Population Graphs and Landscape Genetics

Rodney J. Dyer

Department of Biology and Center for Environmental Studies, Virginia Commonwealth University, Richmond, Virginia 23284-2012; email: rjdyer@vcu.edu

Annu. Rev. Ecol. Evol. Syst. 2015. 46:327–42

First published online as a Review in Advance on October 6, 2015

The *Annual Review of Ecology, Evolution, and Systematics* is online at ecolsys.annualreviews.org

This article's doi: 10.1146/annurev-ecolsys-112414-054150

Keywords

landscape genetics, Population Graphs, spatial structure, connectivity

Abstract

At the heart of the analyses of landscape genetics are isolation models seeking to explain either interindividual or interpopulation connectivity. These models use spatial, ecological, and topographic predictor variables measured between sites in an attempt to explain observed genetic variation. During the past decade, these models have adopted an increasingly sophisticated set of techniques to quantify intervening physical and ecological spaces, although they are restrained by rather mundane approaches to characterizing the genetic components of connectivity. Population Graphs are one approach to improving the quantification of genetic covariance used in models of landscape genetics. I explain the construction of the Population Graph framework, explain its strengths and weaknesses, and provide examples of how it has been used during the past decade within the contexts of landscape and population genetics.

1. INTRODUCTION

The term landscape genetics is commonly applied to studies of population genetics that integrate analyses of genetic connectivity with spatio-ecological data commonly derived from landscape ecology (Dyer 2015). Although originally focused on detecting barriers and identifying gradients (Manel et al. 2003), landscape genetics has grown into a more general term applied to studies that identify features of the intervening landscape that influence either individual-based movement (e.g., Cushman et al. 2006) or interpopulation gene flow (e.g., Murphy et al. 2010). In general, analyses in landscape genetics have the basic form

$$G \sim f(E)$$

where the term G encompasses some translation of genetic structure, diversity, or distance, to be explained by one or more external (E) variables representing characteristics such as topography, space, ecology, or habitat. The relationship between the independent spatial or ecological variables, or both, with the genetic data contained in G is often examined in the context of pairwise distance metrics and isolation models.

Models of genetic differentiation as a function of spatial isolation date to Wright (1943), and their use has been applied broadly to studies based on the sampling of individuals and populations (e.g., Malécot 1948, Slatkin 1993). The use and subsequent extension of classic isolation models are virtually the defining features of contemporary studies of landscape genetics. Extensions to the basic isolation by distance (IBD) model have been proposed that replace Euclidean space with variables representing other potentially isolating features, including isolation by ecological resistance (McRae 2006), isolation by habitat (Nosil et al. 2008), isolation by environment (Wang et al. 2013), and isolation by barrier (Schwartz et al. 2002). Perhaps even more fundamental than the addition of other isolating features is the breadth of objective functions that have been developed to estimate separation under these models. The degree of separation measured as spatial, ecological, or along some other resistance surface, is now commonly measured using not only Euclidean (straight-line) distance but also algorithms that estimate the least-cost-path distance (Foley & Holland 2010), path lengths restricted to specific corridors (Pinto & Keitt 2009), or weighted lengths of all potential paths (circuit theory; McRae et al. 2008).

Although these developments have provided a much broader tool set for understanding connectivity—one that has the potential to more closely describe how organisms interact with the environment through which they move—the genetic component of these analyses has largely remained static. In this review, I highlight one recent extension to the treatment of the genetic component used in analyses of both landscape and population genetics. This approach, called Population Graphs, is an implicitly graph-theoretical representation of multilocus genetic covariance. Although Population Graphs possess characteristics that are analogous to both structure statistics and isolation models commonly employed in population-genetic analyses (Dyer 2007), they are a model-free approach that describes the way in which genetic variation is distributed across the landscape. I begin by introducing the mathematics behind Population Graphs and cover the scope and current limitations of its use. Drawing upon a decade of studies, I show how the graph-theoretical shape of Population Graphs and parameters derived thereof have been used to better understand how landscape features influence genetic connectivity and, by extension, the maintenance of structure. I close by providing an overview of ongoing developments in Population Graph analyses and comment briefly on the current state of studies of landscape genetics.

2. POPULATION GRAPHS

The amount of genetic variation observed among strata, σ_A^2, is a function of both the rate of migration among strata and the genealogical history of the markers and populations being examined. The magnitude of this variation can been quantified using an array of similar statistics, such as Wright's F_{ST} (Wright 1931), Nei's G_{ST} (Nei 1973), and Slatkin's R_{ST} (Slatkin 1995), and the dueling parameters for markers with high allelic diversity from Hedrick (2005, G'_{ST}) and Jost (2008, D_{EST}, although see Whitlock 2011). Despite the plethora of approaches, all of these indices provide a singular inference: the magnitude of variation that exists among strata.

Topology: the pattern of edge connections within a network that define localized and global network structures

Given the centrality of σ_A^2 in these parameters, it should be pointed out that the range of values that these parameters may assume can arise from a wide range of different, and potentially mutually exclusive, demographic histories. This is an unfortunate consequence of summarizing variance into a single index. It is particularly problematic when we are using these tools with the sole intent of uncovering unknown demographic histories and processes. Moreover, statistics such as these are often insensitive to recent changes or perturbations, or both, in mating patterns (e.g., Landguth et al. 2010), making it difficult to resolve the relatively recent or contemporary processes that are the focus of landscape genetics.

The Population Graph framework was designed to get around these problems. Instead of estimating the magnitude of variation, Population Graphs define the way in which genetic variation, independent of magnitude, is distributed among strata. This is done using a two-step process whereby total genetic variation is first decomposed into a geometric interpretation of components within and among strata, and then that graph-theoretical structure is modified using conditional covariance to identify the minimal topological configuration necessary to describe the totality of observed genetic variation. The genetic structure of within-population variance, $\sigma_{W,i}^2$, is represented as a volume surrounding the centroid of the population (**Figure 1*a***). This volume is analogous to the populationwise contribution to error variance from a random effects linear

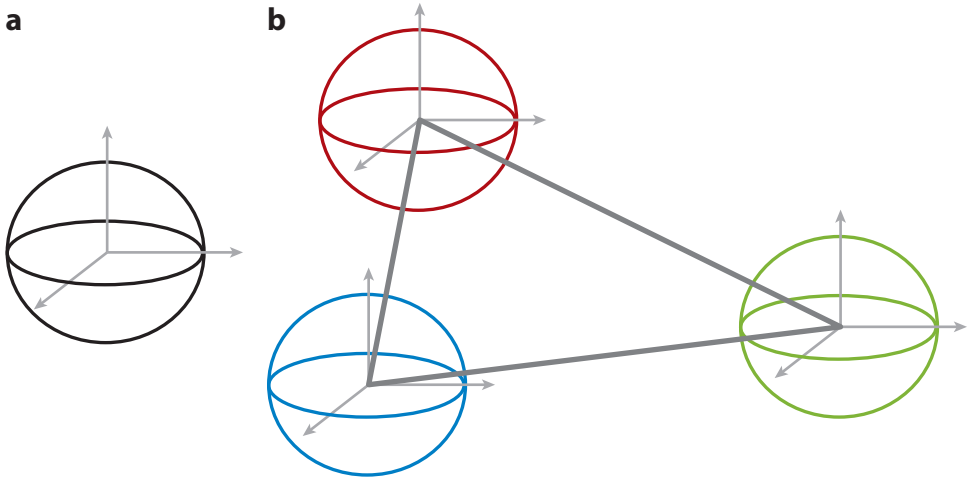

Figure 1

Geometric representation of population-genetic covariance. (*a*) Each individual population is represented geometrically as a node whose volume is proportional to the within-population component of genetic variance, σ_A^2. (*b*) Different populations are connected by edges whose magnitude is proportional to their interpopulation variance $\sigma_{A;i\leftrightarrow j}^2$.

model, such as AMOVA (analysis of molecular variance) (e.g., Dyer et al. 2004). The among-strata component of genetic variation, σ^2_A, is similarly partitioned into the additive components representing variation among all pairs of populations, $\sigma^2_{A;i\leftrightarrow j}$. The centroids of populations are connected with an edge whose length is defined by $\sigma^2_{A;i\leftrightarrow j}$ (**Figure 1*b***). These two components constitute a network structure with nodes (strata) and edges (measures of covariance) identical to how variance is decomposed in the AMOVA analysis (Excoffier et al. 1992), though represented in a graph-theoretical context.

A Population Graph is derived from this network AMOVA structure by applying conditional covariance to the edges that connect all strata. Conditional covariance is used to determine the subset of edges in the network which are redundant, in terms of describing the total pattern of among-strata genetic variation. This process is statistically analogous to building a linear regression model, whereby additional predictor variables are examined for their importance relative to the sufficiency of the existing model (a Type III sums of squares). If, given what is already in the model, the addition of the edge provides significant power to explain the overall topology, then it is kept within the network; otherwise, it is removed. Biologically, conditional covariance functions reveal the demographic covariance network nested within the topology that contains connections among all pairs of strata. This decomposition provides the reverse mapping from σ^2_A back to the underlying demographic model that is unattainable with the use of structure statistics, such as F_{ST} and its many analogs, alone. By way of example, consider a simple three-population model with a source population and two sink populations (**Figure 2**). In this model, both overall and pairwise estimates of genetic structure (F_{ST}) will stabilize and equilibrate through time. However, even if there is no direct migration between islands B and C, their structure estimates [denoted $F_{ST(B,C)}$ in **Figure 2**] will indicate ongoing connectivity. Even without a direct exchange of migrants among islands, differentiation is low because both sink populations receive individuals from the same mainland. From a population-genetics perspective, this pattern would be interpreted as being due to ample interisland gene flow, which is an incorrect interpretation given the underlying demographic model. Ambiguous patterns arising from looking at the magnitude of divergence are not unique

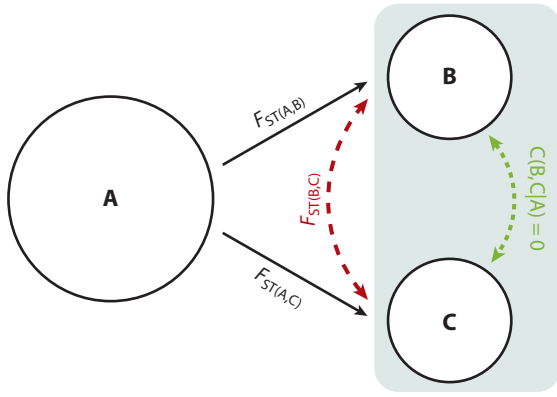

Figure 2

Schematic representing a demographic model with a single-source population (A) and two sink populations (B and C). Black arrows indicate the routes of actual gene flow. If examined using traditional structure statistics, pairwise F_{ST} will be the same across all three comparisons even though no real gene flow is occurring among the islands (*red*). Conditional covariance as implemented in Population Graphs seeks to examine differentiation among populations, given each island's covariance with the entirety of the data (*green*).

to source–sink models alone. Kimura & Weiss (1964) showed that in 1D stepping-stone models, the correlation between allele frequencies among populations decays at a rate of $r \sim e^{-x\sqrt{2\mu/m}}$, where x is the number of steps, μ is the mutation rate, and m is the migration rate. Functionally, this means that for parameters such as $\mu = 10^{-6}$ and $m = 0.01$ (not uncommon in studies of landscape genetics), the correlation of allele frequencies between pairs of populations that are not directly exchanging migrants will exceed $\rho = 0.80$ at distances of up to five steps.

The problem here is that the relationship between the islands, as measured by $F_{ST(B,C)}$ in **Figure 2**, or allele frequency correlations after Kimura & Weiss (1964), is taken without reference to the rest of the data set. Conditional covariance estimates the covariance between the islands, conditional on each individual island's covariance with the mainland. Once the covariance between the mainland and the islands is accounted for, the covariance between the islands, denoted C(B, C|A), will disappear. The resulting covariance topology, which started out as a fully connected graph, will retain connections only between populations with nonzero conditional covariance, which in the example from **Figure 2** resolves the underlying demographic model much better than sets of pairwise estimates.

2.1. Graph-Theoretical Fundamentals

As a graph-theoretical object, a Population Graph can be both manipulated and examined using a host of network-based methods that also have biologically meaningful interpretations. A Population Graph, $G = \{V,E\}$, is composed of two separate sets, nodes and edges, that are estimated from N individuals in k populations using their multilocus genotypes. The node set (also called vertex set in some notations) is denoted as V and represents the sampling strata, with size $|V| = k$. The number of vertices in a graph defines the potential size of the second set, E, the edge set, and can only be as large as $m = k(k-1)/2$. As defined, edges in a Population Graph have a characteristic length equal to $\sigma^2_{A;i\leftrightarrow j}$, and are bidirectional because covariance is a symmetric trait.

Several parameters measured on both node and edge sets can be useful in analyses of population and landscape genetics. Quantitative analyses of graph components can focus on nodes, edges, the clustering of nodes within the graph, the connections between subcomponents in the network, overall network robustness, or a combination of these. The overall size of the graph defines a connection probability, $p_c = |E|/m$. Population Graphs with higher connectivity have less genetic independence and hence lower overall genetic structure. The correlation between these two has been shown via simulations to be exceptionally high (correlation of node centrality with F_{ST} is $\rho = -0.95$; Dyer 2007). The relative importance of individual nodes and edges is also quantifiable using several common graph-theoretical parameters, as discussed in the sidebar, Network Structure.

The way in which nodes are connected within a Population Graph can also provide biologically meaningful inferences. Genetic markers have commonly been used to rank the relative importance of populations based upon the magnitude of diversity or structural divergence, or both (Falk & Holsinger 1991, Milligan et al. 1994). Graph-theoretical parameters can provide an additional set of criteria that quantify the relative position of populations within the functional connectivity network, independently of the amount of resident diversity or relative structure. Measures of connectivity may highlight genetically depauperate populations as being of positional importance based upon their physical location in the gene-flow network. For example, the two populations connecting peninsular Baja California to mainland Mexico (derived from Dyer & Nason 2004) in the Population Graph depicted in **Figure 3** are not overtly diverse or divergent. However, their relative position in the network suggests that they are critical relative to the distribution

NETWORK STRUCTURE

Nodes and edges can be quantified as to their relative importance to overall graph structure using several approaches (Wasserman & Faust 2005). A common set of estimates, grouped under the moniker centrality, seeks to identify the importance of individual components relative to the shape and structure of the entire topology. Metrics used to estimate centrality include the degree (the number of edges attaching to a node), closeness (a weighted measure of distance from the node to all other nodes), and betweenness (a ranking of how many of the shortest paths through the graph go through a specific node or edge). Higher-level network structures can be quantified by identifying the clustering of nodes (referred to as a clique), estimating the modularity within the network for robustness, and calculating relative estimates of network size for comparative purposes (diameter). Within Population Graphs, these approaches have been used to quantify relative population importance in relation to rates of gene flow (Dyer 2007) and differences between stage-stratified cohorts of plants (Herrera-Arroyo et al. 2013) and as measures of network sensitivity (Koen et al. 2013) and robustness (Albert et al. 2013).

of population-genetic covariance as depicted in the connectivity network. Positional importance within a gene-flow network has not been examined in great detail, although it has the potential to provide insights salient to conservation and management decisions.

2.2. Sampling and Scale

At present, networks based upon interindividual distances are not supported, alienating a large fraction of landscape-genetic studies, due to the necessity of estimating within-population variance (see Dyer 2015). At the time of writing, 150 manuscripts had cited the initial Dyer & Nason (2004) Population Graph paper, of which roughly half had applied the approach as a component of their research (the other half citing it parenthetically). Despite being a representation of genetic covariance with discrete strata, Population Graphs have been applied to organisms exhibiting a wide range of natural distributions, from continuous terrestrial (Dyer et al. 2010) and marine (Fitzpatrick et al. 2011) to taxa inhabiting discretely defined populations (e.g., Giordano et al. 2007). More than half of the studies (52%) have applied Population Graphs to plant taxa, with the remainder split among animal systems (38%), fungal studies (8%), and the work on HIV dispersion of Pagán & Holguín (2013) (<1%).

Given its derivation from the AMOVA model, the sample allocation of individuals within strata can be guided by standardized statistical approaches to minimize the variance in Φ_{ST}, an interclass correlation (Smouse et al. 2001). The expected variance in Φ_{ST} depends upon the number of strata (k) and the number of individuals sampled from each population (j). For a fixed sample size ($N = j * k$), the variance in the among-strata component is given by

$$s^2_{\Phi} = 2\frac{[1 + (k-1)\Phi]^2(1-\Phi)^2}{k(k-1)(j-1)} \quad 1.$$

(after Falconer 1981). This relationship should be used to examine a range of potential sampling allocations (e.g., how many individuals and populations are examined) that could minimize the overall variance in genetic-structure parameters. As pointed out by Koen et al. (2013), studies of landscape genetics that use populations as focal sampling units should focus on maximizing the number of sampling locations in preference to increasing within-population precision; it is, after all, the covariance among populations that is created by the patterns of connectivity. This emphasis provides a better characterization of how connectivity is being influenced by the intervening

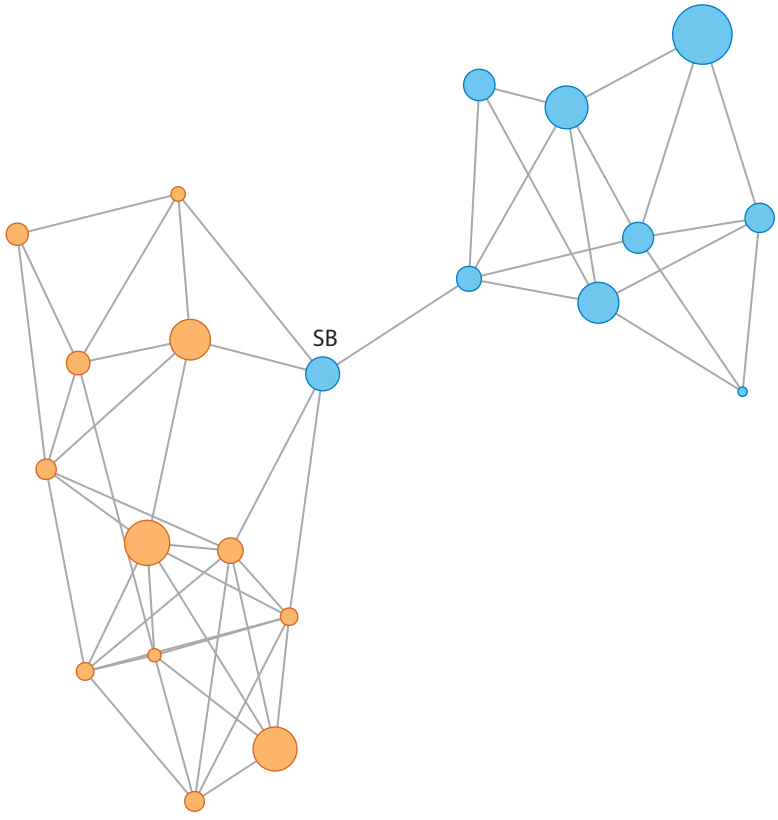

Figure 3

The Population Graph representing the distribution of genetic covariance in the Sonoran Desert–endemic *Lophocereus schottii* (Cactaceae) from Dyer & Nason (2004). Nodes represent individual populations in peninsular Baja California (*orange*) and mainland Mexico (*blue*). The node marked SB is spatially proximate to populations in mainland Mexico, although it is genetically connected to populations in mid-peninsula Baja California. Edges connecting nodes represent the pattern of genetic covariance among all populations.

landscape because there are more comparisons to be examined among populations crossing different landscape features. Individual-based analyses have the opposite problem in that they need more at-site variance, which Landguth et al. (2012) interpreted to mean more loci per individual. Across most levels of among-strata variance seen in studies of landscape genetics, however, the relationship in Equation 1 provides a window of equally powerful sample allocations from which to choose.

A fundamental requirement of estimating covariance is the need for a robust estimate of within-stratum variation because it is used as a denominator in several steps in the calculations. As with any sampling scheme, poor estimation of within-stratum variance will result in reduced statistical power. In general, within-population variance is minimized by the addition of mulitlocus genetic markers. The number of markers required is a function of both the underlying spatial distribution of genetic variation as well as the granularity of the sampling; permutation and subsampling techniques can be used to ensure that parameters have stabilized. Studies have developed Population Graphs from a broad range of marker types including allozymes (Dyer & Nason 2004),

microsatellites (Krafsur et al. 2005), amplified fragment length polymorphisms (Greve et al. 2012), and single nucleotide polymorphisms (Dell'Acqua et al. 2014). Sequence data have yet to be applied to the Population Graph approach, although not because of any inherent incompatibility: Dropping noninformative sites makes variable nucleotides similar to single nucleotide polymorphisms. However, the kind of research questions for which sequence divergence is appropriate often focus on bifurcation and divergence at longer timescales than the more contemporary or relatively recent connectivity analyses that are the focus of most studies of landscape genetics. Depending upon the life-history characteristics of the study organisms and the temporal scale at which the processes are operating, analyses based upon genetic covariance may be more informative than sequence-based markers evaluating sequence variants, even at temporal scales approximating phylogeographic processes (e.g., Dyer & Nason 2004, Dyer et al. 2010, Garrick et al. 2013).

2.3. Statistical Performance

The shape and makeup of a Population Graph, as is true for any analysis of structure, are sensitive to the number of populations being sampled and the amount of variation present. In general, the utility of Equation 1 is that it allows one to determine how to maximize the number of sites (k) across which topological structure is based (e.g., the pattern of edge connections). The effects of not sampling all potential sites, or undersampling individuals within particular sites, in a Population Graph has been examined, thus far, only briefly. Koen et al. (2013) subsampled two empirical data sets to compare the relative statistical stability of pairwise structure statistic (F_{ST}) and distance (e.g., Bray–Curtis distance, D_C) with Population Graph structure. The numerical values estimated for edge weights in the topology were sensitive to both unsampled and undersampled populations, as would be expected given that topology is estimated based upon the totality of the data, and changes in any part of these data will percolate throughout the Population Graph. That is not to say that the statistical inferences obtained from the topology change, but that the lengths of the edges change as the input data set changes. Alternatively, Koen et al. (2013) found that pairwise estimates of genetic structure and distance were unaffected by changes in the entire data set unless they were changed individually within the pair of populations for which the parameter was being estimated (e.g., pairwise parameters are sensitive only to the pairs of populations being manipulated). Despite changes in the edge lengths, subsampling individuals within populations did not negatively impact subsequent model fit with respect to spatial or resistance isolation.

Because the entire data set is used to estimate the shape of a Population Graph rather than using single pairs of populations to estimate individual parameters, topological structure stabilizes faster, after perturbation, than more traditional parameters quantifying genetic structure (F_{ST}, Φ_{ST}), genetic distance (D_C), and isolation (e.g., $\hat{M}$ from Slatkin 1993; Dyer 2007, Dyer et al. 2010, Noutsos et al. 2014). Population Graphs may also be more robust to loci that deviate from optimal expectations. While examining the spatial genetic structure and connectivity in subspecies of the Arabian burnet moth, Klütsch et al. (2011) showed simulation results suggesting that although the presence of null alleles at microsatellite loci increases the size of the edge set (e.g., there are more connections in Population Graphs with null alleles than the same data without them), thereby overestimating connectivity, they do so by increasing the size of the edge set by only 3%, which is insufficient to cause any major inferences about underlying demographic models.

3. INFERENCES BASED UPON GENETIC TOPOLOGIES

Given that Population Graphs focus on how genetic variation is distributed, many of the biologically relevant inferences gained by adopting this approach can be found by analyzing the way in which the network is connected.

3.1. Topological Shape

An almost ubiquitous feature of studies of population and landscape genetics during the past decade has been the utilization of clustering algorithms such as STRUCTURE (Pritchard et al. 2000). These methods are intended to identify clear partitions in the data. In a Population Graph, partitions can result in fragmentation of the topology, thus creating two or more disconnected subgroups (e.g., collections of populations unconnected to the rest of the network). Disconnected groups are statistically independent and represent independently evolving phylogroups. However, clear clusters are not always present in the data, and it is well known that approaches such as STRUCTURE perform relatively poorly when connectivity occurs along gradual gradients (e.g., Pritchard et al. 2007), when structure is nested (Dupanloup et al. 2002, Evanno et al. 2005), or in the presence of autocorrelation (e.g., Schwartz & McKelvey 2009). Although it is possible to adjust prior probabilities to account for some of these problems, it is often the impetus of studies of landscape genetics to specifically identify which features are modifying connectivity, making the approach of adjusting prior probabilities a bit tautological.

Population Graphs are not specifically intended to find partitions but instead to describe how variation is distributed. Their use, however, in conjunction with clustering algorithms may aid in understanding the spatial distribution of genetic variation within species that have heterogeneous or nested structures. Fitzpatrick et al. (2011) leveraged the combination of these tools in the analysis of genetic variance in parrotfish (*Scarus rubroviolaceus*; Scaridae) sampled from throughout the Indian and Pacific Oceans. Both Hawaiian and eastern Pacific populations were clearly partitioned in the STRUCTURE analysis, and populations from the central–western Pacific Ocean and Indian Ocean were both moderately admixed. The Population Graph for this showed a more nuanced interpretation of the data, with the Hawaiian islands being entirely disconnected and the remaining populations being connected only tenuously between eastern Pacific populations and those found in the central–western Pacific and Indian Oceans. It is, perhaps, in these kinds of cases, where genetic structure does not have clean breaks, that Population Graphs may be most helpful in interpreting the processes that produced the observed data. Even without disconnected subcomponents, the pattern of connectivity between subgroups can be evaluated for nested structure using nonparametric approaches. Cases in which this has been particularly informative include identifying endemic groupings in livestock herds (Li et al. 2007), specific landscape features causing genetic vicariance (Dyer & Nason 2004), the genetic consequences of alternate habitat choice (Giordano et al. 2007), subspecies boundaries (Klütsch et al. 2011), and reproductive isolation (Li et al. 2014). The important point here is that the topology itself is not the end result, but it can function as a platform on which subsequent hypotheses can be tested.

3.2. Conditional Genetic Distance

All networks can be represented either graphically for visualization or in matrix format for subsequent analyses. An adjacency matrix, $\mathbf{A} = \{a_{ij}\}; i,j = 1,\ldots,k$, records the pattern of connectivity and the length of the edges that connect nodes. The adjacency matrix for a Population Graph does not define a pairwise distance amenable to analysis directly; it describes only the pattern of connectivity between nodes sharing an edge. However, it is easily manipulated to produce a matrix of shortest-path distances through the network.

Conditional genetic distance (cGD) is defined as the length of the shortest path through the topology connecting populations and is estimated directly from the adjacency matrix (Dyer et al. 2010). The distances between nodes in disconnected topologies (e.g., those with separate components) are infinite and cannot be used in subsequent analyses. The use of distance in a

Population Graph as a metric of isolation was first introduced by Dyer & Nason (2004), who showed that a graph-distance metric provided better-fit models of IBD (calling it isolation by graph distance). The parameter cGD has been used in models of IBD (Noutsos et al. 2014), habitat (Quintela et al. 2014), and resistance (Dyer et al. 2010) as a primary measure of pairwise genetic separation. The parameter cGD has been shown to provide a better fit than traditional pairwise statistics under many conditions due to its property of homoscedasticity (although see Dyer et al. 2010 for when Bray–Curtis distance, D_C, may be preferred).

Extended edges: edges in a topology whose physical separation is much more than would be expected based upon genetic covariance

Compressed edges: edges connecting populations that are physically closer than expected based upon genetic covariance

3.3. Compressed and Extended Edges

Although the pattern of connectivity within a topology is based solely upon genetic information, the nodes represented in the topology depict real populations and can be assigned georeferenced coordinates and mapped. If genetic covariance is determined solely by the physical separation between populations (e.g., a pure IBD model) then we would expect to see the relative lengths of the edges in the Population Graph scaling proportionately to those lengths in the Population Graph. Deviations in edge lengths for mapped Population Graphs arise from one of two potential processes. Long-distance dispersal events and range expansion tend to create physical distances between populations that are much greater than expected given their genetic covariance. In a Population Graph, we can identify these extended edges specifically because their physical length (based upon spatial or ecological distances) will be disproportionately large given their edge weight in the topology defined by genetic covariance. Conversely, landscape features that restrict gene flow, such as inhospitable habitat and other barriers including mountain ranges, rivers, and urban centers, will result in populations that are physically more proximate than expected given their genetic covariance. These compressed edges represent potential spatial locations where genetic differentiation is more than expected given the spatial arrangement of populations: that is, regions where intervening habitat restricts connectivity. Within a single topology, there may be both compressed and extended edges (**Figure 4**), showing that throughout species' distribution, connectivity interacts differentially with specific features of the intervening landscape (e.g., Wegier et al. 2011). For processes such as range expansion, extended edges will not only be present but will also be primarily aligned along the axis of expansion. This was demonstrated in the Sonoran desert plant *Euphorbia lomelii*, where all of the extended edges were aligned with the known route of range expansion following the Pleistocene (Garrick et al. 2009). Alignment and expansion have been found in less obvious cases as well, including in the spatial structure of the invasive grass *Microstegium vimineum*, where two separate invasion lineages have subsequently come into secondary contact (Baker & Dyer 2011). Because the topology itself is created without reference to the spatial arrangement of populations, mapping topologies identify disproportionately scaled edges whose presence gives clues about differential demographic and landscape factors impacting genetic connectivity.

3.4. Topological Congruence

Although individual Population Graphs provide insights into how genetic structure is spatially distributed, collections of Population Graphs may be able to expand these inferences about population dynamics in several ways. There appear to be only a few examples of studies that have compared multiple Population Graphs. Among the first of such studies, Fortuna et al. (2009) analyzed the spatial arrangement of genetic covariance in four co-occurring plant species, with the goal of understanding similarities in the substructure of the graphs. They argued that the underlying modularity in these networks outlines the boundaries of evolutionarily significant units, management units, or both. In a subsequent paper, Albert et al. (2013) extended this work by examining

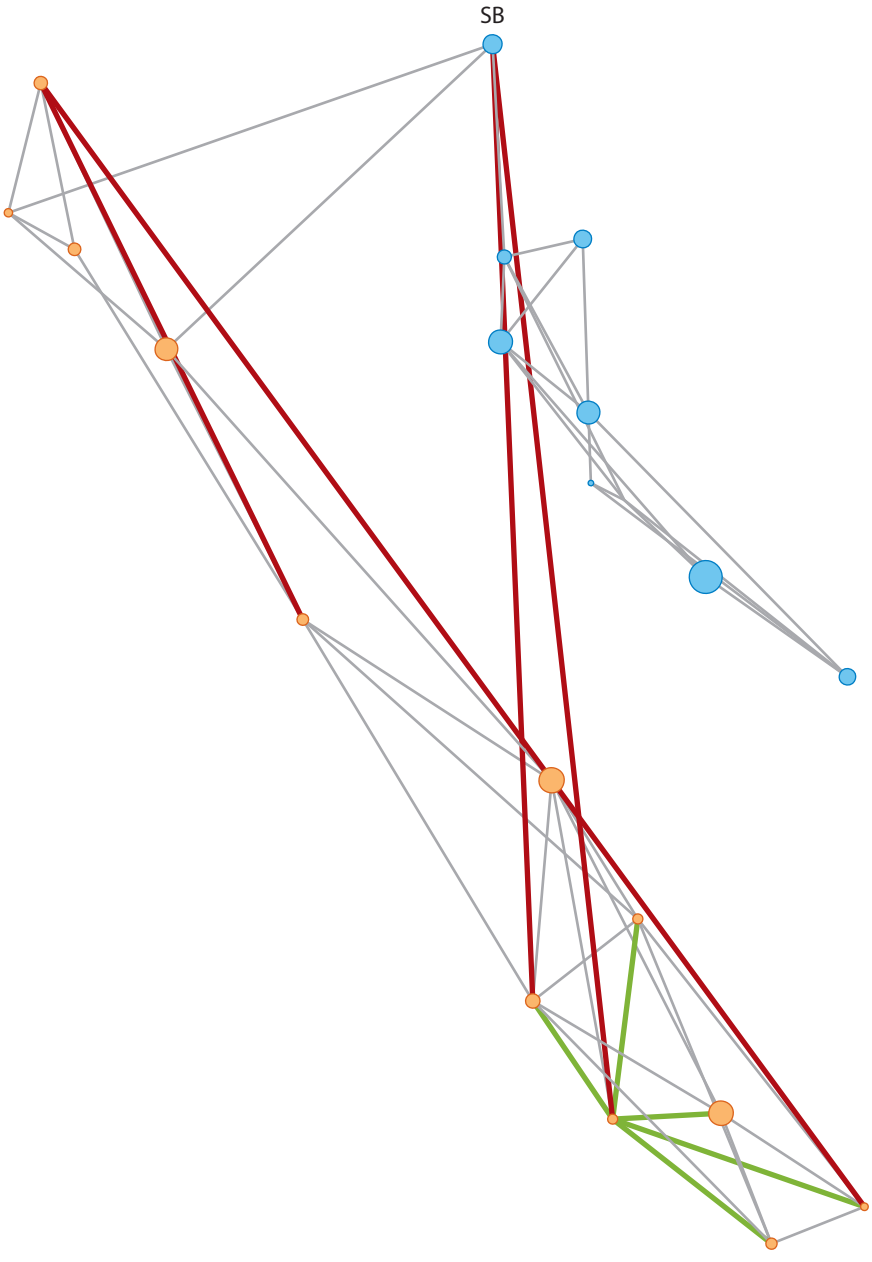

Figure 4

The Population Graph representing the distribution of genetic covariance in the Sonoran Desert–endemic *Lophocereus schottii* (Cactaceae) from Dyer & Nason (2004) mapped onto the spatial coordinates (longitude and latitude) of sampled populations. Population SB (from **Figure 3**) is highlighted for comparison. Edges in this topology are depicted in three categories: normal edges whose length is proportional to that expected under a model of isolation by distance (*gray*); edges whose length is longer than expected (known as extended edges; *red*), which are consistent with long-distance dispersal events; and edges whose length is shorter than expected (known as compressed edges; *green*), which indicate a reduced permeability of intervening landscape features.

Congruence: similarity in the structure of two or more graph topologies

the robustness of network structures, highlighting the importance not only of standing topological structures but also the relative stability of network structures under perturbation. Kutnjak et al. (2014) used a bootstrap approach to determine a similar measure of relative edge stability in the topology created from an alpine plant, although this approach was not applied to collections of topologies. Ranking relative edge sensitivity (or elasticity, as an analog to matrix analyses) has the potential to provide insights into the underlying structure of the topology. Various components of the networks may be differentially robust to population reduction or other demographic processes. This is a component of Population Graphs that needs additional investigation, and there exists a large body of literature about matrix-stability techniques that have been successfully used in life-history analyses and that could be applied to address stability issues within topologies.

Another interesting use of multi-topology comparison came from Widmer et al. (2012), who examined the phylogeographic structure of an epiphytic lichen fungus *Lobaria pulmonaria* and its algal symbiont, both of which were used to estimate Population Graph topologies and examined for congruence. Using Mantel tests to determine correlation in cGD between species, they showed a high degree of symmetry in the two species, which would be expected given the interconnectedness of their life histories. An analysis of symmetry in spatial genetic structure may provide valuable information for a broad range of studies. Coevolutionary hot spots are predicted to exist where there is the greatest asymmetry in gene flow in the interacting taxa (Gomulkiewicz et al. 2000). These areas of asymmetry correspond to Population Graph subcomponents that deviate in their patterns of connectivity. Being able to identify the spatial location of these coevolutionary hot spots by looking at neutral genetic covariance could prove to be a tremendous resource. I am not aware of any published studies that have a sufficient sampling density across populations that is adequate to identify asymmetries in gene flow between coevolving taxa other than those of Widmer et al. (2012), who were not interested in this particular question. Applied in a community context, there are still large gaps in understanding the extent to which co-occurring taxa have similarly responded either to previous demographic shifts or to ongoing gene flow, and how this has influenced spatial genetic structure. Does spatial connectivity for co-occurring plants pollinated by an overlapping suite of generalist pollinators translate into spatial genetic congruence? Although these kinds of questions are still largely unexplored, Population Graphs may make the answers easier to find.

4. PERSPECTIVES

Landscape genetics defines a subset of studies of population genetics that focus on genetic connectivity and the role that spatial and ecological features between sites have in influencing this connectivity (Dyer 2015). Although these particular sets of hypotheses are not new (e.g., Smouse & Wood 1987), the integration of tools largely derived from landscape ecology has increased the specificity with which we can define ecological, topological, and spatial variables. As a consequence, we are able to get a more fine-grained analysis of which features influence gene flow and, perhaps more importantly, their relative significance. However, almost all analytical developments that have bolstered interest in landscape genetics have focused on how we treat the independent variables. The methodological approaches we have used to manipulate and define the genetic aspects of these analyses—the actual answer to the analytical questions that we are using spatial and ecological data to predict—have largely languished in decades-old analytical metrics.

These metrics were derived specifically to estimate evolutionary processes at equilibrium and under ideal conditions (Neigel 1997), neither of which are appropriate standing assumptions for landscape- or many population-genetic studies. What is sorely needed is for population geneticists to embrace questions of landscape genetics and provide additional tools and techniques that may not be readily available or whose importance may not be currently appreciated

among practitioners of landscape genetics. Approaches examining parameters such as coancestry, consanguineous mating, inbreeding, and genetic-effective population size have already been developed. Yet these approaches seem to have been largely ignored by the landscape-genetics community, even though they provide linkages supported by theory about how landscape and ecological effects influence the structure of population genetics. Metrics such as the Bray–Curtis genetic distance have been embraced despite having little theoretical underpinning with respect to population-genetic expectations. When combined with more sophisticated methods for dealing with ecological- and spatial-predictor variables, the Population Graph framework seems able to provide insights that more traditional approaches cannot.

The theory behind the processes that create and maintain genetic covariance among populations is much more developed than those that describe how individual movement translates into population-genetic structure. Given the emphasis in the literature of landscape genetics on how ecological features influence the transient movement of individual animals (Dyer 2015), we are undoubtedly in store for a treasure trove of novel theory linking these two. Although landscape genetics may address only a subset of more common hypotheses about population genetics, uncovering the way in which ecological heterogeneity influences a population's structure can act to solidify connections between ecology and evolution sensu lato.

FUTURE ISSUES

1. Conditional genetic covariance is a symmetric property of a Population Graph. However, underlying networks of genetic connectivity are likely to contain regions of asymmetry (e.g., Fraser et al. 2004, Paz-Vinas et al. 2013). As an iterative approach, asymmetric gene flow can be partitioned out of Population Graphs (Dyer et al. 2010), although the presence and extent of the underlying asymmetry need to be known a priori. The development of methodologies that can identify the presence and relative impact of asymmetric gene flow within a Population Graph will provide information on the net flow of genetic material within the topology. Of particular interest are the application of asymmetric gene flow in studies on hybridization (Noutsos et al. 2014) and the magnitude of gene escape from crop plants (Papa & Gepts 2003).
2. In two cases thus far, Population Graphs have been used to quantify contemporary pollen movement (Dyer et al. 2012, DiLeo et al. 2014) by looking at the spatial structure of sampled pollen pools. Both isolation-by-resistance and gravity-model approaches have been applied to identify which forest features influence pollen-mediated gene flow. In both of these cases, remotely sensed data (obtained using LiDAR and hyperspectral imaging) have provided accurate characterizations of the intervening environment through which gene flow was occurring. The larger the spatial extent of our analyses, the more robust they will be at quantifying the relative permeability of individual landscape features.
3. Given the large volume of studies of landscape genetics that have been conducted on individuals rather than populations, extending the notions of individual-based Population Graphs may bring a corresponding increase in statistical precision to those studies, similar to what has been seen when applying conditional covariance to population-level isolation models. The current stumbling block is deciding how to determine how many loci are necessary to provide sufficiently robust estimates of within-individual variance (e.g., Landguth et al. 2012), and the consequences of mixing within-individual and interindividual variances as they pertain to the estimation of conditional covariance.

DISCLOSURE STATEMENT

The author is not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

ACKNOWLEDGMENTS

I thank Drs. Victoria Sork, John Nason, Ryan Garrick, Andrew Eckert, and Brian Verrelli for ongoing discussions on how Population Graphs can be beneficial to understanding population-genetic processes. The development of Population Graphs was supported by grants from the National Science Foundation (DEB-0543102 and DEB-0543102), as well as grants from the Virginia Commonwealth University Inger and Walter Rice Center for Environmental Life Sciences. This manuscript is the Virginia Commonwealth University Rice River Center Contribution #58.

LITERATURE CITED

Albert EM, Fortuna MA, Godoy JA, Bascompte J. 2013. Assessing the robustness of networks of spatial genetic variation. *Ecol. Lett.* 16:86–93

Baker SA, Dyer RJ. 2011. Invasion genetics of *Microstegium vimineum* (Poaceae) within the James River Basin of Virginia, USA. *Conserv. Genet.* 12:793–803

Cushman SA, McKelvey KS, Hayden J, Schwartz MK. 2006. Gene flow in complex landscapes: testing multiple models with causal modeling. *Am. Nat.* 168:2553–64

Dell'Acqua M, Zuccolo A, Tuna M, Gianfranceschi L, Pé ME. 2014. Targeting environmental adaptation in the monocot model *Brachypodium distachyon*: a multi-faceted approach. *BMC Genom.* 15:801

DiLeo MF, Siu JC, Rhodes MK, López-Villalobos A, Redwine A, et al. 2014. The gravity of pollination: integrating at-site features into spatial analyses of contemporary pollen movement. *Mol. Ecol.* 23:3345–61

Dupanloup I, Schneider S, Excoffier L. 2002. A simulated annealing approach to define the genetic structure of populations. *Mol. Ecol.* 11:2571–81

Dyer RJ. 2007. The evolution of genetic topologies. *Theor. Popul. Biol.* 71:71–78

Dyer RJ. 2015. Is there such a thing as landscape genetics? *Mol. Ecol.* 24:3518–28

Dyer RJ, Chan DM, Gardiakos VA, Meadows CA. 2012. Pollination networks: quantifying pollen pool covariance networks and the influence of intervening landscape on genetic connectivity in the North American understory tree, *Cornus florida* L. *Landsc. Ecol.* 27:239–51

Dyer RJ, Nason JD. 2004. Population Graphs: the graph theoretic shape of genetic structure. *Mol. Ecol.* 13:1713–28

Dyer RJ, Nason JD, Garrick RC. 2010. Landscape modelling of gene flow: improved power using conditional genetic distance derived from the topology of population networks. *Mol. Ecol.* 19:3746–59

Dyer RJ, Westfall RD, Sork VL, Smouse PE. 2004. Two-generation analysis of pollen flow across a landscape V: a stepwise approach for extracting factors contributing to pollen structure. *Heredity* 92:204–11

Evanno G, Regnaut S, Goudet J. 2005. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol. Ecol.* 14:2611–20

Excoffier L, Smouse PE, Quattro JM. 1992. Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics* 131:479–91

Falconer DR. 1981. *Introduction to Quantitative Genetics.* London: Longman. 2nd ed.

Falk DA, Holsinger K. 1991. *Genetics and Conservation of Rare Plants*. New York: Oxford Univ. Press

Fitzpatrick JM, Carlon DB, Lippe C, Robertson DR. 2011. The West Pacific diversity hotspot as a sink for new species? Population genetic insights from the Indo-Pacific parrotfish *Scarus rubroviolaceus*. *Mol. Ecol.* 20:219–34

Foley CJ, Holland JD. 2010. Do flying beetles perceive human-dominated landscapes as complex mosaics or binary patterns? *Landsc. Online* 16:1–18

Fortuna MA, Albaladejo RG, Fernández L, Aparicio A, Bascompte J. 2009. Networks of spatial genetic variation across species. *PNAS* 106:19044–49
Fraser DJ, Lippé C, Bernatchez L. 2004. Consequences of unequal population size, asymmetric gene flow, and sex-biased dispersal on population structure in brook charr (*Salvelinus fontinalis*). *Mol. Ecol.* 13:67–80
Garrick RC, Nason JD, Fernández-Manjarrés JF, Dyer RJ. 2013. Ecological co-associations influence species' response to past climatic change: an example from a Sonoran Desert bark beetle. *Mol. Ecol.* 22:3345–61
Garrick RC, Nason JD, Meadows CA, Dyer RJ. 2009. Not just vicariance: Phylogeography of a Sonoran Desert euphorb indicates a major role of range expansion along the Baja peninsula. *Mol. Ecol.* 18:1916–31
Giordano AR, Ridenhour BJ, Storfer A. 2007. The influence of altitude and topography on genetic structure in the long-toed salamander (*Ambystoma macrodactulym*). *Mol. Ecol.* 16:1625–37
Gomulkiewicz R, Thompson JN, Holt RD, Nuismer SL, Hochberg ME. 2000. Hot spots, cold spots, and the geographic mosaic theory of coevolution. *Am. Nat.* 156:156–74
Greve C, Gimnich F, Hutterer R, Misof B, Haase M. 2012. Radiating on oceanic islands: patterns and processes of speciation in the land snail genus *Theba* (Risso 1826). *PLOS ONE* 7:e34339
Hedrick PW. 2005. A standardized genetic differentiation measure. *Evolution* 59:1633–38
Herrera-Arroyo ML, Sork VL, González-Rodríguez, Rocha-Ramírez V, Vega E, Oyama K. 2013. Seed-mediated connectivity among fragmented populations of *Quercus castanea* (Fagaceae) in a Mexican landscape. *Am. J. Bot.* 100:1663–71
Jost L. 2008. G_{ST} and its relatives do not measure differentiation. *Mol. Ecol.* 17:4015–26
Kimura M, Weiss WH. 1964. The stepping stone model of genetic structure and the decrease of genetic correlation with distance. *Genetics* 49:561–76
Klütsch CFC, Dyer RJ, Misof B. 2011. Combining multiple analytical approaches for the identification of population structure and genetic delineation of two subspecies of the endemic Arabian burnet moth *Reissita simonyi* (Zygaenidae; Lepidoptera). *Conserv. Genet.* 13:21–37
Koen EL, Bowman J, Garroway CJ, Wilson PJ. 2013. The sensitivity of genetic connectivity measures to unsampled and under-sampled sites. *PLOS ONE* 8:e56204
Krafsur ES, Cummings MA, Endsley MA, Marquez JG, Nason JD. 2005. Geographic differentiation in the house fly estimated by microsatellite and mitochondrial variation. *J. Hered.* 96:502–12
Kutnjak DD, Kuttner M, Niketic, Dullinger S, Schøonswetter, Frajman B. 2014. Escaping to the summits: phylogeography and predicting range dynamics of *Cerastium dinaricum*, an endangered high mountain plant endemic to the western Balkan Peninsula. *Mol. Phylogenetics Evol.* 78:365–74
Landguth EL, Cushman SA, Schwartz MK, McKelvey KS, Murphy M, Luikart G. 2010. Quantifying the lag time to detect barriers in landscape genetics. *Mol. Ecol.* 19:4179–91
Landguth EL, Fedy BC, Oyler-McCance SJ, Gearey AL, Emel SL, et al. 2012. Effects of sample size, number of markers, and allelic richness on the detection of spatial genetic pattern. *Mol. Ecol. Resour.* 12:276–84
Li J, Coates BS, Kim KS, Bourguet D, Ponsard S, et al. 2014. The genetic structure of Asian corn borer, *Ostrinia furnacalis*, populations in China: haplotype variance in northern populations and potential impact on management of resistance to transgenic maize. *J. Hered.* 105:642–55
Li MH, Tapio I, Vilkki J, Ivanova Z, Kiselyova T, et al. 2007. The genetic structure of cattle populations (*Bos taurus*) in northern Eurasia and the neighbouring Near Eastern regions: implications for breeding strategies and conservation. *Mol. Ecol.* 16:3839–53
Malécot G. 1948. *Les mathématiques de l'hérédité*. Paris: Masson
Manel S, Schwartz MK, Luikart G, Taberlet P. 2003. Landscape genetics: combining landscape ecology and population genetics. *Trends Ecol. Evol.* 18:189–97
McRae BH. 2006. Isolation by resistance. *Evolution* 60:1551–61
McRae BH, Dickson BG, Keitt TH, Shah VB. 2008. Using circuit theory to model connectivity in ecology, evolution, and conservation. *Ecology* 89:2712–24
Milligan BG, Leebens-Mac J, Strand AE. 1994. Conservation genetics: beyond the maintenance of marker diversity. *Mol. Ecol.* 3:423–35
Murphy MA, Evans JS, Storfer A. 2010. Quantifying *Bufo boreas* connectivity in Yellowstone National Park with landscape genetics. *Ecology* 91:1129–37
Nei M. 1973. Analysis of gene diversity in subdivided populations. *PNAS* 70:3321–23

Neigel JE. 1997. A comparison of alternative strategies for estimating gene flow from genetic markers. *Ann. Rev. Ecol. Syst.* 28:105–28

Nosil P, Egan SP, Funk DJ. 2008. Heterogeneous genomic differentiation between walking-stick ecotypes: "isolation by adaptation" and multiple roles for divergent selection. *Evolution* 62:316–36

Noutsos C, Borevitz JO, Hodges SA. 2014. Gene flow between nascent species: genotypic and phenotypic differentiation within and between *Aquilegia formosa* and *A. pubescens*. *Mol. Ecol.* 23:5589–98

Pagán I, Holguín Á. 2013. Reconstructing the timing and dispersion routes of HIV-1 subtype B epidemics in the Caribbean and Central America. *PLOS ONE* 8:e69218

Papa R, Gepts P. 2003. Asymmetry of gene flow and differential geographic structure of molecular diversity in wild and domesticated common bean (*Phaseolus vulgaris* L.) from Mesoamerica. *Theor. Appl. Genet.* 106:239–50

Paz-Vinas I, Quéméré E, Chikhi L, Loot S, Blanchet S. 2013. The demographic history of populations experiencing asymmetric gene flow: combining simulated and empirical data. *Mol. Ecol.* 22:3279–91

Pinto N, Keitt TH. 2009. Beyond the least-cost path: evaluating corridor redundancy using a graph-theoretic approach. *Land. Ecol.* 24:253–66

Pritchard JK, Stephens M, Donnelly P. 2000. Inference of population structure using multilocus genotype data. *Genetics* 155:945–59

Pritchard JK, Wen X, Falush D. 2007. *Documentation for* structure *software: Version 2.2*. Tech. Doc., Univ. Chicago

Quintela M, Johansson MP, Kristjánsson, Barreiro R, Laurila A. 2014. AFLPs and mitochondrial haplotypes reveal local adaptation to extreme thermal environments in a freshwater gastropod. *PLOS ONE* 9:e101821

Schwartz MK, McKelvey KS. 2009. Why sampling scheme matters: the effect of sampling scheme on landscape genetic results. *Conserv. Gen.* 10:441–52

Schwartz MK, Mills LS, McKelvey K, Ruggiero LS, Allendorf FW. 2002. DNA reveals high dispersal synchronizing the population dynamics of Canada lynx. *Nature* 415:520–22

Slatkin M. 1993. Isolation by distance in equilibrium and nonequilibrium populations. *Evolution* 47:264–79

Slatkin M. 1995. A measure of population subdivision based on microsatellite allele frequencies. *Genetics* 139:457–62

Smouse PE, Dyer RJ, Westfall RD, Sork VL. 2001. Two-generation analysis of pollen flow across a landscape. I. Male gamete heterogeneity among females. *Evolution* 55:260–71

Smouse PE, Wood JW. 1987. The genetic demography of the Gainj of Papua New Guinea: functional models of migration and their genetic implications. In *Mammalian Dispersal Patterns: The Effects of Social Structure on Population Genetics*, ed. DB Chepko-Sade, ZT Halpin. Chicago: Chicago Univ. Press

Wang IJ, Glor RE, Losos JB. 2013. Quantifying the roles of ecology and geography in spatial genetic divergence. *Ecol. Lett.* 16:175–82

Wasserman S, Faust K. 2005. *Social Network Analysis: Methods and Applications*. Cambridge, UK: Cambridge Univ. Press

Wegier A, Piñeyro-Nelson A, Alarcón J, Gálvez-Mariscal A, Álvarez-Buylla ER, Pinero D. 2011. Recent long-distance transgene flow into wild populations conforms to historical patterns of gene flow in cotton (*Gossypium hirsutum*) at its centre of origin. *Mol. Ecol.* 20:4182–94

Widmer I, Del Grande F, Excoffier L, Holderegger R, Keller C, et al. 2012. European phylogeography of the epiphytic lichen fungus *Lobaria pulmonaria* and its green algal symbiont. *Mol. Ecol.* 21:5827–44

Whitlock M. 2011. G'_{ST} and D do not replace F_{ST}. *Mol. Ecol.* 20:1083–91

Wright S. 1931. Evolution in Mendelian populations. *Genetics* 16:97–159

Wright S. 1943. Isolation by distance. *Genetics* 28:114–38