MOLECULAR ECOLOGY

Molecular Ecology (2010) **19**, 3746–3759 doi: 10.1111/j.1365-294X.2010.04748.x

Landscape modelling of gene flow: improved power using conditional genetic distance derived from the topology of population networks

RODNEY J. DYER,* JOHN D. NASON† and RYAN C. GARRICK‡

*Department of Biology, 1000 West Cary Street, Virginia Commonwealth University, Richmond Virginia 23284-2012, USA, †Department of Ecology, Evolution & Organismal Biology, 339 Bessy Hall, Iowa State University, Ames, IA 50011, USA, ‡Department of Ecology and Evolutionary Biology, Environmental Science Center, Yale University, 21 Sachem St, New Haven, CT 06520, USA

Abstract

Landscape genetics is a burgeoning field of interest that focuses on how site-specific factors influence the distribution of genetic variation and the genetic connectivity of individuals and populations. In this manuscript, we focus on two methodological extensions for landscape genetic analyses: the use of conditional genetic distance (*cGD*) derived from population networks and the utility of extracting potentially confounding effects caused by correlations between phylogeographic history and contemporary ecological factors. Individual-based simulations show that when describing the spatial distribution of genetic variation, *cGD* consistently outperforms the traditional genetic distance measure of linearized F_{ST} under both 1- and 2-dimensional stepping stone models and Cavalli-Sforza and Edward's chord distance D_c in 1-dimensional landscapes. To show how to identify and extract the effects of phylogeographic history prior to embarking on landscape genetic analyses, we use nuclear genotypic data from the Sonoran desert succulent *Euphorbia lomelii* (Euphrobiaceae), for which a detailed phylogeographic history has previously been determined. For *E. lomelii*, removing the effect of phylogeographic history significantly influences our ability to infer both the identity and the relative importance of spatial and bio-climatic variables in subsequent landscape genetic analyses. We close by discussing the utility of *cGD* in landscape genetic analyses.

Keywords: gene flow, genetic covariance, landscape genetics

Received 1 December 2009; revision received 4 May 2010; accepted 6 May 2010

Introduction

The goal of landscape genetics is to infer how micro-evolutionary forces operating within and among natural populations are influenced by details of the current environmental context in which they occur. While the integration of landscape ecology and inter-individual relatedness is a fundamentally new endeavour (Manel *et al.* 2003; Storfer *et al.* 2007), understanding the mechanisms by which intervening landscape features influence the genetic connectivity of populations has been a topic of interest for some time (e.g. Spielman & Smouse 1976; Sokal *et al.* 1991; Taylor *et al.* 1993; Baer 1998). In studying the processes that affect inter-population genetic connectivity, neutral markers are the genetic data of choice. These data sets have been useful for investigating the relative influence that ecological and spatial variables, either separately or in combination, have on estimates of genetic differentiation (or realized gene flow). For example, in combination with spatial proximity, biotic and abiotic landscape features such as management history (Holzhauer *et al.* 2006), forest configuration (Cushman *et al.* 2006) and ecological distance (Geffen *et al.* 2004) have all been shown to influence genetic covariance among

Correspondence: Rodney J. Dyer, PhD, Fax: 01 804 828 0503; E-mail: rjdyer@vcu.edu

populations. Insights from neutral genetic markers into how permeable contemporary landscape features are to migration and gene flow also have diverse applications in conservation biology. These include the identification of key dispersal corridors linking fragmented native habitats (Peakall *et al.* 2003; Epps *et al.* 2007) and routes or mechanisms of invasion by exotic species (Estoup *et al.* 2004), as well as distinguishing the effects of metapopulation processes from bottleneck-induced divergence in fragmented landscapes (Zellmer & Knowles 2009).

As a field, landscape genetics has seen rapid methodological progress on two fronts: the formulation of predictive landscape resistance hypotheses and distance metrics (e.g. Cushman *et al.* 2006; McRae 2006; Wang *et al.* 2008) and the development of statistical methods linking genetic and landscape data (Murphy *et al.* 2008; Balkenhol *et al.* 2009; Cushman & Landguth 2010). In this study, we focus on two methodological issues relevant to the predictive power of landscape genetic models: the selection of a genetic distance metric when dealing with data sets based upon populations and the inclusion of scale-appropriate historical demographic hypotheses. Using individual-based stochastic simulations, we examine the relative performance of a genetic distance metric based on the principle of conditional independence derived from Population Graphs (Dyer & Nason 2004) in relation to both linearized F_{ST} and Cavalli-Sforza & Edwards (1967) chord distance D_c under 1- and 2-dimensional stepping stone models of isolation by distance (IBD). These simulated data were analysed using multiple regression on distance matrices (MRDM; after Legendre *et al.* 1994). We then examine an empirical data set from the Sonoran Desert endemic plant *Euphorbia lomelii* V.W. Steinm (Euphorbiaceae), a species for which we have a detailed description of deep-time and post-Pleistocene phylogeographic history across the entire species range (Garrick *et al.* 2009). We show that incorporating phylogeographic information in landscape genetic models can improve the inference concerning the identity and relative importance of spatial and ecological factors acting on genetic connectivity. Finally, we show how estimated genetic covariance can be mapped back onto the landscape to facilitate the development of subsequent, and more specific, landscape genetic hypotheses and analyses.

The case for conditional genetic distance

A classic paradigm for relating gene flow to the spatial separation of populations is that of IBD (Wright 1943, 1946). Under this model, if dispersal is limited relative to the geographic distance between populations then a general increase in genetic differentiation will be associated with increasing inter-population distance. Most traditional approaches for identifying IBD are based on genetic distances calculated in a pairwise manner for all population pairs. Often, transformations are applied (e.g. Slatkin 1993; Rousset 1997) so that the resulting genetic distance has an expected linear relationship with spatial distance under homogeneous 1- or 2-dimensional stepping stone models of migration. Genetic distances are then regressed on spatial distances representing different models of inter-population connectivity. While innumerable studies have successfully investigated the process of IBD using this approach, tests based on pairwise genetic distance, including F_{ST} (and its relatives) and D_c, nevertheless suffer from at least two potential problems that may limit their power and accuracy for landscape genetic inferences.

First, in stepping stone models of IBD (Kimura & Weiss 1964), gene flow between distant populations is likely to involve intervening populations and, in reality, is probably often far more complicated than simple stepping stone frameworks. Indeed, the field of landscape genetics specifically focuses on how landscape heterogeneity influences ongoing gene flow, and so only in the simplest systems will genetic distances calculated for pairs of populations independently of all others directly reflect gene movement. In contrast, a genetic distance metric that simultaneously takes the genetic covariance of all populations into account may provide a more sensitive measure for understanding how gene flow interacts with spatial and ecological landscape variables. In this manuscript, we argue that the sensitivity of analyses linking gene flow to these variables will be improved if pairwise genetic distances are conditioned on the multilocus genetic characteristics of the full set of sample populations.

A second issue potentially limiting the utility of pairwise genetic distance for landscape genetic analysis is that the theoretical relationship between genetic distance (including linearized F_{ST}) and the spatial separation of populations is linear only when gene flow across the landscape is a homogeneous function of spatial distance. This assumed homogeneity conflicts with a basic tenet of landscape genetics that the movement of migrants is influenced by ecological variables (e.g. suitable habitat, dispersal corridors, topography) whose spatial distributions are decidedly heterogeneous. While significant correlations between pairwise genetic distance and spatial distance are nevertheless observed, the predictive power of the spatial distance metric (and consequently other ecological predictor variables in the model) may be enhanced if the genetic distance between population pairs is estimated conditional upon the entire network of populations.

Not previously used in a landscape genetic context is conditional genetic distance (*cGD* or graph distance), a component of Population Graphs (Dyer & Nason 2004). A Population Graph is a network modelled from the conditional genetic covariance structure among populations analysed simultaneously. Population pairs exchanging migrants will exhibit significant conditional covariance and will be connected in the network by edges whose length is inversely proportional to the genetic covariance between the populations. Conversely, populations not directly exchanging migrants are likely to exhibit conditional independence and are not connected to each other in the network by edges, as the absence of direct gene flow allows them to proceed on evolutionary trajectories that are independent given the set of intervening populations through which they exchange migrants. Across the entire Population Graph, *cGD* is estimated as the length of the shortest (geodesic) path connecting pairs of populations. Relative to traditional measures of pairwise genetic distance, including F_{ST} and D_c, *cGD* is expected to be more sensitive as it is calculated based upon the differences in genetic covariation associated with both direct and indirect connectivity (gene flow) among populations, making it potentially better suited for use in landscape genetic modelling.

The case for scale-appropriate historical demographic hypotheses

Landscape genetic studies have focused primarily on spatial and ecological variables influencing migration and gene flow. However, when inference is based on population-level data (cf. individual-based assignment tests), genetic differentiation reflects both historical and contemporary patterns of genetic connectivity and so variables reflecting historical demography too may offer substantial predictive power. For example, in *Pinus flexilis*, Latta & Mitton (1997) showed a high degree of divergence among populations measured by mtDNA and RAPD markers. However, additional genetic information from allozymes and cpDNA markers based upon a larger sample of populations suggested that the populations on the east and west slopes of the Rocky Mountains may have been colonized from separate sources, with the mountains themselves acting as an historical source of vicariance. It just happened that the sampled populations straddled a zone of secondary contact, which was the underlying cause of genetic disequilibrium, and so it was coincidental that the observed differentiation was also arrayed along an ecological gradient.

In addition to illuminating the role of deeper-time processes influencing genetic connectivity, Latta and Mitton's study underscores two perhaps unappreciated benefits of incorporating information on historical demography in landscape genetic studies conducted at large spatial scales. First, when independent geological, paleoclimatic or paleoecological data indicate that biogeographical processes such as vicariance and range expansion have influenced past gene flow, their inclusion as predictor variables can only enhance the precision of landscape genetic models (e.g. Latta 2006; Sork & Smouse 2006). Here, the term precision is used in a statistical context, meaning that models that have taken into consideration these historical processes will provide more accurate landscape genetic inferences. Second, inferences about the effects of specific spatial or ecological variables on genetic connectivity can be more accurate when they are conditioned on historical demographic factors. If ecological and historical effects are independent, then inclusion of historical variables in landscape genetic models should have no effect on the correlation between the ecological variables and the genetic response variable. If, in contrast, the effects of ecological and historical factors are themselves correlated, then we need to ask how much of the correlation between ecology and genetic connectivity remains after accounting for history. Failure to do so is likely to result in spurious inference as to the influence of contemporary ecological factors on gene flow.

Methods

Simulations models: pairwise F_{st} *and* D_c *vs. conditional graph distance*

To determine the relative statistical resolution of different genetic distance metrics for quantifying among-population structure for landscape genetic purposes, we performed a battery of individual-based Monte Carlo simulations following Dyer (2007). During each simulation, three population pairwise genetic distances were estimated: (i) Rousset's (1997) linearized F_{ST}, ($F_{ST}/(1 - F_{ST})$; hereafter denoted simply as F_{ST}), a common metric for landscape genetic analyses, (ii) Cavalli-Sforza & Edwards (1967) chord distance, D_c, which has been shown to perform well with microsatellite data (Nei *et al.* 1983; Takezaki & Nei 1996, 2008), and (iii) conditional genetic distance (*cGD*). Because *cGD* is less well known than F_{ST} and D_c, and a focus of this study, we briefly summarize Dyer & Nason (2004) in describing how a Population Graph is obtained and how *cGD* is calculated from it.

In constructing a Population Graph, the genetic data are initially represented geometrically (following Smouse *et al.* 1982) with each independent allele representing an orthogonal axis in multivariate genetic space.

Each individual's multilocus genotype maps to a point in this space, with a population represented by its geometrical mean (or centroid) over individuals. Within this framework, matrices of squared Euclidean distances among individuals, both within populations and among populations, can be obtained in the normal manner. The matrix of squared inter-population distances is transformed into a matrix of inter-population covariances (following Gower 1966; see also Smouse & Peakall 1999), with the diagonal elements subsequently replaced with intra-population variances calculated from the within-population distance matrices (following Excoffier *et al.* 1992). Utilizing graphical modelling methods (Whittaker 1990; Edwards 2000), this genetic variance-covariance matrix among populations is inverted and standardized to obtain a partial correlation matrix, with the significance of individual partial correlations determined using the simultaneous testing procedure described in Dyer & Nason (2004). If the partial correlation between populations *i* and *j* is significantly greater than expected by chance, then an edge is placed between vertices *i* and *j*. A Population Graph is constructed by applying this procedure to all possible population pairs, and it is from such a graphical model of population genetic structure that *cGD* is obtained as the shortest path between population pairs.

To evaluate the strength of genetic signal provided by F_{ST}, D_c and *cGD*, we simulated populations using the software EASYPOP (version 2.01; Balloux 2001) with the following parameters. Individuals were arrayed into populations of size 100. We specifically modelled hermaphroditic individuals, as is common in plants, but the results apply generally to any breeding system. Representative of microsatellite studies, each individual was randomly assigned diploid genotypes for 12 independently assorting loci, each of which could have up to 20 separate allelic states. All loci were allowed to mutate at a rate of $\mu = 0.0001$ under a single step mutation model. Populations were subject to 1-dimensional (1DSS) and a 2-dimensional (2DSS) stepping stone models of gene migration. In the 1DSS model, 36 populations were simulated, whereas in the 2DSS model 100 populations were simulated in a square lattice, although only the central 36 populations were analysed. Migration was symmetric with $m = 0.05$. Multilocus genotypes from 20 random individuals per population were sampled at 250, 500, 1000, 1500, 2000 and 2500 generations, and then pairwise F_{ST} (after Weir & Cockerham 1984), D_c, and *cGD* were simultaneously estimated from the same data sets.

Currently, there are several analytical approaches that could be used for landscape genetic analysis, many of which were recently reviewed in Balkenhol *et al.* (2009). Of the various methods available, we opted to use the MRDM approach after Legendre *et al.* (1994), which was identified by Balkenhol *et al.* (2009) as one of the best analytical frameworks for predicting genetic connectivity. The MRDM approach provides an intuitive set of methodologies (e.g. regression model estimation) that allowed us to test the relative performance of alternate genetic encoding strategies (see above) not found in approaches based upon canonical correspondence analysis (CCA) or distance-based redundancy analysis (dbRDA; Legendre & Anderson 1999). Moreover, this approach lends itself to removing the influences of historical covariates (e.g. working on the residual variation after partitioning out putative historical factors) much more easily than with alternative multivariate approaches. This is not to say that CCA and dbRDA may not be applicable approaches, indeed they do have their own benefits, however, for the purposes of testing relative fit, extracting potentially confounding factors and building expected response surfaces the MRDM approach was favoured.

We quantified the fit of the three genetic encoding metrics under models of IBD using the MRM function in the R library ecodist (Goslee & Urban 2008). Under the 2DSS model, linearized F_{ST} was fit to the log of physical distance (after Rousset 1997); genetic and physical distance was not transformed in other regressions. A total of 100 replicate runs were performed, and the difference in the proportion of variance explained ($\delta R^2_{F_{ST}} = R^2_{cGD} - R^2_{F_{ST}}$; $\delta R^2_{Dc} = R^2_{cGD} - R^2_{Dc}$) by each genetic encoding metric was recorded. The difference in the performance of linearized F_{ST}, D_c and *cGD* was tested using a *t*-test under a two-tailed null hypothesis H_O:$\delta R^2 = 0$. All statistical analyses were conducted in R (R Development Core Team 2005).

Landscape genetic analysis of Euphorbia lomelii

The study species used here is the stem-succulent euphorb *Euphorbia lomelii* (synon. *Pedilanthus macrocarpus*). The species is long lived, arid adapted and pollinated by hummingbirds (Dressler 1957), with seeds apparently gravity dispersed. For this species, we expect that the vector of gamete exchange with the greatest potential for long distance dispersal is pollen. Its range along the Baja California peninsula extends from the southern tip of the Cape Region, northward to Bahía de Los Angeles occupying all but the upper third of the Baja peninsula. On the mainland, *E. lomelii* populations are known only from a relatively small area of the southwest gulf coast of the state of Sonora, Mexico. In the present study, we focus solely on the peninsular populations, for which we have sampled 33 distinct populations (Fig. 1). For the purpose of subsequent analyses, we will assume that these populations are

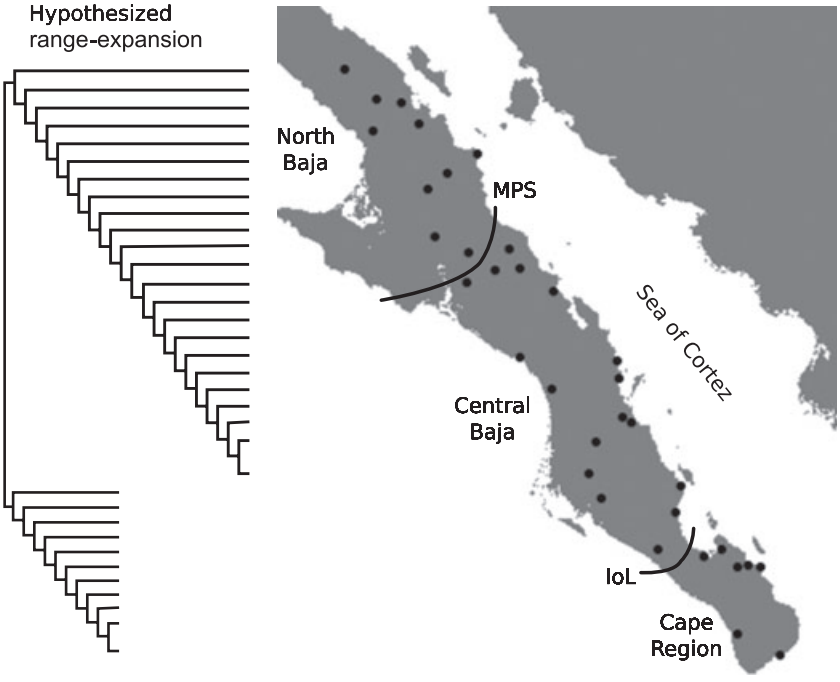

Fig. 1 Sampling distribution of *Euphorbia lomelii* populations in Baja California, Mexico. Right: Sample locations are indicated with respect to three geographical regions (North, Central and Cape Region) defined by two externally identified sources of vicariance: ancient trans-peninsular seaways located at the Isthmus of La Paz (IoL; ca. 3 mya) and mid-peninsula (MPS; ca. 1 mya). Left: A hypothetical population tree topology representing two separate range expansion events as identified in Garrick *et al.* (2009): from northern to central Baja and from south-central Baja into the Cape Region. Four separate vicariance and range expansion hypotheses were represented as phenetic distance matrices to determine the extent to which phylogeographic history influences both the fit and interpretation of contemporary landscape genetic models.

distributed along a 1-dimensional landscape whose main axis along the peninsula, measuring the distance between the most distant populations is 835 km in length. Multilocus genotypic data were generated for 311 individuals based on six co-dominant nuclear loci (mostly introns), with polymorphism screening conducted via PCR-RFLP assays as described in Garrick *et al.* (2008).

Animal phylogeographic studies centred in Baja California have identified two major multi-taxon genetic discontinuities that spatially occur at or near the putative locations of ancient transient seaways (i.e. mid-peninsular seaway, MPS, ca. 1 mya, and the Isthmus of La Paz, IoL, ca. 3 mya; Riddle *et al.* 2000; Lindell *et al.* 2006; Riddle & Hafner 2006). In contrast, some plants show strong signal of northward post-Pleistocene range expansion (Nason *et al.* 2002; Clark-Tapia & Molina-Freaner 2003). Our recent phylogeographic analysis of *E. lomelii* (Garrick *et al.* 2009) revealed (i) significant IBD along the peninsula, (ii) a strong signal of southward post-Pleistocene range expansion, including a distinct, separate relationship between latitude and genetic diversity in the Cape Region, and (iii) a recognizable signal of spatial genetic discontinuities at or near the putative locations of both MPS and IoL vicariance events. Furthermore, coalescent analyses of cpDNA sequences suggested the origin of mid-peninsular vicariance was consistent with the hypothesized ~1 mya time frame derived from mtDNA date estimation in side-blotched lizards (Upton & Murphy 1997). It is these historical phylogeographic patterns that we will attempt to partition from the data prior to examining factors influencing gene flow between populations.

Identifying phylogeographic and landscape genetic effects

We performed GIS least-cost path analysis to identify the set of predictor variables describing genetic covariation among our *E. lomelii* sample populations. Factors used in the construction of these predictor variables were of two varieties; a conditioning set of variables containing the putative set of phylogeographic patterns identified in Garrick *et al.* (2009), and set of contemporary bio-climatic factors predicted to influence either the synchrony of phenology among *E. lomelii* individuals or the movement of pollinators across the peninsular landscape. Using a stepwise regression approach (Draper & Smith 1981) based upon distance matrices (MRDM; Legendre *et al.* 1994), we first removed the effects of phylogeographic history on the distribution of genetic variation among populations and then determined the best model fitting bio-climactic factors to the residual genetic variation. For the analysis of *E. lomelii*, we use the genetic distance metric that showed the greatest power in the simulations (from above) for a 1-dimensional stepping stone model. We selected this migration model based upon the linear distribution of populations along Baja California (Fig. 1). All pairwise genetic distance parameters were estimated using GeneticStudio (Dyer 2009).

Three separate classes of hypothesized phylogeographic processes were used to construct the initial conditioning set of predictor variables: (i) IBD among all peninsular populations, (ii) southward range expansions and (iii) spatial discontinuities consistent with historical sources of vicariance. Prediction matrices for IBD were estimated from the geographic locations of all populations. Range expansion was quantified using a phenetic distance approach. The distribution of genetic diversity in *E. lomelii* suggests that there were two geographically separate post-Pleistocene range expansion events that proceeded to the south, one originating in the northern portion of the Baja California peninsula and the other originating in the region just north of the Isthmus of La Paz (~25.2°N; Fig. 1). To quantify this process as a pairwise distance matrix, we constructed an unrooted bifurcating tree (shown in Fig. 1; also see Nason *et al.* 2002) on which the phenetic distance between populations was calculated. Finally, each of the two hypothesized historical vicariance events (IoL and MPS; Fig. 1) was represented separately in matrix form using dummy variables: the distances between pairs of populations in the same region were assigned no 'cost' (0), whereas populations in different regions were assigned a boundary cost (1). Because these are binary variables, alternative boundary costs were not investigated because there would be no net change in the model coefficient (i.e. it amounts to multiplying by a scaling factor). Analyses proceeded by first fitting *cGD* to models with only these phylogeographic predictor variables in them. The residual variation from the best-fit model was then used as the response variable in estimating models related to present-day landscape ecological variables.

Using GRASS GIS (version 6.2.3; GRASS Development Team 2004), we estimated least-cost path distances for the 12 bio-climatic features listed in Table 1. Bio-climatic layers covering the species range for *E. lomelii* were derived from Tile 22 of the WorldClim data sets (http://www.worldclim.org), with these data resolved at 30-arc seconds of latitude and longitude (~1 km^2). For each landscape factor and population, we defined a cost surface in GRASS using the r.cost function based upon absolute similarity of the feature, as measured at the location of the population. For example, if the feature was elevation, we defined a cost surface whose values were positive in proportion to the deviation away from the elevation for the target population. This cost encoding is consistent with the notion that similarity in local bio-climatic conditions would lead towards greater synchrony in phenology and thereby increase the opportunity for pollen-mediated gene flow. While the selection of a distance metric is critical to the subsequent analyses (see Spear *et al.* 2010), we believe that the variables considered here are a conservative set of factors that influence pollen-mediated gene flow given our understanding of the dispersal ecology of *E. lomelii*.

Using the best metric for quantifying genetic distance as identified by the simulations, we fit linear models containing the phylogeographic predictor variables under a MRDM model (Legendre *et al.* 1994) employing a stepwise approach following Draper & Smith (1981) as above. A total of 10 000 permutations were used to assess both the significance of terms being added to the model as well as the final model significance. The residuals of the candidate model were then used as the response variable in the analysis of contemporary landscape ecological features. The same stepwise procedures were used to estimate the best-fit model for describing the relationship between contemporary landscape features and the distribution of genetic variation after removing the effects of phylogeographic history. Because of autocorrelation in the raw bio-climatic variables, least cost paths were examined for potential multicollinearity as predictor variables in the model estimation. Among the eight bio-climatic variables, only two of the 28 pairs of estimated cumulative least cost paths had an absolute correlation greater than 0.8; the maximum of which was $\rho = 0.81$ between diurnal range and the minimum temperature of the coldest month. All models were fit assuming a *Type I* error rate of $\alpha = 0.05$. To verify that multicollinearity was not a problem in the estimated models, we also estimated the variance inflation factor

Table 1 Phylogeographic, topographic and bio-climatic landscape variables used to estimate cost resistance surfaces for *Euphorbia lomelii* in Baja California

Landscape phylogeographic predictors	
Euclidean distance (isolation by distance)	Southward range expansions
Vicariance at isthmus of la paz (IoL)	Vicariance at mid-peninsular seaway (MPS)
Bio-climatic predictors	
Elevation	Minimum temperature (coldest month)
Maximum temperature (warmest month)	Mean diurnal range
Mean temperature wettest quarter	Mean temperature driest quarter
Precipitation wettest month	Precipitation driest month

(*vif*) for each predictor variable the final model; we assumed that estimated $\text{vif}_i \geq 10$ would indicate potential problems associated with correlated predictor variables (Kutner *et al.* 2004). Finally, all predictors were standardized to mean zero and unit variance so that the coefficients of the terms in the model could be compared with respect to their magnitude and their relative importance in describing the response variable.

Estimating a landscape genetic topology

In addition to identifying putative landscape features that differentially influence genetic covariance, the best-fit model also provides inferences on the spatial locations where the particular combination of landscape features is expected to differentially influence genetic connectivity. This spatial information can be useful for guiding subsequent experimental analysis conducted to test the validity of the fit models. To illustrate, we estimated the predicted response surface for the model fit to genetic distance conditioned on phylogeographic history. The values on this surface provide an estimate of how genetic distance accumulates across the landscape under the least-cost path model. For example, pairs of populations separated by a region of high genetic distance will experience reduced gene flow when compared to population pairs whose intervening landscape has low expected genetic distance. For simplicity, we standardized this response surface to be bound by [0,1] and provide two example transects across the Baja California peninsula showing the slope of this surface.

Results

Simulations models: pairwise F_{st} *and* D_c *vs. conditional genetic distance*

The manner in which genetic data are encoded has significant effects on our ability to explain the spatial distribution of genetic variation in 1- and 2-dimensional stepping stone models. Under a 1DSS model, the difference in the proportion of variance explained by F_{ST} and *cGD* (as measured by $\delta R^2_{F_{ST}}$) ranged from 0.01 to 0.47 (Fig. 2), with a positive value indicating that models using *cGD* performed better than those using linearized F_{ST}. Similarly, *cGD*, outperformed D_c with δR^2_{Dc} ranging from 0.17 to 0.51 (Fig. 2). In all cases of the 1DSS model, *cGD* performed significantly better than models based upon F_{ST} (largest *t*-test $P < 6.6e^{-16}$) or D_c (largest *t*-test $P < 1.1e^{-61}$).

The performance of *cGD* in relation to F_{ST} and D_c under the 2DSS models was both mixed and attenuated. Across all time steps, *cGD* performed significantly better in explaining the spatial distribution of genetic

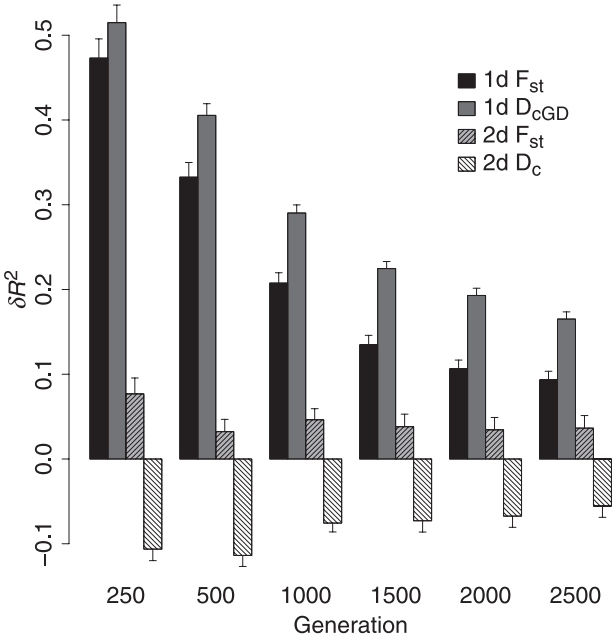

Fig. 2 The difference in the fit of the multiple regression on distance matrices models (as quantified by δR^2) for linearized F_{ST}, chord distance (D_c), or conditional graph distance (*cGD*). Models were fit to both 1-dimensional (dark bars) and 2-dimensional (light bars with cross-hash) stepping stone models of migration. A positive value of δR^2 indicates that models using *cGD* performed better than those using F_{ST} or D_c. Error bars represent 95% confidence limits on the mean difference (δR^2).

variation than F_{ST} ($\delta R^2 F_{ST} = 0.03 - 0.08$; *t*-test $P < 2.6e^{-5}$ in all cases). However, *cGD* did not perform as well as D_c in the 2DSS model ($\delta R^2_{Dc} = -0.11 - (-0.05)$; *t*-test $P < 7e^{-13}$; Fig. 2). For all models in which *cGD* excelled, the observed difference in the explanatory ability of *cGD* vs. the other parameters developed rapidly before decreasing as the systems tended towards genetic equilibrium (Fig. 2).

For comparative purposes, examining δR^2 does not provide insights into which of the parameters stabilizes faster. As a result, we calculated the variance of in R^2 for F_{ST}, D_c, and *cGD* across replicate simulations for each time step. Relative statistical stability in these statistics is based upon the time at which it takes for each of the parameters to reach an asymptote and is directly comparable as identical data sets are being assayed for each parameter during each time step in the simulations. Across time steps, *cGD* stabilizes in considerably fewer generations than either F_{ST} or D_c (Fig. 3).

Landscape genetic analysis of Euphorbia lomelii

Using a stepwise selection procedure, the best model fitting phylogeographic variables to genetic differentiation

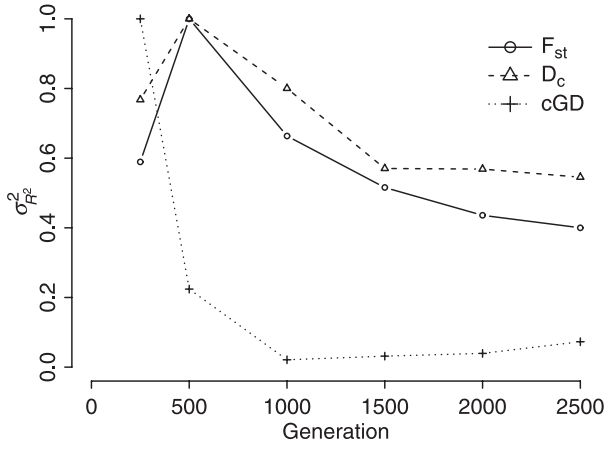

Fig. 3 Estimated variance in the fit of the simulation models for each genetic distance parameter F_{ST}, D_c, and *cGD*. Each estimate was standardized by its maximum value for comparative purposes. The quicker the variance in a parameter reaches its asymptote, the more stable the estimator.

as estimated by *cGD* was highly significant ($F = 22.20$; $P_{perm} = 0.0001$; $R^2 = 0.07$). We selected cGD as the preferred genetic distance metric suggested by the simulations of 1-dimensional landscapes. The initial model contained two significant phylogeographic conditioning variables; IBD and historical vicariance caused by the isthmus of La Paz (Fig. 1, Table 1). The standardized coefficients on the model were $\beta_{IBD} = 0.41$ ($P_{perm} = 0.0007$) and $\beta_{IoL} = 0.27$. ($P_{perm} = 0.0244$) suggesting that IBD was more important in explaining the *cGD* than the vicariance caused by the Isthmus of La Paz. These results suggest that consideration of historical events or processes can be critically important and that they can explain a significant portion of the existing genetic covariance that would typically be left in landscape genetic models. In cases where these factors are not partitioned from the data, it is possible that their effects will be erroneously attributed to contemporary landscape or ecological factors.

The residual variation from the model conditioned on phylogeographic history was then subjected to a stepwise regression approach using MRDM with the bio-climatic variables listed in Table 1. Using the same procedure as above, the best-fit model was highly significant ($F = 13.57$, $P_{perm} = 0.0002$, $R^2 = 0.09$) and included three terms representing both temperature and precipitation (Table 2). In the final models, the maximum variance inflation factor (vif) was 2.2, suggesting that the predictor variables fit to the model do not exhibit multicollinearity to a degree such that we are concerned with the model selection. To determine the extent to which conditioning on phylogeographic history changed the inferences gained, we performed the same model fitting exercises for raw *cGD* (i.e. uncorrected for phylogeography; denoted as cGD_{ucor} below). The best-fit model also contained three bio-climatic variables and was highly significant ($F = 22.27$, $P = 0.0001$), more so than the unconditioned model. Models fit to *cGD* or cGD_{ucor} both contained terms for precipitation in the driest month and mean temperature during the driest quarter. The relative importance of the precipitation variable (as measured by standardized regression coefficients) was 16% higher for the uncorrected than the corrected model (e.g. $\beta_{precip} = 0.47$ vs. $\beta_{precip;\ ucor} = 0.56$) and for the temperature variable was 30% higher for the uncorrected than the corrected model ($\beta_{temp} = 0.53$ vs. $\beta_{temp;\ ucor} = 0.75$; Table 2). In the uncorrected model, elevation was a significant predictor, whereas in the corrected model, maximum temperature in the warmest month was included. Within models, the predictor with the largest standardized effect in the uncorrected model was mean temperature in the driest quarter ($\beta = 0.75$), whereas in the corrected model it was the variable describing the maximum temperature during the warmest month ($\beta = -0.69$). Overall, the conditioning on phylogeographic history not only changed the relative importance of bio-climatic predictor variables, but also changed which subsets of variables were considered

Table 2 Significant slope coefficients (and probabilities in parenthesis) for multiple regression on distance matrices models fit to conditional genetic distance (*cGD*) for *Euphorbia lomelii*. Empty cells represent ecological variables whose inclusion in the model was not found to be significant (at $\alpha = 0.05$). The column labelled *cGD* was the model fitting bio-climatic variables to graph distance after removing the effects of known phylogeographic variables, whereas the column cGD_{ucor} was the model fit to the raw (uncorrected) graph distances. The order in which the factor is listed represents the order of inclusion in the stepwise model. The remaining factors from Table 1 were found to not be significant in the model

Bio-climatic predictors	*cGD*	cGD_{ucor}
Precipitation driest month	0.4720 (0.0007)	0.5641 (0.0001)
Maximum temperature warmest month	−0.6942 (0.0002)	–
Mean temperature driest quarter	0.5316 (0.0022)	0.7546 (0.0001)
Elevation	–	−0.4419 (0.0200)

important in describing the spatial distribution of genetic covariance.

Fitted response surfaces. The subset of significant bio-climatic factors retained in the corrected model (Table 2) was used to estimate a predicted response surface across peninsular Baja. The response surface for the *cGD* model (Fig. 4) indicates the extent to which identified bio-climatic factors are predicted to cause changes in among-population genetic covariance. Two transects across this response surface are provided to illustrate how the identified the spatial locations of putative bio-climatic factors that interact to influence genetic covariance and by extension gene flow.

Discussion

In this study, we show via simulation of IBD processes that conditional genetic distance (*cGD*) represents an improvement over F_{ST}, the pairwise genetic distance measure traditionally used in landscape genetics in both 1- and 2-dimensional stepping stone models. Furthermore, we show that D_c, a distance metric known to perform well with microsatellite loci in other population genetic and phylogenetic applications, outperforms *cGD* in a 2-dimensional model, whereas *cGD* provides better analytical performance under a 1-dimensional model. However, *cGD* does appear to stabilize quicker than either of the remaining parameters following perturbation (also shown in Dyer 2007 in comparison with Φ_{st}). In the context of an empirical data set for the euphorb *E. lomelii* in Baja California, a species whose spatial distribution is generally 1-dimensional, we show how the influence of appropriately parameterized phylogeographic factors can be removed from the data prior to an analysis of landscape factors, ultimately resulting in landscape genetic models with clearer inferences.

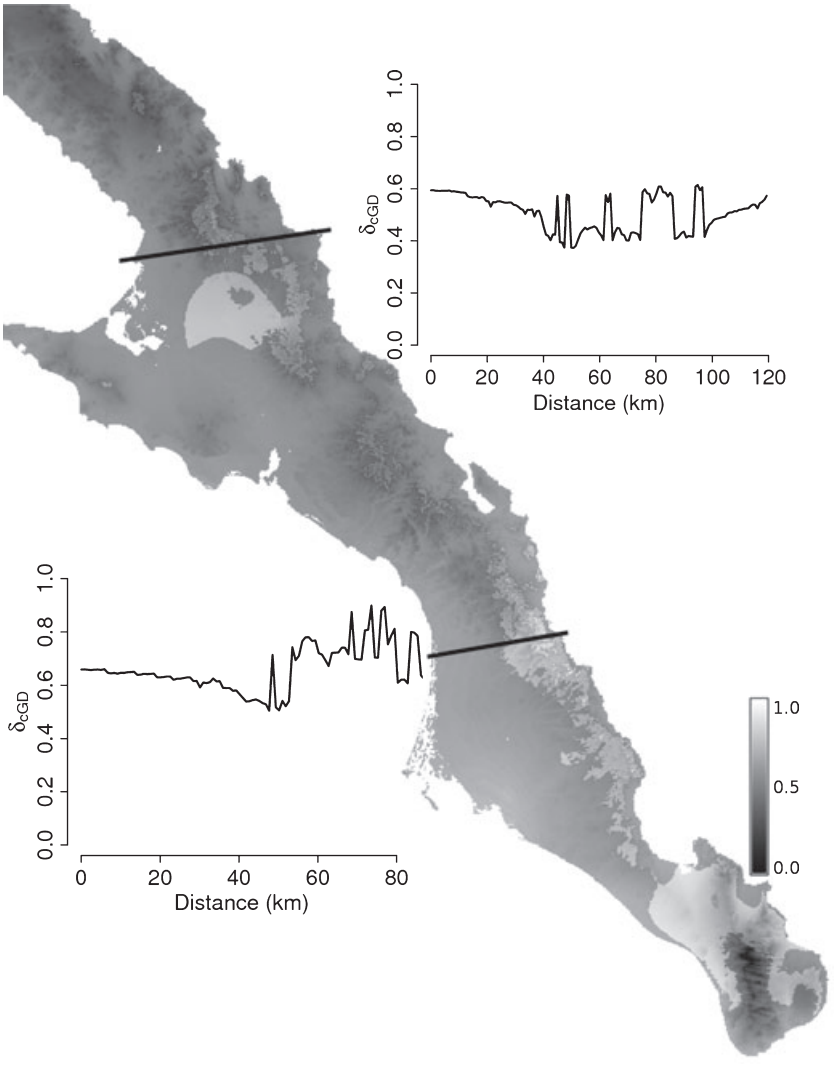

Fig. 4 The predicted surface for the best model describing the distribution of genetic variation in *Euphorbia lomelii* based upon conditional genetic covariance (*cGD*) after removing the influence of phylogeographic history. Lighter regions on the map indicate spatial locations where the combination of ecological factors interact to impede gene flow (and thus increase *cGD* between populations) at a rate greater than darker areas. The insets graphs depict the slope of the response surface (δ*cGD*) across the peninsula and indicate where the change in genetic covariance because of the indicated bio-climactic factors would be greatest. The spatial extent of this map encompasses the entire species distribution of *E. lomelii*.

Benefits of conditional genetic distance for landscape genetic analysis

The results of the IBD simulations performed here provided several interesting insights into our ability to quantify spatial genetic structure. First, in both 1DSS and 2DSS stepping stone models, *cGD* offers a substantial improvement over linearized F_{ST} in statistical power to detect underlying patterns of IBD. This difference in power is especially great when migration follows a 1DSS model (Fig. 2). Similarly, *cGD* outperforms D_c in 1DSS models, however, D_c is better if gene flow proceeds as a 2DSS model. Second, *cGD*, F_{ST} and D_c also differ in rate of approach to their final equilibrium states under a migration-drift process. The simulation results suggest that stabilization of *cGD* under IBD is faster than that for F_{ST} or D_c (Fig. 3 and Dyer 2007), which can be attributed to the fact that cGD is based upon a simultaneous analysis of the entire data set whereas the other parameters are estimated in a pairwise fashion among populations. While the overall F_{ST} for a set of populations has been reported to stabilize in fewer than 100 generations (e.g. Crow & Aoki 1984; Hamilton & Miller 2002), these conclusions are based on simulations of Wright's (1931) Island model where all populations have the potential to exchange genes every generation thereby attaining equilibrium considerably faster than 1DSS and 2DSS models (e.g. Efremov 2005) where migration is limited to neighbouring populations. It should be noted that our simulations represent idealized scenarios in which genetic covariation among populations is a function only of the dispersal process, and this dispersal process is homogeneous in space. We nevertheless expect *cGD* to perform as well or better than other genetic distance metrics under more complicated, real-world conditions in which intervening landscape features exert heterogeneous influences on gene flow. Evaluating this prediction should prove to be a productive area of future research.

An unexpected outcome of our IBD simulations was that across replicate runs, there was a small fraction of data sets for which Population Graphs was unable to provide an estimated topology because of lack of genetic variance within at least one of the populations. The estimation of a Population Graph topology, and by extension *cGD*, requires a nonzero estimate for all within-population variances (Dyer & Nason 2004). In some of the simulations, a population was fixed for alleles, and thus had no variance and the Population Graph could not be estimated. The rate at which *cGD* failed to be estimated was 5.8% for the 1-dimensional model and 1.2% for the 2-dimensional model. The expected probability of homogeneous strata within a data set is determined by the intrinsic genetic variance of the species, the number of loci assayed and the number of individuals sampled per population. Within-population homogeneity is not commonly reported in outbreeding organisms given the number of markers and sample sizes typically used. If encountered in real-world data, one could identify the population or subset of populations that are homogeneous and exclude them from the analysis, or genotype additional individuals to achieve within-population genotypic heterogeneity. In the automated simulations, however, the occasional homogeneous population resulted in *cGD* being undefined, whereas estimates of F_{ST} and D_c could still be obtained.

It is also of note that parameters estimated in a pairwise fashion, such as linearized F_{ST} here, will always have higher variance than a global estimate based upon the simultaneous analysis of all the data. This is because each of the pairwise parameters is estimated from only a subset of the total samples and stochasticity at the level of the individual will play a greater relative role. When using estimates of pairwise parameter values in models such as IBD, we tacitly assume (although with very little basis) that the individual point estimates have no variance. Conversely, parameters fit using the totality of the data, such as *cGD*, do not suffer from this problem.

Conditioning landscape genetic analysis on historical hypotheses

Analysis of the desert euphorb *E. lomelii* provided two valuable insights. First, the results support our contention that the inclusion of predictor variables representing the influence of phylogeographic history can increase the accuracy of the landscape genetic inferences. In particular, when genetic covariation was not conditioned on historical processes the resulting model contained a putatively spurious ecological factor. Admittedly, it is difficult to ascertain the relative precision of this empirical model because the true relationship among the landscape factors is unknown. If a priori information indicates historical processes have significantly influenced the structuring of genetic variation and the effects of these processes are not removed in landscape genetic analyses, the resulting models, independent of their fit, will be incorrect. Moreover, it is also important to note that goodness of fit of a model does not guarantee that the model is correct (e.g. see discussion in Dyer & Nason 2004). As a result, we recommend that one account for information regarding historical factors as a step prior to mounting a landscape genetic analysis. Ultimately, it is the expected response surfaces that we derive from these models (as in Fig. 4) that provide value for subsequent experimental analyses and validation.

In our example using *E. lomelii*, we utilized a least-cost path analysis approach to quantify the bio-climactic distances among sample populations. Currently, there are several different methodological approaches available to quantify ecological distances (e.g. McRae *et al.* 2008; Pinto & Keitt 2009), which could also have been used here. Our selection of a single least cost path was based upon the level of heterogeneity of the bio-climactic variables on our landscape. In our study, bio-climactic factors such as maximum temperature during the warmest month and minimum precipitation during driest month change at a larger granularity than site-specific ecological factors such as forest cover or other habitat suitability measures often used to quantify animal movement. As recently shown by Rayfield *et al.* (2010), single least cost paths show bias in proportion to the heterogeneity of fragmentation and relative habitat suitability. For cases where the landscape factors are more heterogeneous than those used in this study, we recommend the use of multiple path approaches.

Future directions

Population Graphs as a general approach for quantifying how genetic variation is distributed across the landscape provides a framework for analyzing the spatial distribution of genetic variation independent of its overall magnitude. In the present context, components of a Population Graph (e.g. the path length between populations) have been shown to be an effective measure of genetic covariation salient to landscape genetic hypotheses. In addition to identifying putative factors influencing genetic covariance, the results of this work generate a specific set of landscape genetic hypotheses with respect to *E. lomelii*.

Under a homogeneous IBD process, the path lengths (*cGD*) and spatial distances separating neighbouring populations should be approximately proportional. If, however, migration is heterogeneous, then the relationship between expected edge length and spatial separation may be changed in two ways. First, if sporadic long distance migration or colonization events occur then individual populations that are spatially distant will have relatively small *cGD*. We refer to the edges connecting these populations as 'extended edges' because the populations are further apart than expected based on their genetic covariance. In Garrick *et al.* (2009), we used this relationship to identify potential long distance dispersal events associated with recent (post-Pleistocene) range expansion in *E. lomelii*. Conversely, populations may be spatially more proximate than expected by their genetic covariance, which we term 'compressed' edges. Compressed edges are expected to occur when features of the intervening landscape impede migration relative to other similarly separated populations. The predicted genetic response surfaces shown in Fig. 4 are *de facto* null hypotheses for subsequent landscape genetic and experimental studies. Using these expectations to guide sampling allocation, validation of the landscape genetic models estimated here is the logical next step to understanding the interaction of gene flow and landscape features.

Conclusions

In plants, the adult organism is typically sessile with dispersal occurring via two types of propagules: pollen and seed. As sources of gene movement across the landscape, these two modes of dispersal share some properties while also differing in some important regards. In common, pollen- and seed-mediated gene flow are both dependent on dispersal mechanisms extrinsic to the plant, be they biotic or abiotic, such that the spatial dimension of gene flow is determined not only by geographical distance but by features of the intervening landscape that influence the movement of pollen and seed dispersal vectors. Consequently, landscape genetic analysis of the mechanisms underlying genetic connectivity in plants has to consider not only the availability of habitable environments for the plant species itself, but also factors influencing the permeability or resistance of the landscape to dispersal vectors, typically animals or wind.

An important point, if the movement of dispersal vectors is nonrandom with respect to landscape variation in habitat variables, it can generate significant positive correlations between genetic distance and ecological distance in plants that may be difficult to interpret. For example, seed dispersers may move preferentially between habitat patches of a certain type, resulting in habitat-specific patterns of genetic connectivity that reflect the behaviour of the dispersers, not adaptation of the migrants. It is interesting to contrast this against the observation that effective seed migration necessarily requires successful establishment, and hence adaptation to the new environment. As a consequence, seed migration coupled with successful germination should be more likely occur when the environments of the source and receiving populations are similar than when they are dissimilar. In this case, significant correlations between genetic distance and ecological distance may in fact be indicative of local adaptation, not of the genetic markers, but of the migrants themselves. These two alternative explanations for genotype-environment associations (i.e. dispersal biology of the animal vector vs. locally adapted plant populations) may be difficult to

tease apart without conducting reciprocal transplant or laboratory germination experiments. However, biophysiological modelling (cf. correlative species presence/absence-based niche modelling) may provide a framework for removing the potentially confounding influence of animal behaviour (see Kearney & Porter 2009).

Pollen and seed also possess fundamental genetic differences relevant to their affects on gene dispersal (Petit *et al*. 2005). Pollen is haploid and seed diploid so that, given equal migration rates, pollen contributes one-third and seed two-thirds to total nuclear gene flow. Gene flow in maternally inherited markers, in contrast, is related to seed migration alone. These differences in genetic content and dispersal contribute to predictable differences between nuclear and maternally inherited markers in the extent of population genetic differentiation that can be linked to the relative rates of pollen and seed flow (Ennos 1994; Hamilton & Miller 2002). Also important, pollen-mediated gene flow occurs only between established populations so that variables (such as elevation) related to synchrony or asynchrony in flowering across the landscape may often be important predictors of genetic connectivity (as suggested in our data). However, in addition to migration, seed movement also contributes to the colonization of open habitat with effects on genetic connectivity depending on both the number of founders and the genetic correlation among them (Whitlock & McCauley 1990). Through time, these factors can significantly influence how genetic variation is distributed across the landscape.

As demonstrated in the *E. lomelii* data, conditioning landscape analyses on known historical demographic processes aid in our ability to successfully identify and characterize the relationship among landscape features to which the organism is responding. For this plant species, the most important factors were related to temperature extremes and precipitation, both of which contribute to increasing phenological synchrony. In this landscape, and others whose dimensionality tends towards linearity, analytical methods based upon conditional genetic covariance extracted from the Population Graph framework appear to provide more robust model estimation.

Acknowledgements

We thank the Subject Editor and three anonymous reviewers for insightful comments on the initial version of this manuscript. RJD would also like to thank J. Ciminelli and the VCU Center for Environmental Studies for discussions on the GIS modeling. This work was supported by NSF grant DEB-0543102 to JDN and RJD.

References

Baer CF (1998) Species-wide population structure in a southeastern US freshwater fish, *Heterandria formosa*: gene flow and biogeography. *Evolution*, **52**, 183–193.

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

Balloux F (2001) A computer program for the simulation of population genetics. *Journal of Heredity*, **92**, 301–302.

Cavalli-Sforza LL, Edwards AW (1967) Phylogenetic analysis. Models and estimation procedures. *American Journal of Human Genetics*, **19**, 233–257.

Clark-Tapia R, Molina-Freaner F (2003) The genetic structure of a columnar cactus with a disjunct distribution: *Stenocereus gummosus* in the Sonoran desert. *Heredity*, **90**, 443–450.

Crow JF, Aoki K (1984) Group selection for a polygenic behavioral trait: estimating the degree of population subdivision. *Proceedings of the National Academy of Sciences, USA*, **81**, 6073–6077.

Cushman SA, Landguth EL (2010) Spurious correlations and inference in landscape genetics. *Molecular Ecology*, **19**, 3592–3602.

Cushman SA, McKelvey KS, Hayden J, Schwartz MK (2006) Gene flow in complex landscapes: testing multiple hypotheses with causal modeling. *The American Naturalist*, **168**, 486–499.

Draper NR, Smith H (1981) *Applied Regression Analysis*. J Wiley and Sons, New York.

Dressler RL (1957) The genus Pedilanthus (Euphorbiaceae). *Contributions of the Gray Herbarium*, **182**, 1–188.

Dyer RJ (2007) The evolution of genetic topologies. *Theoretical Population Biology*, **71**, 71–79.

Dyer RJ (2009) GeneticStudio: a suite of programs for the spatial analysis of genetic marker data. *Molecular Ecology Resources*, **9**, 110–113.

Dyer RJ, Nason JD (2004) Population graphs: the graph-theoretic shape of genetic structure. *Molecular Ecology*, **13**, 1713–1728.

Edwards D (2000) *Introduction to Graphical Modeling*, 2nd edn. Springer-Verlag, New York.

Efremov VV (2005) Equilibrium between genetic drift and migration at various mutation rates: a simulation analysis. *Russian Journal of Genetics*, **41**, 1283–1288.

Ennos RA (1994) Estimating the relative rates of pollen and seed migration among plant populations. *Heredity*, **72**, 250–259.

Epps C, Wehausen JD, Bleich VC, Torres SG, Brashares JS, McRae BH (2007) Optimizing dispersal and corridor models using landscape genetics. *Journal of Applied Ecology*, **44**, 71–724.

Estoup A, Beaumont M, Sennedot F, Moritz C, Cornuet J-M (2004) Genetic analysis of complex demographic scenarios: spatially expanding populations of the cane toad, *Bufo marinus*. *Evolution*, **58**, 2021–2036.

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**, 479–491.

Garrick RC, Meadows CA, Nicolas AN, Nason JD, Dyer RJ (2008) A set of polymorphic nuclear intron markers for

conservation genetics and phylogeography of Euphorbia species (Pedilanthus clade). *Conservation Genetics*, **9**, 1673–1676.

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. *Molecular Ecology*, **18**, 1916–1931.

Geffen E, Anderson MJ, Wayne RK (2004) Climate and habitat barriers to dispersal in the highly mobile grey wolf. *Molecular Ecology*, **13**, 2481–2490.

Goslee S, Urban D (2008) *Dissimilarity-Based Functions for Ecological Analysis*. CRAN repository. Available from http://cran.r-project.org/web/packages/ecodist/

Gower JC (1966) Some distance properties of latent root and vector methods used in multivariate analysis. *Biometrika*, **53**, 325–338.

GRASS Development Team (2004). *Geographic Resources Analysis Support System (GRASS), GNU General Public License*. Available from http://grass.osgeo.org

Hamilton MB, Miller JR (2002) Comparing relative rates of pollen and seed gene flow in the island model using nuclear and organelle measures of population structure. *Genetics*, **162**, 1897–1909.

Holzhauer SIJ, Ekschmitt K, Sander AC, Dauber J, Wolters V (2006) Effects of historic landscape change on the genetic structure of the bush-cricket. *Landscape Ecology*, **21**, 891–899.

Kearney M, Porter WP (2009) Mechanistic niche modelling: combining physiological and spatial data to predict species' ranges. *Ecology Letters*, **12**, 334–350.

Kimura M, Weiss GH (1964) The stepping stone model of population structure and the decrease of genetic correlation with distance. *Genetics*, **49**, 561–576.

Kutner M, Nachsheim C, Netter J (2004) *Applied Linear Regression Models*. McGraw-Hill, New York, NY.

Latta RG (2006) Integrating patterns across multiple genetic markers to infer spatial processes. *Landscape Ecology*, **21**, 809–820.

Latta RG, Mitton JB (1997) A comparison of population differentiation across four classes of gene marker in limber pine (*Pinus flexilis* James). *Genetics*, **146**, 1153–1163.

Legendre P, Anderson MJ (1999) Distance-based redundancy analysis: testing multispecies responses in multifactorial ecological experiment. *Ecological Monographs*, **69**, 1–24.

Legendre P, Lapointe FJ, Casgrain P (1994) Modeling brain evolution from behaviour: a permutation regression approach. *Evolution*, **48**, 1487–1499.

Lindell J, Ngo A, Murphy RW (2006) Deep genealogies and the mid-peninsular seaway of Baja California. *Journal of Biogeography*, **33**, 1327–1331.

Manel S, Schwartz MK, Luikart G, Taberlet P (2003) Landscape genetics: combining landscape ecology and population genetics. *Trends in Ecology and Evolution*, **18**, 189–197.

McRae BH (2006) Isolation by resistance. *Evolution*, **60**, 1551–1561.

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

Murphy MA, Evans JS, Cushman SA, Storfer A (2008) Representing genetic variation as continuous surfaces: an approach for identifying spatial dependency in landscape genetic studies. *Ecography*, **31**, 685–697.

Nason JD, Hamrick JL, Fleming TH (2002) Historical vicariance and postglacial colonization effects on the evolution of genetic structure in Lophocereus, a Sonoran Desert columnar cactus. *Evolution*, **56**, 2214–2226.

Nei M, Tajima F, Tateno Y (1983) Accuracy of estimated phylogenetic trees from molecular data. II. Gene frequency data. *Journal of Molecular Evolution*, **19**, 153–170.

Peakall R, Ruibal M, Lindenmayer DB (2003) Spatial autocorrelation analysis offers new insights into gene flow in the Australian bush rat, *Rattus fuscipes*. *Evolution*, **57**, 1182–1195.

Petit RJ, Duminil J, Fineschi S, Hampe A, Salvini D, Vendramin GG (2005) Comparative organization of chloroplast, mitochondrial and nuclear diversity in plant populations. *Molecular Ecology*, **14**, 689–701.

Pinto N, Keitt TH (2009) Beyond the least-cost path: evaluating corridor redundance using a graph-theoretic approach. *Landscape Ecology*, **24**, 253–266.

R Development Core Team (2005) *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. Available from http://www.R-project.org.

Rayfield B, Fortin MJ, Fall A (2010) The sensitivity of least-cost habitat graphs to relative cost surface values. *Landscape Ecology*, **25**, 519–532.

Riddle BR, Hafner DJ (2006) A step-wise approach to integrating phylogeographic and phylogenetic biogeographic perspectives on the history of the core North American warm deserts biota. *Journal of Arid Environments*, **66**, 435–461.

Riddle BR, Hafner DJ, Alexander LF, Jaeger JR (2000) Cryptic vicariance in the historical assembly of a Baja California Peninsular Desert biota. *Proceedings of the National Academy of Sciences, USA*, **97**, 14438–14443.

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

Slatkin M (1993) Isolation by distance in equilibrium and non-equilibrium populations. *Evolution*, **47**, 264–279.

Smouse PE, Peakall R (1999) Spatial autocorrelation analysis of individual multiallele and multilocus genetic structure. *Heredity*, **82**, 561–573.

Smouse PE, Spielman RS, Park MH (1982) Multiple-locus allocation of individuals to groups as a function of the genetic variation within and differences among human populations. *American Naturalist*, **119**, 445–463.

Sokal RR, Oden NL, Wilson C (1991) Genetic evidence for the spread of agriculture in Europe by demic diffusion. *Nature*, **351**, 143–145.

Sork VL, Smouse PE (2006) Genetic analysis of landscape connectivity in tree populations. *Landscape Ecology*, **21**, 821–836.

Spear SF, Balkenhol N, Fortin M-J, McRae BH, Scribner K (2010). Use of resistance surfaces for landscape genetic studies: considerations for parameterization and analyses. *Molecular Ecology*, **19**, 3576–3591.

Spielman RS, Smouse PE (1976) Multivariate classification of human populations. I. Allocation of Yanomama Indians to villages. *American Journal of Human Genetics*, **184**, 637–644.

Storfer A, Murphy MA, Evans JS *et al.* (2007) Putting the ''landscape'' in landscape genetics. *Heredity*, **98**, 128–142.

Takezaki N, Nei M (1996) Genetic distances and reconstruction of phylogenetic trees from microsatellite DNA. *Genetics*, **144**, 389–399.

Takezaki N, Nei M (2008) Empirical tests of the reliability of phylogenetic trees constructed with microsatellite DNA. *Genetics*, **178**, 385–392.

Taylor PD, Fahrig L, Henein K, Merriam G (1993) Connectivity is a vital element of landscape structure. *Oikos*, **68**, 571–573.

Upton DE, Murphy RW (1997) Phylogeny of the side-botched lizards (Phrynosomatidae: *Uta*) based on mtDNA sequences support for a mid-peninsular seaway in Baja California. *Molecular Phylogenetics and Evolution*, **8**, 104–113.

Wang Y-H, Yang K-C, Bridgman CL, Lin L-K (2008) Habitat suitability modelling to correlate gene flow with landscape connectivity. *Landscape Ecology*, **23**, 989–1000.

Weir BS, Cockerham CC (1984) Estimating F-statistics for the analysis of population structure. *Evolution*, **38**, 1358–1370.

Whitlock MC, McCauley DE (1990) Some population genetic consequences of colony formation and extinction: genetic correlations within founding groups. *Evolution*, **44**, 1717–1724.

Whittaker J (1990) *Graphical Methods in Applied Multivariate Statistics*. J Wiley and Son, New York.

Wright S (1931) Evolution in Mendelian populations. *Genetics*, **16**, 97–159.

Wright S (1943) Isolation by distance. *Genetics*, **28**, 114–138.

Wright S (1946) Isolation by distance under diverse systems of mating. *Genetics*, **31**, 39–59.

Zellmer AJ, Knowles LL (2009) Disentangling the effects of historic vs. contemporary landscape structure on population genetic divergence. *Molecular Ecology*, **18**, 3593–3602.