

The complex analytical landscape of gene flow inference

Peter B. Marko^{1*} and Michael W. Hart^{2*}

¹Department of Biological Sciences, Clemson University, Clemson, SC 29634, USA

²Department of Biological Sciences, Simon Fraser University, Burnaby, BC V5A 1S6, Canada

Gene flow estimation is essential for characterizing local adaptation, speciation potential and connectivity among threatened populations. New model-based population genetic methods can resolve complex demographic histories, but many studies in fields such as landscape genetics continue to rely on simple rules of thumb focused on gene flow to explain patterns of spatial differentiation. Here, we show how methods that use gene genealogies can reveal cryptic demographic histories and provide better estimates of gene flow with other parameters that contribute to genetic variation across landscapes and seascapes. We advocate for the expanded use and development of methods that consider spatial differentiation as the product of multiple forces interacting over time, and caution against a routine reliance on *post-hoc* gene flow interpretations.

Gene flow rules?

Applications of population genetics have become common at all spatial and temporal scales of analysis in evolutionary ecology, from inferences of mating systems to predictions of the biogeographic consequences of climate change. This wide application of genetic data has come about largely as a result of the democratization of easy-to-use molecular methods and software packages that have allowed many organismal biologists, ecologists and biogeographers to gather, organize and analyze large amounts of genetic data. Because many practicing population geneticists (such as ourselves) have entered the field out of a strong empirical interest (rather than from a theoretical upbringing), there has been a persistent desire for easy-to-use heuristics for interpreting population genetic data. Possibly the most widely used of these rules of thumb is the idea that patterns of spatial genetic differentiation primarily reflect variation in gene flow: genetic homogeneity among populations is generally caused by high rates of gene flow and, conversely, genetic differentiation is best explained by low rates of gene flow. An important corollary of this idea is that gene flow (and associated quantities such as dispersal and immigration rates) can be reliably inferred from measures of strong or weak population genetic differentiation.

A direct causal relationship between gene flow and genetic differentiation is an intuitively appealing idea, particularly for ecologists and biogeographers who routinely think about how plants and animals move across complex

Glossary

Ancestral polymorphism: an allele or allelic lineage shared in common between two populations that was present in both populations before the population divergence occurred, rather than owing to gene flow from one population to the other after the split between the two.

Bayesian inference: a method of statistical inference in which a specified prior probability for a hypothesis is used to determine the likelihood or posterior probability of the hypothesis. In coalescent population genetics, the posterior probability distribution for a population parameter reflects the compatibility of the observed evidence (a gene genealogy) with a range of parameter values specified by the prior probability distribution. If the prior distribution is uniform (same prior probability for all values of a parameter from lowest to highest), then the mode or peak in the posterior probability distribution is also a ML estimate of the parameter value.

Coalescent: formally defined as the stochastic process of mutation and extinction of alleles that generates a gene genealogy; more often thought of as the structure of that genealogy from a retrospective perspective: alleles coalescing backward in time toward a common ancestor. The structure includes both the topology of the gene tree and the temporal distribution of coalescent events; that is, the ages of the nodes or branch points in the gene tree.

Dispersal: physical movement of individuals between geographical locations (populations); can be higher than the rate of gene flow estimated from genetic data if many migrants leave few descendants after they disperse; can be lower than the rate of gene flow inferred from F_{ST} if most alleles are shared between populations as ancestral polymorphisms and not as the descendants of migrants.

Divergence time (t): in the isolation-with-migration model (Figure 1, main text), the time since two populations shared a panmictic common ancestor; can have different interpretations that depend on the biogeographic context; for example, time since one geographic location was colonized by dispersing individuals from the other, or the age of a significant vicariant event (e.g. a geological separation).

Effective population size (N): the number of breeding individuals in an idealized population size that would have the same genetic diversity under random genetic drift as a real population under consideration. Often represented by the symbol N_e .

F_{ST} : Wright's fixation index, a dimensionless measure of differentiation between two or more samples of gene copies that ranges from zero (no differentiation) to 1 (complete differentiation); can be calculated from several different quantities (based on allele frequencies or on genetic distances between alleles), the simplest of which is the standardized variance (var) in allele frequencies where $F_{ST} = \text{var}(p)/[P*(1-P)]$, p is the frequency of an allele in each population, and P is the average frequency of that allele across all populations.

Gene flow (Nm): the effective rate of migration of gene copies into a population; the product of N and m . Similar to F_{ST} , this is formally a dimensionless index, but is more often thought of as a number of immigrant individuals per generation or year, and often calculated as the population migration rate $2Nm$.

Island model of migration: a group of demes, populations, or subpopulations of equivalent size that all exchange migrants at the same rate in both directions. Sewell Wright showed that if the introduction of new alleles to each population from migration was in a dynamic equilibrium with the loss of alleles from each population caused by genetic drift, the degree of differentiation among all populations was a simple function of gene flow.

Lineage sorting: a process following the separation of two populations from a common ancestral population in which the random but differential loss of alleles from each population eventually results in all of the alleles in one population to be more closely related to each other than to any alleles in the other population, a condition known as reciprocal monophyly.

Migration rate (m): the proportion of gene copies in a population that are new immigrants each generation or year. In coalescent population genetics, m is usually scaled by the neutral mutation rate, such that conversion into demographically meaningful units requires an estimate of that mutation rate.

Corresponding author: Marko, P.B. (pmarko@clemson.edu)

* Both authors contributed equally to this work.

Box 1. F-statistics and coalescent gene flow inference

Wright's F_{ST} can be used to calculate gene flow as $Nm = (1/4)F_{ST}^{-1}$ under equilibrium assumptions collectively known as the island model [61]. It is now well known ([62–67], but see [57,58]) that $Nm \neq (1/4)F_{ST}^{-1}$ ([66] has been cited nearly 600 times). Populations might not fit the island model or its variants, particularly with respect to the assumption of evolutionary equilibrium [39,68–70]. Non-equilibrium explanations of spatial differentiation that do not invoke gene flow are similarly well known ([5] and [6] have, together, been cited > 2500 times), and population geneticists are keenly aware of the impact of metapopulation dynamics [71,72] on spatial patterns of differentiation [73–78].

Why, then, have empiricists continued to rely on island model assumptions in their interpretations of genetic differentiation? This practice is probably a lasting legacy of the historically strong influence of early population genetic data (allozymes) on the conceptual orientation of the field. Because the genealogical relationships of allozyme electromorphs could not be inferred from their relative mobilities on a gel, methods were developed (during the 1960s and 1970s) from population genetics models developed during the modern synthesis (during the 1940s) that focused on the prediction of evolutionary change in population frequencies of phylogenetically unordered alleles [79–81].

Methods that do not consider the genealogies (i.e. history) of alleles cannot, however, distinguish allele sharing owing to recurrent gene flow from allele sharing caused by ancestral polymorphism (Figure 1). Such genealogical methods are now in wide use, but not all are yet fully appreciated as a non-equilibrium approach for estimating gene flow [38,82–84]. These methods use the retrospective concept of the coalescent (i.e. the temporal distribution of coalescent events within gene genealogies) to model demographic parameters (gene flow, divergence time or changes in effective population size) that operated in the past to shape observed patterns of genetic variation within and between present-day populations (Figure 1). Other methods use multilocus assignment tests to estimate migration between populations over recent generations without the F_{ST} equilibrium assumption [85–87]. Because our goal here is to evaluate the direct interpretation of population differentiation in terms of gene flow, we have focused on non-equilibrium methods that yield time-averaged estimates of gene flow that are directly comparable to those estimated from F_{ST} and analogous summary statistics typically used as proxies for gene flow.

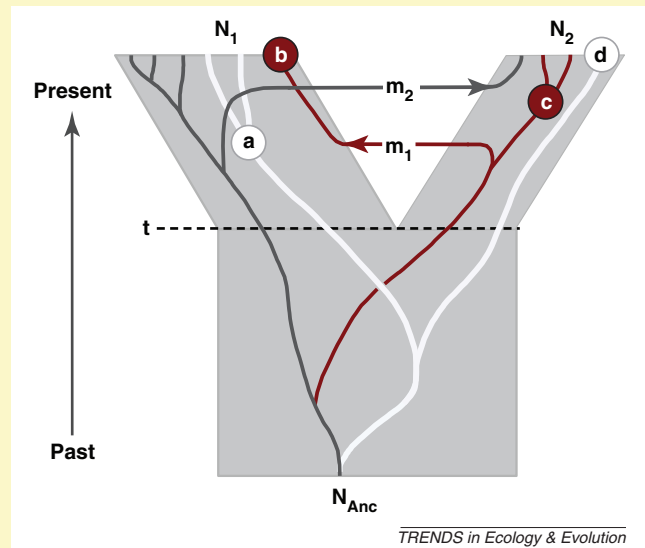


Figure 1. The basic isolation-with-migration model for two populations (modified from [13]). The population demographic quantities are effective population sizes (N_1 , N_2 and N_{Anc}), divergence time (t), and migration (m_1 and m_2 , the per-generation proportion of each population that are migrants). The corresponding model parameters ($\Theta_1 = 4N_1\mu$, $\Theta_2 = 4N_2\mu$, $\Theta_{Anc} = 4N_{Anc}\mu$, t/μ , m_1/μ , m_2/μ) are scaled by the neutral mutation rate (μ); rescaled population parameters in units of individuals (N and m) and years (t) require conversion with an assumed mutation rate. Gene flow (Nm) can be calculated without specifying a mutation rate by multiplying the estimate of Θ for one population by the corresponding estimate of migration and dividing by 4 (and thus is equivalent to Nm estimated from F_{ST} , which makes a corresponding assumption about the negligible effect of mutation on allele frequencies). Estimates of Θ need not be rescaled in terms of individuals because Θ provides information about relative population size. Note that the migration parameters are drawn with respect to time moving forward rather than in terms of the coalescent (time moving backward). By considering population differentiation as a time-dependent process, the model distinguishes between allelic lineages shared owing to gene flow (e.g. b and c, red) from those shared as ancestral polymorphisms (e.g. a and d, white).

landscapes or seascapes. Unfortunately, this idea is too good to be true: the concept is based on the predictive relationship between Wright's index of population genetic subdivision (F_{ST}) and gene flow, an idea widely acknowledged as reliant on too many unrealistic assumptions to estimate gene flow between natural populations reliably (Box 1). Nevertheless, the habit of directly inferring gene flow from F_{ST} or other measures of differentiation has been hard for empiricists (including ourselves) to break: although estimates of gene flow from F_{ST} are nearly absent in the recent literature, most empirical studies of spatial differentiation have continued to interpret F_{ST} variation almost exclusively in terms of gene flow variation. For example, a reading of all 2010 papers in *Molecular Ecology* containing the key words 'gene flow' shows that most make straightforward interpretations of genetic differentiation in terms of gene flow alone. Many of these examples come from landscape genetics studies [1–3] that equate F_{ST} with the inverse of gene flow (but see [4]) 'to integrate the effect of landscape connectivity into gene flow analysis' [1].

This bad habit is surprising because most evolutionary ecologists would probably agree that population differentiation is not caused by gene flow alone, but by a suite of evolutionary forces, including mutation, genetic drift, gene

flow and their interaction over time. The strong tendency to ignore other forces in favor of gene flow has its roots in the conceptual development of the field (Box 1). Throughout the 1980s, population geneticists simply lacked the tools to measure the distinct contributions of each of these interacting evolutionary forces [5–7], but by assuming that those forces are at evolutionary equilibrium (and by assuming populations are of equal size and all have symmetrical exchange of migrants) it was possible to develop rules of thumb in which F_{ST} is a simple function of gene flow (Box 1). Unfortunately, *post-hoc* gene flow interpretations rely on this same important (but fairly cryptic) assumption that patterns of allele sharing among populations reflect a balance between the introduction of alleles by gene flow and the loss of alleles via genetic drift. Many species might be far from equilibrium (Box 2), and their patterns of genetic structure might primarily reflect a combination of colonization history and genetic drift rather than the influence of recurrent gene flow. As a consequence, gene flow might often only be reliably inferred by joint estimation with those additional factors (especially mutation and population size) within a framework that lacks the assumption that all of the forces affecting allele frequencies are in a dynamic evolutionary equilibrium.

Box 2. Slatkin's Paradox

Genetic measures of population differentiation, such as F_{ST} , often conflict with direct observations (e.g. from tagging studies) of individual dispersal between populations or with expectations of high or low dispersal based on the presence or absence of highly dispersive stages (i.e. pollen, gametes, seeds, or larvae) in the life histories of species. In insects, this mismatch between genetic inference and direct observation has long been known as 'Slatkin's Paradox' [5–7,54,88–92]. Analogous inconsistencies between patterns of genetic differentiation and expectations about dispersal are also prominent in studies of all kinds of organisms. Surprisingly, few recent studies have considered Slatkin's hypothesis to explain the paradox: in many populations, gene flow and genetic drift are probably not in an evolutionary equilibrium [6] because processes such as recent range expansions (e.g. in response to warming after the last Pleistocene glaciation) or frequent population extinction and recolonization can homogenize population genetic variation [73]. Such metapopulation dynamics can produce paradoxical patterns, such as low F_{ST} between recently separated populations that share many ancestral alleles in common (ancestral polymorphisms; Box 1) but exchange few migrants. Slatkin noted that better data and methods were needed to distinguish the effects of what he called 'contemporary' recurrent gene flow from the effects of 'historical' connectivity (ancestral polymorphism retained since vicariant events or colonization). Nevertheless, few studies [93] have used non-equilibrium population genetic approaches to distinguish gene flow from other causes of paradoxical patterns of differentiation in a metapopulation context [94].

Instead, most empirical studies make simple interpretations of spatial differentiation almost exclusively in terms of gene flow. This approach ignores the underlying gene genealogy (the coalescent), demographic history and ancestral polymorphisms, and risks significant errors in important conclusions about the ecological and evolutionary causes of present-day genetic variation. Our goal here is to show how isolation-with-migration and other coalescent models can be used alongside conventional F_{ST} -type measures of differentiation to mitigate this risk and resolve puzzles such as Slatkin's Paradox.

Here, we argue that empirical population geneticists can do better to obtain reliable and realistic measures of gene flow that do not depend on simplistic rules of thumb and that methods based on the genealogy of alleles within and between populations (the coalescent) can allow us to do so. These methods are better than F_{ST} (not just different) because they estimate gene flow as one part of a more realistic population model (Box 1). Our success at persuading fellow empiricists that simple gene flow interpretations are inherently flawed depends on the development of convincing examples of the importance of more inclusive approaches and the improved understanding that they offer, including a better understanding of limits to interpretation of spatial variation where the underlying demographic history cannot be estimated from the available data (Box 3). These examples are key because, in our experience, readers and audiences respond weakly to theoretical refutations, such as $F_{ST} \neq 1/(4Nm + 1)$ (Box 1), but respond strongly to case studies in which verbal interpretations of F_{ST} variation are shown to be wrong. We recognize that these examples could be taken only as criticism, which is an implication that we are anxious to avoid, noting that some of our own previously published conclusions [8,9] fall into this broad class of probable errors that were only revealed in subsequent analyses [10,11]. We therefore view the improvements from coalescent analyses as an important source of insight rather than of regret, and hope that readers and colleagues will see our efforts in a similar light.

Non-equilibrium gene flow estimation

The examples developed below characterize gene flow within the isolation-with-migration population divergence model [12–14], which was developed to estimate jointly gene flow, genetic diversity and population divergence time, with the specific goal of distinguishing alleles shared between populations as ancestral polymorphisms (equivalent to Slatkin's 'historical gene flow'; Box 2) from alleles shared as a result of recurrent or contemporary gene flow between separate populations (Box 1). Although most closely associated with the software program IMA [14], many other methods [15–19] now use this conceptual framework to estimate gene flow without the assumption that gene flow and genetic drift are in equilibrium (i.e. lacking the assumption that allele sharing is caused exclusively by recurrent gene flow). These population models, and their use of coalescent gene trees to infer population demographic parameters, have been recently reviewed and compared elsewhere [20–23]. Briefly, the methods use maximum likelihood (ML) to fit the isolation-with-migration model to a ML gene tree for each locus sampled from a pair of populations. Demographic parameters are estimated as Bayesian posterior probabilities across many highly probable gene trees. Other conceptually similar (but computationally different) methods use simulations of DNA sequence evolution under more complex population models (rather than fitting the full likelihood function for simpler models), and are called approximate Bayesian computations (or ABC) [24–27]. The two approaches have different advantages and limitations (Box 4). ABC approaches with complex demographic models are challenging to use, and we look forward to more user-friendly software packages. For simplicity, we use IMA in the examples below, but a combination of IMA and ABC (and other) methods would be desirable.

Divergence time (not gene flow) explains population structure: snails and sea stars

Marine invertebrates provide useful model systems for understanding gene flow, given the many opportunities for comparisons between co-distributed species with different types of larval development (planktonic vs non-planktonic) and vastly different dispersal capabilities [10,28–30]. One study of two co-distributed species, the bat star *Patiria miniata* and the frilled dogwhelk snail *Nucella lamellosa* [11], shows how coalescent methods can be used to estimate recurrent gene flow between populations under conditions that are probably far from equilibrium. Both the bat star and the frilled dogwhelk are common in the cool temperate northeast Pacific, a region containing a mix of species with histories of either recent recolonization from lower latitudes or persistence at both low and high latitudes through multiple Pleistocene glacial cycles ([31,32] and references therein). Across Queen Charlotte Sound (QCS, between northern Vancouver Island and the Alexander Archipelago, see Figure 1 of [11]), the bat star and the frilled dogwhelk show contrasting patterns of mtDNA differentiation: the bat star shows a very large and significant mtDNA population genetic break (pairwise population Φ_{ST} values approximately 0.5) but the frilled dogwhelk shows no significant differentiation (pairwise Φ_{ST} values < 0.05) across the same area [11]. The patterns of

Box 3. When coalescent methods fall 'flat' on their posteriors: not enough data?

The near-shore marine community of Oregon and California has been intensively studied by phylogeographers as a region of significant conservation concern (<http://mlpa.dfg.ca.gov>). A recent, large and comprehensive study [95] of mtDNA from 50 broadly sympatric species highlighted a shared phylogeographic break (non-zero Φ_{ST}) in six species distributed north and south of Cape Mendocino in northern California. We reanalyzed some of these data using IMA: if limited migration (low m) across Cape Mendocino causes this phylogeographic break, then estimates of m should be higher between population pairs to the south or the north of the break compared with m across Cape Mendocino (and other parameters, such as N and t , should not vary between population pairs; Box 1). Instead, we found that migration rates were indistinguishable between differentiated population pairs (red curves in Figure I) and undifferentiated population pairs (black curves).

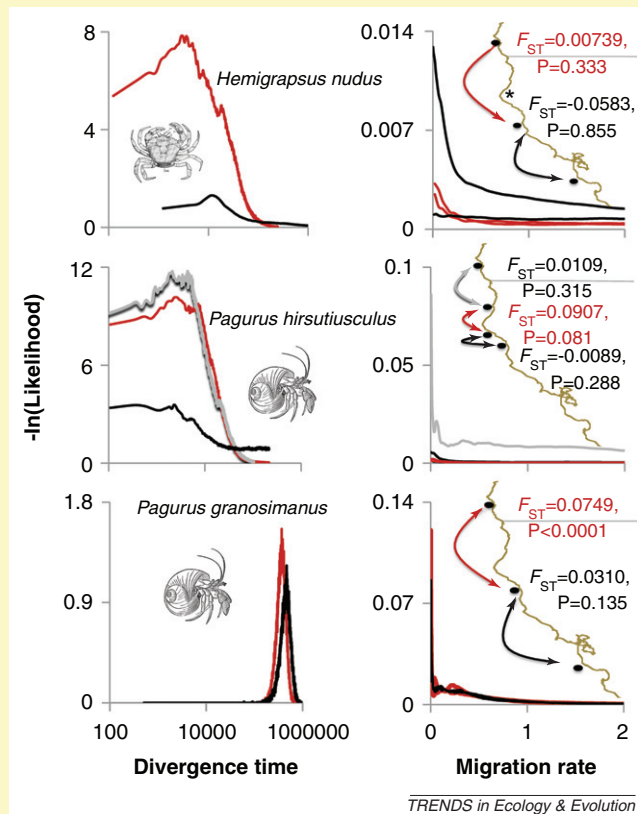


Figure I. Posterior probability distributions for population divergence time [(a) t , years] and migration rate [(b) $m \times 10^{-3}$ per generation] from IMA analyses of crab mtDNA cytochrome *c* oxidase subunit I (COI) sequence variation around Cape Mendocino, California (data from [95]). Sequence alignments were obtained as GenBank popsets with some nucleotide sites and some short sequences excluded to minimize alignment gaps (caused by truncated sequences in the popset). Pairwise Φ_{ST} values (using Kimura two-parameter genetic distances) were calculated between populations for this modified alignment in Arlequin 3.1. We used IMA methods as described in recent papers [10,11] to obtain joint posteriors for parameters (N , m and t) using a published estimate of COI mutation rates of 1×10^{-8} per site per year [96] and generation times of 2 years. As with the original study, we found greater differentiation between pairs of populations on either side of Cape Mendocino (indicated by red posterior distributions and by a star on the inset map in the upper right panel) than between populations immediately to the north (grey) or south (black) of that break. All m posteriors had highest probabilities for the lowest bin in the distribution (i.e. $m = 0$). Divergence times were log-transformed to emphasize the difference between recent population divergences in two species versus Pleistocene divergences in a third (*P. granosimanus*). In none of the three analyses were lower estimates of m or older estimates of t associated with the phylogeographic break at Cape Mendocino. Line drawings from <http://etc.usf.edu/clipart>; vector maps from <http://www.planiglobe.com>.

Other parameter estimates did not help to explain the cause of the phylogeographic break. Assuming similar rates of mutation in each species, estimates of t were post-glacial for *Hemigrapsus nudus* and *Pagurus hirsutiusculus*, but much older (middle Pleistocene) for *Pagurus granosimanus* (Figure I), and not different among population pairs within species. Posterior distributions of N were essentially the same as the prior distributions (uniform or 'flat') in almost all analyses (i.e. the data contained too little coalescent information for joint estimation of N along with other parameters). One can reasonably conclude that, at least in these three species, population differentiation around Cape Mendocino might indeed be caused by restricted gene flow, but the data lack specific evidence for that explanation. Greater genetic drift in some populations with small N (as in our lizard example) remains a plausible hypothesis for differentiation around Cape Mendocino, but more data are needed for a robust test. Alternatively, the two-population demographic model in IMA might simply be a poor match with the actual history of the populations in all three species (Box 4).

Few studies consider how much more data (i.e. loci) might be needed before the data could be confidently used to design and implement a network of protected areas connected by recurrent migration or gene flow [31,51,52]. Analyses of subsets of loci from *Patiria miniata* across the northern phylogeographic break at Queen Charlotte Sound [11] suggest that the accuracy and precision of gene flow estimates can be substantially improved by adding only one or a few loci (Figure II) to an mtDNA data set yielding poor estimates of gene flow.

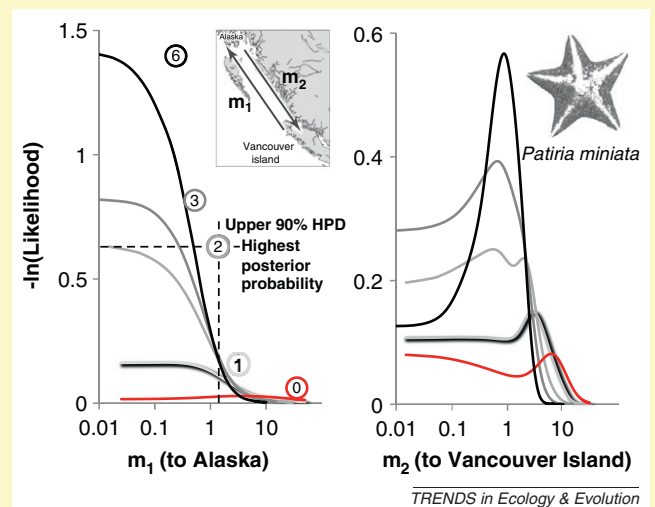


Figure II. Posterior probability distributions for migration rate ($m \times 10^3$ per generation) between northern (Alaska) and southern (Vancouver Island) bat star populations from IMA analyses of mtDNA plus zero (red) to six (grey to black) additional anonymous nuclear loci (ANuLs) (see [11] and Figure I). Posteriors for mtDNA plus 1–3 ANuLs (grey) are averages of five independent IMA runs using different random selections of the six sampled ANuLs. Numbered circles in (a) (northward immigration into Alaska) show the upper confidence limit (the 90% highest posterior density, HPD) for the migration rate estimate plotted against the highest posterior probability (for the ML estimate of the migration rate) for each combination of loci. The two dashed lines show the two values [upper limit of m_1 of approximately 0.0015, with a likelihood score of $-\ln(L)$ of approximately 0.62] for the combination of mtDNA plus two ANuLs. The trend from adding zero to six ANuLs shows lower values for the confidence limit (increased precision) and higher posterior probability of the parameter estimate (greater accuracy) with addition of one or two ANuLs, followed by smaller improvements from additional loci up to six. For southward immigration into Vancouver Island, the trend was similar but less easily shown in (b) because the posterior distributions for two to six ANuLs were broadly overlapping. In both cases, analyses of mtDNA alone (red) produced broad, flat posterior distributions that were more clearly resolved (as zero migration into Alaska, and non-zero migration into Vancouver Island) by the addition of one or a few ANuLs.

Box 4. Limits to isolation-with-migration analyses

Similar to democracies (http://hansard.millbanksystems.com/commons/1947/nov/11/parliament-bill#S5CV0444P0_19471111_HOC_273), isolation-with-migration population models are imperfect but better than all the alternatives, including Wright's island model, which assumes that population history can be adequately summarized by a single parameter (F_{ST}). Coalescent methods that implement an isolation-with-migration model assume that it is a reasonable match to the history of the populations sampled. Therefore, an important specific criticism of two-population divergence models is the assumption that no additional populations ('ghost' populations, see [97,98]) significantly impact the coalescent history of gene copies sampled from the two focal (sampled) populations in each analysis.

In simulation studies [48], ghost gene flow causes inflated estimates of N for the recipient population, older t between the two focal populations, and a bias toward asymmetric gene flow between the focal populations [48]. More simulation studies, especially those investigating bias caused by population structure within each focal population, are clearly needed for assessing the impacts of ghost populations in intraspecific isolation-with-migration analyses [48]. One conclusion from the available studies is that the most robust two-population results will probably come from analyses of relatively large population genetic differences where 'ghost' gene flow effects

will be relatively small. This suggestion should not, however, discourage empiricists from analyzing sample pairs that show no significant F_{ST} differentiation.

A recent elaboration of the method in IMA2 can model the same parameters for multiple sampled and ancestral populations using a specified population phylogeny. Because multi-population gene flow models involve the estimation of many demographic parameters, they can also only be attempted with large numbers of unlinked loci. A combination of ABC methods based on simulations [24,25] (that can compare coalescent models to several probable simulated demographic histories of multiple populations) and methods such as IMA2 that are based on a fully fitted likelihood function (and that can use likelihood ratio tests of hypotheses about parameter estimates), might turn out to be a useful and effective approach to multi-population studies.

Although the limitations of the two-population model are widely appreciated for coalescent methods, it does not seem to be as widely appreciated that *post-hoc* gene flow interpretations of pairwise F_{ST} estimates make the same assumption and are subject to the same biases noted above. Those biases can be compounded by the fact that pairwise F_{ST} inferences typically attribute differentiation exclusively to gene flow (Box 1).

differentiation across QCS are paradoxical because the species with greater differentiation (bat star) has long-lived and highly dispersive planktonic larvae, whereas the species with no significant differentiation (dogwhelk) lacks a dispersive planktonic larval stage altogether.

Of course, many readers of this article will be quick to point out that no experienced phylogeographer would ever conclude that gene flow is greater in the poorly dispersing dogwhelk than in the widely dispersing bat star; most would instead propose a *post-hoc* non-equilibrium explanation, such as recent colonization, for the unexpectedly small Φ_{ST} value across QCS in the dogwhelk. However, to test this hypothesis in a robust way and estimate recurrent gene flow, McGovern *et al.* [11] analyzed a combination of mtDNA and six anonymous nuclear loci from both species with the isolation-with-migration model in IMA. The analyses showed that the estimate of divergence time across QCS was relatively ancient for the bat star (approximately 282 000 years) but far more recent for the dogwhelk (approximately 15 000 years)*. Gene flow estimates for both species also met expectations based on differences in their larval dispersal potential, with significant gene flow across QCS in the bat star but none in the dogwhelk. Taken together, the results from this study show that the absence of a large genetic break in the poorly dispersing dogwhelk is probably explained primarily by a recent population separation (a colonization event), and that the presence of the large genetic break in the bat star reflects the lingering persistent effects of an older vicariant event (strong isolation of northern and southern populations in the middle Pleistocene) rather than unusually restricted gene flow. In fact, between other adjacent populations of the bat star that showed much lower values of Φ_{ST} , gene flow estimates were similar to those across QCS,

emphasizing that the large break in this region is not due to an unusual restriction on recurrent gene flow [11]. Although unconventional patterns of differentiation observed in the frilled dogwhelk hinted at recent colonization (and not gene flow) as the cause of the spatial pattern in that species, this example shows the possibilities for errors in interpreting F_{ST} variation in terms of gene flow alone in other species where indications of more complex demographic histories are less evident, and where such errors might have profound consequences. For example, in the context of marine protected areas (MPAs) and conservation studies, lack of significant spatial differentiation in the sea (such as in the dogwhelk) is often interpreted as (but might not actually be) evidence that populations are 'well-mixed', and that 'a considerable amount of gene flow takes place even among MPAs that are at a great distance from each other' ([33], see also [34,35]).

Population size (not gene flow) explains population structure: endangered lizards

Landscape genetics studies typically rely on equilibrium assumptions and rules of thumb to interpret F_{ST} variation [1]. Notable exceptions include studies such as those of Chan *et al.* [36], who used an isolation-with-migration framework to understand the historical demography that underlies seemingly conventional patterns of strong or weak spatial differentiation (pairwise $F_{ST} = 0.046\text{--}0.347$ for seven microsatellites) among populations of the endangered lizard *Sceloporus arenicolus*, which has a highly restricted geographic range in specific desert habitats in the southwestern USA. The authors' IM (a precursor to IMA) analysis of mtDNA showed that the migration rate has been very low between the northern and central regions of the *S. arenicolus* range. To obtain multilocus estimates of both migration and effective population size, we re-analyzed the combined mtDNA/microsatellite data for all six pairwise combinations among four sites (see Figure 1 of [36]) from the northern region (sites called 'Kenna' and 'Site 20') and central region (called 'Camp' and

* Estimates of divergence time were conservative with respect to the outcome of the analysis: the mutation rate used for the bat star (from geminate species separated by the Isthmus of Panama) is probably an overestimate, whereas the substitution rate for the dogwhelk (from a fossil-calibrated phylogeny) probably underestimates the true mutation rate.

'Site 106') using IMA (see [Supplementary Information](#)). Although estimates of m varied by four orders of magnitude (1×10^{-7} to 2.5×10^{-3}), all are probably demographically trivial as a proportion of effective population size [37–39], a striking result for populations separated by such small distances (as little as 20 km between Camp and Site 106). Because the highest migration rate ($m = 0.00252$ from Camp into Site 106; [Table S1](#)) was into the smallest population (with a genetic effective size of $N = 34$ individuals), the rate of gene flow (Nm) was only 0.087 immigrants per generation. Overall, pairwise F_{ST} was strongly predicted by the smaller Nm for each population pair ($r = -0.75$), but joint estimation of m with N indicates that the demographically significant component of that correlation was N rather than m : effective population sizes varied from 34 to 862 (for Site 20 and Site 106, respectively) to over 100 000 (Kenna); ancestral population sizes were much larger than most extant populations; F_{ST} was highest between the two smallest populations (Site 20 and Site 106), and lowest between the two largest populations (Kenna and Camp).

As Chan *et al.* concluded, coalescent-based analyses suggest that there has been little gene flow between most *S. arenicolus* populations. The IMA results using multiple loci help to clarify this conclusion by showing that the parameter driving strong or weak differentiation was not landscape-based variation in migration rates (m) but rather genetic drift associated with small or large population sizes (N), an effect that Chan *et al.* proposed but were not able to resolve in their IM analyses of mtDNA alone. This insight is important because it suggests that the landscape genetics focus on differentiation and its relation to gene flow (Nm) can be misplaced if the cause of spatial structure is often a combination of trivial differences in m plus vast differences in N ([Table S1](#)). If m truly is demographically trivial, a potentially interesting implication of that result is that maximizing the size or quality of habitat patches might be as or more important a conservation priority than maximizing connectivity between such patches. From an analytical perspective, such a result echoes the view previously articulated [7,37] that m is more likely than Nm to be the quantity of interest for ecologists and conservation biologists. Direct estimates of m might represent crucial improvements in landscape studies because measures of differentiation (such as F_{ST}) give only indirect insight into m in the form of the compound parameter Nm .

Old advice for new analyses: add loci

Good isolation-with-migration analyses require multiple loci [40,41] but many animal studies start from mtDNA sequences because they are relatively easy to collect and are probably 'leading indicators' of population structure [42]. The technical challenges associated with gathering additional unlinked nuclear sequences from non-model organisms (with so-called 'universal' primers, PCR cloning and Sanger sequencing), as well as the appeal of microsatellite hypervariability, resulted in the focus by phylogeographers and landscape geneticists on microsatellites as a nuclear complement to mtDNA [42–44]. Unfortunately, coalescent analyses of microsatellite variation can be complicated by large differences between the simple stepwise mutation model used to infer the microsatellite allele genealogy

versus the complex mutational processes often suggested by observed allelic variation [45–48]. Some empiricists might also have avoided nuclear sequences owing to the perceived problems of low polymorphism [41]. This combination of effects could partly explain the reluctance of some empiricists to use methods that rely on genealogical reconstructions for within-species analyses of gene flow.

For animal studies in particular, mtDNA data represent a good start but not a complete answer to the requirements of coalescent gene flow estimation. The smaller effective population size, higher mutation rate, and ease of sequencing for mtDNA will probably maintain it as a popular choice for initial characterization of spatial patterns of differentiation ([Box 3](#), [42–44]). Given the high variance in the coalescent, however, accurate estimation of demographic parameters from gene trees relies fundamentally on multilocus data [43]. MtDNA is a particularly unusual marker in this respect: although high mutation rates and smaller effective population sizes make it ideal for detecting population differentiation [42], the benefit of rapid lineage sorting and spatial differentiation comes at the expense of information about ancestral polymorphism; with each step towards reciprocal monophyly, sequence markers lose their capability to inform about both ancestral population size and divergence time. Our analyses have estimated six population parameters and, even for this relatively modest problem, it is apparent that single locus (mtDNA) data sets often contain too little information for the joint estimation of all the parameters ([Box 3](#)). The number of loci needed to obtain accurate estimates of parameters will depend on the complexity of the population history and how well that history matches the demographic model in the analytical method ([Box 4](#)), as well as the information content of each individual locus; the few studies that have investigated this issue are encouraging in that they all show that a fairly modest number of nuclear loci can substantially reduce the variance associated with some parameter estimates ([Box 3](#); [49–52]). For other parameters, or for more complex models with many populations and parameters, very large data sets might be necessary to obtain estimates that are both precise and accurate.

Concluding remarks

Despite the growth of phylogenetic biology [53] and its emphasis on historical causation, many empiricists maintain a strong focus on recurrent gene flow as an explanation for population genetic structure without considering the demographic history hidden within the genealogy of sampled alleles. Much of this focus is justified given that gene flow can have a fundamentally important role in shaping patterns of spatial genetic variation [4–7,54–59]. However, the habit of overlooking other factors contributing to population genetic differentiation remains strong in the literature, with recent issues of *Molecular Ecology* (2010) and *Landscape Ecology* (2006) devoted entirely to studies of landscape genetics that generally equate differentiation with the inverse of gene flow.

A significant innovation in landscape genetics is the potential elimination of the population from population genetics by analyzing the distribution of individual genotypes across a landscape characterized by a connectivity

matrix used to interpret the spatial distribution of alleles [2]. This approach helps to avoid what are sometimes problematic definitions of discrete populations (consisting of individuals and their interactions over time) when habitats and individuals are more or less continuously distributed across the landscape or seascape. We argue that this approach might be seriously flawed: although explicit tests of landscape heterogeneity have great potential to identify environmental factors that impede and enhance gene flow, the examples developed here and in other recent studies suggest that the typical landscape genetics approach requires testing and corroboration by other analyses that lack island-model assumptions. Gene flow is a population-level parameter rather than a property of individual organisms, and genotype distributions in space are affected by time-dependent processes (such as genetic drift) on timescales longer than the lifespan of individual organisms.

For better or worse, the inference of gene flow and other population parameters from genetic data seems to require necessarily a population model, preferably one that can account for population genetic patterns that are characteristic of populations far from the drift–mutation–migration equilibrium. We argue that the aspirations of landscape genetics and related areas of phylogeography (to develop mechanistic explanations of genetic variation at the interface between ecological and evolutionary timescales) might go unfulfilled if the population models and methods used in these disciplines continue to rely on *post-hoc* interpretations rooted in island-model assumptions that population geneticists ostensibly rejected more than 10 years ago (Box 1). Although complex coalescent demographic methods and their population models are clearly not perfect (Box 4), we think that their careful use is preferable to the assumption that population history can be adequately described with a single parameter (i.e. F_{ST}). We look forward to future applications, especially those that apply more realistic models to analyze hundreds or perhaps even thousands of loci [60]. Such approaches hint at the potential for a better understanding of gene flow and other demographic parameters as population and landscape genetics enters the genomic age.

Acknowledgments

We thank B. Crespi, S. Edwards, M. Slatkin, R. Waples and three anonymous reviewers for reading and commenting on earlier drafts of this paper. We also thank P. Beerli, R. Burton, L. Chan, S. Edwards, J. Hey, A. Lawton-Rauh and H. Lessios for earlier discussion, comments and criticisms of some of the ideas presented here, and to L. Chan for sharing data and insight. The authors' research, some of which was presented here, has been primarily supported by grants from the National Science Foundation (NSF OCE-0550526 and OCE-0961996 to PBM) and from the Natural Sciences and Engineering Research Council (NSERC 2004-203052 to MWH). This is Technical Contribution No. 5944 of the Clemson University Experiment Station, based upon work supported by CSREES/USDA, under project number SC-1700342. IMA analyses were conducted on the CBSU computer clusters at the Cornell University Theory Center.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at [doi:10.1016/j.tree.2011.05.007](https://doi.org/10.1016/j.tree.2011.05.007).

References

- Manel, S. and Segelbacher, G. (2009) Perspectives and challenges in landscape genetics. *Mol. Ecol.* 18, 1821–1822
- Holderegger, R. and Wagner, H.H. (2008) Landscape genetics. *Bioscience* 58, 199–207
- Sork, V.L. and Waits, L. (2010) Contributions of landscape genetics—approaches, insights, and future potential. *Mol. Ecol.* 19, 3489–3495
- Selkoe, K.A. *et al.* (2010) Taking the chaos out of genetic patchiness: seascape genetics reveals ecological and oceanographic drivers of genetic patterns in three temperate reef species. *Mol. Ecol.* 19, 3708–3726
- Slatkin, M. (1985) Gene flow in natural populations. *Annu. Rev. Ecol. Syst.* 16, 393–430
- Slatkin, M. (1987) Gene flow and the geographic structure of natural populations. *Science* 236, 787–792
- Mallet, J. (2001) Gene flow. In *Insect Movement: Mechanisms and Consequences* (Woiwod, I.P. *et al.*, eds), pp. 337–360, CAB International
- Marko, P.B. and Barr, K.B. (2007) Basin-scale patterns of mtDNA differentiation and gene flow in the Bay Scallop, *Argopecten irradians concentricus* Say. *Mar. Ecol. Prog. Ser.* 349, 139–150
- Keever, C.C. *et al.* (2009) Discordant distribution of populations and genetic variation in a sea star with high dispersal potential. *Evolution* 63, 3214–3227
- Hart, M.W. and Marko, P.B. (2010) It's about time: divergence, demography, and the evolution of developmental modes in marine invertebrates. *Integr. Comp. Biol.* 50, 643–661
- McGovern, T.M. *et al.* (2010) Divergence genetics analysis reveals historical population genetic processes leading to contrasting phylogeographic patterns in co-distributed species. *Mol. Ecol.* 19, 5043–5060
- Nielsen, R. and Wakeley, J. (2001) Distinguishing migration from isolation: a Markov chain Monte Carlo approach. *Genetics* 158, 885–896
- Hey, J. and Nielsen, R. (2004) Multilocus methods for estimating population sizes, migration rates and divergence time, with applications to the divergence of *Drosophila pseudoobscura* and *D. persimilis*. *Genetics* 167, 747–760
- Hey, J. and Nielsen, R. (2007) Integration within the Felsenstein equation for improved Markov chain Monte Carlo methods in population genetics. *Proc. Natl. Acad. Sci. U.S.A.* 104, 2785–2790
- Becquet, C. and Przeworski, M. (2007) A new approach to estimate parameters of speciation models with application to apes. *Genome Res.* 17, 1505–1519
- Anderson, C.N.K. *et al.* (2005) Serial SimCoal: a population genetic model for data from multiple populations and points in time. *Bioinformatics* 21, 1733–1734
- Huang, W. *et al.* (2011) MTML-msBayes: approximate Bayesian comparative phylogeographic inference from multiple taxa and multiple loci with rate heterogeneity. *BMC Bioinform.* 12, 1–14
- Jobin, M.J. and Mountain, J.L. (2008) Rejector: software for population history inference from genetic data via a rejection algorithm. *Bioinformatics* 24, 2936–2937
- Lopes, J.S. *et al.* (2009) PopABC: a program to infer historical demographic parameters. *Bioinformatics* 25, 2747–2749
- Kuhner, M.K. (2009) Coalescent genealogy samplers: windows into population history. *Trends Ecol. Evol.* 24, 86–93
- Beaumont, M.A. *et al.* (2010) In defense of model-based inference in phylogeography. *Mol. Ecol.* 19, 436–446
- Hickerson, M.J. *et al.* (2010) Phylogeography's past, present, and future: 10 years after Avise, 2000. *Mol. Phylogenet. Evol.* 54, 291–301
- Pearse, D.E. and Crandall, K.A. (2004) Beyond F_{ST} : analysis of population genetic data for conservation. *Conserv. Genet.* 5, 585–602
- Fagundes, N.J.R. *et al.* (2007) Statistical evaluation of alternative models of human evolution. *Proc. Nat. Acad. Sci. U.S.A.* 104, 17614–17619
- Evans *et al.* (2008) A coalescent framework for comparing alternative models of population structure with genetic data: evolution of *Celebes* toads. *Biol. Lett.* 4, 430–433
- Lopes, J.S. and Boessenkool, S. (2010) The use of approximate Bayesian computation in conservation genetics and its application in a case study on yellow-eyed penguins. *Conserv. Genet.* 11, 421–433
- Reagon, M. *et al.* (2010) Genomic patterns of nucleotide diversity in divergent populations of U.S. weedy rice. *BMC Evol. Biol.* 10, 180–196

- 28 Grosberg, R.K. and Cunningham, C.W. (2001) Genetic structure in the sea: from populations to communities. In *Marine Community Ecology* (Bertness, M.D. et al., eds), pp. 61–84, Sinauer Associates
- 29 Crisp, D.J. (1978) Genetic consequences of different reproductive strategies in marine invertebrates. *NATO Conf. Ser. IV Mar. Sci.* 2, 257–273
- 30 Burton, R.S. (1983) Protein polymorphisms and genetic differentiation of marine invertebrate populations. *Mar. Biol. Lett.* 4, 193–206
- 31 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
- 32 Marko, P.B. et al. (2010) The expansion–contraction model of Pleistocene demography: rocky shores suffer a sea change? *Mol. Ecol.* 19, 146–169
- 33 Teske, P.R. et al. (2010) Connectivity between marine reserves and exploited areas in the philopatric reef fish *Chrysolephus laticeps* (Teleostei: Sparidae). *Mar. Biol.* 156, 2029–2042
- 34 Cowen, R.K. and Sponaugle, S. (2009) Larval dispersal and marine population connectivity. *Annu. Rev. Mar. Sci.* 1, 443–466
- 35 Pelc, R.A. et al. (2009) Geographical patterns of genetic structure in marine species with contrasting life histories. *J. Biogeogr.* 36, 1881–1890
- 36 Chan, L.M. et al. (2009) The scale of genetic differentiation in the Dunes Sagebrush-Lizard (*Sceloporus arenicolus*), an endemic habitat specialist. *Conserv. Genet.* 10, 131–142
- 37 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. *Mol. Ecol.* 15, 1419–1439
- 38 Lowe, W.H. and Allendorf, F.W. (2010) What can genetics tell us about population connectivity? *Mol. Ecol.* 19, 3038–3051
- 39 Wares, J.P. (2002) Community genetics in the northwestern Atlantic intertidal. *Mol. Ecol.* 11, 1131–1144
- 40 Avise, J.C. (1989) Gene trees and organismal histories: a phylogenetic approach to population biology. *Evolution* 43, 1192–1208
- 41 Hare, M.P. (2001) Prospects for nuclear gene phylogeography. *Trends Ecol. Evol.* 16, 700–706
- 42 Zink, R.M. and Barrowclough, G.F. (2008) Mitochondrial DNA under siege in avian phylogeography. *Mol. Ecol.* 17, 2107–2121
- 43 Edwards, S. and Bensch, S. (2009) Looking forwards or looking backwards in avian phylogeography? A comment on Zink and Barrowclough 2008. *Mol. Ecol.* 18, 2930–2933
- 44 Brito, P.H. and Edwards, S.V. (2009) Multilocus phylogeography and phylogenetics using sequence-based markers. *Genetica* 135, 439–455
- 45 Hedrick, P.W. (1999) Perspective: highly variable loci and their interpretation in evolution and conservation. *Evolution* 53, 313–318
- 46 Hedrick, P.W. (2005) A standardized genetic differentiation measure. *Evolution* 59, 1633–1638
- 47 Ellegren, H. (2004) Microsatellites: simple sequences with complex evolution. *Nat. Rev. Genet.* 5, 435–445
- 48 Strasburg, J.L. and Rieseberg, L.H. (2009) How robust are ‘isolation with migration’ analyses to violations of the IM model? A simulation study. *Mol. Biol. Evol.* 27, 297–310
- 49 Rannala, B. and Yang, Z. (2003) Bayes estimation of species divergence times and ancestral population sizes using DNA sequences from multiple loci. *Genetics* 164, 1645–1656
- 50 Yang, Z. (2002) Likelihood and Bayes estimation of ancestral population sizes in hominoids using data from multiple loci. *Genetics* 162, 1811–1823
- 51 Jennings, W.B. and Edwards, S.V. (2005) Speciation history of Australian grass finches (*Poephila*) inferred from thirty gene trees. *Evolution* 59, 2033–2047
- 52 Carstens, B.C. and Knowles, L.L. (2007) Shifting distributions and speciation: species divergence during rapid climate change. *Mol. Ecol.* 16, 619–627
- 53 Maddison, W.P. (1996) Molecular approaches and the growth of phylogenetic biology. In *Molecular Zoology: Advances, Strategies, and Protocols* (Ferraris, J.D. and Palumbi, S.R., eds), pp. 47–63, Wiley-Liss
- 54 Roderick, G.K. (1996) Geographic structure of insect populations: gene flow, phylogeography, and their uses. *Annu. Rev. Entomol.* 41, 325–352
- 55 Galindo, H.M. et al. (2006) Seascape genetics: a coupled oceanographic-genetic model predicts population structure of Caribbean corals. *Curr. Biol.* 16, 1622–1626
- 56 Loveless, M.D. and Hamrick, J.L. (1984) Ecological determinants of genetic structure of plant populations. *Annu. Rev. Ecol. Syst.* 15, 65–95
- 57 Hamrick, J.L. and Godt, M.J. (1989) Allozyme diversity in plant species. In *Plant Population Genetics, Breeding and Germplasm Resources* (Brown, A.H.D. et al., eds), pp. 43–63, Sinauer Associates
- 58 Bohonak, A.J. (1999) Dispersal, gene flow, and population structure. *Quart. Rev. Biol.* 74, 21–45
- 59 Tatarenkov, A. et al. (2010) Microgeographic population structure of green swordtail fish: Genetic differentiation despite abundant migration. *Mol. Ecol.* 19, 257–268
- 60 Wang, Y. and Hey, J. (2010) Estimating divergence parameters with small samples from a large number of loci. *Genetics* 184, 363–379
- 61 Wright, S. (1969) *Evolution and the Genetics of Populations. Variability Within and Between Populations* (Vol. 4), University of Chicago Press
- 62 Slatkin, M. and Barton, N.H. (1989) A comparison of three indirect methods for estimating average levels of gene flow. *Evolution* 43, 1349–1368
- 63 Neigel, J.E. et al. (1991) Estimation of single generation migration distances from geographic variation in animal mitochondrial DNA. *Evolution* 45, 423–432
- 64 Slatkin, M. (1993) Isolation by distance in equilibrium and nonequilibrium populations. *Evolution* 47, 264–279
- 65 Bossart, J.L. and Prowell, D.P. (1998) Genetic estimates of population structure and gene flow: limitations, lessons and new directions. *Trends Ecol. Evol.* 13, 202–206
- 66 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, 117–125
- 67 Neigel, J.E. (2002) Is F_{ST} obsolete? *Conserv. Genet.* 3, 167–173
- 68 Hellberg, M.E. (1994) Relationships between inferred levels of gene flow and geographic distance in a philopatric coral, *Balanophyllia elegans*. *Evolution* 48, 1829–1854
- 69 Hutchison, D.W. and Templeton, A.R. (1999) Correlation of pairwise genetic and geographic distance measures: inferring the relative influences of gene flow and drift on the distribution of genetic variability. *Evolution* 53, 1898–1914
- 70 Hedgecock, D. et al. (2007) Genetic approaches to measuring connectivity. *Oceanography* 20, 70–79
- 71 Hanski, I. and Simberloff, D. (1997) The metapopulation approach, its history, conceptual domain and application to conservation. In *Metapopulation Biology: Ecology, Genetics, and Evolution* (Hanski, I.A. and Gilpin, M.E., eds), pp. 5–26, Academic Press
- 72 Kritzer, J.P. and Sale, P.F., eds (2006) *Marine Metapopulations*, Academic Press
- 73 Wade, M.J. and McCauley, D.E. (1988) Extinction and recolonization: their effects on the genetic differentiation of local populations. *Evolution* 42, 995–1005
- 74 Whitlock, M.C. and McCauley, D.E. (1990) Some population genetic consequences of colony formation and extinction: genetic correlations within founding groups. *Evolution* 44, 1717–1724
- 75 Giles, B.E. and Goudet, J. (1997) A case study of genetic structure in a plant metapopulation. In *Metapopulation Biology: Ecology, Genetics, and Evolution* (Hanski, I. and Gilpin, M., eds), pp. 429–454, Academic Press
- 76 Pannell, J.R. and Charlesworth, B. (1999) Neutral genetic diversity in a metapopulation with recurrent local extinction and recolonization. *Evolution* 53, 664–667
- 77 Pannell, J.R. (2003) Coalescence in a metapopulation with recurrent local extinction and recolonization. *Evolution* 57, 949–961
- 78 Wakeley, J. (2004) Metapopulation models for historical inference. *Mol. Ecol.* 13, 865–875
- 79 Ewens, W.J. (1990) Population genetics theory—the past and the future. In *Mathematical and Statistical Developments of Evolutionary Theory* (Lessard, S., ed.), pp. 177–227, Kluwer Academic
- 80 Avise, J.C. (2004) *Molecular Markers, Natural History and Evolution*, (2nd edn), Sinauer
- 81 Wakeley, J. (2008) *Coalescent Theory: An Introduction*, Roberts & Co.
- 82 Holsinger, K.E. and Weir, B.S. (2009) Genetics in geographically structured populations: defining, estimating and interpreting F_{ST} . *Nat. Rev. Genet.* 10, 639–650

- 83 Anderson, C.D. *et al.* (2010) Considering spatial and temporal scale in landscape-genetic studies of gene flow. *Mol. Ecol.* 19, 3565–3575
- 84 Storfer, A. *et al.* (2010) Landscape genetics: where are we now? *Mol. Ecol.* 19, 3495–3514
- 85 Cornuet, J.M. *et al.* (1999) New methods employing multilocus genotypes to select or exclude populations as origins of individuals. *Genetics* 153, 1989–2000
- 86 Pritchard, J.K. *et al.* (2000) Inference of population structure using multilocus genotype data. *Genetics* 155, 945–959
- 87 Wilson, G.A. and Rannala, B. (2003) Bayesian inference of recent migration rates using multilocus genotypes. *Genetics* 163, 1177–1191
- 88 Yu, H. *et al.* (2010) Slatkin's paradox: when direct observation and realized gene flow disagree. A case study in *Ficus*. *Mol. Ecol.* 19, 4441–4453
- 89 Jones, A. (2010) Reconciling field observations of dispersal with estimates of gene flow. *Mol. Ecol.* 19, 4379–4382
- 90 McKechnie, S.W. *et al.* (1975) Population genetics of *Euphydryas* butterflies. I. Genetic variation and the neutrality hypothesis. *Genetics* 81, 571–594
- 91 Ahmed, S. *et al.* (2009) Wind-borne insects mediate directional pollen transfer between desert fig trees 160 kilometers apart. *Proc. Natl. Acad. Sci. U.S.A.* 106, 20342–20347
- 92 Ashley, M.V. (2010) Plant parentage, pollination, and dispersal, how DNA microsatellites have altered the landscape. *Crit. Rev. Plant. Sci.* 29, 14–161
- 93 Taneyhill, D.E. *et al.* (2001) Estimating rates of gene flow in endemic butterfly races: the effect of metapopulation dynamics. In *Genes in the Environment: 15th Special Symposium of the British Ecological Society* (Hails, S. *et al.*, eds), pp. 3–25, British Ecological Society
- 94 Marko, P.B. and Hart, M.W. (2011) Retrospective coalescent methods and the reconstruction of metapopulation histories in the sea. *Evol. Ecol.* DOI: 10.1007/s10682-011-9467-9
- 95 Kelly, R.P. and Palumbi, S.R. (2010) Genetic structure among 50 species of the northeastern Pacific rocky intertidal community. *PLoS ONE* 5, pe8594
- 96 Wares, J.P. and Cunningham, C.W. (2001) Phylogeography and historical ecology of the North Atlantic intertidal. *Evolution* 55, 2455–2469
- 97 Beerli, P. (2004) Effect of unsampled populations on the estimation of population sizes and migration rates between sampled populations. *Mol. Ecol.* 13, 827–836
- 98 Slatkin, M. (2005) Seeing ghosts: the effect of unsampled populations on migration rates estimated for sampled populations. *Mol. Ecol.* 14, 67–73