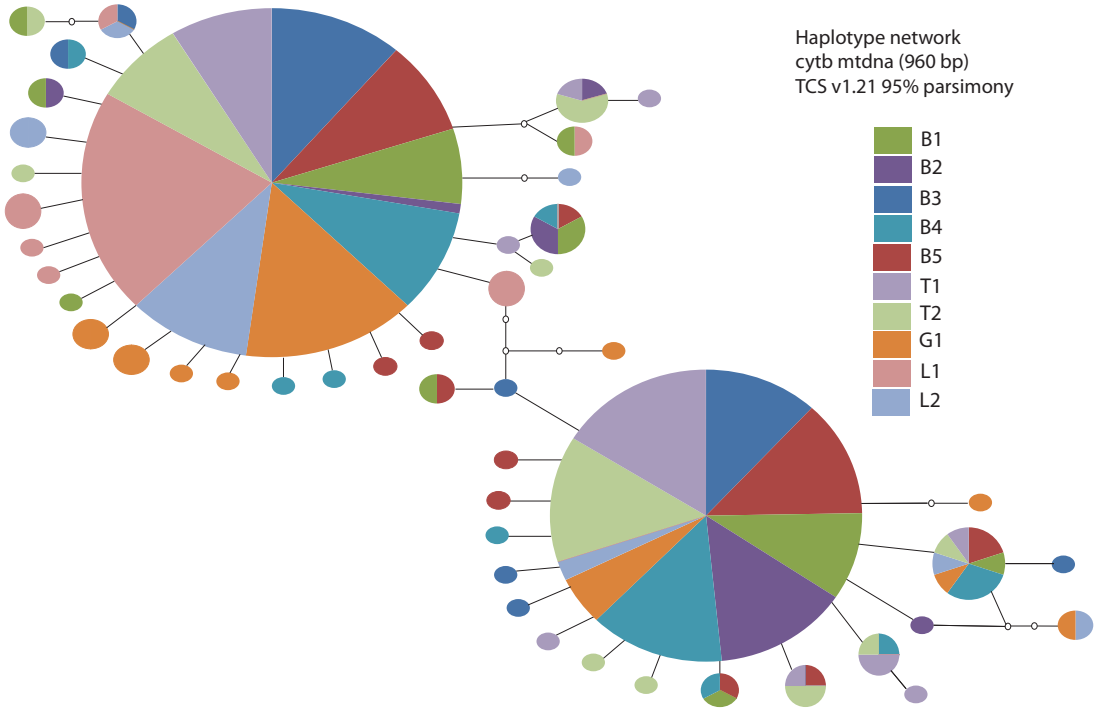

Figure B.2: Haplotype network based on *cytb* mtDNA sequences of all individuals across the entire BBR system. Each haplotype is represented by a circle (n = 48 circles), with size proportional to haplotype frequency. Lines represent one mutation, and small empty circles represent inferred haplotypes. Color coding represents frequency of each haplotype across sampling sites.

Phylogenetic analyses

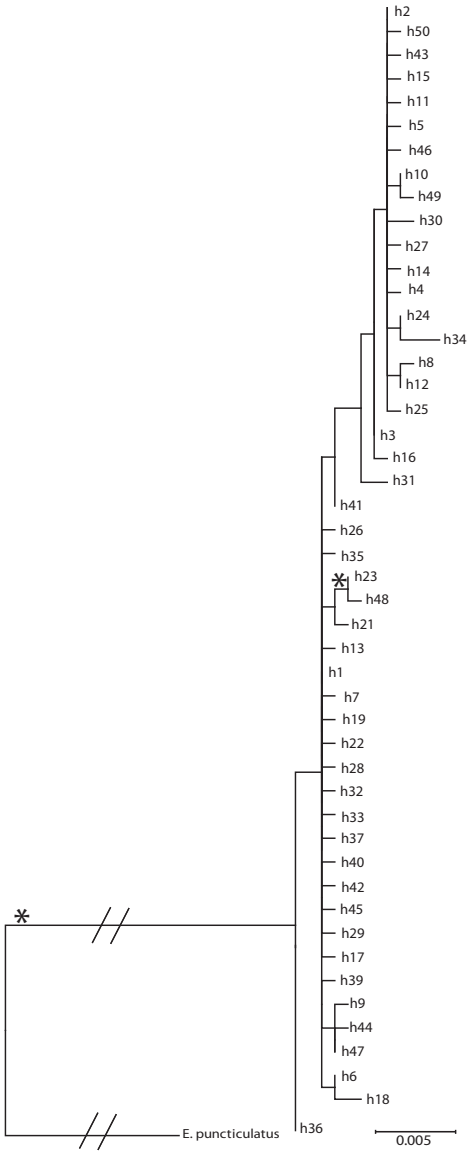

Figure B.3: Maximum likelihood tree of *E. lori cytb* haplotypes (bootstrap replicates = 500) in MEGA 5.1 (Tamura *et al.* 2011). The outgroup *Elacatinus puncticulatus*, has been shown to be basal to the genus (Taylor & Hellberg 2005). Branches with significant (>70%) bootstrap support are marked with a *. There is no resolution of the relationships between *E. lori* haplotypes based on this ML tree, or additional maximum parsimony and neighbor-joining trees based on low sequence divergence.

Robustness checks for statistical analyses

Here, we report results from additional Firth logistic regression and linear mixed model (LMM) analyses that use an alternative distance cut-off for 'oceanic breaks.' We chose to repeat analyses using a cut-off of 8 km, because this is the shortest distance between the third atoll (Turneffe Atoll) and the relatively continuous Belize barrier reef. Differentiation metrics and genetic clustering analyses also suggested that sites on Turneffe atoll were panmictic with the barrier reef sites. Below is a quantitative investigation of this alternative cut-off for breaks:

First, Firth's penalized-likelihood logistic regression revealed that sites separated an oceanic break > 8 km are not significantly more likely to have a positive Φ_{ST} estimate, as compared to sites that are not separated by a break (coefficient = 1.26 (OR = 3.52); s.e. = 0.70; p = 0.06). Furthermore, there was no significant improvement in model fit when 'Euclidian distance' was included as a predictor ($\chi^2 = 0.43$, df = 1, p = 0.51). Similar results emerged using microsatellite-based F_{ST} as the differentiation metric. In this case, the best-fit model also revealed that sites separated by a break were not significantly more likely to have a non-zero value of F_{ST} (coefficient = 1.17 (OR = 3.21); s.e. = 0.87; p=0.17), and that adding Euclidian distance as a predictor did not improve model fit ($\chi^2 = 0.95$; df= 1; p = 0.33). Thus, when the distance cut-off is reduced to 8 km, the Firth logistic regression analyses reject both alternative hypotheses.

Similarly, a linear mixed model analysis of the predictors of positive-only Φ_{ST} values revealed that the best-fit model included only the random effects 'Site 1' and 'Site 2' (H_0; $\Delta AIC_c = 0.00$), compared to a model that added breaks > 8 km (H_1; $\Delta AIC_c = 2.21$), and a model that included both breaks > 8 km and Euclidian distance (H_2; $\Delta AIC_c = 5.32$) (Table A.26). However, a linear mixed model analysis of the predictors of positive-only F_{ST} values revealed that the best-fit model included the random effects 'Site 1' and 'Site 2' and oceanic breaks > 8 km (H_1; $\Delta AIC_c = 0.00$), compared to a model only included the random effects (H_0; $\Delta AIC_c = 1.76$), and a model that included both breaks and

Euclidian distance (H_2; $\Delta AIC_c = 2.41$) (Table A.26). The difference between the ΔAIC_c values is quite small, however, suggesting that there is weak discriminatory power between models. Combined with the Firth logit results, these data suggest, overall, that among both metrics of differentiation (Φ_{ST} for mtDNA and F_{ST} for microsatellites), breaks must be substantively large in space to create gene flow barriers.

Table B.26: Linear mixed model (LMM) output. This analysis used positive-only values of pairwise Φ_{ST} (a) and F_{ST} (b). To meet normality assumptions the dependent variables are logged. 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}				**b) F_{ST}**			
Fixed effects				***Fixed effects***			
Parameter	**Estimate**	**SE**	**t value**	**Parameter**	**Estimate**	**SE**	**t value**
(Intercept)	-1.41	0.22	-6.31	(Intercept)	-2.72	0.15	-17.90
—	—	—	—	Oceanic break >8 km	0.50	0.20	2.53
Random effects				**Random effects**			
Group	S.D.			Group	S.D.		
Site 1	0.22	—	—	Site 1	<0.001	—	—
Site 2	0.65	—	—	Site 2	0.25	—	

Mantel tests comparing geographic distance and genetic distance

The Firth logits and LMMs clearly rejected the isolation-by-distance hypothesis (H_2). Here, we present further evidence that Euclidian distance (km) in not correlated with any metric of pairwise genetic differentiation using mantel tests: mtDNA-based Φ_{ST} and Euclidian distance (spearman's rho = 0.01; permutations = 1,000; p = 0.40), mtDNA-based F_{ST} and Euclidian distance (spearman's rho = 0.20; permutations = 1,000; p = 0.14), and microsatellite-based F_{ST} and Euclidian distance (spearman's rho = -0.04; permutations = 1,000; p = 0.52).

Appendix C

Supplement for Chapter 6

Table C.1: List of primers used for multiplex PCR.

Locus Type	Primer ID	Primer Sequence
Anonymous	Goby1_F	GGCGCACCATTAGAAAGTAGAA
Anonymous	Goby1_R	AAAACTGGGCTATTATGGTCAG
Anonymous	Goby6_F	CAAAAATGGAAATGGTCCAAAT
Anonymous	Goby6_R	CTCCCGCCCTACTATATGAATG
Anonymous	Goby8_F	TGATGGAAAGTAACAGAAAAAGTACA
Anonymous	Goby8_R	GGACTGAAGATCGCAAAAGG
Anonymous	Goby26_F	ACAATGTGGACCGAGAGCTT
Anonymous	Goby26_R	AAAGCTTTTAATCACGCTTAAATTAG
Anonymous	Goby27_F	ACTTGATGTGTGAATGCAGAGG
Anonymous	Goby27_R	GGATGGATAGATGGATAAAACCAG
Anonymous	Goby32_F	ACCTCCAGCAACAGGTAAACAG
Anonymous	Goby32_R	TTGAAGCTTTATAACTGCAACATT
Anonymous	Goby34_F	AAAGCTTGACGACACAAGATCA
Anonymous	Goby34_R	AACACAGCTGGAGAGAGAGAAGA
Anonymous	Goby35_F	ACGCAACGAGTTTAAAGTGTCC
Anonymous	Goby35_R	ACAACGGCTGAAGGAGGAG
Anonymous	Goby41_F	GCTCCAATACTTTTCTAGACAATCC
Anonymous	Goby41_R	GGATCATTACCACCTGAAGCAT
Anonymous	Goby43_F	GCTGCTGTTGAAGAGCTTACC
Anonymous	Goby43_R	CAGCAACAATTGGTCTTGTGTT
Anonymous	Goby44_F	GACCAGCTCCGTGTTCGT
Anonymous	Goby44_R	ATCGTCAGATCCAAACGCTCT
Anonymous	Goby48_F	TTTTGGTTGCTTTTGTTACAATTC
Anonymous	Goby48_R	ACTGCACGGTTTCCTGTTTTAT
Anonymous	Goby50_F	AAAAGTTGCAAAGTGTTGCTTT
Anonymous	Goby50_R	TGGCTACAGAAGGAGCTTATCC
Anonymous	Goby52_F	AGAACAACCAATCAGGGGAAAC
Anonymous	Goby52_R	CCACAGACGGTGGAAACTTTTA
Anonymous	Goby53_F	TCTCACAAAGAAAGGTCCCATTA
Anonymous	Goby53_R	GATGAACATGTGCGAGTCGTAA
Anonymous	Goby55_F	CCTAGCCTGCAGCTTGTAAAGT
Anonymous	Goby55_R	TCCGGACTATAGATAGGTTTCCA
Anonymous	Goby56_F	TAAATGCAGCAGTTCAGACGAT
Anonymous	Goby56_R	GACTCTCTGCTCGTGAAGCTG
Anonymous	Goby57_F	GCTTTCATTCCCGATCAATATC
Anonymous	Goby57_R	TGGATTTAGTCACAGTCAGAGTGAA

Table C.1: List of primers used for multiplex PCR.

Locus Type	**Primer ID**	**Primer Sequence**
Anonymous	Goby59_F	TCACACAATAAAACACTATAACATGC
Anonymous	Goby59_R	TCCGAAAATGCAGTTTGCTT
Anonymous	Goby63_F	CTGTGAAGGAGACCACTTGTTG
Anonymous	Goby63_R	CCGAAAGATTGGTATTTGTTGG
Anonymous	Goby64_F	AAAAAGGAAAGCGGGAGTGT
Anonymous	Goby64_R	CCCCTGGTGAGTGAAGAAAAAT
Anonymous	Goby66_F	TGTTAGTGTGAGCATTAGGAGACT
Anonymous	Goby66_R	TGCATTTGGAAAAATGTTATTCTCT
Anonymous	Goby70_F	AAATGCGATTGGAAAATTCTGT
Anonymous	Goby70_R	TGTGCTGTTGTCAATTGTTGAT
Anonymous	Goby71_F	GCCTGGAGGTTGAGAGAGATTT
Anonymous	Goby71_R	TCATTAAAATCTGAGATGCTCTTCTT
Anonymous	Goby72_F	GCAACAGACAGAACATTGAGGT
Anonymous	Goby72_R	GGTGATGACTGTGCCTCATAGA
Anonymous	Goby76_F	TTAAACTTGATGTGTGAATGCAG
Anonymous	Goby76_R	AAAGGATGGATAGATGGATAAAACC
Anonymous	Goby77_F	ATTACAAACAGCTGAGGGCTGA
Anonymous	Goby77_R	GGTGAACTTTAGTGATAAACTTTGTG
Anonymous	Goby78_F	GGCACTGCCTCTAATCTTTCTG
Anonymous	Goby78_R	ATAAAGGACACGTGGGGTTGT
Anonymous	Goby83_F	CAGAACTGGCATCTCTTGCTT
Anonymous	Goby83_R	GTCTGCTCCGTCATGTTGAAT
Anonymous	Goby87_F	CATACTGCCCAACAAAACAAAT
Anonymous	Goby87_R	AGGACAGAGCTGGACCTCCT
Anonymous	Goby90_F	TGGATATTCTTACGGGAATTTAGC
Anonymous	Goby90_R	ATATTGTGACCGTCAGCGAAG
Anonymous	Goby91_F	TAAAAAGTGCCATCATTCACAG
Anonymous	Goby91_R	TTCAGCCACTTGTATCCACGAT
Anonymous	Goby92_F	GCGCTGTAGACCTCCACATTAT
Anonymous	Goby92_R	GGAATGTATTTCGATTCCATTTG
Anonymous	Goby93_F	TCACGTCTCGTTACATCAAACA
Anonymous	Goby93_R	AAAGCTCCCTCTAGTGGACAAA
Anonymous	Goby95_F	GACGTCCAAGTCAAAGATGATTC
Anonymous	Goby95_R	TTAAGAAACTGAAGAGGATTCTCCAT
Anonymous	Goby96_F	GCAATGTGTCGCTCCTTAATGT
Anonymous	Goby96_R	GCTCCACTTCAATCTCTTAATCA
Anonymous	Goby100_F	AGTGTGTCCTTTCGAGTTTTGG
Anonymous	Goby100_R	TCACAAACTATTGTTCGGGAAA
Anonymous	Goby103_F	CATCACGGTTTGTGTTGCAT
Anonymous	Goby103_R	GGTTTTCGGTAAAGAATGCTGA
Anonymous	Goby106_F	TGTGCTTTTGAGTGAGAGGAGA
Anonymous	Goby106_R	TAATGATGATGATGATGCACGA
Anonymous	Goby107_F	CCAGAATAATGATCTAAGTCTTCTCA
Anonymous	Goby107_R	TGTGATGGTATTTCTTTACAGTTTGG
Anonymous	Goby108_F	ATTGCGGTGATAACTGCAGAAG
Anonymous	Goby108_R	CATGACAGCCACCTACTGGAC
Anonymous	Goby112_F	GAGGTCTTGCCTCTTCTTGG
Anonymous	Goby112_R	CCATGGGTCTTAAAACAGGATAA
Anonymous	Goby113_F	GGAAAGAAGCCTTAGAATTGTG
Anonymous	Goby113_R	TCGATCCATACAACAGTCAACA

Table C.1: List of primers used for multiplex PCR.

Locus Type	**Primer ID**	**Primer Sequence**
Anonymous	Goby117_F	GAGGACACGGCACGAGAG
Anonymous	Goby117_R	ATCACAAAAATGGTACCGACTG
Anonymous	Goby122_F	CGTCTGTGAGCTGTGGAGTAAT
Anonymous	Goby122_R	GAGGAGAAAGCAGAAGCCAAAT
Anonymous	Goby124_F	CTCAGTCTGATCTGAACCCTGA
Anonymous	Goby124_R	GTCGTGCATTCTGTGAACTCTT
Anonymous	Goby129_F	GCCTGAACCCCCTTTTACAC
Anonymous	Goby129_R	ACTGCTGCTGGATCTCCTCCT
Anonymous	Goby131_F	CACACACACACACAAGAGCTTC
Anonymous	Goby131_R	TGGCTTGTTGTGGAGACAATAG
Anonymous	Goby137_F	GGGTCAATCATAGATGTCTCCTTT
Anonymous	Goby137_R	CCAAGCACTTCGACTTTTACAG
Anonymous	Goby141_F	TCACCTCATTATGGAACACGTC
Anonymous	Goby141_R	AGTGTAGGCCTTAAGTGCGAAG
Anonymous	Goby143_F	CATCCGAGACATTTCCCTTTAC
Anonymous	Goby143_R	CTTAGTCCAATTTCTCGGGTTG
Anonymous	Goby151_F	AGACGAGATGTTTTTCGCAGA
Anonymous	Goby151_R	GGCTCTACTATAGGAAGGCCACA
Anonymous	Goby152_F	GATCTGCTTGAGCTTCACCAC
Anonymous	Goby152_R	TCACACTAAATCCCCCACTTTT
Anonymous	Goby153_F	TGCATATAAATAGAAATTCATTCAGC
Anonymous	Goby153_R	CTGCAACAAACACGTTCACCT
Anonymous	Goby154_F	AGGAGCCGTCTTTACCTTCAG
Anonymous	Goby154_R	ACAACTCAGACTCAGGCTTTGA
Anonymous	Goby156_F	GAAATGATGTGGGGATTCATCT
Anonymous	Goby156_R	CCTGGTCTAACAGGGACTTCTG
Anonymous	Goby160_F	GCAAGTGCCCTTTTTCAAATTA
Anonymous	Goby160_R	GGTGAAAAGAAATATGATTCCAACA
Anonymous	Goby162_F	GAACCACGAAGTAAAACATGGA
Anonymous	Goby162_R	TGTCAGTGAACACAAACTAGCAAA
Anonymous	Goby166_F	CAGTTGGTGCAGAAGGACTGTA
Anonymous	Goby166_R	CACAGGATCTCCAAAAGCAGAC
Anonymous	Goby170_F	GGTAATCAGCTGAGGGATGAAG
Anonymous	Goby170_R	TTTCCTTTGTTCAGGTTGTTGTT
Anonymous	Goby174_F	CTCAAAATCATCGAGCAATCAA
Anonymous	Goby174_R	TCATAAATCTCCCTGAGGCAAC
Anonymous	Goby177_F	CCCCTGTACGTCTTCACATTAAC
Anonymous	Goby177_R	GAACCAATTCATTTGCTGTGAG
Anonymous	Goby180_F	TGATTACGCAGGGAAAAGGTAT
Anonymous	Goby180_R	CATTTTGACATGAGGGCATTTA
Anonymous	Goby183_F	GGCTCACTTTCCAAACTCAAAT
Anonymous	Goby183_R	AAAACATAACACTGAGGTGGACTC
Anonymous	Goby186_F	CTCCATGTAAGGATTGGTGCAT
Anonymous	Goby186_R	CACAGAGCAGGAGTCTCTCTCA
Anonymous	Goby187_F	GACAAGCTTCAAAGGCAAAATAA
Anonymous	Goby187_R	AGGGTACCCATCCCCAAAT
Anonymous	Goby194_F	GCTTTTATGGTTTGCTCTTCCA
Anonymous	Goby194_R	GCACCTAATGTTTCCAGTGTCA
Anonymous	Goby198_F	ATTAAAACGAGGCACGTGATTT
Anonymous	Goby198_R	CAGCTAACCCTAAAATGTGATGA

Table C.1: List of primers used for multiplex PCR.

Locus Type	**Primer ID**	**Primer Sequence**
Anonymous	Goby202_F	TTCTGTTTCTCTCTCAGTGAATGC
Anonymous	Goby202_R	AACATGAAACGCGATCAAGAG
Anonymous	Goby205_F	ATCTCCCAAAGACCAAAGTACG
Anonymous	Goby205_R	TTACGACAATGAATGCTGTTTTT
Anonymous	Goby206_F	TTCATGAACCCTTATCGTTTG
Anonymous	Goby206_R	AACGCACGTGTATCTGATGTGT
Anonymous	Goby216_F	TAAACCACGTTGGATCTGGAG
Anonymous	Goby216_R	TCAGCTCCTATAAACAGAATGAGC
Anonymous	Goby228_F	TTCATGAACCTCTGAGAGCAG
Anonymous	Goby228_R	CCCATCCTCGCACCAAAG
Anonymous	Goby231_F	AAGTTTTGGAGTGGACAGGGTA
Anonymous	Goby231_R	TAAGAACTTGGAATGGCTCTCC
Anonymous	Goby237_F	CATTAATTTGAGCACTGCTGCTT
Anonymous	Goby237_R	GAACCGAGCTGAGTGTGAAG
Anonymous	Goby238_F	TGAAACATCAAAAACATTTAAACAA
Anonymous	Goby238_R	GAACGCTCTTCATGCACATTTA
Anonymous	Goby242_F	CACAGCTGCAATTTGTGAGTTT
Anonymous	Goby242_R	CACACTGAGGGAAATACGATGA
Anonymous	Goby243_F	GTGTGAATGCAGAGGGTTGTT
Anonymous	Goby243_R	ACTGGCCTTTTATTGTCTCCAA
Anonymous	Goby258_F	AACTCATTTGCATTCATGTTTCA
Anonymous	Goby258_R	GGACGCTCAAACATACGAGTTTA
Anonymous	Goby259_F	CAGCCTCCACTCAACTCCAG
Anonymous	Goby259_R	GAGCGAGGAGTAAGAACTGTCC
Anonymous	Goby260_F	TCAGTCAATCCAACAAACCTATTG
Anonymous	Goby260_R	CACATGTGAGCAGTGTTTTCAG
Anonymous	Goby264_F	CCCTGACTCTACCTGAGCTACG
Anonymous	Goby264_R	ATCTCCACTGAAGACCAGAGGT
Anonymous	Goby266_F	TCATTGAAATCTGGAAGGTTAAAA
Anonymous	Goby266_R	AACGCCTTATACATAAAATGCTCTC
Anonymous	Goby267_F	TTATAGCCCTCGACTAGCCTTG
Anonymous	Goby267_R	AATGTGTCCAATGTTGGATTTG
Anonymous	Goby270_F	CAATTGCAACAAAACTCCTCCT
Anonymous	Goby270_R	AGAAGAAATATGGGCTCTCTGC
Anonymous	Goby272_F	GCGCTGTTTAAAGGTTCAATTA
Anonymous	Goby272_R	AGCTGATCTATCCCAGGAACTG
Anonymous	Goby280_F	GCAGGTGCAGAACATAGACG
Anonymous	Goby280_R	TTCTTTTGTAGTGTTGGAGCTCAG
Anonymous	Goby285_F	GCGAAAAGGAAGCCATAATACA
Anonymous	Goby285_R	CACCAGATCTCTCAGACCAGTG
Anonymous	Goby287_F	CATGCAGCTAAGGTCTACAACG
Anonymous	Goby287_R	ACCTCTTTGGATAATGGCTTGA
Anonymous	Goby289_F	GTATGAAAGCACTTGCTGTGGA
Anonymous	Goby289_R	ATGCTGTACAATCTGCTCAACG
Anonymous	Goby291_F	TGATCAAAAATACTGTTGCCAAA
Anonymous	Goby291_R	GCATGTTAATAAAGAGGCATGGA
Anonymous	Goby294_F	TGCTCTACAATATGGAACTCCTTG
Anonymous	Goby294_R	CTGGGAGGTACTACCCCACTAC
Anonymous	Goby300_F	GTAAGTCGTGTCCCTGGCTATG
Anonymous	Goby300_R	ACCAGTACGTGAAAGAAAAGCA

Table C.1: List of primers used for multiplex PCR.

Locus Type	**Primer ID**	**Primer Sequence**
Anonymous	Goby307_F	ATGAACACTCACCAGCCTGATT
Anonymous	Goby307_R	TGATGCTCTTTAATTTATGAACCAGA
Anonymous	Goby308_F	TGCATATTCTAGTCAGCCTTCA
Anonymous	Goby308_R	CTGGGATTGAAACGGCATTAT
Anonymous	Goby312_F	AAGGAACACATCAAAACCCTGT
Anonymous	Goby312_R	CACAGAAAGCTGAAACAAAAACA
Anonymous	Goby314_F	GTTTGTTTACCCCAAGAAGCAG
Anonymous	Goby314_R	AAAAATTGCGCAGTGTATTGTG
Anonymous	Goby319_F	GGTCCTGGTTGGTACTCACTGT
Anonymous	Goby319_R	GGTCAAGCACATCCCAAACAT
Anonymous	Goby320_F	AGTTCACTCACGACTCAGGAAA
Anonymous	Goby320_R	AAAGCCACTTGTTCTGTGTTACC
Anonymous	Goby334_F	AACAGGGAAATTCTGGGTTTTT
Anonymous	Goby334_R	GTGACCTCAGTGCTTCCTCTG
Anonymous	Goby335_F	CCATGAGTTCAGAAACATTGTG
Anonymous	Goby335_R	GTTGAAAATGTCCGAAAATCCT
Anonymous	Goby338_F	TGATCACTTCCTACGTGTTGTTTT
Anonymous	Goby338_R	GACATTTTACTGGATAAACTTTGTGA
Anonymous	Goby340_F	AGCTACACAGCACACAGGCTAA
Anonymous	Goby340_R	TACTTCGACCACTGCAGGTATG
Anonymous	Goby342_F	ATCCCAGGTTTCACACTGAGTC
Anonymous	Goby342_R	GAACTGCACAAAACCCTCTGAT
Anonymous	Goby345_F	TTTCATTTGACATTATGAGGCTGT
Anonymous	Goby345_R	TCCCTTTTGGTAAGAGCAATGT
Anonymous	Goby169_F	TTATTCCGTTTCCTTTGTTGCT
Anonymous	Goby169_R	CTGTCACACTCTCCTCCTCTCA
Anonymous	Goby190_F	AACACACAGAATAAACAGAAGAAAAA
Anonymous	Goby190_R	TCCTAAGAAAACATAAAGGAACTTCA
Anonymous	Goby211_F	AAACTAAATGTGTCGGGGAGTG
Anonymous	Goby211_R	GCTGACGTTCGTTTTAATTCTG
Anonymous	Goby336_F	CAGATGAAGATGTCCCTCAGC
Anonymous	Goby336_R	TGTTTGTAAGGTTGATTTTGTTCC
mtDNA	GobycytB_F	GGCCGCCCTACGAAAAACCC
mtDNA	Gobycytb_R	TAGAGGGAAAAAGGCCAAGAAAATAGAAA
mtDNA	Gobycytb2_F	CCAACAGGCCTTAACTCAGATGCAGAC
mtDNA	Gobycytb2_R	AGCCAGTCTGATGTAAGAATAAGAGATGAAGA
Microsatellite	Goby1419_F	GCGACGGGGAGCCTCAAAT
Microsatellite	Goby1419t_R	ATGATTCGGCCGATACGATGGA
Microsatellite	Goby14528_F	GCGCATGAGCCCGTTTTT
Microsatellite	Goby14528_R	GTGCCACCGGTTGCTCTTG
Microsatellite	Goby18144_F	GACCCGGATTAGTCCTGGTTTG
Microsatellite	Goby18144_R	CGGAGTAAATGTTGGCTCAC
Microsatellite	Goby2137_F	CCTCCCTCCCCAGCACA
Microsatellite	Goby21378_R	TTTGTCCAAGTCTAGCAGGTATTC
Microsatellite	Goby23415_F	CACCGTTATTGACTAAAGTGTTCT
Microsatellite	Goby23415_R	ACAGATCCAGATCATCATCCA
Microsatellite	Goby23889_F	GGTGACTCCCGTGTGAAGAGC
Microsatellite	Goby23889_R	AACATGCTATGGCTAACACTGACG
Microsatellite	Goby25362_F	GCTAAAGGCGGTCCATCAAAA
Microsatellite	Goby25362_R	CAGCGTCCAACAGTGTCTTCAGT

Table C.1: List of primers used for multiplex PCR.

Locus Type	**Primer ID**	**Primer Sequence**
Microsatellite	Goby25632_F	TCCAGTGCTATTTGTGGCATTGTT
Microsatellite	Goby25632_R	CTTATTCCCGGTTCCTCCACTGAT
Microsatellite	Goby25745_F	GGGGGAGGCATGTGATGT
Microsatellite	Goby25745_R	AAACTTCCTCTCGCCATAGTGA
Microsatellite	Goby29109_F	AGCCGGGGATTTAGGCAGGAATAA
Microsatellite	Goby29109_R	GAGAGGCTAAATTGATTAAATGCTC
Microsatellite	Goby6231_F	CCAGGGATAGTTTCACGATGTT
Microsatellite	Goby6231_R	ATGGATTTCCCCTTCAGTGTTT
Microsatellite	Goby6266_F	GTAGACGGACGTGTGAGGTTGTT
Microsatellite	Goby6266_R	CCCCCCCAGACTTGTGAATAATGTGA
Microsatellite	Goby6326_F	GGTCCATGGTCCCAAAGAAAC
Microsatellite	Goby6326_R	CAATGTTAGCGCCCAATGTTCG
Microsatellite	Goby985_F	TTTCCCTGCAGCTGTCAGACT
Microsatellite	Goby985_R	AAAGCTCCACATCCGATTTCA

Table C.2: Five multiplex PCR groups (M1 → M5). The name of each locus is provided. The two mitochondrial cytochrome b loci, denoted with an *, were added at 0.5X concentration, given their increased copy number.

M1	**M2**	**M3**	**M4**	**M5**
1419tri	18144tri	21378tri	25362tri	6266tri
14528tet	23889tri	25632tet	29109tet	6326tet
23415tet	985tet	25745tet	6231tet	280
1	70	cytb1*	cytb2*	285
6	71	122	187	287
8	72	124	194	289
26	76	129	198	291
27	77	131	202	294
32	78	137	205	300
34	83	141	206	307
35	87	143	216	308
41	90	151	228	312
43	91	152	231	314
44	92	153	237	319
48	93	154	238	320
50	95	156	242	334
52	96	160	243	335
53	100	162	258	338
55	103	166	259	340
56	106	170	260	342
57	107	174	264	345
59	108	177	266	169
63	112	180	267	190
64	113	183	270	211
66	117	186	272	336

Table C.3: Belizean *E. lori* individuals sequenced at *COI*.

Region	**Individual**
Central Barrier Reef	B15_10
Central Barrier Reef	B15_11
Central Barrier Reef	B15_12
Central Barrier Reef	B15_13
Central Barrier Reef	B15_14
Central Barrier Reef	B15_15
Central Barrier Reef	B15_16
Central Barrier Reef	B15_9
Central Barrier Reef	B18_17

Table C.3: Belizean *E. lori* individuals sequenced at *COI*.

Region	**Individual**
Central Barrier Reef	B18_18
Central Barrier Reef	B18_19
Central Barrier Reef	B18_20
Central Barrier Reef	B18_21
Central Barrier Reef	B18_22
Central Barrier Reef	B18_23
Central Barrier Reef	B18_24
Central Barrier Reef	B19_1
Central Barrier Reef	B19_19
Central Barrier Reef	B19_2
Central Barrier Reef	B19_20
Central Barrier Reef	B19_21
Central Barrier Reef	B19_22
Central Barrier Reef	B19_23
Central Barrier Reef	B19_24
Central Barrier Reef	B19_3
Central Barrier Reef	B19_4
Central Barrier Reef	B19_5
Central Barrier Reef	B19_6
Central Barrier Reef	B19_7
Central Barrier Reef	B19_8
Central Barrier Reef	B20_17
Central Barrier Reef	B20_19
Central Barrier Reef	B20_20
Central Barrier Reef	B20_21
Central Barrier Reef	B20_22
Central Barrier Reef	B20_23
Central Barrier Reef	B20_24
Southern Barrier Reef	B22_1
Southern Barrier Reef	B22_2
Southern Barrier Reef	B22_3
Southern Barrier Reef	B22_4
Southern Barrier Reef	B23_1
Southern Barrier Reef	B23_10
Southern Barrier Reef	B23_11
Southern Barrier Reef	B23_12
Southern Barrier Reef	B23_13
Southern Barrier Reef	B23_14
Southern Barrier Reef	B23_15
Southern Barrier Reef	B23_16
Southern Barrier Reef	B23_17
Southern Barrier Reef	B23_18
Southern Barrier Reef	B23_19
Southern Barrier Reef	B23_2
Southern Barrier Reef	B23_20
Southern Barrier Reef	B23_21
Southern Barrier Reef	B23_22
Southern Barrier Reef	B23_23
Southern Barrier Reef	B23_24
Southern Barrier Reef	B23_3

Table C.3: Belizean *E. lori* individuals sequenced at *COI*.

Region	**Individual**
Southern Barrier Reef	B23_4
Southern Barrier Reef	B23_5
Southern Barrier Reef	B23_6
Southern Barrier Reef	B23_7
Southern Barrier Reef	B23_8
Southern Barrier Reef	B23_9
Southern Barrier Reef	B24_1
Southern Barrier Reef	B24_10
Southern Barrier Reef	B24_11
Southern Barrier Reef	B24_12
Southern Barrier Reef	B24_13
Southern Barrier Reef	B24_14
Southern Barrier Reef	B24_15
Southern Barrier Reef	B24_16
Southern Barrier Reef	B24_2
Southern Barrier Reef	B24_3
Southern Barrier Reef	B24_4
Southern Barrier Reef	B24_5
Southern Barrier Reef	B24_7
Southern Barrier Reef	B24_8
Southern Barrier Reef	B24_9
Southern Barrier Reef	B25_10
Southern Barrier Reef	B25_11
Southern Barrier Reef	B25_12
Southern Barrier Reef	B25_13
Southern Barrier Reef	B25_14
Southern Barrier Reef	B25_15
Southern Barrier Reef	B25_16
Southern Barrier Reef	B25_17
Southern Barrier Reef	B25_18
Southern Barrier Reef	B25_19
Southern Barrier Reef	B25_21
Southern Barrier Reef	B25_22
Southern Barrier Reef	B25_23
Southern Barrier Reef	B25_24
Southern Barrier Reef	B25_9
Glovers	G1_25
Glovers	G1_26
Glovers	G1_27
Glovers	G1_28
Glovers	G1_29
Glovers	G1_30
Glovers	G1_31
Glovers	G1_32
Glovers	G2_25
Glovers	G2_26
Glovers	G2_27
Glovers	G2_28
Glovers	G2_29
Glovers	G2_30

Table C.3: Belizean *E. lori* individuals sequenced at *COI*.

Region	**Individual**
Glovers	G2_31
Glovers	G2_32
Glovers	G3_25
Glovers	G3_26
Glovers	G3_27
Glovers	G3_28
Glovers	G3_29
Glovers	G3_30
Glovers	G3_31
Glovers	G3_32
Lighthouse	L1_25
Lighthouse	L1_26
Lighthouse	L1_27
Lighthouse	L1_28
Lighthouse	L1_29
Lighthouse	L1_30
Lighthouse	L1_31
Lighthouse	L1_32
Lighthouse	L4_25
Lighthouse	L4_26
Lighthouse	L4_27
Lighthouse	L4_28
Lighthouse	L4_29
Lighthouse	L4_30
Lighthouse	L4_31
Lighthouse	L4_32
Lighthouse	L5_25
Lighthouse	L5_26
Lighthouse	L5_27
Lighthouse	L5_28
Lighthouse	L5_29
Lighthouse	L5_30
Lighthouse	L5_31
Lighthouse	L5_32
Turneffe	T1_10
Turneffe	T1_11
Turneffe	T1_12
Turneffe	T1_13
Turneffe	T1_14
Turneffe	T1_15
Turneffe	T1_16
Turneffe	T1_9
Turneffe	T2_10
Turneffe	T2_11
Turneffe	T2_12
Turneffe	T2_13
Turneffe	T2_14
Turneffe	T2_15
Turneffe	T2_16
Turneffe	T2_9

Table C.3: Belizean *E. lori* individuals sequenced at *COI*.

Region	**Individual**
Turneffe	T5_25
Turneffe	T5_26
Turneffe	T5_27
Turneffe	T5_28
Turneffe	T5_29
Turneffe	T5_30
Turneffe	T5_31
Turneffe	T5_32
Turneffe	T6_25
Turneffe	T6_26
Turneffe	T6_27
Turneffe	T6_28
Turneffe	T6_29
Turneffe	T6_30
Turneffe	T6_31
Turneffe	T6_32

Bibliography

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 (2003) Priority effects in coral reef fish communities. *Ecology* 84: 1920–1935.

Almany GR (2004) Differential effects of habitat complexity, predators and competitors on abundance of juvenile and adult coral reef fishes. *Oecologia* 141: 105–113.

Almany GR, Berumen ML, Thorrold SR, Planes S, Jones GP (2007) Local replenishment of coral reef fish populations in a marine reserve. *Science* 316: 742–744.

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.

Anadón JD, Mancha-Cisneros MM, Best BD, Gerber LR (2013) Habitat-specific larval dispersal and marine connectivity: Implications for spatial conservation planning. *Ecosphere* 4: 82.

Andréfouët S, Muller-Karger FE, Robinson JA *et al.* (2005) Global assessment of modern coral reef extent and diversity for regional science and management applications: a view from space. In: Suzuki Y, Nakamori T, Hidaka M *et al.* (eds) *10th International Coral Reef Symposium CDROM*. Japanese Coral Reef Society, Okinawa, Japan, pp 1732–1745.

Andrés JA, Bogdanowicz SM (2011) Isolating microsatellite loci: looking back, looking ahead. *Methods in Molecular Biology* 772: 211–232.

Austerlitz F, Dick CW, Dutech C *et al.* (2004) Using genetic markers to estimate the pollen dispersal curve. *Molecular Ecology* 13: 937–954.

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

Avise, JC, Shapiro DY (1986). Evaluating kinship of newly settled juveniles within social groups of the coral reef fish *Anthias squamipinnis*. *Evolution* 40: 1051–1059.

Baeza JA (2008) Social monogamy in the shrimp *Pontonia margarita*, a symbiont of *Pinctada mazatlanica*, off the Pacific coast of Panama. *Marine Biology* 153: 387–395.

Baeza JA, Stotz W, Thiel M (2001) Life history of Allopetrolisthes spinifrons, a crab associate of the sea anemone Phymactis clematis. *Journal of the Marine Biological Association of the United Kingdom* 81: 69–76.

Balkenhol N, Waits LP, Dezzani RJ (2009) Statistical approaches in landscape genetics: an evaluation of methods for linking landscape and genetic data. *Ecography* 32: 818–830.

Belmaker J, Ziv Y, Shashar N (2009) Habitat patchiness and predation modify the distribution of a coral-dwelling damselfish. *Marine Biology* 156: 447–454.

Belmaker J, Ben-Moshe N, Ziv Y, Shashar N (2007) Determinants of the steep species-area relationship of coral reef fishes. *Coral Reefs* 26: 103–112.

Ben-Tzvi O, Abelson A, Polak O, Kiflawi M (2008) Habitat selection and the colonization of new territories by *Chromis viridis*. *Journal of Fish Biology* 73: 1005–1018.

Ben-Tzvi O, Abelson A, Gaines SD *et al.* (2012) Evidence for cohesive dispersal in the sea. *Plos ONE* 7: e42672.

Berry O, England P, Marriott RJ, Burridge CP, Newman SJ (2012) Understanding age-specific dispersal in fishes through hydrodynamic modelling, genetic simulations and microsatellite DNA analysis. *Molecular Ecology* 21: 2145–2159.

Berthouly-Salazar C, Hui C, Blackburn TM, *et al.* (2013) Long-distance dispersal maximizes evolutionary potential during rapid geographic range expansion. *Molecular Ecology* 22: 5793–5804.

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.

Blowes SA, Connolly SR (2012) Risk spreading, connectivity, and optimal reserve spacing. *Ecological Applications* 22: 311–321.

Booth DJ (1992) Larval settlement-patterns and preferences by domino damselfish *Dascyllus albisella* Gill. *Journal of Experimental Marine Biology and Ecology* 155: 85–104.

Booth DJ (1995) Juvenile groups in a coral-reef damselfish: density dependent effects on individual fitness and population demography. *Ecology* 76: 91–106.

Botsford LW, Hastings A, Gaines SG (2001) Dependence of sustainability on the configuration of marine reserves and larval dispersal distance. *Ecology Letters* 4: 144–150.

Botsford LW, White JW, Coffroth MA *et al.* (2009) Connectivity and resilience of coral reef metapopulations in marine protected areas: matching empirical efforts to predictive needs. *Coral Reefs* 28: 327–337.

Broquet T, Petit EJ (2009) Molecular estimates of dispersal for ecology and population genetics. *Annual Review of Ecology, Evolution, and Systematics* 40: 193–216.

Brothers EB, Thresher RE (1985) Pelagic duration, dispersal, and the distribution of Indo-Pacific coral reef fishes. In: Reaka ML (ed) *The Ecology of Coral Reefs*. NOAA, Washington, D.C., pp 53–69.

Brownstein MJ, Carpten D, Smith JR (1996) Modulation of non-templated nucleotide addi-

tion by Taq DNA polymerase: primer modifications that facilitate genotyping. *BioTechniques* 20: 1004–1010.

Burgess SC, Treml EA, Marshall DJ (2012) How do dispersal costs and habitat selection influence realized population connectivity? *Ecology* 93: 1378–1387.

Burgess SC, Nickols KJ, Griesemer CD *et al.* (2014) Beyond connectivity: how empirical methods can quantify population persistence to improve marine protected-area design. *Ecological Applications* 24: 257–270.

Burnham KP, Anderson DR (2002) *Model selection and multimodel inference: a practical information-theoretic approach.* Springer, New York.

Buston PM (2002) Group structure of the clown anemonefish, *Amphiprion percula*. PhD thesis, Cornell University.

Buston PM (2003) Forcible eviction and prevention of recruitment in the clown anemonefish *Behavioral Ecology* 14: 576–582.

Buston PM (2004) Territory inheritance in the clown anemonefish. *Proceedings of the Royal Society of London, Series B* 271: S252–S254.

Buston PM, Bogdanowicz SM, Wong A, Harrison RG (2007) Are clownfish groups composed of close relatives? An analysis of microsatellite DNA variation in *Amphiprion percula*. *Molecular Ecology* 16: 3671–3678.

Buston PM, Fauvelot C, Wong MYL, Planes S (2009) Genetic relatedness in groups of the humbug damselfish *Dascyllus aruanus*: small, similar-sized individuals may be close kin. *Molecular Ecology* 18: 4707–4715.

Buston PM, Elith J (2011) Determinants of reproductive success in dominant pairs of clownfish: a boosted regression tree analysis. *Journal of Animal Ecology* 80: 528–538.

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.

Christie MR, Tissot BN, Albins MA *et al.* (2010) Larval connectivity in an effective network of marine protected areas. *PLoS ONE* 5: e15715.

Clark JS, Silman M, Kern R, Macklin E, HilleRis-Lambers JH (1999) Seed dispersal near and far: patterns across temperate and tropical forests. *Ecology* 80: 1475–1494.

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

Clobert J, Danchin E, Dhondt AA, Nichols JD, eds (2001) *Dispersal.* Oxford University

Press, New York.

Clobert J, Baguette M, Benton TG, Bullock JM, Ducatez S, eds (2012) *Dispersal ecology and evolution*. Oxford University Press, New York.

Colin PL (1975) *Neon gobies*. T.F.H Publications, Inc., Neptune City, NJ.

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

Colin PL (2010) Fishes as living tracers of connectivity in the tropical western North Atlantic: I. Distribution of the neon gobies, genus *Elacatinus* (Pisces: Gobiidae). *Zootaxa* 2370: 36–52.

Colleye O, Brié C, M Malpot E, Vandewalle P, Parmentier E (2008) Temporal variability of settlement in Carapidae larvae at Rangiroa atoll. *Environmental Biology of Fishes* 81: 277–285.

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.

Cowen RK, Gawarkievicz PJ, Thorrold SR, Werner FE (2007) Population connectivity in marine systems: an overview. *Oceanography* 20: 14–21.

Cowen RK, Sponaugle S (2009) Larval dispersal and marine population connectivity. *Annual Review of Marine Science* 1: 443–466.

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) Selfrecruitment in a Caribbean reef fish: a method for approximating dispersal kernels accounting for seascape. *Molecular Ecology* 22: 2563–2572.

D'Aloia CC, Bogdanowicz SM, Harrison RG, Buston PM (2014) Seascape continuity plays an important role in determining patterns of spatial genetic structure in a coral reef fish. *Molecular Ecology* 23:2902–2913.

Davison AC, Hinkley D (1997) *Bootstrap Methods and Their Application*. Cambridge University Press, Cambridge.

Duckworth AR, Wolff CW (2007) Patterns of abundance and size of Dictyoceratid sponges among neighbouring islands in central Torres Strait, Australia. *Marine Freshwater Research* 58: 204–212.

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.

Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research* 32: 1792–1797.

Elliott JK, Mariscal RN (2001) Coexistence of nine anemonefish species: differential host and habitat utilization, size and recruitment. *Marine Biology* 138: 23–36.

Elliot JK, Elliot JM, Mariscal RN (1995) Host selection, location and association behaviors of anemonefishes in field settlement experiments. *Marine Biology* 122: 377–389.

Emlen ST (1982) The evolution of helping. I. An ecological constraints model. *American Naturalist* 119: 29–39.

Emlen ST (1991) The evolution of cooperative breeding in birds and mammals. In: Krebs JR, Davies NB (eds) *Behavioural ecology: an evolutionary approach.* 3rd ed, Blackwell, Oxford, pp 301–307.

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.

Ezer T, Thattai DV, Kjerfve B, Heyman WD (2005) On the variability of the flow along the Meso-American Barrier Reef system: a numerical model study of the influence of the Caribbean current and eddies. *Ocean Dynamics* 55: 458–475.

Fahrig L, Merriam G (1994) Conservation of fragmented populations. *Conservation Biology* 8: 50–59.

Fautin DG (1992) Anemonefish recruitment: the roles of order and chance. *Symbiosis* 14: 143–160.

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

Foster NL, Paris CB, Kool J *et al.* (2012) Connectivity of Caribbean coral populations: complementary insights from empirical and modelled gene flow. *Molecular Ecology* 21: 1143–1157.

Galindo HM, Olson DB, Palumbi SR (2006) Seascape genetics: A coupled oceanographic-genetic model predicts population structure of Caribbean corals. *Current Biology* 16: 1622–1626.

García C, Jordano P, Godoy JA (2007) Contemporary pollen and seed dispersal in a *Prunus mahaleb* population: patterns in distance and direction. *Molecular Ecology* 16: 1947–1955.

Gardiner NM, Jones GP (2010) Synergistic effects of habitat preference and gregarious behaviour on habitat use in coral reef cardinalfish. *Coral Reefs* 29: 845–856.

González-Sansn G, Aguilar C, Hernández I, Cabrera Y (2009) Effects of depth and bottom communities on the distribution of highly territorial reef fish in the northwestern region of Cuba. *Journal of Applied Ichthyology* 25: 652–660.

Gower JC (1971) A general coefficient of similarity and some of its properties. *Biometrics* 27: 857–871.

Greene DF, Johnson EA (1989) A micrometeorological model of the dispersal of winged or plumed seeds from a point source. *Ecology* 70: 339–347.

Greene DF, Canham CD, Coates KD, Lepage PT (2004) An evaluation of alternative dispersal functions for trees. *Journal of Ecology* 92: 758–766.

Hanski I (1999) *Metapopulation Ecology.* Oxford University Press, New York.

Hanski IA, Gaggiotti OE, eds (2004) *Ecology, genetics and evolution of metapopulations.* Elsevier Academic Press, London.

Hanski IA, Gaggiotti OE (2004) Metapopulation biology: Past, present, and future. In: Hanski I, Gaggiotti OE (eds) *Ecology, genetics and evolution of metapopulations.* Elsevier Academic Press, London, pp 3–22.

Harrison HB, Williamson DH, Evans RD (2012) Larval export from marine reserves and the recruitment benefit for fish and fisheries. *Current Biology* 22: 1023–1028.

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.

Hedgecock D, Barber PH, Edmands S (2007) Genetic approaches to measuring connectivity. *Oceanography* 20: 70–79.

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.

Hixon MA, Jones GP (2005) Competition, predation, and density dependent mortality in demersal marine fishes. *Ecology* 86: 2847–2859.

Hoegh-Guldberg O, Mumby PJ, Hooten AJ *et al.* (2007) Coral reefs under rapid climate change and ocean acidification. *Science* 318: 1737–1742.

Holbrook SJ, Schmitt RJ (2002) Competition for shelter space causes density dependent predation mortality in damselfishes. *Ecology* 83: 2855–2868.

Holbrook SJ, Schmitt RJ (2005) Growth, reproduction, and survival of a tropical sea anemone (Actinaria): benefits of hosting anemonefish. *Coral Reefs* 24: 67–73.

Hosmer Jr, DW, Lemeshow S, Sturdivant RX (2013) *Applied logistic regression.* Wiley, Hoboken, NJ.

Huret M, Petitgas P, Woillez M (2010) Dispersal kernels and their drivers captured with a hydrodynamic model and spatial indices: a case study on anchovy (*Engraulis encrasicolus*) early life stages in the Bay of Biscay. *Progress in Oceanography* 87: 6–17.

IPCC (2014) Climate Change 2014: Impacts, Adaptation, and Vulnerability. Part A: Global and Sectoral Aspects. Contribution of Working Group II to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge.

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 (1982) Chaotic genetic patchiness in an intertidal limpet, *Siphonaria sp. Marine Biology* 70: 157–164.

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.

Jombart T, Devillard S, Balloux F (2010) Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC Genetics* 11: 94.

Jones GP (1986) Food availability affects growth in a coral-reef fish. *Oecologia* 70: 136–139.

Jones GP, Milicich MJ, Emslie MJ, Lunow C (1999) Self-recruitment in a coral reef fish population. *Nature* 402: 802–804.

Jones GP, Planes S, Thorrold SR (2005) Coral reef fish larvae settle close to home. *Current Biology* 15: 1314–1318.

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.

Jordano P, García C, Godoy JA, García-Castaño JL (2007) Differential contribution of frugivores to complex seed dispersal patterns. *Proceedings of the National Academy of Sciences, USA* 104: 3278–3282.

Kalinowski ST, Taper ML, Marshall TC (2007) Revising how the computer program CERVUS accommodates genotyping error increases success in paternity assignment. *Molecular Ecology* 16: 1099–1106.

Kimura M, Weiss GH (1964) The stepping stone model of population structure and the decrease of genetic correlation with distance. *Genetics* 49: 561–576.

Kool JT, Paris CB, Andrefouet S, Cowen RK (2010) Complex migration and the development of genetic structure in subdivided populations: an example from Caribbean coral reef ecosystems. *Ecography* 33: 597–606.

Kritzer JP, Sale PF (eds) (2006) *Marine metapopulations.* Elsevier, Burlington, MA.

Kuwamura T, Yogo Y, Nakashima Y (1994) Population dynamics of goby *Paragobiodon echinocephalus* and host coral *Stylophora pistillata*. *Marine Ecology Progress Series* 103: 17–23.

Lavett Smith CL, Tyler JC, Feinberg MN (1981) Population ecology and biology of the pearlfish (*Carapus bermudensis*) in the lagoon at Bimini, Bahamas. *Bulletin of Marine Science* 31: 876–902.

Legendre P, Lapointe FJ, Casgrain P (1994) Modeling brain evolution from behavior: a permutational regression approach. *Evolution* 48: 1487–1499.

Legendre P, Legendre L (2012) *Numerical Ecology.* 3rd ed, Elsevier, Amsterdam.

Leis JM (1991) The pelagic stage of reef fishes: the larval biology of coral reef fishes. In: Sale P (ed) *The Ecology of Fishes on Coral Reefs.* Academic Press, California, pp. 183–230.

Leis JM, Van Herwerden L, Patterson HM (2011) Estimating connectivity in marine fish populations: what works best? *Oceanography and Marine Biology* 49: 193–234.

Lesser MP (2006) Benthic-pelagic coupling on coral reefs: Feeding and growth of Caribbean sponges. *Journal of Experimental Marine Biology and Ecology* 328: 277–288.

Levins R (1969) Some demographic and genetic consequences of environmental heterogeneity for biological control. *Bulletin of the Entomological Society of America* 15: 237–240.

Levin PS, Tolimieri N, Nicklin M, Sale PF (2000) Integrating individual behavior and population ecology: the potential for habitat dependent population regulation in a reef fish. *Behavioral Ecolology* 11: 565–571.

López E, Britayev TA, Martin D, San Martn G (2001) New symbiotic associations involving Syllidae (Annelida: Polychaeta), with taxonomic and biological remarks on Pionosyllis magnifica and Syllis cf. armillaris. *Journal of the Marine Biological Association of the United Kingdom* 81: 399–409.

Maldonado M, Young CM (1998) Limits on the bathymetric distribution of keratose sponges: a field test in deep water. *Marine Ecology Progress Series* 174: 123–139.

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.

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

Manly BFJ (1986) Randomization and regression methods for testing for associations with geographical, environmental and biological distances between populations. *Researches on Population Ecology* 28: 201–218.

Manly BFJ (1991) Distance matrices and spatial data. In: *Randomization and Monte Carlo Methods in Biology*. Chapman and Hall, New York, pp 112–140.

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.

Matthysen E (2005) Density-dependent dispersal in birds and mammals. *Ecography* 28: 403–416.

Mercado-Molina AE, Yoshioka PM (2009) Relationships between water motion and size-specific survivorship and growth of the demosponge *Amphimedon compressa*. *Journal of Experimental Marine Biology and Ecology* 375: 51–56.

Mitarai S, Siegel DA, Winters KB (2008) A numerical study of stochastic larval settlement in the California Current system. *Journal of Marine Systems* 69: 295–309.

Moffitt EA, White JW, Botsford LW (2011) The utility and limitations of size and spacing guidelines for designing marine protected area networks. *Biological Conservation* 144: 306–318.

Munday PL, Jones GP, Caley MJ (1997) Habitat specialisation and the distribution and abundance of coral-dwelling gobies. *Marine Ecology Progress Series* 152: 227–239.

Nanninga GB, Saenz-Agudelo P, Manica A, Berumen ML (2014) Environmental gradients predict the genetic population structure of a coral reef fish in the Red Sea. *Molecular Ecology* 23: 591–602.

Nathan R (2006) Long-distance dispersal of plants. *Science* 313: 786–788.

Nathan R, Muller-Landau HC (2000) Spatial patterns of seed dispersal, their determinants and consequences for recruitment. *Trends in Ecology and Evolution* 15: 278–285.

Nathan R, Schurr FM, Spiegel O *et al.* (2008) Mechanisms of long-distance seed dispersal. *Trends in Ecology and Evolution* 23: 638–647.

Neigel JE (2002) Is FST obsolete? *Conservation Genetics* 3: 167–173.

Neigel JE, Schmahl GP (1984) Phenotypic variation within histocompatibility-defined clones of marine sponges. *Science* 224: 413–415.

Öhman MC, Munday PL, Jones GP, Caley MJ (1998) Settlement strategies and distribution patterns of coral-reef fishes. *Journal of Experimental Marine Biology and Ecology* 225: 219–238.

Palumbi SR (1984) Tactics of acclimation: Morphological changes of sponges in an unpredictable environment. *Science* 225: 1478–1480.

Paris CB, Helgers J, Van Sebille E, Srinivasan A (2013) Connectivity Modeling System (CMS): A multi-scale tool for the tracking of biotic and abiotic variability in the ocean. *Environmental Modelling & Software* 42: 47–54.

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.

Pinsky ML, Palumbi SR, Andréfouët S, Purkis SJ (2012) Open and closed seascapes: Where does habitat patchiness create populations with high fractions of self-recruitment? *Ecological Applications* 22: 1257–1267.

Planes S, Lenfant P (2002) Temporal change in the genetic structure between and within cohorts of a marine fish, *Diplodus sargus*, induced by a large variance in individual reproductive success. *Molecular Ecology* 11: 1515–1524.

Planes S, Jones GP, Thorrold SR (2009) Larval dispersal connects fish populations in a network of marine protected areas. *Proceedings of the National Academy of Sciences, USA* 106: 5693–5697.

Pflüger FJ, Balkenhol N (2014) A plea for simultaneously considering matrix quality and local environmental conditions when analysing landscape impacts on effective dispersal. *Molecular Ecology* 23: 2146–2156.

Porat D, Chadwick-Furman NE (2004) Effects of anemonefish on giant sea anemones: expansion behavior, growth and survival. *Hydrobiologia* 530/531: 513–520.

Porat D, Chadwick-Furman NE (2005) Effects of anemonefish on giant sea anemones: ammonium uptake, zooxanthellae content and tissue regeneration. *Marine and Freshwater Behaviour and Physiology* 38: 43–51.

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.

Raymond M, Rousset F (1995) GENEPOP (version 1.2): population genetics software for exact tests and ecumenicism. *Journal of Heredity* 8: 248–249.

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.

Ribbens EJ, Silander JA, Pacala SW (1994) Seedling recruitment in forests: calibrating models to predict patterns of tree seedling dispersion. *Ecology*, 75: 1794–1806.

Rieux A, Soubeyrand S, Bonnot F *et al.* (2014) Long-Distance Wind-Dispersal of Spores in a Fungal Plant Pathogen: Estimation of Anisotropic Dispersal Kernels from an Extensive Field Experiment. *PLoS ONE* 9: e103225.

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.

Riginos C, Liggins L (2013) Seascape genetics: populations, individuals, and genes marooned and adrift. *Geography Compass* 7: 197–216.

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

Roberts CM, Ormond RFG (1987) Habitat complexity and coral-reef fish diversity and abundance on red-sea fringing reefs. *Marine Ecology Progress Series* 41: 1–8.

Roberts CM, Andelman S, Branch G *et al.* (2003) Ecological criteria for evaluating candidate sites for marine reserves. *Ecological Applications* 13: S199–S214.

Robledo-Arnuncio JJ, García C (2007) Estimation of the seed dispersal kernel from exact identification of source plants. *Molecular Ecology* 16: 5098–5109.

Robledo-Arnuncio JJ (2008) Seed dispersal estimation without fecundities: a reply to Ralph. *Molecular Ecology* 17: 1883–1884.

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

Roderick GK, Caldwell RL (1992) An entomological perspective on animal dispersal. In: Stenseth NC, Lidicker WZ (eds) *Animal dispersal*, Chapman & Hall, London, pp 274–290.

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

Roth MS, Knowlton N (2009) Distribution, abundance, and microhabitat characterization of small juvenile corals at Palmyra Atoll. *Marine Ecology Progress Series* 376: 133–142.

Rousset F (1997) Genetic differentiation and estimation of gene flow from F-statistics under isolation by distance. *Genetics* 145: 1219–1228.

Rützler K, Macintyre IG (1982) The habitat distribution and community structure of the barrier reef complex at Carrie Bow Cay, Belize. In: Rützler K, Macinture IG (eds) *The Atlantic barrier reef ecosystem at Carrie Bow Cay, Belize, I.* Smithsonian Institution Press, Washington, pp 9–45.

Rützler K, Feller C (1999) Mangrove swamp communities: an approach in Belize. In: Yez-Arancibia A, Lara-Domnquez AL (eds) *Ecosistemas de manglar en América Tropical.* Instituto de Ecología A.C. México, UICN/ORMA, Costa Rica, NOAA/NMFS, Silver Spring, MD, pp 39–50.

Saenz-Agudelo P, Jones GP, Thorrold SR, Planes S (2011) Connectivity dominates larval replenishment in a coastal reef fish metapopulation. *Proceedings of the Royal Society of London, Series B* 278: 2954–2961.

Sale PF, Cowen RK, Danilowicz BS *et al.* (2005) Critical science gaps impede use of no-take fishery reserves. *Trends in Ecology and Evolution* 20: 74–80.

Sandin SA, Pacala SW (2005) Fish aggregation results in inversely density-dependent predation on continuous coral reefs. *Ecology* 86: 1520–1530.

Saura S, Bodin Ö, Fortin MJ (2014) Stepping stones are crucial for species' long-distance dispersal and range expansion through habitat networks. *Journal of Applied Ecology* 51: 171–182.

Sbrocco EJ, Barber PH (2013) MARSPEC: ocean climate layers for marine spatial ecology: Ecological Archives E094-086. *Ecology* 94: 979–979.

Sbrocco, EJ (2014) Paleo-MARSPEC: gridded ocean climate layers for the Mid-Holocene and Last Glacial Maximum. *Ecology* 95: 1710.

Schuelke M (2000) An economic method for the fluorescent labeling of PCR fragments. *Nature Biotechnology* 18: 233–234.

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.

Selkoe KA, Watson JR, White C *et al.* (2010) Taking the chaos out of genetic patchiness: seascape genetics reveals ecological and oceanographic drivers of genetic patterns in three temperate reef species. *Molecular Ecology* 19: 3708–3726.

Selkoe KA, Toonen RJ (2011) Marine connectivity: a new look at pelagic larval duration and genetic metrics of dispersal. *Marine Ecology Progress Series* 436: 291–305.

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

Siegel DA, Kinlan BP, Gaylord B, Gaines SD (2003) Lagrangian descriptions of marine larval dispersion. *Marine Ecology Progress Series* 260: 83–96.

Siegel DA, Mitarai S, Costello CJ *et al.* (2008) The stochastic nature of larval connectivity among nearshore marine populations. *Proceedings of the National Academy of Sciences, USA* 105: 8974–8979.

Smith CL, Tyler JC (1972) Space resource sharing in a coral reef fish community. *Bulletin of the Natural History Museum Los Angeles County* 14: 125–178.

Srinivasan M (2003) Depth distributions of coral reef fishes: the influence of microhabitat structure, settlement, and post-settlement processes. *Oecologia* 137: 76–84.

Storfer A, Murphy MA, Evans JS *et al.* (2007) Putting the 'landscape' in landscape

genetics. *Heredity* 98: 128–142.

Sweatman HPA (1983) Influence of conspecifics on choice of settlement sites by larvae of two pomacentrid fishes (*Dascyllus aruanus* and *D. reticulatus*) on coral reefs. *Marine Biology* 75: 225–229.

Sweatman HPA (1985) The influence of adults of some coral-reef fishes on larval recruitment. *Ecological Monographs* 55: 469–485.

Sweatman HPA (1988) Field evidence that settling coral reef fish larvae detect resident fishes using dissolved chemical cues. *Journal of Experimental Marine Biology and Ecology* 124: 163–174.

Szulkin M, Sheldon BC (2008) Dispersal as a means of inbreeding avoidance in a wild bird population. *Proceedings of the Royal Society of London, Series B* 275: 703–711.

Tamura K, Stecher G, Peterson D, Filipski A, Kumar S (2013) MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0. *Molecular Biology and Evolution* 30: 2725–2729.

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.

Tolimieri N (1998) The relationship among microhabitat characteristics, recruitment and adult abundance in the stoplight parrotfish, *Sparisoma viride*, at three spatial scales. *Bulletin of Marine Science* 62: 253–268.

Truett GE, Heeger P, Mynatt RL *et. al.* (2000) Preparation of PCR-quality mouse genomic DNA with hot sodium hydroxide and tris (HotSHOT). *Biotechniques* 29: 52–54.

van Noordwijk AJ, De Jong G (1986) Acquisition and allocation of resources: their influence on variation in life history tactics. *American Naturalist* 128: 137–142.

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.

Vekemans X, Hardy OJ (2004) New insights from finescale 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.

Victor BC (2015) How many coral reef fish species are there? Cryptic diversity and the new molecular taxonomy. In: Mora C (ed) *Ecology of Fishes on Coral Reefs*. Cambridge

University Press, Cambridge, pp 76–87.

Wagner HH, Fortin MJ (2005) Spatial analysis of landscapes: concepts and statistics. *Ecology* 86: 1975–1987.

Wagner HH, Fortin MJ (2013) A conceptual framework for the spatial analysis of landscape genetic data. *Conservation Genetics* 14: 253–261.

Wang IJ (2013) Examining the full effects of landscape heterogeneity on spatial genetic variation: a multiple matrix regression approach for quantifying geographic and ecological isolation. *Evolution* 67: 3403–3411.

Wang IJ, Bradburd GS (2014) Isolation by environment. *Molecular Ecology* 23: 5649–5662.

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.

White JW, Warner RR (2007) Safety in numbers and the spatial scaling of density-dependent mortality in a coral reef fish. *Ecology* 88: 3044–3054.

White JW, Schroeger J, Drake PT, Edwards CA (2014) The value of larval connectivity information in the static optimization of marine reserve design. *Conservation Letters* 7: 533-544.

Whiteman EA, Cotê IM (2002) Cleaning activity of two Caribbean cleaning gobies: intra- and interspecific comparisons. *Journal of Fish Biology* 60: 1443–1458.

Whitlock MC, McCauley DE (1999) Indirect measures of gene flow and migration: $F_{ST} \neq 1/(4Nm+ 1)$. *Heredity* 82: 117–125.

Wilkinson CR, Chesire AC (1989) Patterns in the distribution of sponge populations across the central Great Barrier Reef. *Coral Reefs* 8: 127–134.

Wilkinson CR, Evans E (1989) Sponge distribution across Davies Reef, Great Barrier Reef, relative to location, depth, and water movement. *Coral Reefs* 8: 1–7.

Williams DM, Sale PF (1981) Spatial and temporal patterns of recruitment of juvenile coral reef fishes to coral habitats within One Tree Lagoon, Great Barrier Reef. *Marine Biology* 65: 245–253.

Wilson JA (2005) Age class interactions in a marine goby, *Elacatinus prochilos* (Bohlke and Robins, 1968). *Journal of Experimental Marine Biology and Ecology* 327: 144–156.

Wilson SK, Burgess SC, Cheal AJ *et al.* (2008) Habitat utilization by coral reef fish: implications for specialists vs. generalists in a changing environment. *Journal of Animal Ecology* 77: 220–228.

Wong MYL, Buston PM, Munday P, Jones GP (2007) The threat of punishment enforces peaceful cooperation and stabilizes queues in a coral reef fish. *Proceedings of the Royal Society of London, Series B* 274: 1093–1099.

Wong MYL, Munday PL, Buston PM, Jones GP (2008) Monogamy when there is the potential for polygyny: tests of multiple hypotheses in a group-living fish. *Behavioral Ecology* 19: 353–361.

Wong MYL (2010) Ecological constraints and benefits of philopatry promote group-living in a social but non-cooperatively breeding fish. *Proceedings of the Royal Society of London, Series B* 277:353–358.

World Resources Institute (2005) *Belize Coastal Data CD*. WRI, Washington, DC.

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

Zar JH (1984) *Biostatistical analysis*, 2nd ed, Prentice Hall, Englewood Cliffs, NJ.

Curriculum Vitae