Genetic analysis of larval dispersal, gene flow, and connectivity

Peter B. Marko* and Michael W. Hart*

Keywords: demographic connectivity, differentiation, coalescence, relatedness

Introduction

"We have to remember that what we observe is not nature in itself but nature exposed to our method of questioning." Werner Heisenberg (1958)

One of the major hallmarks of marine species is that many produce large numbers of small pelagic larvae that drift in the ocean for varying periods of time. For these species, establishing the degree to which different populations are connected by larval dispersal is a fundamental goal for larval ecologists interested in understanding the influence of planktonic processes and larval supply on ecological and evolutionary processes within populations. Assessing and predicting local population and community dynamics, spread of invasive species, patterns of local adaptation, spread of advantageous alleles, maintenance of local biodiversity, sustainability of fisheries, and effective marine reserve design, all require some knowledge of rates and patterns of larval exchange among populations.

However, the tiny size of most marine larvae and the variable length of time they spend in the plankton present obvious and significant obstacles for identifying the geographic origins and destinations of dispersing larvae. The fate of marine larvae in the plankton may be likened to a black box (Buston and D'Aloia 2013): for any local population we can estimate its contribution to the pool of individuals in the planktonic darkness (many dispersing larvae), and its harvest of individuals that emerge into the light (fewer settling larvae), but we cannot easily describe the processes that affect the destination of larvae that disperse from a particular source, or the source of larvae that settle or recruit into a particular destination.

As with the study of all unobservable processes, the methods of inquiry will determine, to some extent, the apparent properties of the process. For example, not long ago, observations of marine larvae far offshore (Scheltema 1986) combined with widespread genetic homogeneity at allozyme loci (Buroker 1983; Saunders et al. 1986; Rosenblatt and Waples 1986), led many marine ecologists to the reasonable conclusion that marine larvae regularly travelled vast distances, such that many marine populations were likely well mixed on spatial scales of thousands to tens of thousands of kilometers (Palumbi 1992). When Palumbi (1995) reviewed the evidence for associations between variation in larval dispersal potential (such as among species with long or short pelagic larval duration) and the geographic distribution of genetic variation, nearly all comparative studies analyzed small numbers of populations and loci (typically allozymes and mtDNA) with a limited number of analytical approaches. In the intervening years, the size of data sets and the diversity of methods of analysis have grown dramatically, and significant progress has been made towards understanding the scope and scale of larval dispersal. Many tools have been used, but much of this progress has come from using genetic methods. In this chapter, we describe some of the most commonly used analytical genetic approaches and then discuss how these methods have improved our understanding of larval dispersal.

*Both authors contributed equally to this work.

Genetic approaches to the study of larval dispersal

Rates and patterns of larval dispersal have been studied using a wide variety of methods, including direct observations of larval movement (e.g., Gerrodette 1981; Olson 1985; Knowlton and Keller 1986), mark-and-recapture (e.g., Jones et al. 1999; Thorrold et al. 2001), natural environmental markers (e.g., DiBacco and Levin 2001), oceanographic modeling (Largier 2003; Siegel et al. 2003), hybrid zones (e.g., Gilg and Hilbish 1996), and the expansion of geographic ranges caused by anthropogenic introductions (e.g., Kinlan et al. 2005) or by climate change (e.g., Sunday et al. 2015). Techniques that employ artificial tags (i.e., to otoliths or calcified structures) or that estimate dispersal from inferred parent-offspring relationships may be considered direct measurements of net dispersal, but these methods do not necessarily measure larval dispersal alone, but instead some combination of larval and adult movement. We consider only direct observations of advection or diffusion of individual larvae in the plankton as yielding direct measurements of larval dispersal. Although direct observations give immediate insight into dispersal, they are limited to species with large, short-lived larvae that can be followed on small spatial scales. Indirect methods based on experimental or natural marking of larvae are similarly limited to species with larval structures that can be marked and that are retained in adults, and they share with direct observations some other important limitations (especially the inability to infer average effective rates of dispersal integrated over longer periods of time into the ecological or geological past). Consequently, indirect methods that use genetic data have become the most widely used approach for inferring patterns of larval dispersal.

Although many methods have been developed to infer patterns of larval dispersal from genetic data, we think a more useful and important categorization of genetic approaches is based on the theoretical framework used to infer patterns of larval dispersal. The oldest and most familiar framework is based on data in the form of allele frequencies and gene genealogies in explicit population genetic models. These population models include one or more parameters that represent **migration (*m*)**, the *proportion* of individuals or gene copies (i.e., alleles or haplotypes) in a population that are new immigrants each generation (and that successfully reproduce), and use a diverse range of assumptions and calculations, and either optimization (i.e., maximum likelihood) methods or simulations to find the best estimate of migration and other model parameter values fitted to the genetic data. When combined with an estimate of the effective population size (N_e or simply N), the product of the two parameters together can be used to characterize the population migration rate (***Nm***), usually interpreted as the *number* of immigrant individuals (i.e., organisms) per generation or the number of immigrant gene copies ($2Nm$, for diploid loci) per generation (Wright 1969). Because these **population-based** estimates represent the number of gene copies moving between populations, Nm is often called **gene flow** (Slatkin 1987, 1993; but see Mallet 2001).

The second, newer framework for these approaches is based on data in the form of multilocus genotypes for individual organisms without specifying a demographic model without explicit parameters that represent migration or gene flow. Instead, these **individual-based** approaches use multivariate clustering, similarity measures, or inferred parentage to assign sampled individuals to groups (i.e., nominal populations consisting of similar genotypes, or families consisting of parents, offspring, and siblings). Simple Mendelian inheritance rules or population parameters are used to identify the most-likely number of groups or clusters that

minimizes within-group differences (e.g., Pritchard et al. 2000; Wilson and Rannala 2003), or to assign individuals to likely families including parent-offspring pairs or groups of siblings (e.g., Marshall et al. 1998; Wang 2004). These individual-based approaches infer larval dispersal from *counts of migrant individuals*, including individuals that are strongly clustered with individuals from a different sample, and individuals that have a high likelihood of being closely related to parents or siblings in a different sample.

One of the most important distinctions between population-based methods and individual-based methods is the timescale over which they provide information about migration. All population-based approaches assume that the spatial distribution of alleles, the frequency of alleles in populations, and the structure of gene genealogies evolve slowly, and that this long timescale has an important effect on the observed patterns of genetic similarity or differences among populations. Therefore, population-based approaches estimate migration and gene flow as the time-averaged cumulative effects of variation in larval dispersal, larval retention, population growth, and other demographic or ecological processes integrated over relatively long evolutionary timescales (on the order of hundreds to thousands of generations). Individual-based approaches also make an important but different assumption about the temporal scale of genetic variation that can be used to detect migration. Strong allele- or genotype-sharing between members of the same genetic cluster or members of the same family is expected to be rapidly broken down by random mating with local mates (from other clusters or families) after migration. Consequently, recombination among immigrant and local alleles will degrade the signal of group membership or family identity among the descendants of an immigrant after one or a few generations. Counts of immigrants based on such approaches can only identify new migrants or their recent descendants on short ecological timescales, on the order of one or two generations (e.g., Wilson and Rannala 2003). A corollary of this limitation is that individual-based methods may not easily distinguish between recent immigrants that reach sexual maturity and reproduce (and contribute to gene flow) and those that do not.

With the recent incorporation of individual-based approaches into the genetic toolkit of larval biologists, researchers can potentially compare counts of immigrant individuals to historical patterns of gene flow inferred from more complex population-based approaches. In cases where both approaches give similar estimates of genetic connectivity, larval dispersal may have been consistently high (or low) over both long and short (recent) timescales, and those concordant measures of dispersal on both timescales may give reliable insight into the strength (or weakness) of population connectivity (e.g., Pinsky et al. 2016). Numerous factors can potentially explain contradictory results from these two approaches (e.g., Palstra et al. 2007), such as natural year-to-year variability, recent changes in larval dispersal caused by human impacts (and an opportunity for conservation ecologists to mitigate that change), genotyping errors, or violation of assumptions of either method. Distinguishing among these hypotheses may be difficult and will likely require repeated studies.

As with other approaches to the study of larval dispersal, those that use genetic techniques have strengths and weaknesses, and the choice of methods will depend on what a researcher wants to know about larval dispersal: an understanding of migration rates over relatively long evolutionary timescales (population-based methods) or documentation of extremely recent migration events (individual-based methods). In the next section, we briefly discuss the theory and practice of studying larval dispersal of marine species using genetic data

in either population- or individual-based methods. We then highlight some specific areas of progress in applying both individual- and population-based methods, and consider how those approaches give concordant or discordant insight into genetic and demographic connectivity based on larval dispersal among marine populations.

How to estimate larval dispersal from genetics

Population-based Methods

The use of genetic methods to estimate larval dispersal requires a realistic model of the processes that cause allele frequency changes and the evolution of allelic differences between populations. The primary processes are mutation, genetic drift, and natural selection. By contrast, gene flow is a homogenizing evolutionary force that slows, erodes, or prevents the buildup of genetic difference between populations. Most of the population genetic theory that has been developed to understand the movement of genes and individuals does so by focusing on the interaction between gene flow (which introduces alleles to populations) and genetic drift (which eliminates alleles from populations). Because genetic drift is a stochastic evolutionary force caused by random mating in a finite-sized population, it will work randomly and independently in different populations, such that, in the absence of gene flow between populations, genetic drift is expected to cause allele frequencies to diverge. Given sufficient time (generations), drift will cause the fixation of different alleles in different populations, meaning that different alleles will reach a frequency of 1.0 in each of the individual populations.

Neutral genetic markers. To concentrate exclusively on gene flow and genetic drift, population geneticists focus on neutral polymorphisms: allelic differences that are expected to have no (or few) direct effects on fitness. To focus on neutral loci, population geneticists can study genes or nucleotide sites that appear to have few functional constraints, such as microsatellite loci, anonymous DNA, and synonymous 3rd codon polymorphisms in protein-coding DNA (e.g., Karl and Avise 1992). One important consideration in identifying candidate neutral polymorphisms is their possible linkage to other polymorphisms under selection. More importantly, however, population geneticists can identify neutral polymorphisms for analysis using data from multiple, unlinked genetic loci. Unlike migration and genetic drift, natural selection is expected to cause idiosyncratic patterns of differentiation at individual loci, such that loci affected by selection can be identified and potentially excluded as outliers with respect to a larger sample of loci from across the genome (Schopf 1974; Koehn et al. 1976; Johnson and Black 1984). The important corollary of this idea is that similar spatial patterns of differentiation across multiple loci are best explained by the action of gene flow and genetic drift, forces that are expected to affect all loci across the genome in the same way. The statistical power for identifying outlier loci (influenced by selection) has increased significantly as data from very large numbers of loci can now be gathered and analyzed, potentially within the framework of a linkage map (e.g., Bradbury et al. 2013).

Classical Population Genetics. With respect to estimates of dispersal and gene flow, an important point of departure for classical population genetic methods is the Hardy-Weinberg Equilibrium (HWE) principle: in the absence of any evolutionary forces acting on a genetic locus, allele frequencies at that locus will remain constant over time. This deterministic theory also predicts that, for a locus with two alleles A and a with frequencies p and q, respectively, the

proportion of AA homozygotes, aa homozygotes, and Aa heterozygotes is expected to be p^2, q^2, and 2pq, respectively. Deviations from the expected genotype proportions in natural populations provide evidence that at least one force of evolution is influencing allele frequencies. In the absence of selection, Hardy-Weinberg deviations can be caused by several processes that lead to non-random mating among individuals, and can provide insight into the genetic structure of populations and, potentially, patterns gene flow.

For example, consider a neutral genetic locus in two isolated populations (**Fig. 1**). Because there is no larval dispersal (and no gene flow) between eastern and western populations, genetic drift has caused allele frequencies to diverge, such that they have become fixed for different alleles. In this situation, the HWE Principle provides a null model of high gene flow by predicting that, if eastern and western individuals *were* freely exchanging migrants and completely interbreeding with one another, half of the individuals (i.e., 2pq) should be heterozygotes. Although an extreme example, any divergence in allele frequencies between populations will result in a deficit of heterozygotes compared to expected HWE genotypic proportions under the null model of high gene flow between populations.

Detection of heterozygote deficiencies forms the basis for the most common measure of population genetic divergence, Wright's (1978) $F_{ST} = (H_T - H_S)/ H_T$, the difference between the expected HWE heterozygosity for the "total" population (H_T) and the average expected heterozygosity among "subpopulations" or individual populations (H_S) scaled by H_T. Other measures of genetic differentiation have been developed, but are like F_{ST} in that they describe how genetic variation is partitioned among populations or samples from different geographic locations (e.g., Excoffier et al. 1992). Genetic differentiation measured as F_{ST} and its analogs can also be used to infer the rate of gene flow among populations; in the simplest case, $F_{ST} = 1/(4Nm + 1)$ under a set of assumptions known collectively as Wright's (1951) "island model." The island model assumes many equally-sized populations (each of size N) in which a fixed proportion (m) of every population are immigrants each generation, and that m is relatively small. The model also assumes each population has been separated for long enough that gene flow among populations and genetic drift within populations have reached an evolutionary equilibrium. An important consequence of these assumptions is that F_{ST} can only be used to estimate the compound parameter Nm, a useful parameter for understanding the impact of gene flow on allele frequencies, but of less value in understanding the demographic impact of migration (see below).

Because many real populations likely do not conform to island-model assumptions (Whitlock and McCauley 1999; Beerli and Felsenstein 1998; Neigel 2002), estimates of Nm from F_{ST} are now rare in the literature (Marko and Hart 2011). However, understanding the theoretical relationship between Nm and F_{ST} highlights several important facts. First, Wright's equation $F_{ST} = 1/(4Nm + 1)$ demonstrates that a single migrant per generation has the same impact on allele frequencies in a large population as a single migrant per generation has in a small population. This counter-intuitive result is explained by the fact that, in a large population, the relatively small impact of a single migrant on allele frequencies is opposed by relatively weak genetic drift; by contrast, in a small population, a single migrant has a much larger effect on allele frequencies, but the impact of gene flow in a small population is opposed by much stronger genetic drift. Second, the relationship between F_{ST} and Nm emphasizes how little gene flow is necessary to keep allele frequencies similar among populations (**Fig. 2**) and how difficult

it can be to precisely measure migration when $Nm > 10$ migrants per generation, especially considering that the error associated with estimates of F_{ST} is often as large as the estimate itself when F_{ST} is small (Waples 1998). This "gene flow problem" (Waples 1998) makes real population differentiation very difficult to distinguish from random noise in marine species with high gene flow, and makes gene flow that is sufficient to homogenize allele frequencies very difficult to distinguish from panmixia (i.e., all individuals are potential mating partners).

Coalescent Population Genetics. Rather than modeling how allele frequencies are expected to change moving forward in time, coalescent population genetics focuses on the genealogical history (i.e., a gene tree) for a sample of gene copies moving backward in time: if two individual gene copies have the same common ancestor in a previous generation, those two copies are said to have coalesced. The earliest applications of coalescent theory were used to make inferences about demographic parameters for a single population, but the theoretical framework of the coalescent has been expanded to incorporate other demographic parameters. For example, gene flow between populations can be estimated with gene trees by inferring the rate at which gene copies in one population coalesce in an ancestor in another population (Nath and Griffiths 1993; Beerli and Felsenstein 1999; Nielsen and Wakeley 2001).

The primary advantage of coalescent gene flow estimators is that they often employ a much more realistic model of gene flow and population history than Wright's Island model (Beerli and Felsenstein 1999). Most coalescent gene flow estimators use a Bayesian statistical framework, in which a posterior distribution is estimated for each demographic parameter by simultaneously "sampling" (searching among) tree topologies and parameter values. Coalescent methods typically use computationally intensive Markov Chain Monte Carlo (MCMC) samplers, in which small random changes are repeatedly applied to gene trees. The likelihood of each gene tree and parameter estimate is calculated at each step in the search until the search converges on a sample of highly likely gene trees and associated parameter estimates. This capability to calculate likelihoods for multiple individual parameter values (including those associated with gene flow and other demographic processes) is an important source of the increased realism of coalescent gene flow estimators, in contrast to the insights gained from single summary statistics (such as F_{ST}). Either Nm or m can be estimated with coalescent samplers, but estimates of m are typically scaled by mutation and can only be converted into demographically meaningful values with an estimate (or assumption) about the mutation rate of the markers.

Coalescent methods have important limitations (Marko and Hart 2011). First, despite lacking several of the unrealistic assumptions of Wright's Island model, each coalescent estimator has an underlying demographic model that still makes some assumptions about the history and structure of the sampled populations that may or may not match reality. Second, given the high among-locus variance in the coalescent, robust and consistent answers from coalescent estimators require data from multiple loci (Karl et al. 2012). Third, even with high-performance computing clusters, it is impractical to apply coalescent methods that use MCMC samplers to genomic datasets consisting of thousands of individual Single Nucleotide Polymorphisms (SNPs) in which each SNP has its own gene tree (although many population genetic questions often do not require thousands of loci). Lastly, coalescent theory is based on mathematical approximations that assume that effective population size is large and that migration rates are low. So, like F_{ST}, coalescent methods are not expected to perform well when Nm is very large (>10) or when N is very small (<100).

Approximation Methods. A less computationally-intensive alternative, Approximate Bayesian Computation (ABC) methods (Beaumont 2002; Lopes and Beaumont 2009; Csilléry et al. 2010) estimate posterior distributions of demographic parameters from simulated data sets and summary statistics (e.g., number of alleles, genetic diversity, genetic distances) rather than from samples of likely gene trees. Instead of assuming a single demographic model (as with Nm from F_{ST} and from coalescent estimators), ABC typically starts by simulating data under several alternative demographic models (e.g., with and without migration) and then either accepting or rejecting models by comparing summary statistics for each simulated data set to the observed values. The posterior distributions for demographic parameters of interest are then approximated from the distribution of parameters values from the accepted models (like model selection and model averaging approaches used in some coalescent methods based on likelihoods). Like other population-based methods, even though the posterior for *m* is estimated in demographic quantities in most ABC methods, the value of m depends on the priors for mutation rates used in the simulations. Because ABC methods do not make full use of sequence data (i.e., the coalescent), they typically do not provide estimates of demographic parameters as precise as those from MCMC methods (Beaumont et al. 2002). However, the practical advantages of ABC lie in the capability to consider very large genome-wide data sets and to make direct comparisons among complex demographic models defined by the investigator (e.g., Rougemont et al. 2016). Alternatively, when estimating migration rates between two or more populations, the joint site frequency spectrum (SFS) can also be used instead of summary statistics. The SFS is more informative than any single summary statistic (all summary statistics can be calculated from allele frequencies, but allele frequencies cannot be calculated from summary statistics) and is advantageous in that increasing the number of SNPs or individuals does not proportionally increase the computational time, but greatly increases the power of the analysis (Excoffier et al. 2013). Another relatively new computationally efficient approach for inferring demographic parameters (including migration) combines diffusion approximation (e.g., Fisher 1922; Kolmogorov 1931; Kimura 1964) of the expected SFS under alternative demographic models with maximization of the similarity between the observed SFS and the simulated SFS across simulated parameter values in the demographic model (Gutenkunst et al. 2009).

Individual-based Methods

In contrast to population-based methods, the rate or direction of ongoing dispersal can be estimated by the classification of individual genotypes as immigrants (or the recent descendants of immigrants). This approach has great appeal because these insights into the dispersal history of individual organisms provide estimates of net dispersal and therefore come closer to the kind of direct insight into connectivity that would be gained from direct observations of larval movements in the plankton. Here we focus on three methods and their implementation that have been used by empiricists studying marine larvae. All three of these methods use transient, short-lived effects of immigration on inter-individual genetic variation to detect recent or ongoing migration events. This feature also sets the individual-based methods apart from population-based methods that draw inferences from allele frequencies, coalescent times, or other population genetic variables that are associated with changes in genetic variation on long timescales. Thus, individual-based methods give insights into connectivity among populations that are complementary to the results from population-based methods, but like population-based methods, they have their own distinctive limitations on those insights (e.g., Jones et al. 2005;

Planes et al. 2009; Christie et al. 2010; Harrison et al. 2012; Saenz-Agudelo et al. 2012; Pusack et al. 2014).

Clustering Individuals from Different Samples. The most intuitively appealing individual-based methods use clustering of individual genotypes from multiple samples (e.g., from several geographic locations) into one or more genetically similar groups; immigrants can then be identified as genotypes from one sample that are more confidently clustered with genotypes from some other sample or location. Clustering methods can potentially be used to estimate immigration (m) by dividing the number of immigrant genotypes identified in a population sample by the size of that sample, if the number of genetic clusters is known and correctly specified.

The most widely cited clustering method is the suite of algorithms in the program called STRUCTURE described by Pritchard et al. (2000). Such clustering methods are sometimes described as making fewer assumptions about demographic structure and history in comparison to population-based methods (Pearse and Crandall 2004). In the case of STRUCTURE, each cluster of genetically similar individuals is assumed to have its own demographic history in which genotypes at a single locus are in HWE (due to random mating in a large population), and alleles at different loci are expected to be in linkage equilibrium (due to recombination and independent assortment). Optimization by MCMC is used to find the individual assignments to clusters that minimize linkage disequilibrium (LD) and maximize HWE within each cluster. Thus, although clustering and other individual-based methods do not directly estimate migration rates or population size in an explicit population model, some of those model parameters enter the individual-based methods under the guise of the quantities to be optimized in the search for immigrant genotypes.

Under these assumptions, recent immigration events are expected to cause both transient LD among loci (by adding unusual combinations of alleles at different loci) and deviations from HWE for single loci (by adding unusual genotypes). These effects are expected to be short-lived because random mating and recombination will break up linkage groups and restore HWE within one or a few generations after each immigration event. Consequently, STRUCTURE results are sensitive mainly to the genetic signal from recent or ongoing immigration. In the original STRUCTURE model, first-generation immigrant individuals could be identified as those multilocus genotypes from one geographic region (or destination) that had a high probability of assignment to a cluster that was common in a different geographic region (or source). In subsequent versions of the model, recent descendants of immigrants could also be identified as those individuals with an admixture of alleles characteristic of both the source and destination populations (Falush et al. 2003).

Although STRUCTURE can be thought of as a method to "let the data define the populations" (Pearse and Crandall 2004), an important assumption of STRUCTURE is that the the total sample consists of one or more genetically discrete clusters each of which is internally homogeneous, and that the true number of clusters (K) is known and specified by the researcher. In other words, K is not a model parameter value estimated from the data by optimization, but rather a variable in the optimization algorithm. Incorrectly specifying K can lead to errors in assignment, and thus errors in inferring dispersal from the distribution of clusters among different geographic locations. The limitations on inferring both number of clusters and assignment of genotypes to clusters in the same optimization are well known (Pritchard et al.

2000), and several heuristic solutions to estimating K have been proposed (Evanno et al. 2005; Kalinowski 2011; Puechemaille 2016).

A second important limitation of the STRUCTURE method is the assumption that gene flow is low (Pritchard et al. 2000) and immigrants are rare. This assumption is inherent in all individual-based methods (and like the assumptions underlying population-based methods), which depend on the occurrence of recognizable clusters or differentiated populations that could be the source of distinctive immigrant genotypes. This leads to the surprising expectation that as the true migration rate (and the expected occurrence of immigrants in population samples) increases, the sensitivity of individual-based methods to count immigrants and quantify migration rates may greatly decline. Given this constraint, can clustering methods be used to discover ecologically meaningful gene flow, or can these methods discover only gene flow that is trivially low? Simulations (e.g., Waples and Gagiotti 2006) suggest that there may be "situations where *Nm* is high enough that a realistic population sample would contain enough immigrants to shed light on immigration patterns, yet where there remained enough differentiation between populations to endow genetic assignment methods with adequate power for F_0 [first-generation] immigrant detection" (Paetkau et al. 2004). It seems uncertain whether such situations are common among systems of marine animal populations. However, this sensitivity to the homogenizing effects of long-term gene flow suggests caution in the interpretation of individual-based estimates of ongoing gene flow when those estimates are high (Saenz-Agudelo et al. 2009; Lowe and Allendorf 2010).

Assigning Immigrant Individuals to Source Populations. This second approach includes some of the same model parameters from population-based methods (including the migration rate, *m*), but estimates those model parameters by identifying recent immigrants and their population of origin (rather than by characterizing long-term rates of gene flow). The most widely used of these methods is called BayesAss (Wilson and Rannala 2003). Each sampling location is assumed to constitute a population that may include some first-generation immigrants from one or more source populations, as well as recent (second or third generation) descendants of immigrants; allele and genotype frequencies at each locus can vary among populations (and to vary away from HWE conditions); and different pairs of populations may exchange migrants at different rates. Unlike clustering methods, BayesAss starts with the assignment (Paektau et al. 1995) of each individual genotype to the sample where that genotype's expected frequency is the greatest (based on the observed distributions of alleles). Then, like clustering methods, the fit of the model to multilocus genotype assignments is evaluated by maximum likelihood, and optimization is used to find the most likely values of *m* that can account for the number and source of immigrant genotypes (or descendants of recent immigrants one or two generations into the past). As with clustering methods, the assignment tests are sensitive only to recent immigration because additional generations of mating with non-immigrant genotypes will erode the signal of immigrant ancestry beyond the second-generation descendants of immigrants.

A significant limitation of this assignment approach may be its ability to resolve complex patterns of migration among a biologically realistic (large) number of populations (Faubet et al. 2007; Mardulyn et al. 2008). Meirmans (2014) showed that estimates of migration rates from BayesAss analyses may be biased by computational limitations on the ability to optimize model parameter values (migration rates into many populations) from limited data (small numbers of

individuals and loci). In general, the quality of BayesAss results improves with larger samples of organisms and deeper sampling of genomes, but declines with larger numbers of populations.

Assigning Individuals to Families. A third – and conceptually distinct – approach to counting migration events includes the fewest population model parameters. This approach uses genealogical methods to infer parent-offspring relationships among sampled genotypes, and infers migration from the discovery of close family members in different population samples. One highly cited method is called CERVUS; the original version was designed to assign paternity to offspring given genotype data for those offspring and their known mothers (Marshall et al. 1998); extensions of the method allowed for the effects of genotyping errors, and for more accurate assignment of parentage given only genotypes of offspring and candidate parents (and without a known maternal or paternal genotype; Kalinowski et al. 2007). Some candidate parents can be excluded by allelic mismatches with offspring genotypes; like clustering and assignment methods, CERVUS then uses likelihood scores to assess non-excluded candidate parents and identify the most likely parent for each parent-offspring pair (based on the frequencies of the shared alleles, and heterozygosity of the parental genotypes). Unlike other methods, which fit population model parameters (and characterize confidence in the parameter value estimates) by optimization, parentage methods are based on simple Mendelian inheritance rules rather than on a formal population model; instead of optimization, the confidence in the identification of a specific parent-offspring pair in CERVUS is assessed by comparison to simulations that use empirical allele frequencies from the sampled populations.

Although methods like CERVUS are designed to assign parentage and identify parent-offspring pairs, some studies that identify the same parent(s) for more than one offspring can thus also identify full- or half-sibling pairs, including siblings that were collected in different population samples. This is a significant but largely untapped strength of parentage methods: they provide the only individual- or population-based genetic approach that can quantify migration specifically caused by advection of offspring away from their parents (e.g., a planktonic cohort of sibling larvae that disperses away from the parental population in an ocean current), and distinguish this from migration caused by diffusion of siblings away from each other (e.g., spread of siblings of the same cohort or different cohorts due to spatial or temporal variation in current speed and direction). Both advection and diffusion in the plankton contribute to observed levels of migration and gene flow, but the two modes of dispersal have different ecological and evolutionary consequences (Palmer and Strathmann 1981) on both small (Grosberg 1991) and large spatial scales (Largier 2003).

Parentage methods (especially those that use exclusion to screen out most candidate parent-offspring pairs) also have a significant weakness: they may be sensitive to the effects of genotyping errors that cause non-parental alleles to be observed in true offspring of a parent (e.g., so-called stuttering of microsatellite allele sizes; Bonin et al. 2004); conversely, as the size of studies grow, a large number of pair-wise comparisons can cause unrelated individuals to share alleles by chance. A counter-intuitive effect of the sensitivity to genotyping error is that the number of mistakes in parentage assignment may increase with the number of sampled loci in assignment methods based on exclusion (because non-parental alleles observed in offspring may be sufficient to mistakenly exclude a true parent-offspring pair), in contrast to other individual- or population-based methods where confidence in clustering or population assignment should increase with the number of loci sampled (Sobel et al. 2002). Genotyping errors can be included

in probabilistic models for identifying parent-offspring pairs (Kalinowski et al. 2007), but decreasing the sensitivity of the models to genotyping error also decreases the accuracy with which true parent-offspring pairs may be distinguished from other genetic similarities between individuals (Christie 2009). This sensitivity to genotyping errors may impose a significant limitation on the use of some parentage methods in genome-scale studies of larval dispersal and gene flow.

Improved understanding of larval dispersal and gene flow

Here we highlight several areas of advancement since Palumbi's (1995) review, especially those areas that have benefited from the application of new coalescent population models or individual-based methods.

Biological correlates of larval dispersal: planktonic larval duration

Because most marine larvae cannot be followed directly in the plankton, ecologists have long searched for useful proxies for dispersal potential. One common and accessible proxy – the duration of the planktonic larval stage (PLD) – can be estimated by rearing larvae in the laboratory or by observing calibrated growth marks (such as daily increments in growth of fish otoliths) in larvae collected from the plankton. Estimates of dispersal potential based on PLD vary among species from several minutes to several years (Strathmann and Strathmann 2007). Palumbi (1995) reviewed the early evidence for variation in PLD. He asked whether PLD covaries with (and statistically accounts for) realized dispersal measured as differentiation itself (e.g., F_{ST}) or as a pattern of increased differentiation among populations separated by larger geographic distances known as isolation-by-distance (IBD), in which the strength of IBD is characterized by the slope of a regression of F_{ST} against geographic distance between pairs of populations (Slatkin 1993; Rousset 1997). This definition of IBD among populations in a stepping-stone model of multiple habitat patches is slightly different from Wright's (1943) original definition of IBD among individuals in a single habitat patch in which typical dispersal distances are less than the dimensions of the habitat (what Wright called "local inbreeding in a continuous area"). However, the two definitions share a similar concept of limited or localized dispersal leading to greater genetic differences on larger spatial scales.

Several early and important comparative studies of congeneric marine gastropod species with or without planktonic larvae established support for the specific and intuitive idea of an inverse relationship between PLD and population genetic differentiation (e.g., Berger 1973, Snyder and Gooch 1973, Gooch 1975), and the for the general idea that PLD can be used as a proxy for realized or typical larval dispersal distances (Crisp 1978). Other comparative genetic studies of related and co-distributed molluscs (e.g., Hoagland 1986; Kyle and Boulding 2000; Collin 2001), fishes (e.g., Waples 1987; Doherty et al. 1995), echinoderms (e.g., McMillan et al. 1992; Arndt et al. 1998), corals (Hellberg 1996), and crustaceans (e.g., Duffy 1993) provide additional evidence of a strong negative correlation between PLD and population genetic differentiation.

However, a steady accumulation of counterexamples cast some doubt on the generality of this pattern (Burton 1983; Palumbi 1995). Some of the exceptions are more difficult to evaluate given that they are generally not comparative studies of either closely-related or co-distributed taxa, but are instead studies of population structure in single species that revealed unexpectedly high or low genetic differentiation relative to the authors' expectations based on PLD estimates

(Saunders et al. 1985; Watts et al. 1990; France et al. 1992; Todd and Lambert 1993; Planes 1993; Wares et al. 2001; Taylor and Hellberg 2003; Rocha et al. 2005; Baums et al. 2006; Bowen et al. 2006; Marko et al. 2007). Several literature reviews and meta-analyses have attempted to resolve this issue, but have come to substantially different conclusions, with some studies reporting a strong relationship between PLD and metrics of population differentiation (Bohonak 1999; Siegel 2003; Shanks et al. 2003) and others reporting a much weaker relationship (Bradbury et al. 2008; Ross et al. 2009; Weersing and Toonen 2009; Kelly and Palumbi 2010; Selkoe and Toonen 2011; Riginos et al. 2011).

Several factors may account for this uncertain or contentious relationship between PLD and population genetic differentiation. First, although meta-analyses of large numbers of studies have great power, they also confound variation in the biology of the study organisms with variation in the methodological approaches and shortcomings of the individual studies (Selkoe and Toonen 2011; Dawson 2014). Sample sizes, spatial scales of sampling, biogeographic region, genetic marker choice, and the metric of realized larval dispersal (especially the choice of F_{ST} versus the IBD slope) can affect the apparent relationship between PLD and population genetic differentiation (Weersing and Toonen 2009; Selkoe and Toonen 2011).

Second, the strength of early comparative studies lay in phylogenetically-controlled comparisons of co-distributed taxa. However, these studies focused on comparisons between species with relatively large, qualitative differences in dispersal potential (i.e., planktonic larvae versus non-planktonic larvae). Although species that lack planktonic larvae are relevant to predictions about the effect of PLD on dispersal, a strongly bimodal distribution of PLD (with one mode at zero for species without a planktonic larva) biases the perceived strength of the overall relationship between time spent in the plankton and genetic differentiation (Bay et al. 2006; Ross et al. 2009; Weersing and Toonen 2009; Kelly and Palumbi 2010; Riginos et al. 2011).

Third, the use of PLD as a proxy for dispersal potential is itself fraught with difficulty. Laboratory measures of PLD do not easily account for seasonal and annual variation (especially in temperature and food availability) in nature. The capabilities of some larvae to greatly extend their time in the plankton, by the uptake of dissolved organic matter (e.g., Moran and Manahan 2004) or by developmental arrest (e.g., Pradillon et al. 2001), and the abilities of other larvae to enhance or limit their advection by active swimming (e.g., Kough et al. 2014) or by orientation to physical and chemical cues in the ocean (e.g., Mouritsen et al. 2013), may also contribute to a mismatch between laboratory measurements of PLD, actual time spent in the plankton, and the realized effects of larval duration on dispersal and gene flow.

Lastly, F_{ST}-based metrics of realized dispersal invoke Wright's Island Model assumptions. Are these assumptions justified, or would other measures of migration and gene flow (e.g., McGovern et al. 2010; Crandall et al. 2012) more closely reflect the dispersal capabilities of larvae? Most of these assumptions are difficult to test. The assumption of a drift-migration equilibrium is supported by (but not a requirement for) observations of significant IBD in a system of populations (Wares 2002). Gene flow and genetic drift are also more likely to be near equilibrium at small spatial scales between neighboring populations (Slatkin 1993; Hellberg 1995). Consistent with that view, Selkoe and Toonen (2011) found a stronger relationship between realized dispersal distances (the IBD slope) and PLD when analyzing F_{ST} measured on relatively small spatial scales (<650 km). It was surprising, then, that the relationship between

PLD and F_{ST} was not improved by analyzing only datasets with significant IBD (Selkoe and Toonen 2011).

Given these conflicting results, it seems fair to ask: does planktonic larval duration predict the scale and magnitude of marine larval dispersal as estimated from measures of population genetic differentiation? Clearly, species that lack planktonic larvae typically show much stronger patterns of population genetic differentiation than species with planktonic larvae. However, the overall quantitative relationship between PLD and genetic differentiation is much less strong, probably because dispersal potential of planktonic larvae is difficult to estimate from laboratory PLD, F_{ST} is sometimes a poor proxy for gene flow, and F_{ST} itself is sensitive to the overall level of polymorphism of the genetic markers (e.g., Meirmans and Hedrick 2010), and because many other factors probably affect the realized dispersal of marine larvae.

One proposed solution to that unsatisfying conclusion is to ask the same question in a more specific and restricted form. Evidence of a fundamental mechanistic relationship between PLD and realized dispersal is most often clear in comparisons between co-distributed species that have occupied the same seascape for the same amount of time (as measured by similarly aged coalescents), so that genetic drift has had the same opportunity to cause the evolution of population differentiation proportional to PLD differences among species (Dawson 2014; Dawson et al. 2014). Focusing only on such synchronously diverging and co-distributed taxa may give the clearest view of the mechanistic relationship between PLD and F_{ST}. Unfortunately, this approach is less useful for interpreting population differentiation in studies of single or idiosyncratic species (lacking a co-distributed species of a similar coalescent age for comparison), for which population genetic structure may be important and interesting but for which a comparative context is unavailable for interpreting the geographic extent or magnitude of differentiation relative to PLD.

A second proposed solution is to ask the question in a broadly expanded form that includes PLD plus the many other factors that influence realized dispersal, such as the regional oceanography or "seascape" (e.g., Barshis et al. 2012; Sunday et al. 2014; Liggins et al. 2015), habitat availability (e.g., Crandall et al. 2012), population size and fecundity (e.g., Saenz-Agudelo et al. 2010; Dawson et al. 2014), temperature (e.g., Kelly and Eernisse 2007; Bradbury et al. 2008), adult behavior during spawning (e.g., Carson et al. 2010), and larval behavior in the plankton (e.g., Gerlach et al. 2007). This multifactorial approach seems promising. For example, numerous studies have reported highly variable genetic differentiation on small spatial scales, much less than the expected geographic scale of larval dispersal. This pattern is in some ways the opposite of IBD, and has been dubbed "chaotic genetic patchiness" (Johnson and Black 1982; David et al. 1997 Selkoe et al. 2010; Cornwell et al. 2016). One proposed cause of such patterns is strong temporal variation in both the sources of planktonic larvae and in availability of recruitment sites among destination populations (Johnson and Black 1984). A second proposed cause of such patterns is high variation in reproductive success among broadcast-spawning adults, leading to low genetic diversity within cohorts of offspring (relative to diversity among all adults) and strong differentiation among cohorts (Hedgecock 1994; Hedgecock and Pudovkin 2011). This type of sweepstakes reproduction (with few winners, and many losers) can reduce local genetic effective population size within each reproductive event and may promote population differentiation. However, migration of larvae – and especially the diffusion of larvae from a single cohort away from each other – should lead to the homogeneous sharing of

offspring of the sweepstakes 'winners' among many populations, and should prevent the evolution of genetic differentiation among populations with overlapping generations derived from multiple cohorts of larvae (Hedgecock 1994). Broquet et al. (2013) used simulations to show that a simple additional factor – a small reduction in diffusion, or a small tendency for larvae to aggregate and to disperse together – could produce strong local population differentiation that was quantitatively like empirical measures of chaotic patchiness in nature (also see Eldon et al. 2016). Broquet et al. (2013) conceded that aggregation and collective dispersal of cohorts seems unlikely, but is consistent with greater than expected kinship among some larvae and new recruits (Selkoe et al. 2006; Hedgecock et al. 2007; Iacchei et al. 2013; Ottman et al. 2016; Sewlyn et al. 2016), and could be readily tested using data from the quantification of dispersal kernels (**Fig. 3**) using individual-based methods.

Estimates of dispersal distances

Given that few generalities can be made about larval dispersal from large meta-analyses, the attention of some marine ecologists has shifted towards understanding how far individual larvae travel between fertilization and metamorphosis, and to characterizing the frequency distribution of those dispersal distances. That frequency distribution is also known as the dispersal kernel (Largier 2003; Botsford et al. 2009), and it is expected to be related to many of the most important ecological and evolutionary quantities associated with pelagic larval development.

Isolation-by-distance. One approach to measuring characteristic larval dispersal distances has used parameter estimates from population-based methods such as F_{ST}. However, this fixation index is an unreliable measure of genetic differentiation in species with high gene flow (**Fig. 2**; Waples 1998). The IBD slope is likely a better proxy for realized dispersal in high-dispersal marine species (Palumbi 2003) because there should be no relationship between F_{ST} and geographic distance if error associated with F_{ST} is as large as the estimates of F_{ST}.

In an influential study, Kinlan and Gaines (2003) used a population genetic simulation of larval dispersal (from Palumbi 2003) to characterize the relationship between variation in simulated dispersal distance and variation in the resulting IBD slope among simulated populations. They then applied that function to estimate dispersal distances from empirical studies of IBD patterns among population genetic samples of marine organisms. Kinlan and Gaines (2003) used this simulation approach to circumvent the difficulties in obtaining direct estimates of larval dispersal distances for species with prolonged planktonic development and large population sizes (Cowen and Sponaugle 2009). As with Palumbi (2003), they inferred surprisingly short dispersal distances of only approximately 0.01-100 km per generation, and proposed that a large part of that extraordinary variation was related to PLD and other intrinsic biological differences in dispersal potential among marine animal species.

Many other marine animals have geographic ranges that greatly exceed the dispersal distances estimated from IBD patterns. A plausible mechanism leading to that difference is stepping-stone migration (Kimura and Weiss 1964) over many generations of short-distance dispersal events that link far-flung populations (e.g., Planes and Fauvelot 2002). Crandall et al. (2012) reviewed examples of IBD, and noted that genetic evidence for stepping-stone migration (and its implications for larval dispersal) is rare relative to many examples of spatial genetic variation that are more consistent with the effects of recent range expansion, rare long-distance dispersal events, or other nonequilibrium processes. Crandall et al. (2012) used a clever

comparative approach to distinguish the effects of stepping-stone gene flow from the contributions of other processes to population genetic variation. They sampled mtDNA population variation across Indo-Pacific island archipelagos for neritid snails in which adults of some species live in marine habitats, but adults of some other species have evolved to live only in freshwater streams on high islands. Both types of species have swimming planktonic marine larvae, and adults of both types can live on high islands (in freshwater or marine habitats), but the two types of species differ in their opportunities for stepping-stone gene flow: species with marine adults can disperse between high islands over several generations by using the marine habitats of low atolls as stepping stones; in contrast, amphidromous species can only disperse between high islands directly in a single generation (because adults are unable to live on atolls that lack freshwater habitats). By using a coalescent population-based method to model migration rates independent of other population model parameters, Crandall et al. (2012) showed that stepping-stone dispersal led to greater gene flow between high islands for one marine species but not for amphidromous species. A physical circulation model of ocean currents provided important context for that comparison: in some cases, high islands were connected by strong currents that allowed direct dispersal between populations of both marine and amphidromous species; only in cases without a strong direct oceanographic connection did the two types of species differ in migration rates (due to their different abilities to make use of stepping-stone dispersal via atolls).

This example has several interesting implications. First, it shows that careful dissection of multiple factors (including ocean currents and adult habitat requirements) may often be needed to reveal the specific circumstances that lead to a correlation between genetic differentiation and geographic distance, and to estimate the contribution of stepping-stone larval dispersal to the direction and magnitude of gene flow. Second, the neritid example suggests that coalescent population models (rather than summary statistics such as F_{ST} from Wright's Island Model) may be needed to estimate migration rates from genetic data in a complex biogeographical setting where many other processes also affect larval movements in the plankton. Third, the example illustrates a potentially important limitation on the use of individual-based methods and dispersal kernels to understand larval migration, gene flow, and population structure. Because individual-based methods detect only the effects of ongoing or recent immigration over one or a few generations, they may be inherently incapable of capturing a complete view of larval migration if stepping-stone dispersal over many generations is an important mechanism linking far-flung parts of a metapopulation. Distant populations in a stepping-stone model may show little genetic connectivity on short time scales (measured by individual-based methods), but may show substantial connectivity on longer time scales (measured by population-based methods), especially if stepping-stone gene flow is episodic and influenced by temporally varying patterns of ocean circulation. In this sense, the characterization of dispersal kernels from data using clustering methods or assignment tests (see below) may be a necessary but insufficient basis for linking larval biology and dispersal to ecological or evolutionary processes, and may require the addition of population-based methods that integrate the effects of gene flow over many generations via stepping-stone migration.

Dispersal kernels. In contrast to the IBD approach that uses population-based methods, a second approach to measuring characteristic larval dispersal distances is the dispersal kernel or the frequency distribution of distances between parents and their offspring. This distribution can be characterized using individual-based methods to identify and count immigrants, assign them

to source populations, and describe the frequency distribution of larval dispersal distances. An elegant example from a coral reef fish illustrates both the promise and the limits of this approach. D'Aloia et al. (2014) used microsatellite polymorphisms and mitochondrial DNA sequences to characterize population genetic variation in the neon goby *Elacatinus lori* from several populations sampled on the continuous barrier reef system and on isolated atolls along ~150 km of the coast of Belize in the western Caribbean. Adults of this species live in large sponges where males brood developing embryos; hatched larvae leave the sponge (and the parents) to spend three weeks or more in the plankton before recruiting to a new sponge host. Despite the potential for long-distance larval dispersal during that long planktonic period, population differentiation (e.g., mtDNA F_{ST} up to 0.46) was strong on both small (< 20 km) and large spatial scales; differentiation was especially strong between atolls and the barrier reef; and there appeared to be little evidence of IBD.

To better understand those patterns, D'Aloia et al. (2015) then intensively sampled $>10^5$ single-locus microsatellite genotypes from thousands of *E. lori* adults and newly recruited offspring along a 41 km transect of the barrier reef. The parentage method in CERVUS identified 120 parent-offspring pairs, which allowed D'Aloia et al. (2015) to characterize the dispersal kernel from that high-resolution frequency distribution: the most common dispersal distance was ≤1 km, the mean distance (estimated by model-fitting; **Fig. 3**) was <3 km, and the largest distance traveled by any offspring was only ~16 km. Many (24) offspring were members of 11 different full- or half-sib families, including eight families in which siblings diffused away from each other and were advected different distances away from their parents; two families in which siblings were advected the same distance from the parent(s) (without diffusion away from each other); and one family in which all siblings recruited to the same location with the parent (and without advection or diffusion). A single family included both the longest observed dispersal distance from the parent (16.4 km) for one of the half-sibs, as well as the modal dispersal distance from the parent (0.8 km) for the other half-sib. That variation within and among families suggests that advection (away from parents), diffusion (away from siblings), and local retention of larvae all contribute to patterns of larval dispersal and gene flow in *E. lori*.

Additional modeling (D'Aloia et al. 2015) suggested that population genetic differentiation along the barrier reef was caused by a form of isolation-by-distance among mating groups (associated with different host sponges) that were separated by distances greater than a few kilometers, like Wright's (1943) original definition of IBD. Although D'Aloia et al. (2014) emphasized a lack of isolation-by-distance among populations, a relatively strong pattern of IBD was evident among the barrier reef populations only (**Fig. 4**), and that pattern was consistent with the expected effects of short-distance larval dispersal along the continuously distributed barrier reef habitat. In contrast, D'Aloia et al. (2015) argued that populations on atolls are demographically isolated from the barrier reef (and from each other) by deeper water and the absence of stepping stones (the host sponge species) in the intervening habitat. That interpretation was supported by the absence of IBD on any geographic scale for pairwise differentiation among atoll populations or between atolls and the barrier reef (**Fig. 4**).

Comparable patterns of spatial genetic variation have been found in other *Elacatinus* species (Taylor and Hellberg 2003, 2006), and similar dispersal kernels have been found in a handful of other studies that used individual-based genetic methods to characterize dispersal distances (e.g., clownfish; Buston et al. 2012; Almany et al. 2016; Pinsky et al. 2016). These

results seem important because they suggest that realized dispersal distances of many coral reef fish may be considerably shorter than expected (Kinlan and Gaines 2003; Cowen & Sponaugle 2009; Shanks 2009). They implicate larval behaviour in the plankton (e.g., swimming activity, orientation, vertical migration) and stepping-stone gene flow between habitat patches that are separated by distances less than the typical larval dispersal distance (e.g., Crandall et al. 2012) as critically important factors that mediate the relationship between potential and realized dispersal over both short timescales (reflected in the dispersal kernel) and over long timescales (reflected in population model parameters such as pairwise population F_{ST}).

This detailed treatment of a specific example from gobies illustrates both the potential importance of individual-based genetic methods and the limits on their application for understanding marine larval dispersal. In a pilot study focused on a single *E. lori* population from a 0.5 km section of barrier reef, D'Aloia et al. (2013) were able to assign almost 5% of offspring to parents; when the geographic scale of the study was extended to capture the full dispersal kernel, the assignment rate was reduced by about half (D'Aloia et al. 2015), in spite of the large sample size of adults and offspring (>7000), because the broader study area included many more individuals of which a smaller proportion were sampled, genotyped, and assigned to families. That difference suggests that it may be realistic to use parentage (or other individual-based) methods to characterize dispersal kernels only on the smallest spatial scales in small populations of organisms with long-lived adults. For example, Stockwell et al. (2016) sampled thousands of SNP loci from Indo-Pacific parrotfish (*Scarus niger*) collected at several population locations in the Philippines separated by >100 km. They discovered convincing evidence of larval dispersal and gene flow on that larger spatial scale (two pairs of siblings in different populations, identified using an assignment method for identifying family members called ML-RELATE; Kalinowski et al. 2006). That discovery suggests that the method can in principle be applied to widely separated populations connected by planktonic larval dispersal on scales much greater than observed in species like *E. lori*. However, the low assignment rates for *S. niger* individuals grouped into families (~2%) implies that a vast collecting and genotyping effort might be needed to find large numbers of sibling pairs (and dispersal distances) and to fully characterize the dispersal kernel to the same precision achieved by D'Aloia et al. (2015; **Fig. 3**). Moreover, such studies can provide only individual snapshots of the dispersal kernel; accounting for temporal or spatial variation in the frequency distribution of dispersal distances would require spatial replication and temporal repetition. Those sampling constraints seem to put the dispersal kernel for *S. niger* and many other species (with broad geographic ranges, extensive larval dispersal, and large populations subject to spatial and temporal variation in dispersal) out of the reach of individual-based methods, even for studies like that of Stockwell et al. (2016) that use automated high-throughput genomic data collection.

Consequences of larval dispersal: genetic connectivity versus demographic connectivity

The distinction between population- and individual-based methods, including the diversity of specific parameters that each may estimate (e.g., *Nm* versus *m*), has become increasingly important for marine ecologists, who have emphasized understanding patterns of marine "connectivity" (**Fig. 5**) via larval dispersal to resolve fisheries and conservation problems such as the design of effective networks of marine protected areas (MPAs; Gaines et al. 2010, Fig. 1). In this context, it is useful to distinguish among several types of connectivity that reflect different population processes and can be estimated from genetic data under different model assumptions.

First, marine ecologists and resource managers are typically interested in characterizing **demographic connectivity**, the degree to which population persistence or growth rates are affected over **ecological timescales** by larval dispersal among populations (Waples 1998; Waples and Gaggioti 2006; Lowe and Allendorf 2010). How much dispersal is necessary to provide significant demographic linkage among individual populations is not well understood and highly context-dependent. One proposed standard is that a migration rate of ~10% (i.e., m = 0.1) is needed to maintain demographic synchrony (similar population dynamics and growth rates) among separate populations linked by dispersal (Hastings 1993; Waples and Gaggiotti 2006).

By contrast, biogeographers and population geneticists have emphasized the degree to which larval dispersal affects evolutionary processes within populations and evolutionary divergence among them (Lowe and Allendorf 2010). These effects can include both positive (e.g., preventing inbreeding depression) and negative consequences (e.g., preventing the fixation of locally adapted alleles via gene swamping) of gene flow. Wright (1951) concluded that Nm>1 was sufficient to prevent the harmful effects of inbreeding and genetic drift within populations. As a consequence, many population geneticists have used this one-migrant-per-generation (OMPG) "rule" as a threshold value defining **genetic connectivity**, the amount of gene flow necessary to homogenize allele frequencies among populations over **evolutionary timescales**. However, close inspection of the relationship between F_{ST} and Nm (**Fig. 2**) reveals that much greater gene flow (Nm >10) is necessary to prevent divergence in allele frequencies. The confusion in the literature over the OMPG rule and the expected effects of gene flow led Lowe and Allendorf (2010) to recognize three distinct levels of genetic connectivity: *drift connectivity* (gene flow that prevents divergence in allele frequencies, Nm >10); *inbreeding connectivity* (gene flow that prevents the harmful loss of genetic diversity, Nm >1); and *adaptive connectivity* (gene flow that is sufficiently high to promote the spread of adaptive alleles, Nm >0.1).

The differences among population- and individual-based genetic methods for analyzing larval dispersal have obvious parallels with these distinctions between demographic connectivity (focused on per capita migration rates, m) and genetic connectivity (focused on gene flow, Nm), as well as implications for the choice of methods to meet different study objectives or goals. Clustering and assignment methods characterize migration distances between populations, and parentage methods characterize migration distances away from parents; if sampling is sufficiently intensive, those methods can estimate the migration rate as the proportion of migrants entering a destination population (m) through the identification of migrant individuals, but not necessarily the evolutionary impact of immigration in the destination population (Nm) unless the effective population size is known. Therefore, almost by definition, individual-based methods have the potential to characterize demographic connectivity on short (ecological) timescales (Christie et al. 2010); few give insight into effective population size, such that individual-based methods offer little insight into genetic connectivity on longer (evolutionary) timescales.

By contrast, most population-based methods (F_{ST} and coalescent methods) provide estimates of the compound parameter Nm (the number of migrants per generation), which provides insight into the evolutionary impact of larval dispersal. This is an important limitation because estimates of Nm provide little insight into demographic connectivity unless N or the population growth rate is known (Lowe and Allendorf 2010). This shortcoming of population-

based methods may often be moot for some marine species with large population sizes because Nm can be most reliably estimated when it is small ($Nm < 10$), and for many marine species with large population sizes, $Nm < 10$ implies $m << 0.1$ and is likely to be demographically insignificant. The same cannot be said for cases in which $Nm > 10$, because population-based methods have little power to distinguish drift connectivity from strong demographic connectivity (**Fig. 2**). However, even in cases where Nm can be reasonably interpreted in the context of demographic connectivity (or if m can be calculated with a known mutation rate), the parameter estimates from most population-based methods are time-averaged over long (evolutionary) timescales, and may not be similar to present-day migration rates, particularly for species whose populations may be declining or growing, or whose distributions may be fragmented and rapidly changing in response to anthropogenic effects.

Conclusions

For the last 20 years, population-based genetic approaches have been the go-to analytical methods for studying marine connectivity, whereas the development and use of individual-based methods is in its infancy. Do the differences among the two classes of methods and their variants matter for genetic analyses of larval migration? Population-based methods have proven to be most useful for identifying very strong genetic breaks that often reflect a long-term lack of both genetic and demographically significant connectivity among populations, and genetic data can provide crucial complementary data for ecological studies that find evidence of demographic asynchrony (Peterson et al. 1996; Marko and Barr 2007). However, if gene flow is high, population-based methods are not expected to return clear and reliable answers about either genetic or demographic connectivity. Individual-based methods may have greater potential and show considerable promise to uncover evidence of demographic connectivity among populations, but, like population-based methods, are not expected to perform well when migration rates are high (Waples and Gaggioti 2006; Paetkau et al. 2004; Lowe and Allendorf 2010) and rely fundamentally on intensive sampling of all potential source and sink populations. In short, genetic methods can provide strong inferences about either an absence of (either genetic or demographic) connectivity or high genetic connectivity, but they are, at present, less useful for demonstrating strong demographic connectivity.

Clearly, no single approach will always tell a complete story of realized dispersal and connectivity. Comparing results from different analyses may generate insights or reveal problems inherent in any single approach, particularly for predicting marine connectivity and its implications for fisheries management or the design of networks of MPAs in a rapidly changing environment. However, even though individual- and population-based estimates of connectivity can be inferred from the same data, direct comparisons of migration rates over evolutionary timescales to migration on more recent ecological timescales have been rare (e.g., Pusack et al. 2014; D'Aloia et al. 2015; Pinsky et al. 2016), possibly because gene flow is often too high to expect informative population-based results on the spatial scale at which individual-based methods are most often applied (but see Pinsky et al. 2016), and because the types of genetic markers that are best for individual-based methods (microsatellites and SNPs, but see Rougemont et al. 2016) are also not ideal for some of the most powerful population-based methods (DNA sequences). In addition to inferences about larval dispersal from genetic data, direct measurements of the impacts of dispersal on demographic processes are likely crucial for

demonstrating meaningful demographic connectivity for the purposes of resource management (Waples 1998; Lowe and Allendorf 2010). The complementary combination of these methods seems likely to illuminate both the scope and scale of larval dispersal in marine systems.

Summary

- Genetic data can be used to characterize the scale or magnitude of connectivity via larval dispersal in the plankton as the per capita migration rate (m), the rate of gene flow (Nm), or counts of immigrant individuals.
- Population-based methods infer average effective rates of connectivity on long time scales (hundreds to thousands of generations), and those estimates will influenced by many processes (including larval dispersal).
- Individual-based methods based on clustering or assignment of individual genotypes to populations or families are suitable for estimating connectivity on short timescales.
- The typical or characteristic larval dispersal distance for any one system of populations may best be characterized by isolation-by-distance patterns (using population model methods) or by the dispersal kernel (using parentage-based methods)
- Migration rates estimated from individual-based methods may be more relevant to ecological studies of demographic connectivity (e.g., among demes in a network of marine protected areas) compared to rates of gene flow estimated from population-based methods.

Acknowledgements

The ideas expressed in this chapter have been developed and honed through conversations with many individuals. We are especially grateful to Mark Christie, Joe Neigel, and an anonymous reviewer for numerous suggestions and constructive criticism. Thanks to Cassidy D'Aloia for sharing data and unpublished analyses of gobies.

References

Almany, G. R., Hamilton, R. J., Bode, M., Matawai, M., Potuku, T., Saenz-Agudelo, P., Planes, S., Berumen, M. L., Rhodes, K. L., Thorrold, S. R., Russ, G. R., Jones, J. P. 2013. Dispersal of grouper larvae drives local resource sharing in a coral reef fishery. *Current Biology* 23, 626–630.

Arndt, A., Smith, M. J. 1998. Genetic diversity and population structure in two species of sea cucumber: differing patterns according to mode of development. *Molecular Ecology* 7: 1053-64.

Barshis, D. J., Sotka, E. E., Kelly, R. P., Sivasundar, A., Menge, B. A., Barth, J. A., Palumbi, S. R. 2011. A link between coastal upwelling and temporal genetic variability in the acorn barnacle *Balanus glandula*. *Marine Ecology Progress Series* 439: 139-150.

Baums, I. B., Paris, C. B., Chérubin, L. M. 2006. A bio-oceanographic filter to larval dispersal in a reef-building coral. *Limnology and Oceanography* 51: 1969–1981

Bay, L. K., Crozier, R. H., Caley, M. J. 2006. The relationship between population genetic structure and pelagic larval duration in coral reef fishes on the Great Barrier Reef. *Marine Biology* 149: 1247–1256.

Beaumont, M. A., Zhang, W., Balding, D. J. 2002. Approximate Bayesian computation in population genetics. *Genetics* 162: 2025–2035.

Beerli, P., Felsenstein, J. 1999. Maximum-likelihood estimation of migration rates and effective population numbers in two populations using a coalescent approach. *Genetics* 152: 763-73

Berger, E. 1973. Gene-enzyme variation in three sympatric species of *Littorina*. *Biological Bulletin* 145: 83-90.

Bohonak, A. J. (1999) Dispersal, gene flow, and population structure. *Quarterly Review of Biology* 74: 21-45.

Bonin, A., Bellemain, E., Eidesen, P. B., Pompanon, F., Brochmann, C., and Taberlet, P. 2004. How to track and assess genotyping errors in population genetics studies. *Molecular Ecology* 13: 3261-3273.

Botsford, L. W., White, J. W., Coffroth, M.-A., Paris, C. B., Planes, S., Shearer, T. L., Thorrold, S. R., Jones, G. P. 2009. Connectivity and resilience of coral reef metapopulations in marine protected areas: matching empirical efforts to predictive needs. *Coral Reefs* 28: 327-337.

Bowen, B. W, Bass, A. L., Muss, A., Carlin, J., Robertson, D. R. 2006.Phylogeography of two Atlantic squirrelfishes (family Holocentridae): exploring links between pelagic larval duration and population connectivity. *Marine Biology* 149: 899–913,

Bradbury, I. R., Laurel, B., Snelgrove, P. V. R., Bentzen P., Campana, S. E. 2008. Global patterns in marine dispersal estimates: the influence of geography taxonomic category and life history. *Proceedings of the Royal Society B* 275: 1803-1809.

Bradbury, I. R., Hubert, S., Higgins, B., Bowman, S., Borza, T., Paterson, I. G., Snelgrove, P. V. R., Morris, C. J., Gregory, R. S., Hardie, D., Hutchings, J. A., Ruzzante, D. E., Taggart, C. T., Bentzen, P. 2013. Genomic islands of divergence and their consequences for the resolution of spatial structure in an exploited marine fish. *Evolutionary Applications* 6: 450–461.

Broquet, T., Viard, F., Yearsley, M. 2013. Genetic drift and collective dispersal can result in chaotic genetic patchiness. *Evolution* 67: 1660-1675.

Buroker, N. E. 1983. Population genetics of the American oyster *Crassostrea virginica* along the Atlantic coast and Gulf of Mexico. *Marine Biology* 75:99-112.

Burton, R. S. 1983. Protein polymorphisms and genetic differentiation of marine invertebrate populations. *Marine Biology Letters* 4: 193–206.

Buston, P. M. & D'Aloia, C. C. 2013. Marine Ecology: Reaping the benefits of local dispersal. *Current Biology* 23: R351-R353.

Buston, P. M., Jones, G. P., Planes, S., Thorrold, S. R. 2012. Probability of successful larval dispersal declines fivefold over 1 km in a coral reef fish. *Proceedings of the Royal Society B - Biological Sciences* 279: 1883-1888.

Carson, H. S., López-Duarte, M. P., Rasmussen, L., Wang, D., and Levin, L. A. 2010. Reproductive timing alters population connectivity in marine metapopulations. *Current Biology* 20: 1926–1931.

Christie, M. R. 2009. Parentage in natural populations: novel methods to detect parent-offspring pairs in large data sets. *Molecular Ecology Resources* 10: 115-128.

Christie, M. R., Tissot, B. N., Albins, M. A., Beets, J. P., Jia, Y., Ortiz, D. M., Thompson, S. E., Hixon, M. A. 2010. Larval connectivity in an effective network of marine protected areas. *PLoS One* 5: e15715.

Collin, R. 2001. The effects of mode of development on phylogeography and population structure of North Atlantic *Crepidula* (Gastropoda: Calyptraeidae). *Molecular Ecology* 10: 2249-2262.

Cornwell, B. H., Fisher, J. L., Morgan, S. G., Neigel, J. E. 2016. Chaotic genetic patchiness without sweepstakes reproduction in the shore crab *Hemigrapsus oregonensis*. *Marine Ecology Progress Series* 548: 139-152.

Cowen, R. I., Sponaugle, S. 2009. Larval dispersal and marine population connectivity. *Annual Review of Marine Science* 1: 443-466.

Crandall, E. D., Treml, E. A., and Barber, P. H. 2012. Coalescent and biophysical models of stepping-stone gene flow in neritid snails. *Molecular Ecology* 21: 5579-5598.

Crisp, D. J. (1978) Genetic consequences of different reproductive strategies in marine invertebrates. *NATO Conference Series IV Marine Sciences* 2: 257-73

Csilléry, K., Blum, M. G. B., Gaggiotti, O. E., François, O. 2010. Approximate Bayesian computation in practice. *Trends in Ecology and Evolution* 25: 410–418.

D'Aloia, C. C., Bogdanowicz, S.M., Francis, R. K., Majoris, J. E., Harrison, R. G., Buston, P. M. 2015. Patterns, cause, and consequences of marine larval dispersal. *Proceedings of the National Academy of Sciences* 112: 13940-13945.

D'Aloia, C. C., Bogdanowicz, S. M., Harrison, R. G., Buston, P. M. 2014. Seascape continuity plays an important role in determining patterns of spatial genetic structure in a coral reef fish. *Molecular Ecology* 23: 2902-2913.

D'Aloia, C. C., Bogdanowicz, S. M., Majoris, J. E., Harrison, R. G., Buston, P. M. 2013. Self-recruitment in a Caribbean reef fish: a method for approximating dispersal kernels accounting for seascape. *Molecular Ecology* 22: 2563-2572.

David, P., P. Berthou, P. Noel, Jarne, P. 1997. Patchy recruitment patterns in marine invertebrates: a spatial test of the density-dependent hypothesis in the bivalve *Spisula ovalis*. *Oecologia* 111: 331–340.

Dawson, M. N. 2014. Natural experiments and meta-analyses in comparative phylogeography. *Journal of Biogeography* 41: 52– 65.

Dawson, M. N., Hays, C. G., Grosberg, R. K., Raimondi, P. T. 2014. Dispersal potential and population genetic structure in the marine intertidal of the eastern North Pacific. *Ecological Monographs* 84: 435-456.

DiBacco, C., Levin, L. A. 2000. Development and application of elemental fingerprinting to track the dispersal of marine invertebrate larvae. *Limnology and Oceanography* 45: 871–80.

Doherty, P. J., Planes, S, Mather, P. 1995. Gene flow and larval duration in seven species of fish from the Great Barrier Reef. *Ecology* 76:2373–2391

Duffy, J. E. 1993. Genetic population structure in two tropical sponge-dwelling shrimps that differ in dispersal potential. *Marine Biology* 116: 459-70.

Eldon, B., Riquet, F., Yearsley, J., Jollivet, D., Broquet, T. 2016. Current hypotheses to explain genetic chaos under the sea. *Current Zoology* 62: 551-566.

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., Smouse, P. E., Quattro, J. M. 1992. Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics* 131: 479-491.

Excoffier, L., Dupanloup, I., Huerta-Sánchez, E., Sousa, V. C., Foll, M. 2013. Robust demographic inference from genomic and SNP data. *PLoS Genetics* 9: e100390.

Falush, D., Stephens, M., and Pritchard, J. K. 2003. Inference of population structure using multilocus genotype data: Linked loci and correlated allele frequencies. *Genetics* 164: 1567-1587.

Faubet, P., Waples, R. S., and Gaggiotti, O. E. 2007. Evaluating the performance of a multilocus Bayesian method for the estimation of migration rates. *Molecular Ecology* 16: 1149-1166.

Faurby, S., Barber, P. H. 2012. Theoretical limits to the correlation between pelagic larval duration and population genetic structure. *Molecular Ecology*. 21: 3419-3432.

Felsenstein, J. 1992. Estimating effective population size from samples of sequences: inefficiency of pairwise and segregating sites as compared to phylogenetic estimates. *Genetical Research* 59: 139-147.

Felsenstein, J. 2006. Accuracy of coalescent likelihood estimates: Do we need more sites, more sequences, or more loci? *Molecular Biology and Evolution* 23: 691-700.

Fisher, R. A. 1922. On the dominance ratio. Proceedings of the Royal Society of Edinburgh 55: 399–433.

Gaines, S. D., White, C., Carr, M. H., Palumbi, S. R. 2010. Designing marine reserve networks for both conservation and fisheries management. *Proceedings of the National Academy of Sciences of the USA* 107: 18286-18293.

Gerlach, G., Atema J., Kingsford M. J., Black K. P., Miller-Sims, V. 2007 Smelling home can prevent dispersal of reef fish larvae. *Proceedings of the National Academy of Sciences of the USA* 104: 858–863.

Gerrodette, T. 1981. Dispersal of the solitary coral *Balanophyllia elegans* by demersal planular larvae. *Ecology* 62: 611–619.

Gilg, M. R., Hilbish, T. J. 2003. Geography of marine larval dispersal: Coupling genetics with fine-scale physical oceanography. *Ecology* 84: 2989-2998.

Gooch, J. L. 1975. Mechanisms of evolution and population genetics. In: O. Kinne (ed.). *Marine Ecology*, Vol. 2. Part I, p. 349. London: Wiley.

Grosberg, R. K. 1991. Sperm-mediated gene flow and the genetic structure of a population of the colonial ascidian *Botryllus schlosseri. Evolution* 45: 130-142.

Gutenkunst, R. N., Hernandez, R. D., Williamson, S. H., Bustamante, C. D. 2009. Inferring the joint demographic history of multiple populations from multidimensional SNP frequency data. *PLoS Genetics* 5: e1000695.

Hare, M. P. 2001. Prospects for nuclear gene phylogeography. *Trends in Ecology and Evolution* 16: 700–706.

Hare, M. P., Nunney, L., Schwartz, M. K., Ruzzante, D. E., Burford, M., Waples, R. S., Ruegg, K., Palstra, F. 2011. Understanding and estimating effective population size for practical application in marine species management. *Conservation Biology* 25: 438-449.

Harrison, H. B., Williamson, D. H., Evans, R. D. 2012. Larval export from marine reserves and the recruitment benefit for fish and fisheries. *Current Biology* 22: 1023–1028.

Hartl, D. L., Clark, A. 1997. *Principles of population genetics*. Sunderland: Sinauer Associates.

Hastings, A. 1993. Complex interactions between dispersal and dynamics: lessons from coupled logistic equations. *Ecology* 74: 1362–1372.

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

Hedgecock, D., Launey, S., Pudovkin, A. I., Naciri, Y., Lapegue, S., Bonhomme, F.. 2007. Small effective number of parents (N-b) inferred for a naturally spawned cohort of juvenile European flat oysters *Ostrea edulis*. *Marine Biology* 150:1173–1182.

Hedgecock, D., and A. I. Pudovkin. 2011. Sweepstakes reproductive success in highly fecund marine fish and shellfish: a review and commentary. *Bulletin of Marine Science* 87: 971–1002.

Heisenberg, W. 1958. *Physics and Philosophy: The Revolution in Modern Science.* London: Penguin.

Hellberg, M. E. 1995. Stepping stone gene flow in the solitary coral *Balanophyllia elegans:* equilbrium and nonequilibrium at different spatial scales. *Marine Biology* 123: 573–81

Hellberg, M. E. 1996. Dependence of gene flow on geographic distance in two solitary corals with different larval dispersal capabilities. *Evolution* 50: 1167-75.

Hickerson, M. J. and Cunningham, C. W. 2005. Contrasting Quaternary histories in an ecologically divergent sister pair of low-dispersing intertidal fish (*Xiphister*) revealed by multilocus DNA analysis. *Evolution* 59: 344-360

Hoagland, K. E. 1986. Patterns of encapsulation and brooding in the Calyptraeidae (Prosobranchia: Mesogastropoda). *American Malacological Bulletin* 4: 173–183.

Johnson, M. S., Black, R. 1982. Chaotic genetic patchiness in an intertidal limpet, *Siphonaria* sp. *Marine Biology* 70: 157–164.

Johnson, M. S., Black, R. 1984. Pattern beneath the chaos: the effect of recruitment on genetic patchiness in an intertidal limpet. *Evolution* 38: 1371–1383.

Jones, G. P., Milicich, M. J., Emslie, M. J, Lunow, C. 1999. Self-recruitment in a coral reef fish population. *Nature* 402: 802–804.

Jones, G. P., Planes, S., Thorrold, S. R. 2005. Coral reef fish larvae settle close to home. *Current Biology* 15: 1314–1318.

Kalinowski, S. T. 2011. The computer program STRUCTURE does not reliably identify the main genetic clusters within species: simulations and implications for human population structure. *Heredity* 106: 625-632.

Kalinowski, S. T., Taper, M. L., and Marshall, T. C. 2007. Revising how the computer program CERVUS accommodates genotyping error increases success in paternity assignment. *Molecular Ecology* 16: 1099-1106.

Kalinowski, S. T., Wagner, A. P., and Taper, M. L. 2006. ML-RELATE: a computer program for maximum likelihood estimation of relatedness and relationship. *Molecular Ecology Notes* 6: 576-579.

Karl, S. A., and Avise, J. C. 1992. Balancing selection at allozyme loci in oysters: implications from nuclear RFLPs. *Science* 256: 100-102.

Karl, S. A., Toonen, R. J., Grant, W. S., Bowen, B. W. 2012. Common misconceptions in molecular ecology: echoes of the modern synthesis. *Molecular Ecology* 21: 4174-4189.

Kelly, R. P., Eernisse, D J. 2007. Southern hospitality: a latitudinal gradient in gene flow in the marine environment. *Evolution* 61: 700–707.

Kelly, R. P., Palumbi, S. R. 2010. Genetic structure among 50 species of the northeastern pacific rocky intertidal community. *PLoS One* 5: e8594.

Kingman, J. F., 1982. The coalescent. *Stochastic Processes and Applications* 13: 235–248.

Kimura, M. 1964. Diffusion models in population genetics. Journal of Applied Probability 1: 177–232.

Kimura, M., Weiss, G. H. 1964. The stepping stone model of population structure and the decrease of genetic correlation with distance. *Genetics* 49: 561-576.

Kinlan, B. P., and Gaines, S. D. 2003. Propagule dispersal in marine and terrestrial environments: a community perspective. *Ecology* 84: 2007-2020.

Kinlan, B. P., Gaines, S. D., Lester, S. E. 2005. Propagule dispersal and the scales of marine community process. *Diversity and Distributions* 11: 139-148.

Komolgorov, A. 1931. Über die analttischen Methoden in der Wahrscheinlichkeitsrechnung. *Mathematische Annalen* 104: 415–458.

Knowlton, N., Keller, B.D. 1986. Larvae which fall short of their potential: highly localized recruitment in an alpheid shrimp with extended larval development. *Bulletin of Marine Science* 39: 213-223.

Koehn, R. K., Milman, R., Mitton, J. B. 1976. Population genetics of marine pelecypods. IV. Selection, migration, and genetic differentiation in the blue mussel *Mytilus edulis*. *Evolution* 30: 2-32.

Kough, A. S., Paris, C. B., Staaterman, E. 2014. In situ swimming and orientation behavior of spiny lobster (*Panulirus argus*) postlarvae. *Marine Ecology Progress Series* 504: 207-219.

Kritzer, J. P., Sale, P. F. 2004. Metapopulation ecology in the sea: from Levins' model to marine ecology and fisheries science. *Fish and Fisheries* 5: 131-140.

Kuhner, M. K. 2009. Coalescent genealogy samplers: Windows into population history. *Trends in Ecology and Evolution* 24: 86-93.

Kuhner, M. K., Yamato, J., Felsenstein, J. 1998. Maximum likelihood estimation of population growth rates based on the coalescent. *Genetics* 149: 429-434.

Kyle, C. J., Boulding, E. G. 2000. Comparative population genetic structure of marine gastropods (*Littorina* spp.) with and without pelagic larval dispersal. *Marine Biology* 137: 835-45.

Largier, J. L. 2003. Considerations in estimating larval dispersal distances from oceanographic data. *Ecological Applications* 13: S71-S89.

Liggins, L. , Treml, E. R., Possingham, H. P, Riginos, C. 2016. Seascape features, rather than dispersal traits, predict spatial genetic patterns in co-distributed reef fishes. *Journal of Biogeography* 43: 256-267.

Lopes, J. S., Beaumont, M. A. 2009. ABC: A useful Bayesian tool for the analysis of population data. *Infection, Genetics, and Evolution* 10: 826-833.

Lowe, W. H. and Allendorf, F. W. 2010. What can genetics tell us about population connectivity? *Molecular Ecology* 19: 3038-3051.

Mallet, J. 2001. Gene flow. In: Woiwod, I.P., Reynolds, D.R. & Thomas, C.D. (eds.) *Insect Movement: Mechanisms and Consequences*, pp. 337-360. Wallingford: CAB International.

Mardulyn, P., Vaesen, M.-A., and Milinkovitch, M. C. 2008. Controlling population evolution in the laboratory to evaluate methods of historical inference. *PLoS One* 3: e2960.

Marko, P. B. 2004. 'What's larvae got to do with it?' Disparate patterns of post-glacial population structure in two benthic marine gastropods with identical dispersal potential. *Molecular Ecology* 13:597–611.

Marko, P. B., Barr, K. B. 2007. Basin-scale patterns of mtDNA differentiation and gene flow in the Bay Scallop, *Argopecten irradians concentricus* Say. *Marine Ecology Progress Series* 349: 139-150.

Marko, P. B., Hart, M. W. 2011. The complex analytical landscape of gene flow inference. *Trends in Ecology and Evolution* 26: 448-456.

Marko, P. B., Rogers-Bennett, L., Dennis, A. B. 2007. MtDNA population structure and gene flow in lingcod (*Ophiodon elongatus*): limited connectivity despite long-lived pelagic larvae. *Marine Biology* 150: 1301-11.

Marshall, T. C., Slate, J., Kruuk, L. E. B., Pemberton, J. M. 1998. Statistical confidence for likelihood-based paternity inference in natural populations. *Molecular Ecology* 7: 639-655.

McGovern, T. M, Keever C. A., Hart, M. W., Saski, C., Marko, P. B. 2010. Divergence genetics analysis reveals historical population genetic processes leading to contrasting phylogeographic patterns in co-distributed species. *Molecular Ecology* 19: 5043–5060.

McMillan, W. O., Raff, R. A., Palumbi, S. R. 1992. Population genetic consequences of developmental evolution in sea urchins (Genus *Heliocidaris*). *Evolution* 46: 1299-312.

Meirmans, P. G. 2014. Non-convergence in Bayesian estimation of migration rates. *Molecular Ecology Resources* 14: 726-733.

Meirmans, P. G., Hedrick, P. W. 2010. Assessing population structure: F_{ST} and related measures. *Molecular Ecology Resources* 11: 5-18.

Mills, L. S., Citta, J. J., Lair, K. P., Schwartz, M. K., Tallmon, D. A. 2000. Estimating animal abundance using noninvasive DNA sampling: promise and pitfalls. *Ecological Applications* 10: 283-294.

Moran, A. L., Manahan, D. T. 2004. Physiological recovery from prolonged starvation in larvae of the Pacific oyster *Crassostrea gigas*. *Journal of Experimental Marine Biology and Ecology* 306: 17-36.

Mouritsen, H., Atema, J., Kingsford, M. J., Gerlach, G. 2013. Sun compass orientation helps coral reef fish larvae return to their natal reef. *PLoS One* 8:e66039

Nath, H. B., Griffiths, R. C. 1993. The coalescent in two colonies with symmetric migration. *Journal of Mathematical Biology* 31: 841-852.

Neigel, J. E. 2002. Is F_{ST} obsolete? *Conservation Genetics* 3: 167-173.

Nielsen, R., Wakeley, J. 2001. Distinguishing migration from isolation: A Markov chain Monte Carlo approach. *Genetics* 158, 885-896

Olson, R. R. 1985. The consequences of short distance larval dispersal in a sessile marine invertebrate. *Ecology* 66: 30-39.

Ottmann, D., Grorud-Colvert, K., Sard, N. M., Huntington, B. E., Banks, M. A., Sponaugle, S. 2016. Long-term aggregation of larval fish siblings during dispersal along an open coast. *Proceedings of the National Academy of Sciences USA*. 113: 14067-14072.

Paetkau, D., Slade,R., Burden, M., Estoup, A. 2004. Genetic assignment methods for the direct, real-time estimation of migration rate: a simulation-based exploration of accuracy and power. *Molecular Ecology* 13: 55-65.

Palmer, A. R., Strathmann, R. R. 1981. Scale of dispersal in varying environments and its implications for life histories of marine invertebrates. *Oecologia* 48: 308-318.

Palstra, F. P., O'Connell, M. F., Ruzzante, D. E. 2007. Population structure and gene flow reversals in Atlantic salmon (*Salmo salar*) over contemporary and long-term temporal scales: effects of population size and life history. *Molecular Ecology* 16: 4504-4522.

Palumbi, S. R. 1992. Marine speciation on a small planet. *Trends in Ecology and Evolution* 7: 114-118.

Palumbi, S. R. 1995. Using genetics as an indirect estimator of larval dispersal. In: L. McEdward (ed.), *Ecology of Marine Invertebrate Larvae*, pp. 369-388. Cambridge: Cambridge University Press.

Palumbi, S. R. 2003. Population genetics, demographic connectivity, and the design of marine reserves. *Ecological Applications* 13: S146-S158.

Pearse, D. E., Crandall, K. A. 2004. Beyond F_{ST}: Analysis of population genetic data for conservation. *Conservation Genetics* 5: 585-602.

Pelc, R. A., Warner, R. R., Gaines, S. D. 2009. Geographical patterns of genetic structure in marine species with contrasting life histories. *Journal of Biogeography* 36: 1881-1890.

Peterson, C. H., Summerson, H. C., Luettich, R. A. Jr. 1996. Response of bay scallops to spawner transplants: a test of recruitment limitation. *Marine Ecology Progress Series* 132: 93-107.

Pinsky, M. L., Saenz-Agudelo, P., Salles, O. C., Almany, G. R., Bode, M., Berumen, M. L., Andrefouet, S., Thorrold, S. R., Jones, G. P., Planes, S. 2016. Marine dispersal scales are congruent over evolutionary and ecological time. *Current Biology* (in press). http://dx.doi.org/10.1016/j.cub.2016.10.053.

Planes, S. 1993. Genetic differentiation in relation to restricted larval dispersal of the convict surgeonfish *Acanthurus triostegus* in French Polynesia. *Marine Ecology Progress Series* 98: 237-246.

Planes, S., Fauvelot, C. 2002. Isolation by distance and vicariance drive genetic structure of a coral reef fish in the Pacific Ocean. *Evolution* 56: 378-399.

Planes, S., Jones, G. P., Thorrold, S. R. 2009. Larval dispersal connects fish populations in a network of marine protected areas. *Proceedings of the National Academy of Sciences of the USA* 106: 5693-5697.

Pluzhnikov, A., Donnelly, P. 1996. Optimal sequencing strategies for surveying molecular genetic diversity. *Genetics* 144: 1247-1262.

Pradillon, F., Shillito, B., Young, C., Gaill, F. Developmental arrest in vent worm embryos. *Nature* 413: 698-699.

Pritchard, J. K., Stephens, M., Donnelly, P. 2000. Inference of population structure using multilocus genotype data. *Genetics* 155: 945-959.

Puechmaille, S. J. 2016. The program STRUCTURE does not reliably recover the correct population structure when sampling is uneven: subsampling and new estimators alleviate the problem. *Molecular Ecology Resources* 16: 608-627.

Pusack, T. J., Christie, M. R., Johnson, D. R., Stallings, C. D., Hixon, M. A. 2014. Spatial and temporal patterns of larval dispersal in a coral-reef fish metapopulation: evidence of variable reproductive success. *Molecular Ecology* 23: 3396-3408.

Riginos, C., Douglas, K. E., Shanahan, D. F., Treml, E. A. 2011. Effects of geography and life history traits on genetic differentiation in benthic marine fishes. *Ecography* 34: 566-575.

Rocha, L. A., Robertson, D. R., Roman, J., Bowen, B. W. 2005. Ecological speciation in tropical reef fishes. *Proceedings of the Royal Society London, Series B* 272: 573–579.

Rougemont, Q., Gagnaire, P.-A., Perrier, C., Genthon, C., Besnard, A.-L., Launey, S. Evanno, G. 2016. Inferring the demographic history underlying parallel genomic divergence among pairs of parasitic and nonparasitic lamprey ecotypes. *Molecular Ecology* (in press).

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

Ross, P. M., Hogg, I. D., Pilditch, C. A., Lundquist, C. J. 2009. Phylogeography of New Zealand's coastal benthos. *New Zealand Journal of Marine and Freshwater Research* 43: 1009-1027.

Saenz-Agudelo, P., Jones, G. P., Thorrold, S. R., 2016. Estimating connectivity in marine populations: an empirical evaluation of assignment tests and parentage analysis under different gene flow scenarios. Molecular Ecology 18: 1765-1776.

Saenz-Agudelo, P., Jones, G. P., Thorrold, S. R., 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.

Saunders, N. C., Kessler, L. G., Avise, J. C. 1986. Genetic variation and geographic differentiation in mitochondrial DNA of the horseshoe crab, *Limulus polyphemus*. *Genetics* 112: 613–627.

Scheltema, R. S. 1971. Larval dispersal as a means of genetic exchange between geographically separated populations of shallow-water benthic marine gastropods. *Biological Bulletin* 140: 284-322.

Schopf, T. J. M. 1974. Survey of genetic differentiation in a coastal zone invertebrate: the ectoproct *Schizoporella errata*. *Biological Bulletin* 146: 78-87.

Selkoe, K. A., Gaines, S. D., Caselle, J. E., Warner, R. R. 2006. Current shifts and kin aggregation explain genetic patchiness in fish recruits. *Ecology* 87:3082–3094.

Selkoe, K. A, Toonen, R. J. 2011. Marine connectivity: a new look at pelagic larval duration and genetic metrics of dispersal. *Marine Ecology Progress Series* 436: 291-305.

Selkoe, K. A., Watson, J. R., White, C. W., Horin, T. B., Iacchei, M., Mitari, S., Siegel, D. A., Gaines, S. D., Toonen, R. J. 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: 3708-3426.

Shanks, A. L. 2009. Pelagic larval duration and dispersal distances revisited. *Biological Bulletin* 216: 373-385.

Selwyn, J. D., Hogan, J. D., Downey-Wall, A. M., Gurski, L. M., Portnoy, D. S., Heath, D. D. 2016. Kin-Aggregations Explain Chaotic Genetic Patchiness, a Commonly Observed Genetic Pattern, in a Marine Fish. *PLoS ONE* 11: e0153381.

Shanks, A. L., Grantham, B. A., Carr, M. H. 2003. Propagule dispersal distance and the size and spacing of marine reserves. *Ecological Applications* 13:1 59–169.

Siegel, D. A., Kinlan, B. P., Gaylord, B., Gaines, S. D. 2003. Lagrangian descriptions of marine larval dispersion. *Marine Ecology Progress Series* 260: 83-96.

Sinclair, M. 1988. *Marine Populations: an Essay on Population Regulation and Speciation.* Seattle: University of Washington Press.

Slatkin, M. 1987. Gene flow and the geographic structure of natural populations. *Science* 236: 787-792.

Slatkin, M. 1993. Isolation by distance in equilibrium and nonequilibrium populations. *Evolution* 47: 264-279.

Snyder, T. P., Gooch, J. L. 1973. Genetic differentiation in *Littorina saxatilis* (Gastropoda). *Marine Biology* 22:177-182.

Sobel, E., Papp, J. C., Lange, K. 2002. Detection and integration of genotyping errors in statistical genetics. *American Journal of Human Genetics* 70: 496-508.

Starrs, D., Ebner, B. C., Fulton, C. J. 2016. All in the ears: unlocking the early life history biology and spatial ecology of fishes. *Biological Reviews* 91: 86-105.

Stockwell, B. L., Larson, W. A., Waples, R K., Abesamis, R. A., Seeb, L. W., and Carpenter, K. E. 2016. The application of genomics to inform conservation of a functionally important reef fish (*Scarus niger*) in the Philippines. *Conservation Genetics* 17: 239-249.

Strathmann, M. F., and Strathmann, R. R. 2007. An extraordinarily long larval duration of 4.5 years from hatching to metamorphosis for teleplanic veligers of *Fusitriton oregonensis. Biological Bulletin* 213: 152-159.

Sunday, J. M., Pecl, G. T., Frusher, S., Hobday, A. J., Hill, N., Holbrook, N. J., Edgar, G. J., Stuart-Smith, R., Barrett, N., Wernberg, T., Watson, R. A., Smale, D. A., Fulton, E. A., Slawinski, D., Feng, M., Radford, B. T., Thompson, P. A., and Bates, A.E. 2015. Species traits and climate velocity explain geographic range shifts in an ocean-warming hotspot. *Ecology Letters* 18: 944-953.

Sunday, J. M., Popovic, I., Palen, W. J., Foreman, G. G., Hart, M. W. 2014. Ocean circulation model predicts high genetic structure observed in a long-lived pelagic developer. *Molecular Ecology* 23: 5036-5047.

Taylor, M. S, Hellberg, M. E. 2003. Genetic evidence for local retention of pelagic larvae in a Caribbean reef fish. *Science* 299: 107-109.

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

Thorrold, S. R., Latkoczy, C., Swart, P. K., Jones, C. M. 2001. Natal homing in a marine fish metapopulation. *Science* 291: 297–299.

Todd, C. D., Lambert, W. J., Thorpe, J. P. 1998. The genetic structure of intertidal populations of two species of nudibranch molluscs with planktotrophic and pelagic lecithotrophic larval stages: are pelagic larvae "for" dispersal? *Journal of Experimental Marine Biology and Ecology* 228: 1-28.

Wade, M. J., McCauley, D. E. 1988. Extinction and recolonization: their effects on the genetic differentiation of local populations. *Evolution* 42: 995-1005.

Wakeley, J. 2008. *Coalescent theory: an introduction*. Greenwood Village: Roberts and Company Publishers.

Wang, J. (2004) Sibship reconstruction from genetic data with typing errors.*Genetics* **166**: 1963-1979.

Wang, J. Y., Frasier, T. R., Yang, S. C., White, B. N. 2008. Detecting recent speciation events: the case of the finless porpoise (genus *Neophocaena*). *Heredity* 101: 145-155.

Waples, R. S. 1987. A multispecies approach to the analysis of gene flow in marine shore fishes. *Evolution* 41: 385-400.

Waples, R. S. 1998. Separating the wheat from the chaff: patterns of genetic differentiation in high gene flow species. *Journal of Heredity* 89: 438-450.

Waples, R. S. and Gaggiotti, O. 2006. What is a population? An empirical evaluation of some genetic methods for identifying the number of gene pools and their degree of connectivity. *Molecular Ecology* 15: 1419-1439.

Wares, J. P. 2002. Community genetics in the Northwestern Atlantic intertidal. *Molecular Ecology* 11: 1131-1144.

Watts, R. J., Johnson, M. S., Black, R. 1990. Effects of recruitment on genetic patchiness in the urchin *Echinometra mathaei* in Western Australia. *Marine Biology* 105: 141-151.

Weersing, K., Toonen, R. J. 2009. Population genetics, larval dispersal, and connectivity in marine systems. *Marine Ecology Progress Series* 393: 1-12.

Whitlock, M. C. and McCauley, D. E. 1999. Short review: Indirect measures of gene flow - F_{ST} does not equal 1/(4Nm+1). *Heredity* 82: 117-25.

Williams, P. D., and Hastings, A. 2013. Stochastic dispersal and population persistence in marine organisms. *American Naturalist* 182: 271-282.

Wilson, G. A., and Rannala, B. 2003. Bayesian inference of recent migration rates using multilocus genotypes. *Genetics* 163: 1177-1191.

Wright, S. 1943. Isolation by distance. *Genetics* 28: 114-138.

Wright, S. 1951. The genetical structure of natural populations. *Annals of Eugenics* 15: 323–354.

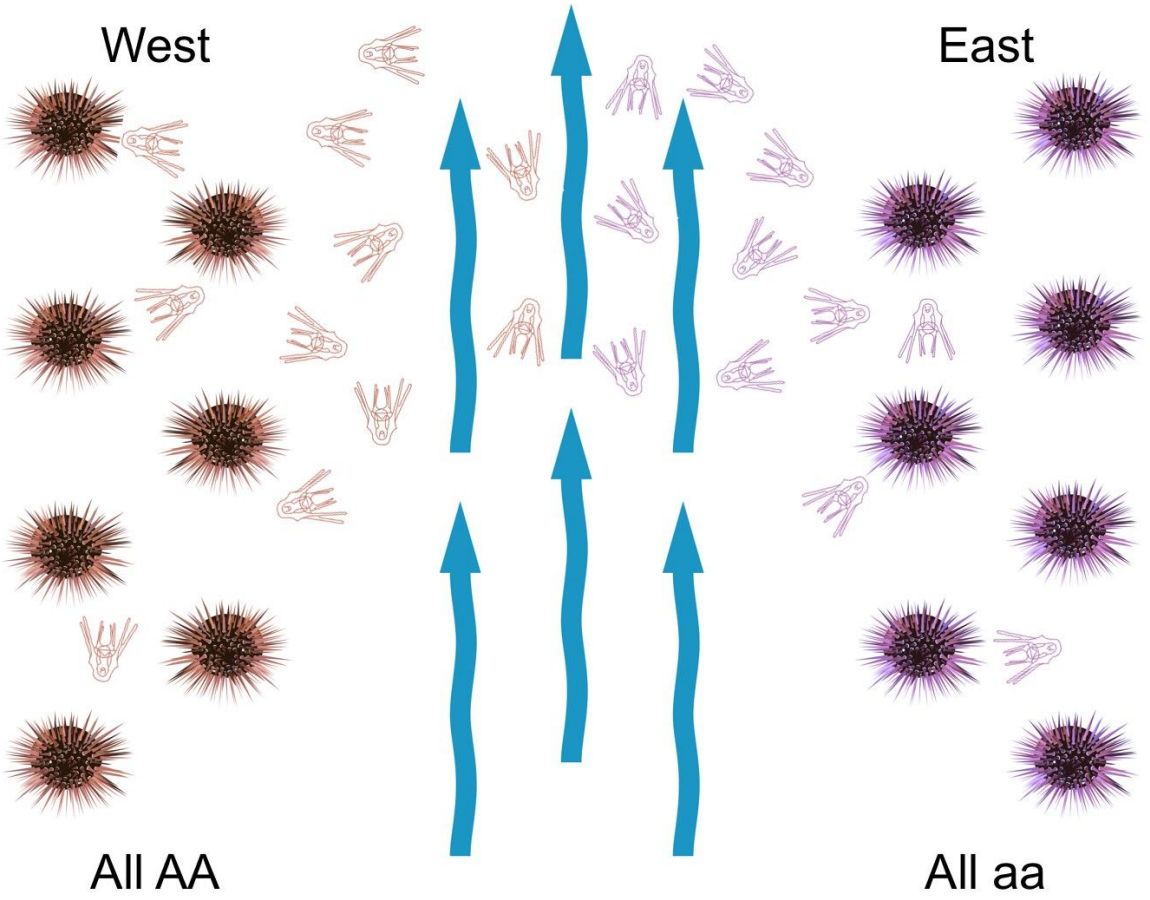

Fig. 1. A hypothetical example in which a large difference in allele frequency between two populations results in a deficiency of heterozygotes if all individuals are assumed to be members of a single population with p = 0.5 and q = 0.5. The eastern and western populations are completely isolated by a strong northward flowing current that prevents larvae from being exchanged between east and west. Expected heterozygosity for the total population is $H_T = 2pq = 0.5$ but the average expected heterozygosity for each of the two individual populations is $H_S = 0$. Therefore, $F_{ST} = (H_T - H_S)/H_T = 1.0$. Adapted from Hartl and Clark (1997).

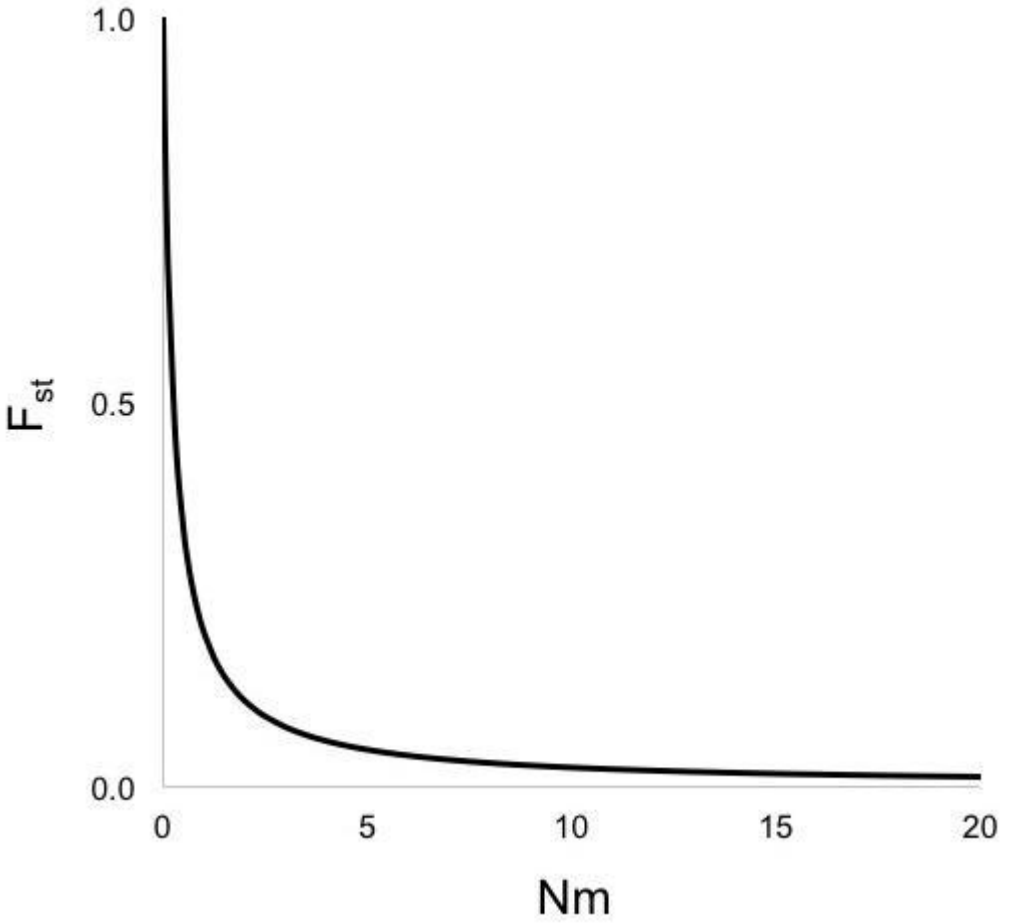

Fig. 2. The relationship between F_{ST} and Nm based on Wright's (1978) equation $F_{ST} = 1/(4Nm + 1)$ that assumes Wright's Island Model.

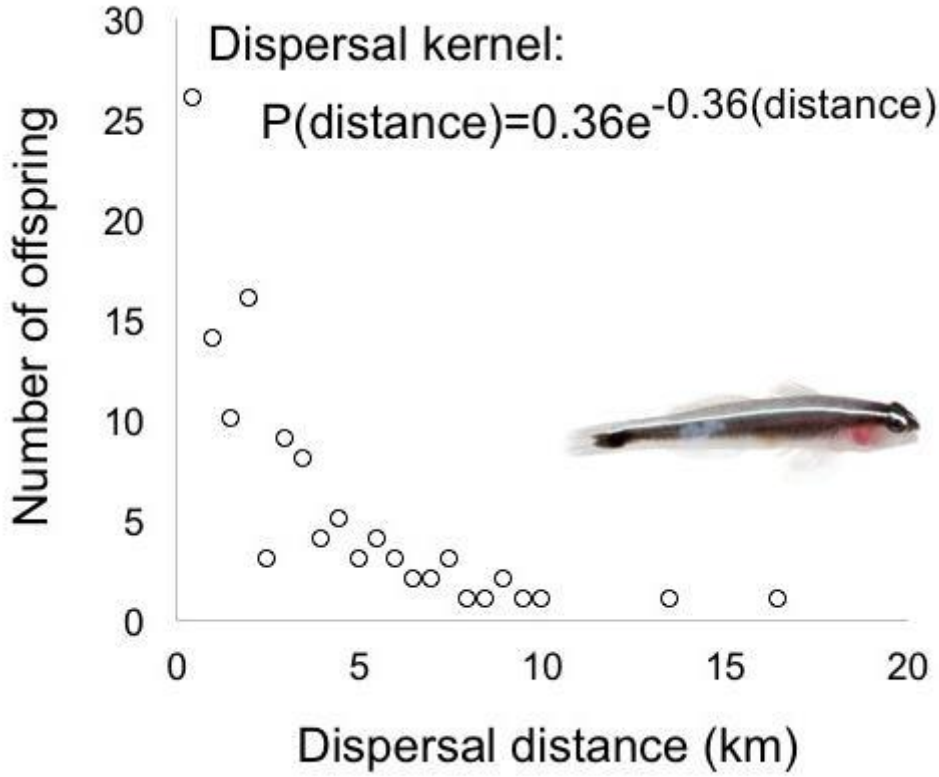

Fig. 3. Frequency distribution of dispersal distances by larvae away from parents for 120 parent-offspring pairs of the neon goby *Elacatinus lori* from the western Caribbean. The dispersal kernel was estimated by fitting a negative exponential function to the frequency distribution, with a best estimate of the decay parameter λ=0.36; the inverse of the decay parameter is the average dispersal distance (~2.8 km). Data and analysis from D'Aloia et al. (2015); larval goby image from Smithsonian Belize Larval-Fish Group 2002, image ID C2-19.

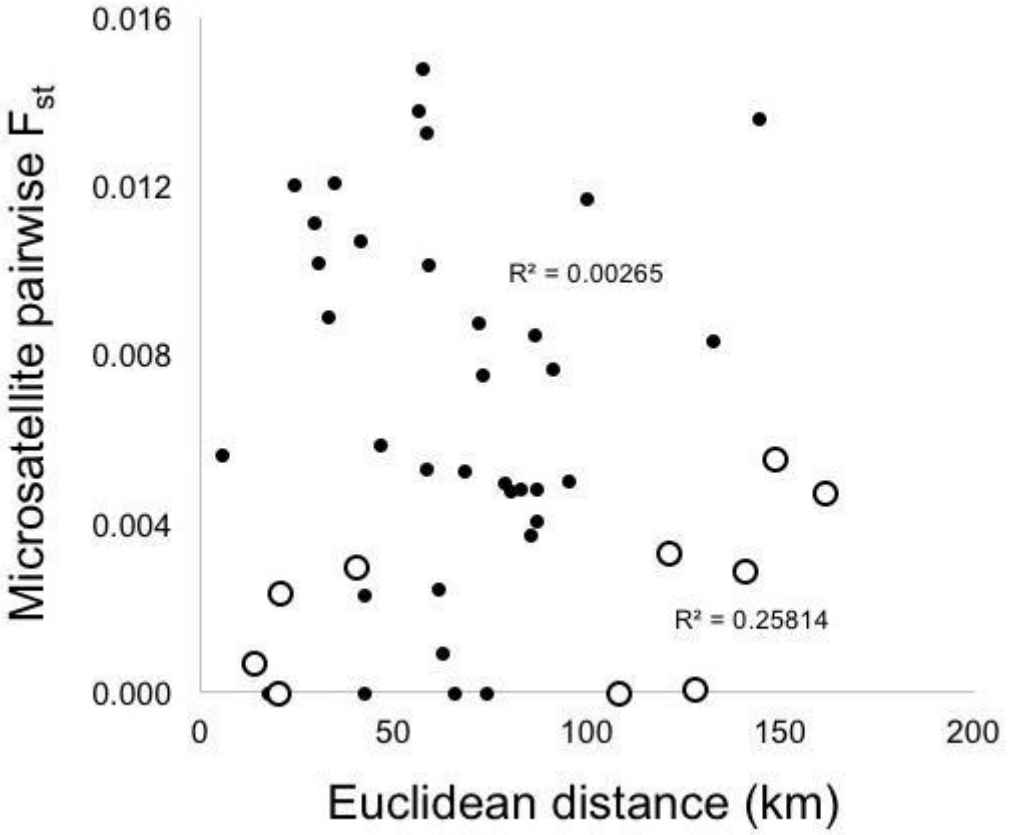

Fig. 4. Patterns of isolation-by-distance (IBD) in the neon goby *Elacatinus lori* from the western Caribbean (D'Aloia et al. 2014) based on 13 microsatellite loci sampled for 20-30 adults from five populations along a continuous barrier reef and five populations from isolated atolls. Relatively strong IBD is detectable among barrier reef populations (open symbols; high coefficient of determination R^2 ~0.26) that are connected by stepping-stone gene flow. D'Aloia et al. (2014) found no IBD among populations from atolls, or between atoll and barrier reef populations (closed symbols; R^2~0.0026).

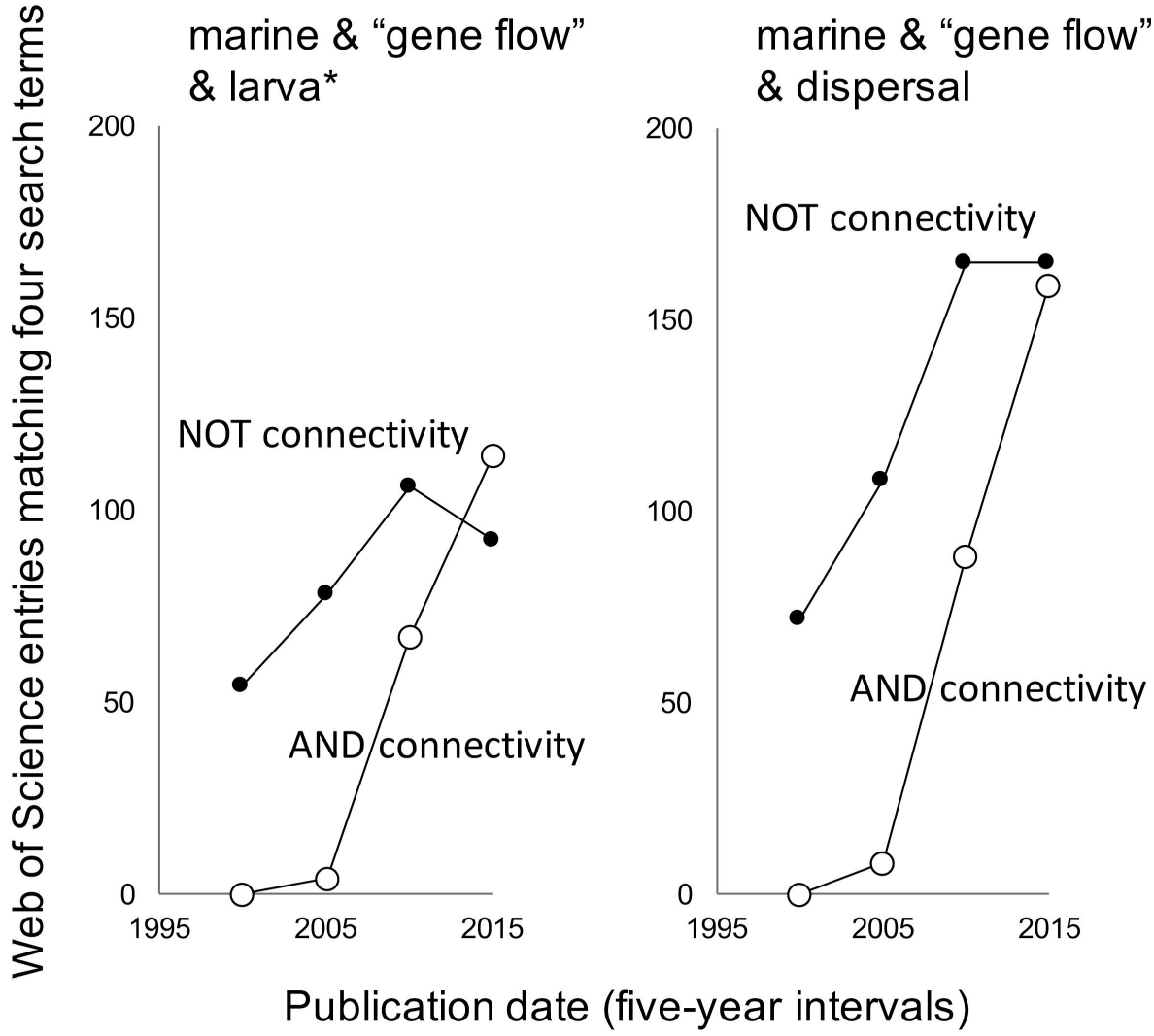

Fig. 5. Trends in the use of the keyword ‘connectivity’ in studies of marine larval dispersal and gene flow; data are counts of citations that use different keyword combinations in searchable fields of records in the Web of Science database for the years 1996 through 2015 in five-year increments following the review by Palumbi (1995). Results are shown for two alternative keyword searches using the terms ‘larva*’ or ‘dispersal’; in each alternative search the results for items without ‘connectivity’ (closed symbols) are contrasted with results for items including ‘connectivity’ (open symbols). In both cases the use of ‘connectivity’ has dramatically increased since 2005.