Vol. 607: 143–154, 2018
https://doi.org/10.3354/meps12792

MARINE ECOLOGY PROGRESS SERIES
Mar Ecol Prog Ser

Published December 6

Limited dispersal explains the spatial distribution of siblings in a reef fish population

C. C. D'Aloia[1,2,*], A. Xuereb[2], M.-J. Fortin[2], S. M. Bogdanowicz[3], P. M. Buston[4]

[1]Biology Department, Woods Hole Oceanographic Institution, Woods Hole, MA 02543, USA
[2]Department of Ecology & Evolutionary Biology, University of Toronto, Toronto, ON M5S 3B2, Canada
[3]Department of Ecology and Evolutionary Biology, Cornell University, Ithaca, NY 14853, USA
[4]Department of Biology and Marine Program, Boston University, Boston, MA 02215, USA

ABSTRACT: Extensive larval dispersal and a high degree of planktonic cohort mixing were long presumed to disrupt kin aggregations in marine environments. Yet, recent genetic studies of diverse marine taxa have suggested that kin may be found in close proximity to each other after settlement, raising interesting questions about the ecological and behavioral processes that could generate these patterns. We drew on sibship reconstruction to test whether kin cohesion and/or the scale of dispersal could explain patterns of relatedness in the coral reef fish *Elacatinus lori*. We genotyped 4074 recently settled individuals along a 41 km transect on the Belize Barrier Reef. Because most individuals in the population were unrelated, we found that high-confidence sibling assignments required a large number of microsatellites (≥55). Using 71 microsatellites, we documented 371 sibling pairs which were non-randomly distributed on the reef: 50% were ≤3 km apart and 99% were ≤18 km apart. The spatial distribution of sibling pairs was congruent with predictions from the limited dispersal hypothesis, and we found no evidence that siblings disperse cohesively. These results underscore the importance of (1) accounting for the relative abundance of different relationship types within a population to accurately identify siblings and (2) carefully applying spatial analyses to discriminate between alternative ecological kin structuring mechanisms. More broadly, this study provides a framework for linking spatial distributions of siblings to the processes that generate them, highlighting the potential for sibship data to provide new insights into marine larval dispersal.

KEY WORDS: Genetic relatedness · Kinship · Collective dispersal · Larval dispersal · Coral reef · Spatial ecology · Microsatellite sequencing

INTRODUCTION

The spatial arrangement of kin has far-reaching effects on the potential for ecological interactions between related conspecifics. When kin are near to each other and interactions are common, inbreeding may occur, with subsequent effects on individual fitness (Crnokrak & Roff 1999), the genetic makeup of populations (Charlesworth & Charlesworth 1999), and, ultimately, the evolution of social behaviors (Hamilton 1964, West et al. 2002). While the importance of spatial kin structure is widely acknowledged, it has been historically understudied in the marine environment due to assumptions about larval dispersal (Kamel & Grosberg 2013). Specifically, larval cohort mixing within small-scale turbulence and extensive dispersal distances were presumed to disrupt most kin associations and prevent the spatial clustering of kin post-settlement (Victor 1984, Leis 1991).

Yet genetic evidence is mounting that relatives are sometimes found in close spatial proximity after larval dispersal. Putative relatives have been identified within recruitment pulses (Grosberg & Quinn 1986, Selkoe et al. 2006, Christie et al. 2010, Bernardi et al.

*Corresponding author: cassidy.daloia@gmail.com

2012), social groups and shoals (Buston et al. 2009, Selwyn et al. 2016), and discrete reefs or sampling sites (Iacchei et al. 2013, Salles et al. 2016, Adrian et al. 2017). Moreover, these genetic studies encompass species that differ in key ecological traits. For example, kin have been documented in species with varying larval durations, ranging from the colonial ascidian *Botryllus schlosseri* that disperses for just 0–48 h (Grosberg & Quinn 1986) to the spiny lobster *Panulirus interruptus* that disperses for 8–11 mo (Iacchei et al. 2013). Kin have also been observed in populations inhabiting diverse marine ecosystems, such as coral reefs (Bernardi et al. 2012), oyster reefs (Adrian et al. 2017), and kelp forests (Selkoe et al. 2006). Finally, kin have been found in species exhibiting a broad range of post-settlement movement capacities, including sessile invertebrates (Adrian et al. 2017), sessile invertebrates living atop moving hosts (Hart & Grosberg 1999), site-attached fishes (Buston et al. 2009), and highly mobile fishes (Horne et al. 2016). Combined with the spatial distribution of kin at settlement, these movement capacities will influence the probability of kin interactions. Collectively, this growing body of empirical work suggests that relatives can be found after settlement in diverse marine species.

These genetic kinship data offer a unique opportunity to make inferences about ecological and behavioral processes that occur during the larval phase — an appealing prospect because the full larval phase is not yet directly observable *in situ*. However, simply documenting the co-occurrence of close kin post-settlement is insufficient to discriminate between alternative generating processes (where 'close' kin are defined as first- and second-order relatives). Any effort to link observed kinship patterns to larval processes must include appropriate spatial analyses to test for deviations from random expectations (Dale & Fortin 2014). Essential steps include quantifying the spatial pattern of kin structure and comparing the observed pattern to a set of predictions from alternative hypotheses.

Several hypotheses about larval dispersal have been put forth to explain the nearby occurrence of close kin after settlement (D'Aloia & Neubert 2018) (Fig. 1). As a baseline for comparison, we can consider a null hypothesis of well-mixed dispersal (H_0), wherein individuals have an equal probability of traveling to any site in the study domain. This leads to the expectation that sibling pairs will be equally frequent at all distances between individuals. A first alternative hypothesis posits that dispersal is limited (H_1). There is generally some spatial limit to dispersal, even though marine species are known to exhibit strong heterogeneity in dispersal scale (Kinlan & Gaines 2003). When dispersal is limited, a larva's movement away from its origin point is restricted, and so sibling pairs will be less frequent, on average, as the distance between individuals increases. Generating species-specific predictions of the limited dispersal hypotheses hinges upon an *a priori* measure of dispersal capacity, which may be estimated directly through genetic parentage analyses (D'Aloia et al. 2015, Williamson et al. 2016, Almany et al. 2017), inferred from isolation by distance slopes (Puebla et al. 2012, Pinsky et al. 2017), or predicted by biophysical models (Siegel et al. 2003). Both H_0 and H_1 assume that larvae disperse independently.

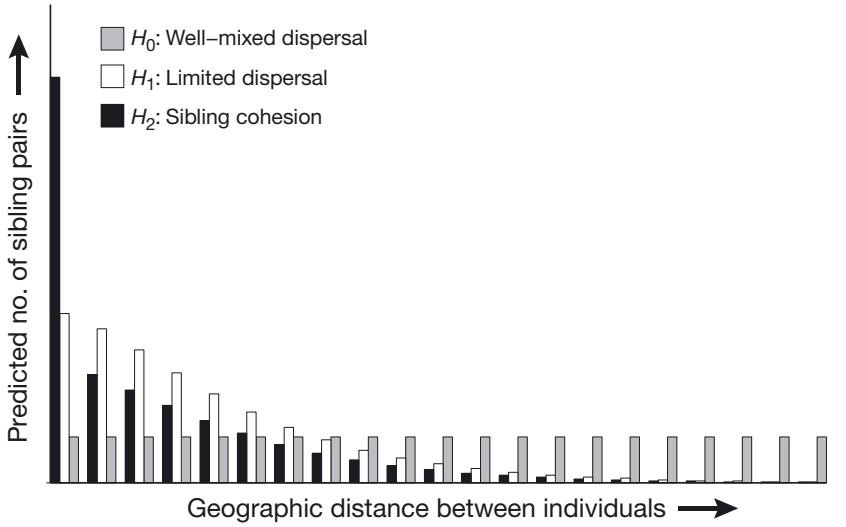

Fig. 1. Predictions of 3 alternative hypotheses regarding mechanisms governing the spatial distribution of siblings. H_0: null hypothesis, where dispersal is well-mixed and sibling pairs are evenly distributed across space; H_1: limited dispersal hypothesis, where the number of sibling pairs declines as the distance between individuals increases; and H_2: sibling cohesion hypothesis, where, in addition to exhibiting limited dispersal, siblings have a tendency to disperse cohesively. For illustrative purposes, axes are intentionally unitless and singular examples are used for dispersal decay rate and sibling cohesion strength (n = 1000 sibling pairs hypothesis^{-1})

A second alternative, but not mutually exclusive, hypothesis posits that larvae may not disperse independently (H_2). This process, in which larval trajectories are correlated, has been termed cohesive dispersal, collective dispersal, aggregated dispersal, and group dispersal (Siegel et al.

2008, Broquet et al. 2013, Riquet et al. 2017, Burgess et al. 2018), and could be driven by passive retention in water packets or active behaviors to recognize and stay with conspecifics. For the purpose of this study, which focuses on close kin, we use the term 'sibling cohesion,' and consider the specific effect of siblings traveling together for the duration of the larval phase. If siblings tend to have non-independent dispersal trajectories, there should be a substantial increase in the number of siblings found nearby after settlement (above and beyond the expectation from well-mixed or limited dispersal alone) (D'Aloia & Neubert 2018) (Fig. 1). Although some genetic kinship papers have discussed the potential effects of cohesion versus limited dispersal verbally (e.g. Selwyn et al. 2016, Adrian et al. 2017), few have explicitly tried to discriminate between them (but see Grosberg & Quinn 1986).

The aforementioned hypotheses are not intended to capture the full spectrum of marine kinship drivers; indeed, other relevant factors are known to influence the overall level of kinship in marine populations. One example is the well-studied phenomenon of sweepstakes reproductive success (Hedgecock 1994), which predicts that, in highly fecund marine species, a small fraction of adults can have disproportionately high reproductive success due to stochastic processes. This phenomenon can decrease genetic diversity within cohorts, increase the absolute number of siblings in a population, and generate skew in the size distribution of family groups, but does not, on its own, leave a distinct spatial signature (D'Aloia & Neubert 2018). Because we are interested in linking kin structure patterns to larval processes, our focus is on hypotheses that generate testable spatial predictions (Fig. 1).

We tested alternative hypotheses for the formation of marine kin structure in a population of the sponge-dwelling coral reef fish *Elacatinus lori* (Colin 2002). Two lines of evidence suggest the species is a good candidate for disentangling the drivers of kin structure. First, *E. lori* has a measured dispersal kernel based on a previous genetic parentage study (D'Aloia et al. 2015). The species exhibits very limited dispersal, with most offspring traveling less than 2 km from their parents despite having a 26 d larval duration. Because the dispersal kernel has been estimated, we can generate the expected pattern of sibling structure from limited dispersal (H_1; Fig. 1). Second, *E. lori* has the potential for kin cohesion: it is a demersal spawner and males guard clutches of eggs, composed of sibling groups, until they hatch into competent larvae (Majoris et al. 2018a). After controlling for predicted dispersal distances, we can test for evidence of sibling cohesion (H_2; Fig. 1), thereby distinguishing the potential effects of philopatry and cohesion. Here, we begin by exploring how genetic relatedness estimates are sensitive to assumptions about the demographic makeup of populations and attributes of the genetic marker panel. These explorations are essential first steps in reliably identifying siblings. Next, we describe the observed distribution of siblings and use spatial analyses to determine which ecological processes influence the pattern of *E. lori* kin structure.

MATERIALS AND METHODS

Tissue collection

In summer 2013, we collected *Elacatinus lori* tissue samples along a 41 km stretch of the Belize Barrier Reef, centered on Carrie Bow Cay (16° 48′ 13″ N, 88° 04′ 37″ W) (Fig. 2). Every km, we collected tissue from approximately 100 individuals for a total of n = 41 collection sites and n = 4112 individuals (Table S1 in the Supplement at www.int-res.com/articles/suppl/m607p143_supp.pdf). We treated each site as a point location by recording waypoints at the beginning and end of each SCUBA dive and taking their midpoint as the location. Samples were collected along a single depth profile where fish density is high (z_{mean} ± SD = 16.03 ± 2.19 m). Sampling focused on recently settled individuals, i.e. 'settlers' (<18 mm standard length), which can remain on the outside of *Aplysina fistularis* sponges for several weeks post-settlement (Majoris et al. 2018b). Fish were collected using slurp guns, euthanized with MS-222 at the surface, and tissue was stored in 95% EtOH. Because we were interested in first- and second-order relationships and we sampled within a single generation, this sampling regime enabled us to identify half- and full siblings.

Multiplex PCR and sequencing

All individuals were sequenced at 71 microsatellite loci using a multiplex PCR protocol for targeted amplicon sequencing. For full details on multiplex PCR and Nextera barcoding methods, see D'Aloia et al. (2017). Here, we briefly outline the main steps and highlight modifications made for this data set.

The 71 microsatellites were chosen from a genomic DNA library enriched for simple repeats (D'Aloia et al. 2013). Using primers listed in Table S2, we amplified these loci with multiplex PCR reactions in 384-

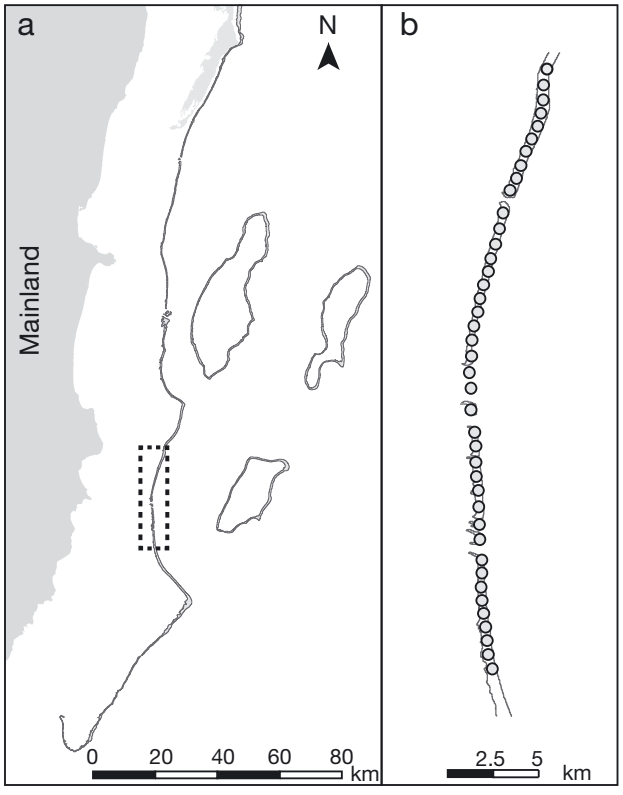

Fig. 2. Schematic of *Elacatinus lori* sampling transect from 2013; (a) the Belize Barrier Reef. Dashed box: the 41 km transect location; (b) zoom-in on the transect. Grey circles: sampling locations (n = 41), spaced ~1 km apart. Forereef habitat is outlined in the background

well plates using QIAGEN Multiplex PCR kits. Next, we pooled samples across multiplex groups and ran a barcoding PCR using llumina's S5 and N7 Nextera primers. To prepare the sequencing library, all barcoded individuals were pooled and cleaned with Ampure XP (Beckman Coulter). The library was diluted to a 2 nM concentration and sequenced on an Illumina MiSeq with paired 250 bp reads at Cornell University's BioResource Center.

Post-sequencing data processing

We used a custom Perl script to trim reads for quality, sort loci by the forward PCR primer, and assign individual genotypes at each locus (see D'Aloia et al. 2017 for full details). We used a minor allele proportion of 0.2 and a quality score of Q20. To filter PCR artefacts and paralogs while retaining true variants, we set a minimum read length of 225 bp and used a matching command that required 90% of the first 40 bp to match a reference contig at each locus.

To retain only high-quality markers and individuals with little missing data, we applied 3 additional filters after running the script. First, all individuals with >20% missing data were excluded. Given the large marker panel, individuals with missing data were still sequenced at ≥56 loci after this filter, which was sufficient to confidently assign sibling relationships (see Results). Second, to reduce the presence of false haplotypes, we recoded any locus with ≤5 total reads as missing data. We also recoded putative heterozygotes with 6 to 10 total reads as homozygotes for the allele with the higher number of reads. Third, for compatibility with genetic relatedness software, we removed enough singleton haplotypes (i.e. haplotypes found in only one sample) at 4 loci to ensure that the number of alleles per locus was below 127 (an arbitrary but fixed threshold). After all filters were applied there were n = 4074 individuals distributed evenly across the transect. We used these individuals for all subsequent analyses.

Tests of linkage disequilibrium

Relatedness estimates assume that loci behave independently. As the number of markers, n, increases, the number of unique pairwise comparisons between them increases substantially as a binomial coefficient, i.e. $\binom{n}{2}$, and some pairwise linkage becomes inevitable. With our large panel of 71 microsatellites there were 2485 pairwise comparisons. We used a multilocus index of association ($\bar{r}_d$), which summarizes linkage disequilibrium (LD) while accounting for the number of loci, to assess whether linkage would influence our results (Agapow & Burt 2001).

Identifying siblings

We used the R package 'related' to estimate pairwise genetic relatedness (Pew et al. 2015). We began by simulating pairs of individuals with known relationships based on the observed allele frequencies. We used these simulated pairs to test alternative relatedness estimators and found that multiple estimators, both likelihood and moment-based, performed equally well, i.e. correlation coefficients with expected relatedness values exceeded 0.97. Because all estimators were accurate and strongly correlated (Table S3), and estimating relatedness in large populations is computationally intensive due to the pairwise nature of the data, we proceeded with the Queller and Goodnight moment-based estimator (Queller & Goodnight 1989).

We used it to estimate empirical pairwise relatedness (r) between all pairs of individuals with a 5% sequencing error rate. This mean allelic error rate was calculated from n = 282 individuals that we sequenced twice (sensu Pompanon et al. 2005). We performed a sensitivity analysis to ensure that the error rate did not affect the results.

Because we were specifically interested in the spatial structure of close kin, we adopted a threshold approach to assign pairs as either 'siblings' (inclusive of both half-siblings and full siblings) or 'nonrelatives' (also inclusive of more distant relatives). Assigning pairs to relationship categories based on pairwise r is notoriously difficult due to factors including the overlapping expected r distributions of different relationship types, the true proportion of relationship types within the population, and the number and diversity of genetic markers (Csilléry et al. 2006, Taylor 2015). To identify an appropriate threshold, we used the simulated data to visualize the expected overlap in the pairwise relatedness distributions for unrelated, half-sibling, and full-sibling pairs (Fig. 3a). Next, we focused on the overlapping region between nonrelatives and half-siblings, and accounted for the proportion of these different relationships within the population. A previous genetic parentage study conducted along the same transect revealed that for every 1 pair of half-siblings in the *E. lori* population, there were 550 pairs of nonrelatives (D'Aloia et al. 2015). This unequal proportion of relationships generates the expectation that pairs of half-siblings are rare compared to pairs of nonrelatives at all overlapping relatedness values (even in the upper tail of the nonrelative distribution) (Fig. 3b). Thus, in a large population with a relatively high proportion of nonrelatives, the misclassification of nonrelatives as half siblings will be higher than the reverse.

We also used empirical pairwise r data to test how the estimated number of sibling pairs changed as a function of the number of genetic markers used in the relatedness estimates. The number of putative siblings declined as the size of the marker panel increased (Fig. 3c) (note: this effect was also evident using alternative likelihood approaches to categorical relationship assignments [Table S4]). Beyond 55 markers, the number of sibling pairs based on r thresholds of 0.3 and higher was relatively consistent. However, at the lower threshold of 0.25 (the expected value for half-siblings), the number of pairs continued to decline as more markers were added, up to the full panel. Thus, to accurately discriminate *E. lori* half-siblings, large panels of ≥70 markers were required. We proceeded with a conservative $r \geq 0.30$ threshold and all 71 microsatellites to discriminate siblings from non-relatives or more distantly related relatives. While this conservative threshold will misclassify some half-siblings as nonrelatives, it was necessary to account for the substantially higher frequency of nonrelatives in the *E. lori* population.

Identifying sibling groups

We used network analyses to look for connections among sibling pairs (Handcock et al. 2008). Nodes represented individual fish and links between nodes represented the presence of a sibling relationship. Because every individual was connected to one or more individuals in the network through a sibling relationship, but not all individuals in each network component were necessarily siblings (e.g. 2 unrelated individuals sharing a half-sibling), we referred to the resulting components as 'extended sibling groups'.

Testing for spatial patterns in relatedness

After identifying siblings, we quantified the spatial pattern of sibling structure and compared the observed pattern to predictions from alternative hypotheses (Fig. 1). We generated the expected distribution for each hypothesis by simulating the dispersal of sibling pairs, with the number of simulated pairs equal to the total number of observed pairs. Individuals dispersed according to their underlying kernel (e.g. uniform for H_0; Laplace for H_1) and the distance between siblings, after settlement, was calculated. We took the average of 10 000 simulations to be the expected distribution. Next, we used bootstrapped K-S tests to determine whether the frequency of observed versus expected sibling pairs differed across distance classes (n = 10 000 bootstraps). The bootstrapped version of the K-S test can test for differences between 2 distributions when data are drawn from discrete distributions and ties exist (Abadie 2002).

RESULTS

Genetic summary information

The 71 microsatellite markers were highly polymorphic (Table S2). The number of alleles per locus ranged from 13–122, with a mean (±SD) of 46.77 ± 28.91. Using all 71 microsatellites, mean pairwise relatedness within the population was −0.0001 ± 0.05,

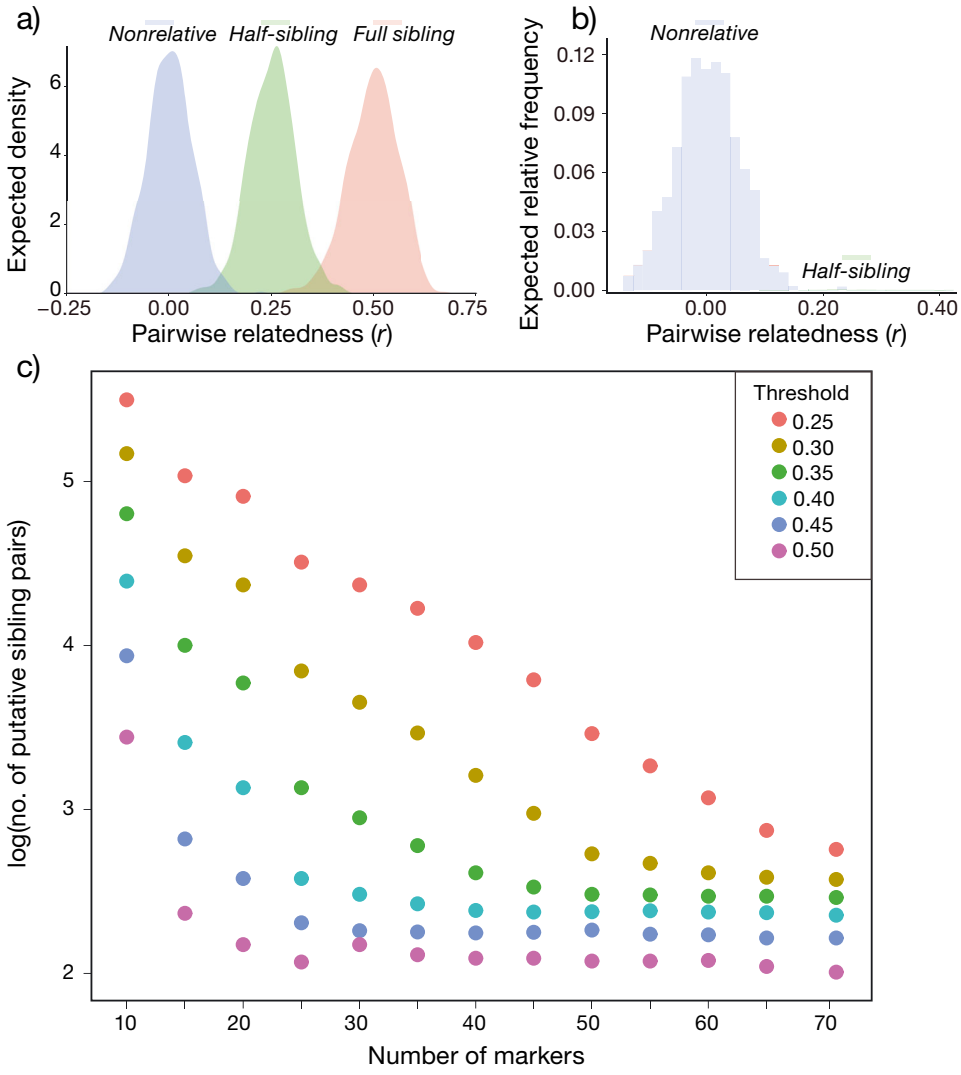

Fig. 3. Sensitivity of pairwise relatedness estimates for *Elacatinus lori*. (a) Expected density of relatedness values for each relationship type given the observed allele frequencies (note that nonrelative and half-sibling distributions overlap); (b) accounting for the high frequency of nonrelative pairs in the population predicts that half-siblings will be relatively rare (here, not even visible) among all pairs with relatedness values in the overlapping region with nonrelatives; (c) number of putative sibling pairs (log scale) declines as a function of the number of markers used to estimate relatedness for all 4074 genotyped individuals. The pattern is shown at different relatedness thresholds used to distinguish nonrelatives from siblings (where colors reperesent unique thresholds)

revealing that the majority of pairs were not siblings. This result was insensitive to the sequencing error rate used in the relatedness estimator. There was no significant level of multilocus LD ($\bar{r}_d$ = 0.00367, n = 999 permutations, p = 0.982; Fig. S1).

Number of siblings

Considering all 4074 genotyped individuals, 624 of them (15.3%) had at least one sibling in the population. The 4074 genotyped individuals yielded 8 296 701 unique pairs, of which 371 (0.004%) were pairs of siblings. Thus, even though a substantial percentage of individuals had at least one sibling on the transect, the fraction of total pairs that were siblings was small.

Network connections between siblings revealed that in this population, pairs of siblings did not frequently link up to other pairs (Fig. 4a). The extended sibling groups were predominantly dyads or triads, with just 3 larger groups composed of 4, 5, and 6 individuals, respectively (Fig. 4b), suggesting a lack of extreme reproductive variance.

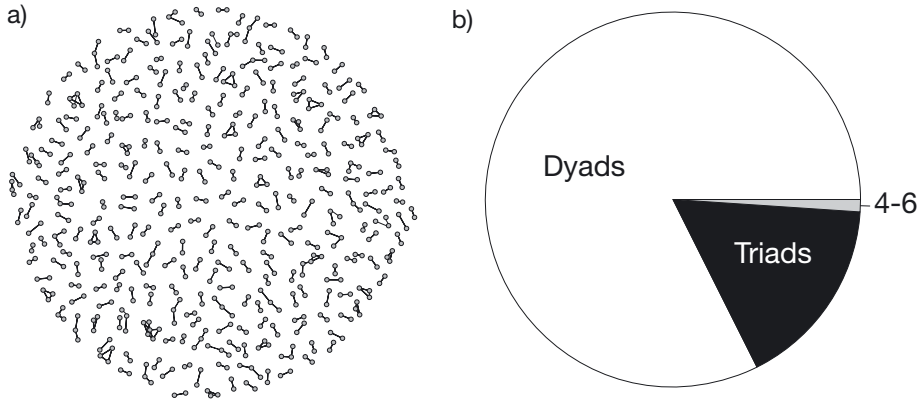

Fig. 4. Network of *Elacatinus lori* siblings and size of sibling groups. (a) Nodes represent individual fish and lines represents sibling relationships (note: line length does not indicate pairwise *r* magnitude; node positions are not spatially explicit); (b) proportion of discrete network components (i.e. 'extended sibling groups') that are dyads, triads, or larger (4–6 ind.)

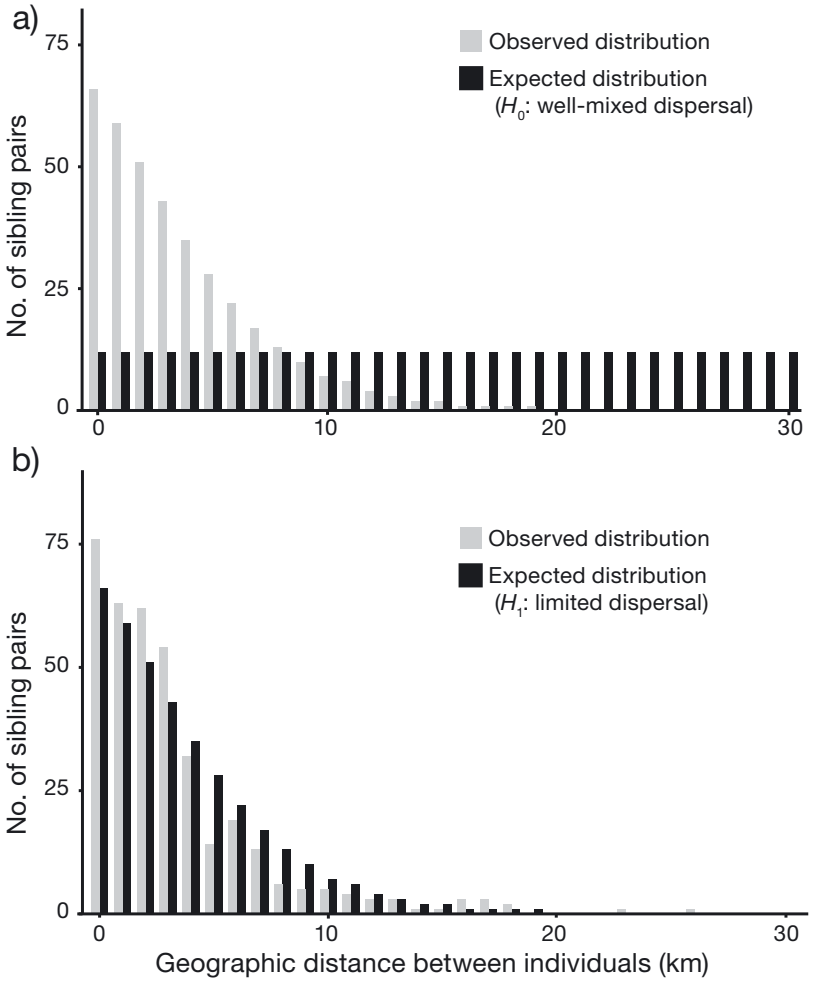

Fig. 5. Comparison of observed versus expected spatial distributions of sibling pairs, based on alternative hypotheses. (a) Observed distribution is significantly different than the expected well-mixed dispersal distribution (H_0); (b) observed distribution is not significantly different from the expected limited dispersal distribution (H_1)

Testing for spatial patterns in relatedness

The observed sibling distribution was right-skewed, with small distances between most sibling pairs (50th percentile = 2.8 km apart; 99th percentile = 18.0 km apart). There was a significant difference between the observed distribution and the expected distribution based on a null hypothesis of well-mixed dispersal (H_0; bootstrapped K-S test: $D = 0.7419$, $n_{boot} = 10\,000$, $p < 0.001$). The expected well-mixed distribution was uniform across all distances, whereas the observed distribution had a relative overabundance of short distances between siblings (≤5 km) and a relative lack of long distances between siblings (≥10 km) (Fig. 5a).

In contrast, the observed sibling distribution was not significantly different than the expected distribution based on limited dispersal (H_1), which was generated from the species' Laplace dispersal kernel (bootstrapped K-S test: $D = 0.0968$, $n_{boot} = 10\,000$, $p = 0.968$) (Fig. 5b). Because limited dispersal explained the spatial pattern, there was no need to invoke sibling cohesion (H_2), which predicts a substantially higher frequency of siblings at short distances than was observed (see Fig. 1). Together, these results revealed that *E. lori* sibling pairs tend to cluster at the scale of a few km, and that their full spatial pattern was predicted by the species' dispersal kernel.

DISCUSSION

The spatial distribution of kin within populations, i.e. kin structure, has fundamental implications for competition, cooperation, and mating between relatives (Hamilton 1964, West et al. 2002). The far-reaching consequences of kin structure have motivated empirical investigations of genetic relatedness in populations that span taxa and ecosystems (Bourke & Franks 1995, Hatchwell 2010, Kamel & Grosberg 2013). In marine populations, there has been additional interest in studying kinship because of the potential to draw inferences about processes that may occur during the larval phase. We used sibship reconstruction to explore the patterns and ecological causes of kin structure within a large population of the neon goby *Elacatinus lori*. We showed that a large microsatellite marker panel was needed to reliably identify siblings, then documented hundreds of sibling pairs along a 41 km transect. A careful analysis of their spatial distribution revealed that the pattern of kin structure could be parsimoniously explained by the species' dispersal kernel alone, without invoking any larval kin cohesion. Below, we consider the challenges of reliably identifying siblings within large benthic marine populations, and discuss the role of dispersal in shaping spatial patterns of post-settlement relatedness.

The challenge of identifying siblings

All downstream ecological analyses of the patterns, causes, and consequences of kin structure hinge upon accurate relatedness estimates, which are known to be influenced by multiple factors including the proportion of different relationship types within the population and various attributes of the genetic marker panel. Our estimate of the relative frequency of nonrelative to half-sibling *E. lori* pairs (550:1) underscores the importance of accounting for the population's relationship-type structure. Specifically, the high abundance of nonrelatives means that most pairs with relatedness values falling within the overlapping region between nonrelatives and half-siblings (Fig. 3b) will actually be nonrelatives. This effect has been largely overlooked in previous relatedness studies (but see Csilléry et al. 2006, Buston et al. 2009).

Our results also align with previous studies focusing on the effects of locus type, locus polymorphism, and locus number (Blouin et al. 1996, Glaubitz et al. 2003, Santure et al. 2010, Kopps et al. 2015, Kaiser et al. 2017). A key issue is that a small marker panel can identify true relatives such as half- and full siblings, but will also misclassify many nonrelatives as half-siblings, thereby inflating overall sibling estimates (Kopps et al. 2015) (Fig. 3). In our own data set, for example, we found that at least 55 highly variable microsatellites (mean number of alleles = 46) were needed to prevent nonrelatives from being identified as siblings, consistent with a simulation study which suggested that, depending on the breeding system, a minimum of 40 to 80 microsatellites (mean number of alleles = 10) are needed to accurately identify categories of relatives (Kopps et al. 2015).

Of course, as the number of markers increases, some pairwise LD is expected, leading to a potential trade-off between the benefits of a large marker panel and the effect of LD. Because relatedness estimators assume that loci are independent, LD may result in overconfident relatedness estimates, though linked loci are thought to perform 'approximately' the same as unlinked loci (Csilléry et al. 2006, Wang 2007). In *E. lori*, we found no significant background level of multilocus LD and documented a clear trend of a declining number of siblings with an increasing number of markers (Fig. 3). These lines of evidence suggest that, at least in this data set, the benefit of having many markers outweighs modest levels of pairwise LD.

In combination with earlier empirical and simulation studies, our results have important implications for marine relatedness studies: assuming a 1:1 ratio of nonrelatives to relatives and/or using small to moderately sized microsatellite panels will likely lead to the misclassification of many nonrelatives (or distant relatives) as half-siblings in large populations. Moving forward, if microsatellite markers are used (even highly polymorphic ones), we suggest a minimum of ~50 markers to identify siblings. While this number is high, the amplicon sequencing approach that we employ offers an efficient and cost-effective method for obtaining multilocus genotypes (D'Aloia et al. 2017).

Sibling pairs and groups within the population

Our finding that at least 371 pairs of *E. lori* siblings occur within a single population is consistent with previous kinship studies of diverse marine taxa, including ascidians (Grosberg & Quinn 1986), barnacles (Veliz et al. 2006), lobster (Iacchei et al. 2013), and numerous species of fish (Selkoe et al. 2006, Buston et al. 2009, Bernardi et al. 2012, Selwyn et al. 2016, Salles et al. 2016). We emphasize that 371 pairs

may be an underestimate, as our conservative *r* threshold likely excluded some true half-siblings. Our results contribute to the growing body of evidence that siblings can be found within marine populations after a dispersive larval phase.

The *E. lori* network of 'extended sibling groups' revealed that most discrete network components were in fact unconnected pairs (dyads) or trios (triads), and the largest sibling group consisted of just 6 individuals (Fig. 4). While network theory has not yet been widely applied to marine kinship, a few studies have used networks to identify 'extended families' within sibling cohorts in invertebrate (Veliz et al. 2006) and fish species (Selwyn et al. 2016). For *E. lori*, the absence of large intra-generational sibling groups demonstrates that many distinct parents are contributing offspring to settlement cohorts and there is relatively little variance in parental reproductive success. We also found that relatedness values were consistent across settler size classes (Fig. S2), contrasting with other studies that have found related individuals tend to be similarly sized and more prevalent in younger age classes (Buston et al. 2009, Riquet et al. 2017). To the extent that size is a reliable proxy for age, our results suggest that *E. lori* patterns of relatedness do not change throughout this early life stage in response to processes such as post-settlement selection or kin competition.

Disentangling the causes of spatial kin structure

Given the spatial pattern of *E. lori* kin structure, what can be said about the causes? Two main hypotheses regarding processes that increase genetic relatedness in marine populations have been proposed: limited dispersal and larval cohesion. Considering dispersal distance first, the observed sibling distribution clearly deviated from a null hypothesis of well-mixed (i.e. uniform) dispersal over the scale of the study area. Instead, the spatial pattern was consistent with the limited dispersal hypothesis. Genetic parentage analysis from this same study area provided strong evidence of a Laplace dispersal kernel for *E. lori* (median distance = 1.7 km; max. distance = 16.4 km; D'Aloia et al. 2015). Based on this kernel, distances between siblings were predicted to be up to ~16 km apart if they happened to disperse in the same direction, or up to ~32 km apart if they happened to disperse in opposite directions. Indeed, we found the distances between siblings to be congruent with these predictions, suggesting that restricted dispersal alone can explain the spatial pattern of relatedness.

Limited dispersal has long been recognized in the theoretical literature as a mechanism that can generate 'viscous' populations where kin live in close proximity. In turn, population viscosity can play an important role in facilitating kin mating and the evolution of altruism (Hamilton 1964, Queller 1994). However, given the scale of our sampling design (collection locations were spaced every 1 km) and the species' limited mobility post-settlement (individuals are site-attached and do not move far beyond their primary host sponge), we cannot directly address the probability of kin mating from this study.

The ability to detect dispersal-influenced kin structure in other species will depend upon the relative spatial scales of larval dispersal and field sampling efforts. Ideally, empirical studies of sibling structure will be designed to capture distances between siblings that match the decline in the species' dispersal probability over space. For species lacking a measured dispersal kernel, it will be challenging to design a sampling strategy commensurate with dispersal capacity *a priori*, but biophysical models (Siegel et al. 2003) or population genetic data (Kinlan & Gaines 2003, Pinsky et al. 2017) can serve as a starting point.

In addition to limited dispersal, recent relatedness studies, in conjunction with studies of chaotic genetic patchiness (Broquet et al. 2013, Iacchei et al. 2013), have fueled interest in an alternative mechanism: the potential for cohesive (or collective) larval dispersal. However, our data do not support that hypothesis in *E. lori*. The cohesion hypothesis posits that individuals stay together throughout the larval phase through active behavior to stay with conspecifics (Berenshtein et al. 2018) and/or passive retention in the same water packets (Siegel et al. 2008, Harrison et al. 2013). The strongest evidence for this hypothesis comes from otolith microchemistry studies that reveal some individuals share dispersal pathways for at least part of the larval phase (Ben-Tzvi et al. 2012, Shima & Swearer 2016). If the cohesive individuals are kin, siblings should be found closer together than expected by chance. At first pass, the finding of hundreds of sibling pairs in one population may seem suggestive of kin cohesion. Yet sibling cohesion, as defined by D'Aloia & Neubert (2018), is predicted to have a specific spatial effect on kin structure: there should be a substantial overabundance of sibling pairs at short distances because cohesive siblings settle together. This sibling cohesion effect is predicted to persist even when siblings break up into multiple cohesive groups that each disperse independently, and when other mechanisms (e.g. limited dispersal, variable reproductive success) operate simultane-

ously. Thus, the spatial organization of *E. lori* sibling pairs clearly deviates from the expected pattern predicted by sibling cohesion. Future studies may elucidate the prevalence and strength of larval cohesion generally, and kin cohesion specifically, across marine taxa.

Sibship-based insights into dispersal distances

The strong match between the limited dispersal prediction and the observed sibling distribution provides further indirect support that larval dispersal is spatially restricted in this species. Because of the challenges of measuring marine dispersal (Jones et al. 2009) and the stochastic nature of the environment in which the larvae disperse (Siegel et al. 2008), congruent lines of evidence from multiple approaches build confidence in empirical estimates. The body of dispersal work conducted on this particular species highlights the complementarity between multigenerational population genetic analysis (D'Aloia et al. 2014), intergenerational parentage analysis (D'Aloia et al. 2015), and intra-generational sibship analysis (this study), which has been previously highlighted in at least one other fish species (Schunter et al. 2014). While both the parentage and sibship *E. lori* studies were conducted along the same transect, the sample size increased over 200%, from 120 parent–offspring pairs in the parentage study to 371 sibling pairs in this sibship study. These additional data support the initial estimate of the Laplace kernel. More broadly, these types of studies exemplify the promise of individual-based genetic methods to provide new insights in recent dispersal events and larval ecology (Marko & Hart 2018).

CONCLUSIONS

Genetic kinship studies have the potential to provide new insights into the ecological dynamics of the larval phase. However, an essential first step towards achieving this goal is accurately identifying kin. We demonstrated how relatedness estimates are sensitive to both the genetic marker panel (a well-studied phenomenon; Santure et al. 2010, Kopps et al. 2015) and the proportion of different relationship types in the population (a relatively understudied phenomenon; Csilléry et al. 2006, Buston et al. 2009). After controlling for these factors, we identified hundreds of sibling pairs within this large reef fish population. Their spatial distribution was parsimoniously explained by the species' dispersal kernel alone, rather than by more complex cohesive dynamics among siblings. In turn, these results support prior empirical evidence that *E. lori* exhibits spatially restricted dispersal despite a month-long larval phase, highlighting the complementarity of parentage and sibship analyses in studying larval dispersal. Taken together, these findings have broad implications for continued progress in the study of marine kinship. Most importantly, the careful application of spatial analysis is necessary when inferring alternative behavioral and/or ecological processes from post-settlement patterns of kin structure.

Acknowledgements. Much of this work was inspired by our friend, mentor, and colleague Richard G. Harrison. We are grateful to Diana Acosta, Alben David, Kevin David, Robin Francis, John Majoris, Alissa Rickborn, and Derek Scolaro for assistance in the field; to Eliana Bondra for assistance in the lab; and to Jose Andrés and 2 anonymous reviewers for feedback. C.C.D. was supported by a Weston Howland Junior postdoctoral scholarship from Woods Hole Oceanographic Institution. Research was funded by a start-up award to P.M.B. from the Trustees of Boston University and NSF Grant OCE-1260424 to P.M.B. and colleagues. All work was approved by the Boston University IACUC (#13-008) and Belize Fisheries.

LITERATURE CITED

Abadie A (2002) Bootstrap tests for distributional treatment effects in instrumental variable models. J Am Stat Assoc 97:284–292

Adrian A, Lack C, Kamel S (2017) Kin aggregations occur in eastern oyster *Crassostrea virginica* reefs despite limited regional genetic differentiation. Mar Ecol Prog Ser 584: 79–90

Agapow PM, Burt A (2001) Indices of multilocus linkage disequilibrium. Mol Ecol Notes 1:101–102

Almany GR, Planes S, Thorrold SR, Berumen ML and others (2017) Larval fish dispersal in a coral-reef seascape. Nat Ecol Evol 1:0148

Ben-Tzvi O, Abelson A, Gaines SD, Bernardi G and others (2012) Evidence for cohesive dispersal in the sea. PLOS ONE 7:e42672

Berenshtein I, Paris CB, Gildor H, Fredj E, Amitai Y, Kiflawi M (2018) Biophysical simulations support schooling behavior of fish larvae throughout ontogeny. Front Mar Sci 5:254

Bernardi G, Beldade R, Holbrook SJ, Schmitt RJ (2012) Full-sibs in cohorts of newly settled coral reef fishes. PLOS ONE 7:e44953

Blouin MS, Parsons M, Lacaille V, Lotz S (1996) Use of microsatellite loci to classify individuals by relatedness. Mol Ecol 5:393–401

Bourke AF, Franks NR (1995) Social evolution in ants. Princeton University Press, Princeton, NJ

Broquet T, Viard F, Yearsley JM (2013) Genetic drift and collective dispersal can result in chaotic genetic patchiness. Evolution 67:1660–1675

Burgess SC, Snyder RE, Rountree B (2018) Collective dispersal leads to variance in fitness and maintains offspring size variation within marine populations. Am Nat 191: 318–332

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. Mol Ecol 18:4707–4715

Charlesworth B, Charlesworth D (1999) The genetic basis of inbreeding depression. Genet Res 74:329–340

Christie MR, Johnson DW, Stallings CD, Hixon MA (2010) Self-recruitment and sweepstakes reproduction amid extensive gene flow in a coral-reef fish. Mol Ecol 19: 1042–1057

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

Crnokrak P, Roff DA (1999) Inbreeding depression in the wild. Heredity 83:260–270

Csilléry K, Johnson T, Beraldi D, Clutton-Brock T and others (2006) Performance of marker-based relatedness estimators in natural populations of outbred vertebrates. Genetics 173:2091–2101

D'Aloia CC, Neubert MG (2018) The formation of marine kin structure: effects of dispersal, larval cohesion, and variable reproductive success. Ecology 99:2374–2384

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. Mol Ecol 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. Mol Ecol 23:2902–2913

D'Aloia CC, Bogdanowicz SM, Francis RK, Majoris JE, Harrison RG, Buston PM (2015) Patterns, causes, and consequences of marine larval dispersal. Proc Natl Acad Sci USA 112:13940–13945

D'Aloia CC, Bogdanowicz SM, Harrison RG, Buston PM (2017) Cryptic genetic diversity and spatial patterns of admixture within Belizean marine reserves. Conserv Genet 18:211–223

Dale MR, Fortin MJ (2014) Spatial analysis: a guide for ecologists, 2nd edn. Cambridge University Press, Cambridge

Glaubitz JC, Rhodes OE, DeWoody JA (2003) Prospects for inferring pairwise relationships with single nucleotide polymorphisms. Mol Ecol 12:1039–1047

Grosberg RK, Quinn JF (1986) The genetic control and consequences of kin recognition by the larvae of a colonial marine invertebrate. Nature 322:456–459

Hamilton WD (1964) The genetical evolution of social behaviour. II. J Theor Biol 7:17–52

Handcock MS, Hunter DR, Butts CT, Goodreau SM, Morris M (2008) statnet: software tools for the representation, visualization, analysis and simulation of network data. J Stat Softw 24:1–11

Harrison CS, Siegel DA, Mitarai S (2013) Filamentation and eddy–eddy interactions in marine larval accumulation and transport. Mar Ecol Prog Ser 472:27–44

Hart MW, Grosberg RK (1999) Kin interactions in a colonial hydrozoan (*Hydractinia symbiolongicarpus*): population structure on a mobile landscape. Evolution 53:793–805

Hatchwell BJ (2010) Cryptic kin selection: kin structure in vertebrate populations and opportunities for kin-directed cooperation. Ethology 116:203–216

Hedgecock D (1994) Does variance in reproductive success limit effective population sizes of marine organisms? In: Beaumont A (ed) Genetics and evolution of aquatic organisms. Chapman and Hall, London, p 122–134

Horne J, Bradbury I, Paterson I, Hardie D and others (2016) Complex post-larval dispersal processes in Atlantic cod revealed by age-based genetics and relatedness analysis. Mar Ecol Prog Ser 556:237–250

Iacchei M, Ben-Horin T, Selkoe KA, Bird CE, García-Rodríguez FJ, Toonen RJ (2013) Combined analyses of kinship and F_{ST} suggest potential drivers of chaotic genetic patchiness in high gene-flow populations. Mol Ecol 22:3476–3494

Jones GP, Almany GR, Russ GR, Sale PF, Steneck RS, van Oppen MJH, Willis BL (2009) Larval retention and connectivity among populations of corals and reef fishes: history, advances and challenges. Coral Reefs 28:307–325

Kaiser SA, Taylor SA, Chen N, Sillett TS, Bondra ER, Webster MS (2017) A comparative assessment of SNP and microsatellite markers for assigning parentage in a socially monogamous bird. Mol Ecol Resour 17:183–193

Kamel SJ, Grosberg RK (2013) Kinship and the evolution of social behaviours in the sea. Biol Lett 9:20130454

Kinlan BP, Gaines SD (2003) Propagule dispersal in marine and terrestrial environments: a community perspective. Ecology 84:2007–2020

Kopps AM, Kang J, Sherwin WB, Palsbøll PJ (2015) How well do molecular and pedigree relatedness correspond, in populations with diverse mating systems, and various types and quantities of molecular and demographic data? G3 Genes Genom Genet 5:1815–1826

Leis JM (1991) The pelagic stage of reef fishes: the larval biology of coral reef fishes. In: Sale PF (ed) The ecology of fishes on coral reefs, 1st edn. Acadmic Press, San Diego, CA, p 183–230

Majoris JE, Francisco FA, Atema J, Buston PM (2018a) Reproduction, early development, and larval rearing strategies for two sponge-dwelling neon gobies *Elacatinus lori* and *Elacatinus colini*. Aquaculture 483: 286–295

Majoris JE, D'Aloia CC, Francis RK, Buston PM (2018b) From individual decisions to population patterns: the distribution of a coral reef fish is explained by natural selection for behavioral preferences. Behav Ecol 29:429–439

Marko PB, Hart MW (2018) Genetic analysis of larval dispersal, gene flow, and connectivity. In: Carrier TJ, Reitzel AM, Heyland A (eds) Evolutionary ecology of marine invertebrate larvae. Oxford University Press, Oxford, p 164–188

Pew J, Muir PH, Wang J, Frasier TR (2015) related: an R package for analysing pairwise relatedness from codominant molecular markers. Mol Ecol Resour 15:557–561

Pinsky ML, Saenz-Agudelo P, Salles OC, Almany GR and others (2017) Marine dispersal scales are congruent over evolutionary and ecological time. Curr Biol 27:149–154

Pompanon F, Bonin A, Bellemain E, Taberlet P (2005) Genotyping errors: causes, consequences and solutions. Nat Rev Genet 6:847–859

Puebla O, Bermingham E, McMillan WO (2012) On the spatial scale of dispersal in coral reef fishes. Mol Ecol 21: 5675–5688

Queller DC (1994) Genetic relatedness in viscous populations. Evol Ecol 8:70–73

Queller DC, Goodnight KF (1989) Estimating relatedness using genetic markers. Evolution 43:258–275

Riquet F, Comtet T, Broquet T, Viard F (2017) Unexpected collective larval dispersal but little support for sweepstakes reproductive success in the highly dispersive brooding mollusc *Crepidula fornicata.* Mol Ecol 26: 5467–5483

Salles OC, Pujol B, Maynard JA, Almany GR and others (2016) First genealogy for a wild marine fish population reveals multigenerational philopatry. Proc Natl Acad Sci USA 113:13245–13250

Santure AW, Stapley J, Ball AD, Birkhead TR, Burke T, Slate J (2010) On the use of large marker panels to estimate inbreeding and relatedness: empirical and simulation studies of a pedigreed zebra finch population typed at 771 SNPs. Mol Ecol 19:1439–1451

Schunter C, Pascual M, Garza JC, Raventos N, Macpherson E (2014) Kinship analyses identify fish dispersal events on a temperate coastline. Proc R Soc B 281:20140556

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

Selwyn JD, Hogan JD, Downey-Wall AM, Gurski LM, Portnoy DS, Heath DD (2016) Kin-aggregations explain chaotic genetic patchiness, a commonly observed genetic pattern, in a marine fish. PLOS ONE 11:e0153381

Shima JS, Swearer SE (2016) Evidence and population consequences of shared larval dispersal histories in a marine fish. Ecology 97:25–31

Siegel DA, Kinlan BP, Gaylord B, Gaines SD (2003) Lagrangian descriptions of marine larval dispersion. Mar Ecol Prog Ser 260:83–96

Siegel DA, Mitarai S, Costello CJ, Gaines SD, Kendall BE, Warner RR, Winters KB (2008) The stochastic nature of larval connectivity among nearshore marine populations. Proc Natl Acad Sci USA 105:8974–8979

Taylor HR (2015) The use and abuse of genetic marker-based estimates of relatedness and inbreeding. Ecol Evol 5:3140–3150

Veliz D, Duchesne P, Bourget E, Bernatchez L (2006) Genetic evidence for kin aggregation in the intertidal acorn barnacle (*Semibalanus balanoides*). Mol Ecol 15: 4193–4202

Victor BC (1984) Coral reef fish larvae: patch size estimation and mixing in the plankton. Limnol Oceanogr 29: 1116–1119

Wang J (2007) Triadic IBD coefficients and applications to estimating pairwise relatedness. Genet Res 89:135–153

West SA, Pen I, Griffin AS (2002) Cooperation and competition between relatives. Science 296:72–75

Williamson DH, Harrison HB, Almany GR, Berumen ML and others (2016) Large-scale, multidirectional larval connectivity among coral reef fish populations in the Great Barrier Reef Marine Park. Mol Ecol 25:6039–6054

Editorial responsibility: Philippe Borsa, Montpellier, France

Submitted: April 27, 2018; Accepted: October 16, 2018
Proofs received from author(s): November 14, 2018