Conserv Genet (2013) 14:253–261
DOI 10.1007/s10592-012-0391-5

RESEARCH ARTICLE

A conceptual framework for the spatial analysis of landscape genetic data

Helene H. Wagner · Marie-Josée Fortin

Received: 6 February 2012 / Accepted: 18 July 2012 / Published online: 28 July 2012

Abstract Understanding how landscape heterogeneity constrains gene flow and the spread of adaptive genetic variation is important for biological conservation given current global change. However, the integration of population genetics, landscape ecology and spatial statistics remains an interdisciplinary challenge at the levels of concepts and methods. We present a conceptual framework to relate the spatial distribution of genetic variation to the processes of gene flow and adaptation as regulated by spatial heterogeneity of the environment, while explicitly considering the spatial and temporal dynamics of landscapes, organisms and their genes. When selecting the appropriate analytical methods, it is necessary to consider the effects of multiple processes and the nature of population genetic data. Our framework relates key landscape genetics questions to four levels of analysis: (i) node-based methods, which model the spatial distribution of alleles at sampling locations (nodes) from local site characteristics; these methods are suitable for modeling adaptive genetic variation while accounting for the presence of spatial autocorrelation. (ii) Link-based methods, which model the probability of gene flow between two patches (link) and relate neutral molecular marker data to landscape heterogeneity; these methods are suitable for modeling neutral genetic variation but are subject to inferential problems, which may be alleviated by reducing links based on a network model of the population. (iii) Neighborhood-based methods, which model the connectivity of a focal patch with all other patches in its local neighborhood; these methods provide a link to metapopulation theory and landscape connectivity modeling and may allow the integration of node- and link-based information, but applications in landscape genetics are still limited. (iv) Boundary-based methods, which delineate genetically homogeneous populations and infer the location of genetic boundaries; these methods are suitable for testing for barrier effects of landscape features in a hypothesis-testing framework. We conclude that the power to detect the effect of landscape heterogeneity on the spatial distribution of genetic variation can be increased by explicit consideration of underlying assumptions and choice of an appropriate analytical approach depending on the research question.

Keywords Spatial statistics · Neutral genetic variation · Adaptive genetic variation · Node analyses · Link analyses · Neighborhood analyses · Boundary analyses

H. H. Wagner (✉)
Department of Ecology and Evolutionary Biology, University of Toronto, 3359 Mississauga Road, Mississauga, ON L5L 1C6, Canada
e-mail: helene.wagner@utoronto.ca

M.-J. Fortin
Department of Ecology and Evolutionary Biology, University of Toronto, 25 Harbord St, Toronto, ON M5S 3G5, Canada
e-mail: mariejosee.fortin@utoronto.ca

Introduction

Landscape genetics investigates neutral or adaptive genetic variation in spatially heterogeneous landscapes (Holderegger and Wagner 2008; Storfer et al. 2007; Storfer et al. 2010). Landscape genetic studies may have either (i) an evolutionary focus aiming to infer the effects of microevolutionary processes on genetic data while accounting for landscape effects such as isolation by distance or matrix resistance (Storfer et al. 2010), or (ii) an ecological focus aiming to quantify and test landscape effects on organism

dispersal that potentially results in gene flow and spread of adaptive genes (Manel et al. 2010; Spear et al. 2010). In the former case, spatial and landscape effects are largely a nuisance for statistical hypothesis testing (Diniz-Filho et al. 2009; Legendre and Legendre 1998), whereas in the latter case, the same effects may be the primary interest (Fortin and Dale 2005). In both cases, landscape genetic studies are faced with conceptual and statistical challenges related to inferring processes, which are not directly observable, from the resulting spatial genetic structure (Anderson et al. 2010). For instance, the observed population genetic structure represents a single realization of the underlying process, and independent replication at the landscape scale is difficult to achieve (Wagner and Fortin 2005). The same pattern may result, however, from different alternative processes, e.g., different dispersal scenarios could explain the same population genetic structure or different series of historical events can generate similar spatial patterns of genetic structure (Epperson 2003). Observed genetic patterns are likely the result of several, potentially interacting processes, such as habitat change (hereafter landscape change), population dynamics, habitat selection, dispersal mode, matrix resistance to dispersal, mating and reproduction system, mutation, selection and genetic drift (Fig. 1). Clearly, a single study is unlikely to address this full complexity but will identify which processes to prioritize while making implicit or explicit assumptions on others.

In addition, methods for the spatial analysis of landscape heterogeneity (Fortin and Dale 2005; Wagner and Fortin 2005) and of population genetic structure (Epperson 2003; Slatkin and Arter 1991; Sokal 1979) have largely been developed independently, drawing on different theories and analytical methods, and thus lack a coherent framework for their joint analysis. In order to apply and further develop methods for relating genetic data to landscape features, landscape ecologists and spatial statisticians need to understand the specific nature of population genetic data and population genetic methods, and population geneticists need to understand the nature of landscape ecological data and spatial statistical methods.

Here, we stress how the choice of analysis method should be guided first by the research question, then by the methods that can address the research question, and finally by the nature of the data at hand. Consequently, this paper provides a framework for categorizing research questions based on the processes considered (Fig. 1, Table 1). We then classify statistical methods that relate genetic data to landscape data according to four analysis levels that imply different data models (Fig. 2, Table 2). We thus aim to: (1) clarify where and how spatial and temporal processes matter in landscape genetics, and (2) point out common misconceptions and prevent misunderstandings of the complexity of issues that arise when moving from one field to another or in interdisciplinary research collaborations between population geneticists, landscape ecologists, spatial statisticians and other experts from related fields (Balkenhol et al. 2009a).

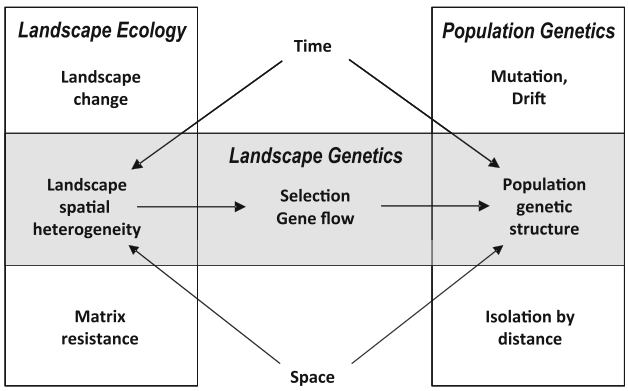

Fig. 1 Conceptual model of how space and time affect relevant processes in the fields of landscape ecology (*left box*) and population genetics (*right box*). The primary focus of landscape genetics is how landscape spatial and temporal heterogeneity shape the spatial distribution of genetic variation by modifying environmental selection and gene flow (*grey box*). By linking concepts from landscape ecology and population genetics, landscape genetics embraces both spatial and temporal dimensions and needs to consider the relevant ecological and evolutionary processes, either explicitly or through assumptions

Processes to consider

Most ecological and evolutionary processes are inherently spatially dynamic. Therefore, landscape genetic studies need to consider the effects of processes both in space and time (Fig. 1).

Landscape change occurs due to various processes that create different types of spatial dynamics: (i) natural or anthropogenic disturbance may temporarily or permanently alter the amount, quality and connectivity of habitat. Ecosystems and their organisms may be adapted to natural disturbances, such as wildfires or insect outbreaks, and many natural disturbance regimes result in a shifting mosaic with dynamic equilibrium at a larger spatial scale (Folke et al. 2004; Holling 1992). However, anthropogenic disturbances may interact with natural disturbances in complex ways, altering ecological system dynamics. (ii) Succession, e.g., triggered by a disturbance event, changes species' composition and species' spatial distribution over time. (iii) Climate change will gradually alter habitat quality, though extreme events such as drought may trigger abrupt changes. (iv) Land-use change typically occurs as a discrete change from one land-use type to another, though changes in land-use intensity such as fertilizer application may be more gradual.

Table 1 Examples of landscape genetics research questions and common assumptions

Micro-evolutionary processes	Analysis context	*Static landscape* Assumption: mutation—drift—migration equilibrium	*Changing landscape* Assumption: contemporary gene flow is measured
Gene flow Assumption: markers are not affected by selection	*Ecological*: landscape connectivity	What is the resistance of different land-cover types to the dispersal of a species of interest?	How does a new barrier affect the dispersal of a species of concern?
	Evolutionary: genetic connectivity	Is there enough gene flow to maintain a viable population?	Does a new barrier disrupt gene flow for a species of concern?
Selection Assumption: loci are under selection or linked to adaptive genes	*Ecological*: adaptation to selective environment	Which loci are potentially linked with adaptive genes?	Does a change in habitat quality trigger rapid evolutionary change?
	Evolutionary: cohesion of a spatially structured population	How does gene flow affect the frequency of adaptable genes?	How do beneficial mutations spread across a spatially structured population?

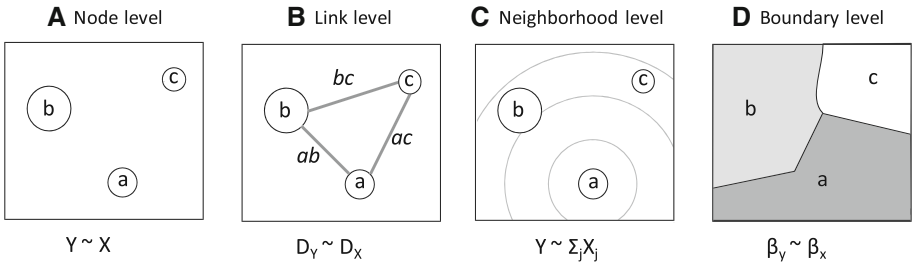

Fig. 2 Four analysis levels at which landscape genetic data can be analyzed: **A** at the node level, the alleles (**Y**) at each sampling location *a*, *b* or *c* are related to environmental conditions or landscape features (**X**) observed at the same location. Alternatively, **Y** may represent an aggregate measure of genetic diversity such as allelic richness. Nodes of different sizes refer either to different genetic values or habitat patch sizes. **B** at the link level, genetic distance $\mathbf{D_Y}$ between pairs of sampling locations *ab*, *ac* and *bc* is related to distance-based landscape data $\mathbf{D_X}$ describing the intervening matrix along each link. **C** at the neighborhood level, the alleles **Y** (or diversity measure, see above) at sampling location *a* are related to a patch-level connectivity measure $\Sigma_j X_j$ that quantifies local landscape context (see text). **D** at the boundary level, spatially contiguous, discrete populations *a*, *b* and *c* are inferred from genetic data and overlaid with landscape features to identify barriers to gene flow. This may be achieved by relating spatial rates of change in genetic data, β_Y, to spatial rates of change in landscape predictors β_X. Alternatively, β may denote between-cluster components of variance in **X** or **Y** (see text)

Table 2 Examples of spatial analysis methods that can be used to describe or model spatial patterns in one or more response variables Y at each of the four analysis levels

	Spatial pattern description of Y	Spatial modeling of Y
Node	Global spatial autocorrelation (Moran's *I*, Geary's *c*, variography)	Spatial regression
	Spatial interpolation (Kriging)	Gravity model
	sPCA	Partial direct ordination using MEM
Link	Mantel correlogram	Partial Mantel tests
		Partial regression of distance matrices
Neighborhood	Local spatial autocorrelation (Moran's *I*, Geary's *c*)	Geographical weighted regression
Boundary	Boundary detection (Wombling, Monmonier)	Correlation between boundaries (Boundary overlap statistics) or clusters with environmental/ landscape features (POPS)
	Spatial Bayesian clustering (TESS)	

Similarly, key micro-evolutionary processes contribute to genetic diversity at neutral loci or adaptive genes (Fig. 1). Even in a homogeneous environment, *mutation* will over time change genetic variation, *genetic drift* will reduce genetic diversity due to the stochastic loss of alleles from one generation to the next, and restricted dispersal can create spatial genetic structure (*isolation by distance, IBD*; Lande 1991; Wright 1943, 1948). Other micro-evolutionary

processes are directly affected by landscape spatial heterogeneity: *environmental selection* depends on local habitat quality (Garant et al. 2007), whereas *dispersal* (movement from natal site to new site) and resulting *migration* (dispersal to new site followed by reproduction) may depend on characteristics of the natal or new site as well as the physical distance between sites (IBD) and the nature of the intervening landscape (*matrix resistance*), including the presence of complete barriers to movement or more gradual differences in traversability or mortality (Spear et al. 2010).

In ecological and evolutionary processes, both spatial and temporal constraints need to be considered (Fig. 1). For instance, the study of selection in spatially heterogeneous landscapes may be compromised by spatial autocorrelation introduced by restricted dispersal due to isolation by distance or matrix resistance (Epperson 2003; Landguth et al. 2010). The fact that genetic processes operate within time units of generations may introduce important time lags in the response of spatial genetic structure to landscape change. Such time lags, referred to as ghost landscape effects (Anderson et al. 2010), need to be differentiated from the contemporary landscape effects. Dyer et al. (2010) showed that removing the effect of known phylogeographic history may significantly improve our ability to assess landscape effects. Our conceptual framework (Fig. 1) may help researchers identify the main processes of interest for their study, those processes that are accounted for in the study design and statistical analysis, and any processes that are not studied and for which the underlying assumptions should be made explicit. For instance, a specific study focusing on landscape resistance to dispersal and gene flow would explicitly investigate the processes of gene flow and landscape resistance while accounting for IBD in the statistical analysis. It may ignore landscape change and thus assume that landscape has been constant over many generations. It may further ignore evolutionary processes of selection, mutation and drift, with underlying assumptions that the molecular markers studied are neutral and not linked to genes under selection, mutation is negligible at the temporal scale studied, and local populations have not undergone recent bottlenecks.

Research questions

Landscape genetic studies may address a wide range of hypotheses and research questions, depending on the focal micro-evolutionary process (gene flow vs. selection), the ecological or evolutionary perspective (Storfer et al. 2010), and whether the intent is to study equilibrium conditions or transient dynamics after landscape change (Table 1). Studies of gene flow commonly assume that the molecular markers used are not affected by selection (Holderegger and Wagner 2008). Studies of selection may rely on known adaptive genes (e.g., in model organisms) or the identification of outlier loci as putatively adaptive loci or assumedly neutral markers that may be linked to unknown adaptive genes due to proximity on the genome (Holderegger and Wagner 2008; Holderegger et al. 2010).

The assumption that the amount, quality and connectivity of habitat remained constant for a period long enough to reach an equilibrium between mutation, drift and migration may rarely be met in real landscapes. Violations of underlying assumptions (such as Hardy–Weinberg equilibrium) may compromise the validity of interpretation and generalization. The problem thus is how to assess and account for the potential effect of such violations on the results. Other studies explicitly focus on changes in dispersal, gene flow or adaptation in response to a known landscape change or manipulation of amount, quality or connectivity of habitat, such as the adding or removal of a putative barrier or the loss or addition of corridors and stepping stone habitat. In such cases, it is important to measure contemporary gene flow that reflects current connectivity, which may be contrasted with estimates of gene flow prior to the landscape change (Anderson et al. 2010).

In landscape genetics, it is often difficult to tease apart ecological and evolutionary perspectives. The distinction is relevant, though, as misunderstandings may arise if one researcher implicitly takes an ecological perspective, another an evolutionary perspective, on what may appear to be the same issue. When studying gene flow, an ecologist may want to use genetic data as a proxy to infer dispersal rates among habitat patches with the goal of testing hypotheses on matrix resistance that ultimately may lead to effective planning of habitat networks (Méndez et al. 2011). Such studies often implicitly assume that all dispersal events will (or are equally likely to) result in gene flow, but few studies explicitly investigate this discrepancy (Greenwald 2010).

From an evolutionary perspective, we are directly interested in inferring gene flow as a measure of genetic connectivity, which may lead to the identification of populations at risk of loss of viability and ultimately extinction. Small isolated populations are likely to experience higher levels of inbreeding or Allee effects, which may result in depression of population fitness, and reduction of genetic diversity through drift, which may reduce the potential for evolutionary adaptation (Frankham 2005; Spielman et al. 2004). Such assessment of genetic connectivity often assumes that new alleles arise in a population by immigration, not by mutation. Landscape genetics however typically relies on markers with relatively high mutation rates, and estimates of mutation rates for specific marker types carry high levels of uncertainty. The

assumption that mutation may be ignored at the population level may not hold, as illustrated by a recent study that found a strong signal of mutation for both symbionts (fungus and alga) within replicate populations of the lichen *Lobaria pulmonaria* (DalGrande et al. 2012).

Currently, most landscape genetic studies of adaptive variation take an ecological perspective (Manel et al. 2010), such as aiming to identify outlier loci. A major challenge in this context is accounting for spatial autocorrelation due to restricted dispersal or landscape spatial heterogeneity. An evolutionary perspective may focus on the cohesion of spatially structured populations in the context of speciation (typically assuming a static landscape) or the spatial dynamics of rapid evolutionary change in a changing environment (Holderegger and Wagner 2008). In the latter case, assumptions may be needed regarding the probability that the same beneficial mutation has multiple independent origins. Note that the time lag in the genetic signal is related to the rate of landscape change as compared to generation time.

Approaches linking genetic and landscape data

Our framework is based on four analysis levels for relating neutral or adaptive genetic data to landscape pattern (Fig. 2), depending on the research questions. These analysis levels can be characterized by adopting terminology from graph theory, where *nodes* refer to habitat patches and *links* to lines that connect any two patches (Dale and Fortin 2010). The four analysis levels imply different data models, and in some cases it may be appropriate to convert data from one data model to another to best match the analysis to the research question.

Node-based methods (Fig. 2a) relate the presence of adaptive genes or the genetic diversity of local populations to environmental site conditions at sampling locations or to patch attributes such as area or age. They thus address the question of what determines the presence and abundance of alleles, or the genetic diversity, at a spatial location. Note that the question relating to the presence and abundance of alleles is only meaningful for adaptive genetic variation, i.e., genes under selection or loci linked to such genes.

At the node level, the analysis of allele frequency data from n individuals, or predefined local populations (e.g., demes, home ranges, nodes), is typically based on a response matrix $\mathbf{Y}$ with n rows and m columns representing alleles. The explanatory matrix $\mathbf{X}$ with n rows and p columns represents the environmental and landscape predictors. Multivariate statistical methods such as ordination (e.g., redundancy analysis RDA, canonical correspondence analysis CCA; Legendre and Legendre 1998) can be used to predict the frequency of each allele y from local site conditions x. These methods can be used to identify outlier loci that are empirically associated with landscape predictors. IBD may create spatial autocorrelation in the residuals, thus violating the assumption of independent error of non-spatial linear models. If statistical tests (e.g., Moran's I) indicate autocorrelated residuals, the analysis should be modified to account for spatial autocorrelation due to IBD (Dray et al. in press; Wagner and Fortin 2005). This may be accomplished by including an additional set of predictors $\mathbf{W}$ that model spatial structure in the data at multiple scales based on the spatial locations of sampling points. The matrix $\mathbf{W}$ can be used to partial out spatial variation by including it in the linear model as $\mathbf{Y} \sim \mathbf{X} + \mathbf{W}$. Methods for creating matrix $\mathbf{W}$ has been developed as Principal Components of Neighborhood Matrices (PCNM; Borcard et al. 2011) and generalized to Moran's Eigenvector Map (MEM; Dray et al. 2006) analysis. We can thus partial out all spatial variation when assessing the response of allele frequencies $\mathbf{Y}$ to local site conditions $\mathbf{X}$ (Manel et al. 2010). However, this blanket approach may also eliminate spatial variation unrelated to gene flow (Lichstein et al. 2002; Wagner and Fortin 2005). As an alternative, spatial regression could be used to explicitly build isolation by distance into the model (Diniz-Filho et al. 2009) (Table 2).

Link-based methods (Fig. 2b) relate pair-wise genetic distance between individuals or demes to their landscape distance (e.g., geographic distance, cost distance, presence or number of barriers) hypothesized to be related to the probability of dispersal and migration. They thus address the question of how likely gene flow is between two patches. This analysis level may be particularly relevant for assessing genetic connectivity, testing hypotheses on landscape resistance, and identifying corridors for conservation applications.

At the link level, the analysis of allele frequency data from n individuals, or demes, is typically based on a $n \times n$ matrix $\mathbf{D_Y}$ of pairwise genetic distances, and a set of p matrices $\mathbf{D_{Xp}}$ of size $n \times n$. Each matrix $\mathbf{D_{Xp}}$ provides a measure of landscape distance d_{ij} between two sampling locations i and j. In the simplest case, d_{ij} is equal to the geographic distance, representing isolation by distance. The effect of a barrier (e.g., mountain, river, road) can be represented with a binary matrix. Similarly, the intervening landscape can be described in terms of the presence (binary), absolute distance (e.g., in km) or relative distance (in percent of total distance) of each cover type p that an organism would have to traverse along a transect between sampling locations i and j (e.g., Angelone et al. 2011). Alternatively, d_{ij} may represent a potential cumulative cost for an organism to traverse the intervening landscape between sampling locations i and j. Cost values of different cover types may be derived from field data on organism movement behavior, from expert

opinion, or through optimization given the genetic data (Spear et al. 2010). Cumulative cost d_{ij} is commonly calculated either (i) along the shortest physical distance (transect) between locations i and j, (ii) along a corridor of specified width, (iii) along the least-cost path, or (iv) as total resistance (McRae et al. 2008) integrating over all possible paths (Spear et al. 2010).

Mantel tests have been used to test the association between two distance matrices $\mathbf{D_Y}$ and $\mathbf{D_{X1}}$, and partial Mantel tests allow accounting for a third matrix $\mathbf{D_{X2}}$. These methods evaluate the (linear or rank) correlation between vectors of pairwise distances d_{ij}. Appropriate permutation tests need to be used for significance testing to account for the inflated sample size, as each vector has length $n(n-1)/2$ (Legendre and Fortin 2010). (Partial) Mantel tests have been shown to have low statistical power in general (Legendre and Fortin 2010), thus likely to fail to detect an effect especially with small data sets. On the other hand, landscape genetic studies often experience the opposite problem that many candidate models show statistically significant effects and criteria are needed to identify the most important factors (Cushman and Landguth 2010; Cushman et al. 2006). Multiple regression of distance matrices (Legendre and Legendre 1998; Legendre and Fortin 2010) rephrases the problem from correlation to regression and allows modeling of the response matrix $\mathbf{D_Y}$ by a number of matrices $\mathbf{D_{X1}}$, $\mathbf{D_{X2}}$, …, $\mathbf{D_{Xp}}$ simultaneously. The power of link-based analysis may be improved by reducing the full distance matrix to a subset of links, either based on generic spatial graph models (e.g., planar graph that connects only adjacent sampling locations; Dale and Fortin 2010) or using population graphs based on conditional genetic distance (Dyer et al. 2010; Garroway et al. 2011). More research is needed to develop and test approaches at model selection, e.g., using criteria such as AIC or BIC (Ward 2008) taking into account the inflated sample size in regression of distance matrices as well as inferential problems related to spatial autocorrelation in predictor and response variables, which in turn affect the effective sample size.

Neighborhood-based methods (Fig. 2c) relate genetic diversity of demes, or their genetic differentiation, to attributes of the local landscape context, either within a certain radius around the sampling location or as a function of the neighboring patches, with weights inversely proportional to their landscape distance scaled by the dispersal ability of the organism. They thus address the question of how connected each patch is to all nearby patches, and how such connectivity affects genetic diversity or differentiation at the level of demes (Keyghobadi et al. 2005). Such neighborhood-level analysis is compatible with a patch-based metapopulation perspective and often uses a patch connectivity model S_i (Balkenhol et al. 2009b; Bulman et al. 2007; James et al. 2011) to quantify local landscape context in a node-based connectivity measure. While link-based methods model connectivity between pairs of nodes, neighborhood-level analysis integrates connectivity for a focal patch across all other patches in its neighborhood. Here, a $n \times m$ matrix $\mathbf{Y}$ of allele frequencies, or a single vector y of an aggregate measure such of genetic diversity or differentiation, is related to a $n \times p$ matrix $\mathbf{S}$ of patch connectivity measures, where each variable S_p contains the connectivity value for each patch i with respect to landscape predictor p. Each patch connectivity measure may be calculated for a single environmental factor, or for a resistance surface that integrates multiple factors (Balkenhol et al. 2009b; Foll and Gaggiotti 2006). Note that summarizing landscape features, e.g., road density or the proportions of different habitat types, within a given radius around each sampling location, irrespective of the other sampling locations, does not provide connectivity measures $\mathbf{S}$ but node-based variables $\mathbf{X}$ that quantify the local landscape context.

Technically, this type of analysis brings information about the pairwise links back into a node-based data structure. This means that matrix $\mathbf{S}$, containing variables modeling dispersal from neighboring patches (potentially affecting gene flow), may be combined with a matrix $\mathbf{X}$ that contains additional variables relating to local site conditions (potentially affecting selection) in the same analytical framework. It is important to keep in mind what exactly is being modeled. In the multivariate case, allele frequencies in matrix $\mathbf{Y}$ are predicted by patch connectivity. For instance, this could mean that we predict the presence of an allele from an overall patch connectivity measure or from the amount of woodland cover in the local landscape around the patch. Although conceptually questionable for neutral markers, simulation results indicate that such neighborhood-based methods may perform comparably well under a range of landscape genetic scenarios (Balkenhol et al. 2009b). More research is needed to clarify the applicability of the approach for different research questions, including the application of distance-based RDA where $\mathbf{Y}$ is replaced by a genetic distance matrix $\mathbf{D}_y$. Gravity models provide a promising avenue for the joint modeling of node- and link-related processes (Murphy et al. 2010).

Boundary-based methods (Fig. 2d), which include boundary detection algorithms and spatial Bayesian clustering methods, aim to delineate discrete or admixed populations in space (Barbujani et al. 1989; Safner et al. 2011). The determination of whether individual genetic data ($\mathbf{Y}$) collected at different locations stem from a panmictic population or from several geographically and genetically distinct populations is an important goal in conservation genetics (Frankham 2009). When spatially distinct populations are identified based on genetic data, the next step in

boundary-level methods consists in determining which landscape features separate them. Boundary-based methods thus address the question of what landscape features constitute a barrier to gene flow.

Spatially distinct populations can be delineated using boundary detection methods such as Monmonier's algorithm (Monmonier 1973) or Womble's bilinear algorithm (Barbujani et al. 1989). In essence, using quantitative data (e.g., gene frequencies, environmental data) these algorithms identify boundaries between adjacent locations (based on Delaunay links) as highest rates of change β. Hence for each link d_{ij} between directly adjacent sampling locations i and j (nodes or grid cells), a rate of change value β_Y indicates how likely the link is to cross a genetic boundary. This can be reduced to a binary classification of links as boundary or non-boundary, either based on significance testing or an arbitrary threshold value. Similarly, boundaries in environmental conditions can be identified by assessing spatial rate of change β_X.

Bayesian clustering methods (Dawson and Belkhir 2001) without genetic admixture result in a categorical vector y of length n that specifies for each observation (individual or deme) to which of k discrete genetic populations it has been assigned. If the method allows admixture, the response is a $n \times k$ matrix $\mathbf{M}$ with probabilities of membership of each of n observations for each of k genetic populations, with row sums adding to 1. The presence of isolation by distance may lead to spurious boundary detection (Safner et al. 2011). Spatial Bayesian clustering as proposed by François and Durand (2010) adds a spatial graph constraint to determine spatial cluster memberships.

Once genetic boundaries, or spatial cluster memberships, are identified, the next step is to relate them to environmental or landscape features. This can be achieved by testing the degree of spatial overlap (Jacquez et al. 2000; Oden et al. 1993) between genetic boundaries and landscape features ($\mathbf{X}$) hypothesized to restrict dispersal. The spatial coincidence between the genetic (β_Y) and landscape boundaries (β_X) can be assessed using boundary overlap statistics (Barbujani and Sokal 1991; Fortin et al. 1996; Oden et al. 1993; St-Louis et al. 2004) to test specific hypotheses about landscape effects on the genetic data. Alternatively, the relationship between the cluster memberships and environmental conditions $\mathbf{X}$ can be tested using ancestry distribution models as implemented in POPS software (Prediction of Population genetic Structure; Durand et al. 2009; Jay 2011).

Conclusions

Landscape genetics research questions require analysis at different spatio-temporal scales and thus different data models, analysis levels and statistical methods (Table 2). Node-based methods are the natural choice for studying selection, link-based methods for studying gene flow. The barrier effect of new landscape features may best be detected by testing the overlap between genetic boundaries in contemporary gene flow assessed before and after establishment of the hypothesized barrier. Novel questions relating to the spread of adaptive variation in changing landscapes will require methods that integrate node- and link-based analysis, for which neighbourhood-level analysis may be best suited. Statistically sound and powerful methods are needed for each analysis level, as questions are not easily transferable to a different analysis level.

The presence of multiple processes, and the absence of equilibrium conditions, may compromise many methods. Landscape genetics needs a true integration of population genetic and ecological analysis methods and the development of new methods that take into account the complexity of landscape genetic data.

Acknowledgments This work resulted from a Distributed Graduate Seminar (Developing Best Practices for Testing Landscape Effects on Gene Flow), conducted through the National Center for Ecological Analysis and Synthesis, a center funded by the National Science Foundation grant #EF-0553768, the University of California, Santa Barbara, and the State of California. It was also supported by NSERC discovery grants to HHW and MJF. We thank Michelle DiLeo and Yessica Rico for their thoughtful comments.

References

Anderson CD, Epperson BK, Fortin MJ, Holderegger R, James PMA, Rosenber MS, Sribner KT, Spear SF (2010) Considering spatial and temporal scale in landscape-genetic studies of gene flow. Mol Ecol 19:3565–3575

Angelone S, Kienast F, Holderegger R (2011) Where movement happens: scale-dependent landscape effects on genetic differentiation in the European tree frog. Ecography 34:714–722

Balkenhol N, Gugerli F, Cushman SA, Waits L, Coulon A, Arntzen J, Holderegger R, Wagner HH (2009a) Identifying future research needs in landscape genetics: where to from here? Landsc Ecol 24:455–463

Balkenhol N, Waits LP, Dezzani RJ (2009b) Statistical approaches in landscape genetics: an evaluation of methods for linking landscape and genetic data. Ecography 32:818–830

Barbujani G, Sokal RR (1991) Genetic population structure of Italy. II. Physical and cultural barriers to gene flow. Am J Hum Genet 48:398–411

Barbujani G, Oden NL, Sokal RR (1989) Detecting regions of abrupt change in maps of biological variables. Syst Zool 38:376–389

Borcard D, Gillet F, Legendre P (2011) Numerical ecology with R. Springer, New York

Bulman CR, Wilson RJ, Holt AR, Bravo LG, Early RI, Warren MS, Thomas CD (2007) Minimum viable metapopulation size, extinction debt, and the conservation of a declining species. Ecol Appl 17:1460–1473

Cushman SA, Landguth EL (2010) Spurious correlations and inference in landscape genetics. Mol Ecol 19:3592–3602

Cushman S, McKelvey K, Hayden J, Schwartz M (2006) Gene flow in complex landscapes: testing multiple hypotheses with causal modeling. Am Nat 168:486–499

Dale MRT, Fortin MJ (2010) From graphs to spatial graphs. Annu Rev Ecol Evol Syst 41:21–38

DalGrande F, Widmer I, Wagner HH, Scheidegger C (2012) Vertical and horizontal photobiont transmission within populations of a lichen symbiosis. Mol Ecol. doi:10.1111/j.1365-294X.2012.05482.x

Dawson KJ, Belkhir K (2001) A Bayesian approach to the identification of panmictic populations and the assignment of individuals. Genet Res 78:59–77

Diniz-Filho JAF, Nabout JC, Telles MPdC, Soares TN, Rangel TFLVB (2009) A review of techniques for spatial modeling in geographical, conservation and landscape genetics. Genet Mol Biol 32:203–211

Dray S, Pélissier R, Couteron P, Fortin MJ, Legendre P, Peres-Neto PR, Bellier E, Bivand R, Blanchet FG, De Cáceres M, Dufour A, Heegard E, Jombart T, Munoz F, Oksanen J, Thioulouse J, Wagner HH. Community ecology in the age of multivariate multiscale spatial analysis. Ecol Monogr. doi:10.1890/11-1183.1

Dray S, Legendre P, Peres-Neto PR (2006) Spatial modelling: a comprehensive framework for principal coordinate analysis of neighbour matrices (PCNM). Ecol Model 196:483–493

Durand E, Jay F, Gaggiotti OE, François O (2009) Spatial inference of admixture proportions and secondary contact zones. Mol Biol Evol 26:1963–1973

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 19:3746–3759

Epperson BK (2003) Geographical genetics. Princeton University Press, Princeton

Folke C, Carpenter S, Walker B, Scheffer M, Elmqvist T, Gunderson L, Holling CS (2004) Regime shifts, resilience, and biodiversity in ecosystem management. Annu Rev Ecol Evol Syst 35:557–581

Foll M, Gaggiotti OE (2006) Identifying the environmental factors that determine the genetic structure of populations. Genetics 174:875–891

Fortin MJ, Dale MRT (2005) Spatial analysis: a guide for ecologists. Cambridge University Press, New York

Fortin MJ, Drapeau P, Jacquez GM (1996) Quantification of the spatial co-occurrences of ecological boundaries. Oikos 77:51–60

François O, Durand E (2010) Spatially explicit Bayesian clustering models in population genetics. Mol Ecol Resour 10:773–784

Frankham R (2005) Genetics and extinction. Biol Conserv 126:131–140

Frankham R (2009) Introduction to conservation genetics. Cambridge University Press, New York

Garant D, Kruuk LEB, McCleery RH, Sheldon BC (2007) The effects of environmental heterogeneity on multivariate selection on reproductive traits in female great tits. Evolution 61:1546–1559

Garroway CJ, Bowman J, Wilson PJ (2011) Using a genetic network to parameterize a landscape resistance surface for fishers, Martes pennanti. Mol Ecol 20:3978–3988

Greenwald KR (2010) Genetic data in population viability analysis: case studies with ambystomatid salamanders. Anim Conserv 13:115–122

Holderegger R, Wagner HH (2008) Landscape genetics. Bioscience 58:199–207

Holderegger R, Buehler D, Gugerli F, Manel S (2010) Landscape genetics of plants. Trends Plant Sci 15:675–683

Holling CS (1992) Cross-scale morphology, geometry, and dynamics of ecosystems. Ecol Monogr 62:447–502

Jacquez GM, Maruca S, Fortin MJ (2000) From fields to objects: a review of geographic boundary analysis. J Geogr Syst 2:221–241

James PMA, Coltman DW, Murray BW, Hamelin RC, Sperling FAH (2011) Spatial genetic structure of a symbiotic beetle-fungal system: toward multi-taxa integrated landscape genetics. PLoS One 6:e25359

Jay F (2011) PoPS: prediction of population genetic structure—program documentation and tutorial. University Joseph Fourier, Grenoble

Keyghobadi N, Roland J, Matter SF, Strobeck C (2005) Among- and within-patch components of genetic diversity respond at different rates to habitat fragmentation: an empirical demonstration. Proc Biol Sci 272:553–560

Lande R (1991) Isolation by distance in a quantitative trait. Genetics 128:443–452

Landguth EL, Cushman S, Murphy MA, Luikart G (2010) Relationships between migration rates and landscape resistance assessed using individual-based simulations. Mol Ecol Resour 10:854–862

Legendre P, Fortin MJ (2010) Comparison of the Mantel test and alternative approaches for detecting complex multivariate relationships in the spatial analysis of genetic data. Mol Ecol Resour 10:831–844

Legendre P, Legendre L (1998) Numerical ecology. Elsevier, New York

Lichstein JW, Simons TR, Shriner SA, Franzreb KE (2002) Spatial autocorrelation and autoregressive models in ecology. Ecol Monogr 72:445–463

Manel S, Joost S, Epperson BK, Holderegger R, Storfer A, Rosenberg MS, Scribner KT, Bonin A, Fortin MJ (2010) Perspectives on the use of landscape genetics to detect genetic adaptive variation in the field. Mol Ecol 19:3760–3772

McRae BH, Dickson BG, Keitt TH, Shah VB (2008) Using circuit theory to model connectivity in ecology, evolution, and conservation. Ecology 89:2712–2724

Méndez M, Tella JL, Godoy JA (2011) Restricted gene flow and genetic drift in recently fragmented populations of an endangered steppe bird. Biol Conserv 144:2615–2622

Monmonier MS (1973) Maximum-difference barriers: an alternative numerical regionalization method. Geogr Anal 5:245–261

Murphy MA, Dezzani RJ, Pilliod DS, Storfer A (2010) Landscape genetics of high mountain frog metapopulations. Mol Ecol 19:3634–3649

Oden NL, Sokal R, Fortin MJ, Goebl H (1993) Categorical wombling: detecting regions of significant change in spatially located categorical variables. Geogr Anal 25:315–336

Safner T, Miller MP, McRae BH, Fortin MJ, Manel S (2011) Comparison of bayesian clustering and edge detection methods for inferring boundaries in landscape genetics. Int J Mol Sci 12:865–889

Slatkin M, Arter HE (1991) Spatial autocorrelation methods in population genetics. Am Nat 138:499–517

Sokal RR (1979) Testing statistical significance of geographic variation patterns. Syst Zool 28:227–232

Spear SF, Balkenhol N, Fortin MJ, McRae BH, Scribner KT (2010) Use of resistance surfaces for landscape genetic studies: considerations for parameterization and analysis. Mol Ecol 19:3576–3591

Spielman D, Brook BW, Frankham R, Schaal BA (2004) Most species are not driven to extinction before genetic factors impact them. Proc Natl Acad Sci USA 101:15261–15264

St-Louis V, Fortin MJ, Desrochers A (2004) Spatial association between forest heterogeneity and breeding territory boundaries of two forest songbirds. Landsc Ecol 19:591–601

Storfer A, Murphy MA, Evans JS, Goldberg CS, Robinson S, Spear SF, Dezzani R, Delmelle E, Vierling L, Waits LP (2007) Putting the 'landscape' in landscape genetics. Heredity 98:128–142

Storfer A, Murphy MA, Spear SF, Holderegger R, Waits LP (2010) Landscape genetics: where are we now? Mol Ecol 19:3496–3514

Wagner HH, Fortin MJ (2005) Spatial analysis of landscapes: concepts and statistics. Ecology 86:1975–1987
Ward EJ (2008) A review and comparison of four commonly used Bayesian and maximum likelihood model selection tools. Ecol Model 211:1–10
Wright S (1943) Isolation by distance. Genetics 28:114–138
Wright S (1948) On the roles of directed and random changes in gene frequency in the genetics of populations. Evolution 2:279–294