Patterns, causes, and consequences of marine larval dispersal

Cassidy C. D’Aloia1,2, Steven M. Bogdanowicz3, Robin K. Francis4, John E. Majoris5, Richard G. Harrison6, and Peter M. Bustona7

1Department of Biology and Marine Program, Boston University, Boston, MA 02215; 2Department of Ecology and Evolutionary Biology, Cornell University, Ithaca, NY 14853; and 3Department of Ecology, Evolution, and Marine Biology, University of California, Santa Barbara, CA 93106

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 16.4 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 and conserve biodiversity broadly, then reserves that are close in space (<10 km) will accommodate those members of the community that are short-distance dispersers.

Significance

Networks of marine protected areas (MPAs) have been widely implemented to combat global population declines. Although their efficacy largely depends on larval exchange between populations, quantitative analyses of dispersal have been limited due to the difficulties of tracking larvae. Here, we systematically measure dispersal in the fish Elacatinus lori, producing the first robust estimate of a dispersal kernel. We find that dispersal declines exponentially, with most larvae traveling less than 2 km from their parents. Remarkably, the distance an individual travels is unrelated to the number of days it spends in the larval phase. These results suggest that simple distance-based models may be useful conservation tools and that MPAs that are close in space will accommodate short-distance dispersers.

Author contributions: C.C.D., S.M.B., R.G.H., and P.M.B. designed research; C.C.D., S.M.B., R.K.F., J.E.M., and P.M.B. performed research; C.C.D. analyzed data; and C.C.D. wrote the paper.

The authors declare no conflict of interest.

1Present address: Department of Ecology and Evolutionary Biology, University of Toronto, Toronto, ON, Canada M5S3G5.

2To whom correspondence should be addressed. Email: cassidy.dalio@gmail.com.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1513754112/-//DCSupplemental.
until they hatch. Individuals then spend ~26 d in a larval phase—a duration that approximates the median value in reef fishes (30). Upon completing the larval phase, individuals settle onto sponges. The close association with sponges makes the fish easy to locate and capture. Additionally, a pilot study of local dispersal in *E. lori* (≤500 m) hypothesized that a rapid spatial decline in the probability of dispersal might explain the observed level of self-recruitment within a single patch of reef (31). 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.

### Results and Discussion

#### 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. 1). We collected 3,033 tissue samples from potential parents and 4,112 samples from potential offspring (*SI Appendix*, Table S1). We then genotyped individuals using a two-step protocol for 20 microsatellite loci (*Methods* and *SI Appendix*, Table S2). 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 p.d.f.s to the observed distribution of dispersal distances (*SI Appendix*, Table S3), controlling for variation in sampling effort along the transect. The best-fit functional form of the dispersal kernel was exponential as follows:

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

where $f(x) =$ probability density, $x =$ distance (in kilometers), and the decay parameter $\lambda = 0.36$ (95% CI = 0.30, 0.43) (Fig. 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 16.4 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-d larval phase, and therefore the potential to disperse far via ocean currents (22), *E. lori* exhibits a spatially restricted leptokurtic pattern of dispersal.

#### 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 (16). We found subtle evidence of spatial variation in the magnitude of the decay rate (Fig. 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. 2C). Here, PLD is defined as the number of days an individual spends in the larval phase and is measured by counting the daily incremental growth rings deposited in the sagittal otolith, from the hatching mark out to the settlement mark (32). Likewise, no variables related to the microhabitat at origin influenced the kernel shape (*SI Appendix*, Fig. S1). 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 (*SI Appendix*, Table S4). Taken together, these results demonstrate that distance is the primary explanatory variable of the dispersal pattern.

These results lend 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. 2C), despite reported directionality in ocean currents around the Belize barrier reef (BBR) (33). This suggests that dispersal is either isotropic despite anisotropic currents, or that currents were isotropic over the time frame of our study. We also found no evidence to support the hypothesis that long-distance dispersers experience elevated postsettlement mortality rates. If long-distance migrants suffer higher postsettlement mortality rates, the dispersal kernel for larger, older settlers should have a significantly larger decay rate ($\lambda$) 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. 2D). This result should be treated with some caution because size is a coarse proxy for postsettlement age in fishes (see *Methods* for details), although it is reasonable to assume that the small settlers are younger, on average, than the large settlers. Finally, we show that PLD does not influence the shape of the kernel (Fig. 2E).

Given the widespread use of PLD as a proxy for dispersal distance, we conducted additional analyses to test the robustness of our findings. Although we observed substantial intraspecific variation in PLD (*SI Appendix* (34) (mean ± SD = 26 ± 5.6 d; range = 15–42 d), we found that individuals’ PLDs are not correlated with the net distance traveled regardless of whether we treat PLD as a continuous (Fig. 3B) or categorical variable (Fig. 3C) (“short” mean = 22 d; “medium” mean = 26 d; “long” mean = 30 d). These results call into question the pervasive use of PLD as a proxy for dispersal potential (e.g., refs. 22 and 23).

Next, to move beyond description toward a predictive dispersal model, we adopted a logistic regression approach and explored the predictors of the probability of successful dispersal (11). 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 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 \((SI\ Appendix,\ Table\ S5)\). 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 \((34)\). Notably, when the regional predictor was removed from the model, there was no substantive reduction in predictive accuracy \((SI\ Appendix,\ Table\ S5)\), revealing that distance is the primary predictor of successful dispersal between any two locations.

**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 \((Fig. 4A)\). The continuity of reef habitat provides a structural basis for multigenerational stepping-stone dispersal along the barrier reef and within each atoll. Although 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 \((ESUs)\) \((Fig. 4B)\). This prediction is qualitatively consistent with previous estimates of spatial genetic structure for \(E.\ lori\) in Belize \((35)\). 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 \((Fig. 4C)\). Although 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 similarly restricted dispersal patterns.

**Conclusions**

Here, we address a critical knowledge gap in the field of marine ecology—the measurement of a statistically robust empirical dispersal kernel. Our findings demonstrate that, for E. lori, the probability of successful dispersal declines exponentially and predictably as a function of distance from source. Given this result, a relevant question becomes, is this pattern representative of other marine taxa? One hypothesis is that E. lori is an average disperser [the logic being that it is a member of the most speciose marine fish family (28) and its larval duration is close to the median for reef fishes (30)]. An alternative hypothesis is that E. lori lies at the extreme short end of a dispersal continuum [the logic being that we observed short-distance dispersal events and it is an endemic to the Mesoamerican reef (28)]. To test these alternative hypotheses and determine the generality of our results, empirical studies of multiple species will be required.

Turning to consider the causes of variation in dispersal, the spatially restricted kernel, and its consistency with respect to direction and PLD, suggest that ocean currents are not the sole driver of E. lori dispersal. Instead, these patterns suggest that larval behavior may have a strong effect (36). Plausible hypotheses are that larvae either (i) behave in ways that minimize their transport away from natal reefs or (ii) navigate in ways that allow them to return to natal reefs following transport. Future integrative research on dispersal patterns, ocean currents, and larval behavior will be necessary to determine the relative contribution of oceanography and behavior to realized dispersal patterns.

Finally, considering the consequences of dispersal, we conclude that some marine fish populations are strongly demographically connected on the scale of only 0–10 km—an order of magnitude less than current estimates (19). Furthermore, based on observations of spatial genetic structure (35), we conclude that for some marine fish populations, there is strong congruence between the scale of demographic and genetic connectivity. Thus, we suggest that if networks of marine reserves have broad biodiversity targets that aim to protect whole communities of reef fishes, individual reserves may need to be close in space (<10 km apart) to accommodate short-distance dispersers.

**Methods**

**Data.** All genotyping and dispersal data are provided as Excel spreadsheets as Datasets S1–S3.

**Transact.** 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 BBR, centered on Carrie Bow Cay, Belize. The sampling design included three regions for adult tissue collection, spaced roughly 10 km apart, and 41 sites for settler collection, spaced roughly 1 km apart (Fig. 1). The transect was designed to test the hypothesis of a dispersal kernel that we generated in a pilot study of self-recruitment within a 0.5 × 0.125-km plot (31) (see SI Appendix, SI Methods for details). All sampling was conducted using SCUBA at an average (±SD) depth of 16.03 ± 2.19 m. Sampling was focused at those depths where E. lori density is the highest (29). The total length of the three adult collection regions varied slightly (northern, 4.3 km; central, 5.0 km; southern, 3.4 km). 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. All spatial analyses described below were conducted using the midpoint of each dive.

**Tissue Collection.** All methods were approved by the Boston University Institutional Animal Care and Use Committee (protocol #13-008). Adults were found on the inside of yellow tube sponges, captured using slurp guns, and restrained in aquarium nets while a nonlethal fin clip was obtained from the caudal fin. Fin clips were transferred to 95% EtOH at the surface. Approximately 1,000 adult fin clips were collected in each region (SI Appendix, Table S1). Settlers were found on the outside of yellow tube sponges and captured using slurp guns. Because settlers were too small for nonlethal fin-clipping, individuals were collected in plastic bags underwater and euthanized with MS-222 at the surface. At every site (n = 41 sites), ~100 settlers were obtained (SI Appendix, Table S1).

**Genotyping.** Genomic DNA was extracted with a rapid HotSHOT protocol following the methods described by Truett et al. (37). Briefly, tissue samples were lysed through an incubation in hot sodium hydroxide and treated with a neutralizing agent (Tris-HCl). For efficiency, individuals were genotyped following a two-step protocol. First, all individuals (n = 17,141) were genotyped at 14 microsatellite loci using the Type-It Microsatellite PCR Kit (QIagen), following the methods of D’Aloia et al. (31). Second, following an initial parentage analysis that identified potential parent–offspring matches based on 14 loci (see below), the individuals (n = 567) from these putative matches were genotyped at six additional loci to reduce the possibility of false-positive assignments (SI Appendix, Table S2). We calculated summary statistics (number of alleles, observed heterozygosity, expected heterozygosity) in CERVUS, version 3.0 (38), identified deviations from Hardy–Weinberg equilibrium in GENEPOP, version 4.2 (39), and tested for the presence of null alleles in MICRO-CHECKER, version 2.2.3 (40) (SI Appendix, Table S2).
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 natural logorithm of the likelihood ratio (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 \(n\) 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 \((31)\). Individuals from parent-offspring assignments with an LOD score exceeding the critical value at 95% confidence were selected for additional genotyping \((n = 313\) offspring; \(n = 254\) parents) and run through a second parentage analysis based on the expanded genotypes at 20 loci. Based on the simulation parameters defined above, CERVUS estimated 0% type I and 4% type II error rates. Final parent-offspring assignments were selected if their LOD score exceeded the critical value and met additional filtering criteria (see SI Appendix, SI Methods for details). We conducted additional parentage analyses to confirm that our results were robust to the inclusion of loci with moderate frequencies of null alleles (see SI Appendix, SI Methods for details).

Using Size as a Proxy for Age. To test whether postsettlement processes influenced the shape of the observed kernel, we used settler size as a proxy for age by measuring the standard length (SL) of each offspring collected. Size is a coarse proxy for postsettlement trajectory; \(9 \text{–} 11 \text{~mm}\) individuals are younger, on average, than category 2 individuals.

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 Euclidean 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 traveled 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 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, and then mounted on a slide and immersed in oil for 2-7 d. 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 (see SI Appendix, SI Methods for details).

Statistical Analysis 1—Descriptive Dispersal Models. We estimated an empirical dispersal kernel by fitting alternative 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 Akaike information criterion and Bayesian information criterion (SI Appendix, Table S3) \((43)\). 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 (see SI Appendix, SI Methods for details). 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: \(i\) northern/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 (SI Appendix, Fig. S1). We also considered all of the above-mentioned...
variables in a multivariate analysis to investigate the relationship between these predictors and the mean and variance of dispersal distance (see SI Appendix, SI Methods for details).

Statistical Analysis 2—Predictive Dispersal 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 trajectories included distance (in kilometers), 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, with construction for binomial data (44). We also calculated the AUC with 95% CIs (n = 10,000 bootstrap replications) as a metric of the predictive accuracy of the model (34).

Connectivity Matrix Analyses. To investigate patterns of population connectivity across the BBR system, we generated a connectivity matrix between all potential source sites (N) and all potential settlement sites (N). To define this set of sites, we obtained a map of the BBR system from the iMARS Millennium Coral Reef Mapping Project (45). Focusing only on the “forereef” zone where E. lori resides, we converted the forereef polygon into 1-km2 grid cells using a fishnet tool in ArcMap 10.2 (ESRI). Next, we manually removed excess reef patches to ensure that included distance never exceeded 1 km, given that E. lori are only abundant on the outer reef slope (29). This led to a total of 516 sites. We generated a distance matrix between all sites by calculating the Euclidean distance between the centroids of all grids. We then used the logistic model to predict the probability of successful dispersal between all sites. To scale up from the 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 (SI Appendix, Table S5). 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 (see SI Appendix, SI Methods for details).

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, Colleen Webb, and two anonymous reviewers for comments on this manuscript. This work was supported by National Science Foundation (NSF) Grant OCE-1260424, and C.C.D. was supported by NSF Graduate Research Fellowship DGE-1247312. All work was approved by Belize Fisheries and Boston University Institutional Animal Care and Use Committee.