Conserv Genet
DOI 10.1007/s10592-012-0415-1

RESEARCH ARTICLE

Sample design effects in landscape genetics

Sara J. Oyler-McCance · Bradley C. Fedy · Erin L. Landguth

Received: 30 January 2012 / Accepted: 24 September 2012

Abstract An important research gap in landscape genetics is the impact of different field sampling designs on the ability to detect the effects of landscape pattern on gene flow. We evaluated how five different sampling regimes (random, linear, systematic, cluster, and single study site) affected the probability of correctly identifying the generating landscape process of population structure. Sampling regimes were chosen to represent a suite of designs common in field studies. We used genetic data generated from a spatially-explicit, individual-based program and simulated gene flow in a continuous population across a landscape with gradual spatial changes in resistance to movement. Additionally, we evaluated the sampling regimes using realistic and obtainable number of loci (10 and 20), number of alleles per locus (5 and 10), number of individuals sampled (10–300), and generational time after the landscape was introduced (20 and 400). For a simulated continuously distributed species, we found that random, linear, and systematic sampling regimes performed well with high sample sizes (>200), levels of polymorphism (10 alleles per locus), and number of molecular markers (20). The cluster and single study site sampling regimes were not able to correctly identify the generating process under any conditions and thus, are not advisable strategies for scenarios similar to our simulations. Our research emphasizes the importance of sampling data at ecologically appropriate spatial and temporal scales and suggests careful consideration for sampling near landscape components that are likely to most influence the genetic structure of the species. In addition, simulating sampling designs a priori could help guide filed data collection efforts

Keywords Partial Mantel test · CDPOP · Causal modeling · Simulation modeling · Isolation-by-distance · Isolation-by-landscape resistance · Isolation-by-barrier · Cluster sampling · Linear sampling · Systematic sampling · Random sampling

S. J. Oyler-McCance (✉) · B. C. Fedy
U.S. Geological Survey,
Fort Collins Science Center, Fort Collins CO 80526, USA
e-mail: soyler@usgs.gov

B. C. Fedy
Department of Environment and Resource Studies,
University of Waterloo, Waterloo, ON N@L 3G1, Canada

E. L. Landguth
Division of Biological Sciences, University of Montana,
Missoula, MT 59812, USA

Introduction

The field of landscape genetics aims to integrate population genetics, landscape ecology, and spatial statistics (Manel et al. 2003; Storfer et al. 2007) with the goal of quantifying the impact of landscape composition, configuration, and matrix quality on the spatial distribution of genetic variation (Holderegger and Wagner 2008; Balkenhol et al. 2009). Previous approaches to landscape genetics focused on describing and mapping populations (e.g., Pritchard et al. 2000; Dupanloup et al. 2002; Francois et al. 2006) and on identifying factors that influence rates and patterns of gene flow within and among populations (e.g., Coulon et al. 2004; Cushman et al. 2006; McRae and Beier 2007; Schwartz et al. 2009). More recent work has greatly expanded the field to include research investigating functional connectivity and landscape resistance to gene flow (Thomassen et al. 2010; Galindo et al. 2010; Selkoe et al. 2010), linking genetic pattern to ecological processes

Published online: 09 October 2012

(Bruggeman et al. 2010), comparing historical and contemporary landscape processes (Dyer et al. 2010; Knowles and Alvarado-Serrano 2010), and examining how environmental variation impacts adaptive genetic variation (Freedman et al. 2010; Manel et al. 2010; Eckert et al. 2010).

Identifying how landscapes facilitate or deter gene flow (functional connectivity) is a high priority for managers and conservation biologists charged with the management of viable populations in an ever changing world that is driven by anthropogenic forces (Agee and Johnson 1987; Trombulak and Baldwin 2010, Sork and Waits 2010). Some even advocate a paradigm shift that includes managing for change and embracing resilience-based ecosystem stewardship (e.g., managing for ecosystems that have the ability to change and adapt while remaining within critical thresholds, Chapin et al. 2009). Efforts to explicitly quantify the impact of landscape features on connectivity have provided a range of statistical approaches toward this goal (Murphy et al. 2008; Spear et al. 2010; Cushman and Landguth 2010; Shirk et al. 2010). In such analyses of functional connectivity and landscape resistance to gene flow, one area that remains largely unexplored is the sensitivity of landscape genetic analyses to variation in the sampling design used to obtain the genetic data (Balkenhol et al. 2009; Segelbacher et al. 2010; Epperson et al. 2010; Balkenhol and Landguth 2011). While the effects of the spatial sampling design on landscape genetic inference has received some attention (Murphy et al. 2008; Schwartz and McKelvey 2009), much less is known about the effects of the study design in terms of the number of sampled individuals, number of loci analyzed per individual, and number of alleles per locus on the ability to correctly and reliably identify the generating process (but see Landguth et al. 2011). Several recent papers identified this topic as among the most pressing methodological issues to address in landscape genetics (Balkenhol et al. 2009; Segelbacher et al. 2010; Epperson et al. 2010; Balkenhol and Landguth 2011).

Landguth et al. (2011) investigated the effect of study design on landscape genetics inference using a spatially-explicit, individual-based program to simulate genetic differentiation in a spatially continuous population inhabiting a landscape with gradual changes in resistance to movement. They simulated a wide range of combinations varying the number of loci, alleles per locus, and individuals sampled from the population. The authors assessed how those three aspects of study design influenced the statistical power to successfully identify the generating process among competing hypotheses of isolation-by-distance (IBD), isolation-by-barrier (IBB), and isolation-by-landscape resistance (IBR) using a causal modeling approach with partial Mantel tests (Mantel 1967). Further, they modeled the statistical power to identify the generating process as a response surface for equilibrium and non-equilibrium conditions after introduction of IBR. However, their study used a spatial random sample drawn from a continuously distributed underlying population to test their ability to correctly identify the generating process. In reality, a truly random sampling design may be very difficult to achieve in the field. Building on the work of Landguth et al. (2011), we used the same spatially continuous population inhabiting a landscape with gradual changes in resistance to movement, yet in addition to a random sample design, we also investigated alternative sampling designs that emulated more realistic conditions that might be considered for implementation in a field study of a continuously distributed organism. The sampling designs we consider here include random, linear (sampling that would be associated with a linear transect, such as a river, road, or trail), systematic (sampling that involves attempting to cover the landscape in a systematic grid), cluster (sampling several groups of individuals where groups include individuals that are close together), and single study site (sampling one group of individuals that are all close to one another as might be reflective of a single study site) sampling designs. Our objectives were to evaluate and compare the performance of the five different sampling designs in terms of their capacity to correctly identify the generating landscape process. In addition, we examined how these designs were influenced by variation in sample sizes, number of alleles per locus, number of loci, and generational time after IBR (i.e., the generating landscape process) was introduced.

Models and methods

Study area, population, and genetic data

Our goal was to assess the sensitivity of landscape genetic inference to sampling design, rather than assessing sensitivity to characteristics of the landscape (e.g., barriers), while comparing the sampling design sensitivity to the results obtained by Landguth et al. (2011). Therefore, we used the same landscape resistance model, population, and genetic data generated for all simulations by Landguth et al. (2011). The landscape resistance surface was adopted from an empirically tested model of landscape resistance to American black bear (*Ursus americanus*) movement in Northern Idaho, USA, from Cushman et al. (2006) to ensure that the simulated scenario mimicked a realistic system (Fig 1a; extent of approximately 3,000-km^2 with resistance values ranging from 1 to 62 in 90-m grid cells). The landscape pattern is represented through a resistance surface with grid cell values representing costs of movement through the landscape. The resistance surface is a

combination of forest cover, elevation, and roads. On this surface, 1,000 individual locations in the Universal Transverse Mercator (UTM) coordinate system were initialized by populating grid cell values in a hexagonal pattern at 1.6-km spacing unless the cell value was greater than 6 (Fig. 1a). This was done to place individuals in habitat that was relatively suitable for the species given this landscape resistance hypothesis.

Genetic exchange across 500 non-overlapping generations among the 1,000 individuals as a function of individual-based movement through mating and dispersal on the given landscape was simulated with a spatially-explicit, landscape genetic program (CDPOP v0.85, Landguth and Cushman 2010). In CDPOP, mating and dispersal are modeled as probabilistic functions of cumulative cost between individual locations across these resistance surfaces (i.e., least-cost path or step-wise summed resistance values between locations). These movement (mating and dispersal) cost functions are scaled to a user-specified maximum dispersal distance. Movement (mating and dispersal) was simulated between these individuals as a function of the inverse-square of cost scaled to a maximum movement distance of 39,200 cost-units, which is ~22 % of the total cost distance on the landscape suggesting a moderate-range dispersing organism and corresponding to the range of positive spatial autocorrelation of genetic

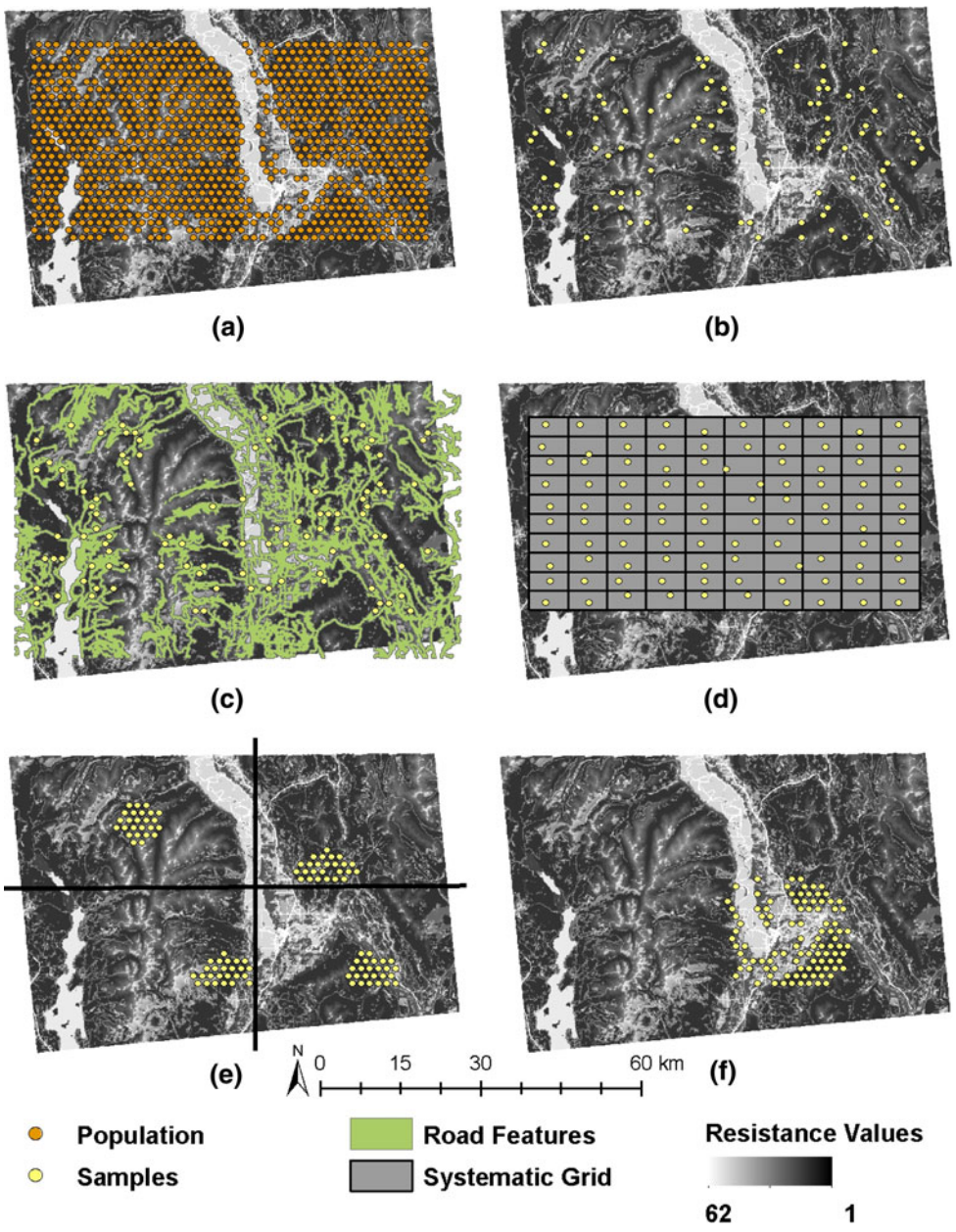

Fig. 1 Sample designs on an isolation-by-landscape resistance surface. **a** The total population of 1,000 individuals, **b** Random sampling design—samples were randomly chosen (an example with sample size 100 is shown). **c** Linear sampling design—327 possible individuals to sample from chosen 250 m from all road features in the study area. An example with sample size 100 is shown. **d** Systematic sampling design—A systematic grid was placed on the study area and samples nearest the center of each cell were taken. 3 × 3, 10 × 10, 14 × 14, and 18 × 18 grids were placed resulting in sample sizes of 9, 99, 195, and 305 (some grid pixels did not have an individual within) (the 10 × 10; a sample size of 99 is shown). **e** Cluster sampling design—the study site was divided into four quadrats and four random individuals were selected. Samples consisted of those four chosen individuals and their nearest neighbors. A sample size of 100 is shown. **f** Single study site sampling design −10, 100, 200, and 300 clustered samples were taken from a randomly chosen individual (an example of 100 samples is shown). (Color figure online)

relatedness among individuals as a function of cost distance in the Cushman et al. (2006) data set. This maximum cost distance value constrains all mate choices and dispersal distances to be less than or equal to 39,200 cost-units apart with probability of mating or dispersal distance within that limit specified by an inverse-square probability function (Landguth and Cushman 2010).

The genotypes were initialized for the 1,000 individuals by randomly assigning allelic states across the initial population with a random sex assignment that contained 25 loci (which were subsequently sub-sampled down to two levels; 10 and 20) with the k-allele mutation rate set to 0. Fifty replicate simulations were conducted for two levels of alleles per locus (5 and 10 maximum alleles at the beginning of each simulation run, thus simulating a panmictic initial population with maximum allelic diversity). CDPOP simulates spatially-referenced genotypes for all individuals at each generation with independent assortment and no linkage disequilibrium in Mendelian inheritance. Mating parameters were set in CDPOP to represent a population that was heterosexual with a polygamous structure (females mated without replacement and males mated with replacement). Offspring parameters were set such that each female had a number of offspring with random sex assignment following a Poisson process with mean of 4. This guaranteed a positive lambda value that ensures that all spatial locations were filled through dispersal movement at each generational time step and avoids empty locations that require immigrants from an outside population. This maintained a constant population of 1,000 at every generation and the remaining offspring were discarded once all the 1,000 locations were occupied by a dispersing individual. This is equivalent to forcing emigrants out of the study area once all available home ranges are occupied (Landguth and Cushman 2010).

Once simulations were completed, individuals were sampled following five study designs: (1) random, (2) linear, (3) systematic, (4) cluster, and (5) study site. We also varied the number of alleles per locus (5 or 10) and the number of loci used (10 or 20) to emulate more realistic scenarios that are common to current landscape genetic studies. Simulations were sampled at two time periods representing non-equilibrium (generation 20) and equilibrium (generation 400) conditions.

Random sampling design

For the random sampling design, we randomly sequentially sub-sampled from the entire population 10, 100, 200, and 300 individuals (Fig. 1b). We took a unique sample across each of the 50 replicate simulations, initial number of alleles per locus (5 and 10), and number of loci (10 and 20) at generations 20 and 400 resulting in 1,600 data sets.

Linear sampling design

The linear sampling design was used to emulate sampling along a linear feature such as a river, trail, or road, or line transect sampling as has been implemented in studies of plants and small mammals (Gamache et al. 2003; Gauffre et al. 2008). We used the original road feature (TIGER 2007; http://www.census.gov/geo/www/tiger/) in the Cushman et al. (2006) resistance surface as linear features from which to sample. We buffered all the roads in the study area at a distance of 250-m. 327 out of the 1,000 individuals fell within this buffered distance (Fig. 1c). We then sampled sequentially random individuals (10, 100, 200, and 300) with a unique draw across each of the 50 simulations, initial number of alleles per locus (5 and 10), and number of loci (10 and 20) at generations 20 and 400 resulting in 1,600 data sets.

Systematic sampling design

The systematic sampling design aims to obtain samples spread evenly (at regular intervals) throughout the study area. Such sampling designs are implemented often in non-invasive genetic sampling studies (Kendall et al. 2008; Barba et al. 2010). A uniform grid design was used for the systematic sampling approach (Fig. 1d). Four uniform grids were placed on the study area and the individuals that were closest to the center of each grid cell were sampled. If an individual did not fall within a grid cell, then that grid was skipped. Grids that were sized 3×3, 10×10, 14×14, and 18×18 were placed resulting in sample sizes of 9, 99, 217, and 305, respectively. These sample sizes were used to resample the simulated population for each of the 50 simulations, initial number of alleles per locus (5 and 10), and number of loci (10 and 20) at generations 20 and 400 resulting in 1,600 data sets.

Cluster sampling design

The cluster sampling design was used to emulate the situation where samples are collected in concentrated areas due to logistics (limited access), scale (i.e., including multiple study sites), or opportunistically (e.g., obtaining samples from hunters) (Martinez et al. 2002; Oyler-McCance et al. 2005; Cegelski et al. 2006; Pernetta et al. 2011). The study area was divided into four areas shown in Fig. 1e and four random individuals were chosen for each of the 50 replicates and sample sizes of 8, 100, 200, and 300 (clusters of 2, 25, 50, and 75, respectively) were selected based on the nearest neighbor to the four individuals. Fifty different random four individuals and respective sample size cluster were used to resample the simulated population for each of the 50 replicates, initial

number of alleles per locus (5 and 10), and number of loci (10 and 20) at generations 20 and 400 resulting in 1,600 data sets.

Study site sampling design

The study site sampling design was used to emulate the situation where a single study site is chosen for sampling (Fig. 1f). Field research is often conducted at a single study site. Our intention with this sampling design was to address potential concerns and model performance when extrapolating results from a single study site to novel areas, outside of where models were developed (Miller et al. 2004). Fifty random individuals were chosen for each of the 50 replicates and sample sizes of 10, 100, 200, and 300 were selected based on nearest neighbor. The 50 initial random individuals and respective sample sizes were used to resample the simulated population for each of the 50 replicates, initial number of alleles per locus (5 and 10), and number of loci (10 and 20) at generations 20 and 400 resulting in 1,600 data sets.

Statistical analysis of simulation results

Each simulation was evaluated at two time steps (20 and 400 generations). Equilibrium partial Mantel r was the value of r once spatial genetic equilibrium reached an approximate asymptote. Preliminary analysis showed that results for partial Mantel tests were most variable after 20 generations, whereas after 400 generations, spatial genetic equilibrium was achieved in all simulations and the association between landscape pattern and genetic structure had stabilized.

Inter-individual genetic distance was calculated as the proportion of shared alleles (Bowcock et al. 1994), and landscape-cost distance model (IBR distance) was calculated for each pair of sampling locations as the cumulative cost associated with traversing the least cost path from one sampling location to the other using COSTDISTANCE in ArcGIS v9.0 (ESRI 1999–2008). Euclidean distance (IBD distance) was calculated from the Universal Transverse Mercator coordinates between all pairs of individuals. The barrier-cost distance (IBB distance) was represented as a model matrix similar to Legendre and Legendre (1998), with pair-wise distance equal to 1 for two individuals from opposite sides of a complete barrier separating half of the 1,000 individuals, and pair-wise distance equal to 0 for two individuals from the same side of the barrier (panmixia).

For each scenario, we performed a partial Mantel test to correlate genetic distance to IBR distance accounting for IBD distance using the library ecodist v1.1.3 (Goslee and Urban 2007) in the statistical software package R (R Development Core Team 2009). Due to the highly correlated hypotheses of IBR, IBD, and IBB (Mantel $r = 0.938$ for IBD to IBR, Mantel $r = 0.984$ for IBD to IBB, and Mantel $r = 0.972$ for IBB to IBR), we used causal modeling, which involves a series of diagnostic Mantel and partial Mantel tests (Legendre and Legendre 1998). These tests included a simple Mantel test to correlate genetic distance to IBR distance and partial Mantel tests to correlate genetic distance to IBR distance accounting for IBD distance, IBR distance accounting for IBB distance, IBB distance accounting for IBR distance, and IBD distance accounting for IBR distance. For all tests, we calculated Mantel's r and P value based on 1,999 permutations, corresponding to a 0.005 precision for the cutoff value, $\alpha = 0.05$.

Successful identification of the generating process (i.e., IBR) required a combination of three significant correlations with genetic distance (Mantel test of IBR distance, partial Mantel test of IBR distance accounting for IBD distance, and partial Mantel test of IBR distance accounting for IBB distance) and two non-significant correlations with genetic distance (partial Mantel test of IBD distance accounting for IBR distance and partial Mantel test of IBB distance accounting for IBR distance).

Based on the 50 replicate simulations for each parameter combination, we examined the probability of successfully identifying the correct landscape resistance scenario for each sampling design. We included three covariates with values common in field-based landscape genetic analyses: sample size, number of loci, and number of alleles per locus. All analyses were conducted separately for two time steps at 20 and 400 generations. The modeled response variable was the probability of correctly identifying the generating process (IBR) as described above, and this probability was skewed towards success (i.e., probability = 1). We were interested in the differences among sampling designs (random, linear, systematic, clustered) and their probability of successfully identifying the underlying process. The effects of sampling design on the probability of success were tested using a general linear model ANCOVA in which the response variable was the probability of success. Probability of success was influenced by the three covariates: number of alleles, loci, and sample size (Landguth et al. 2011). Therefore, these variables were entered as covariates in the modeling process in order to remove the predictable variance associated with the covariates from the error term in model estimation.

Results

Overall, the random, linear, and systematic sampling designs produced relatively predictable results. As we

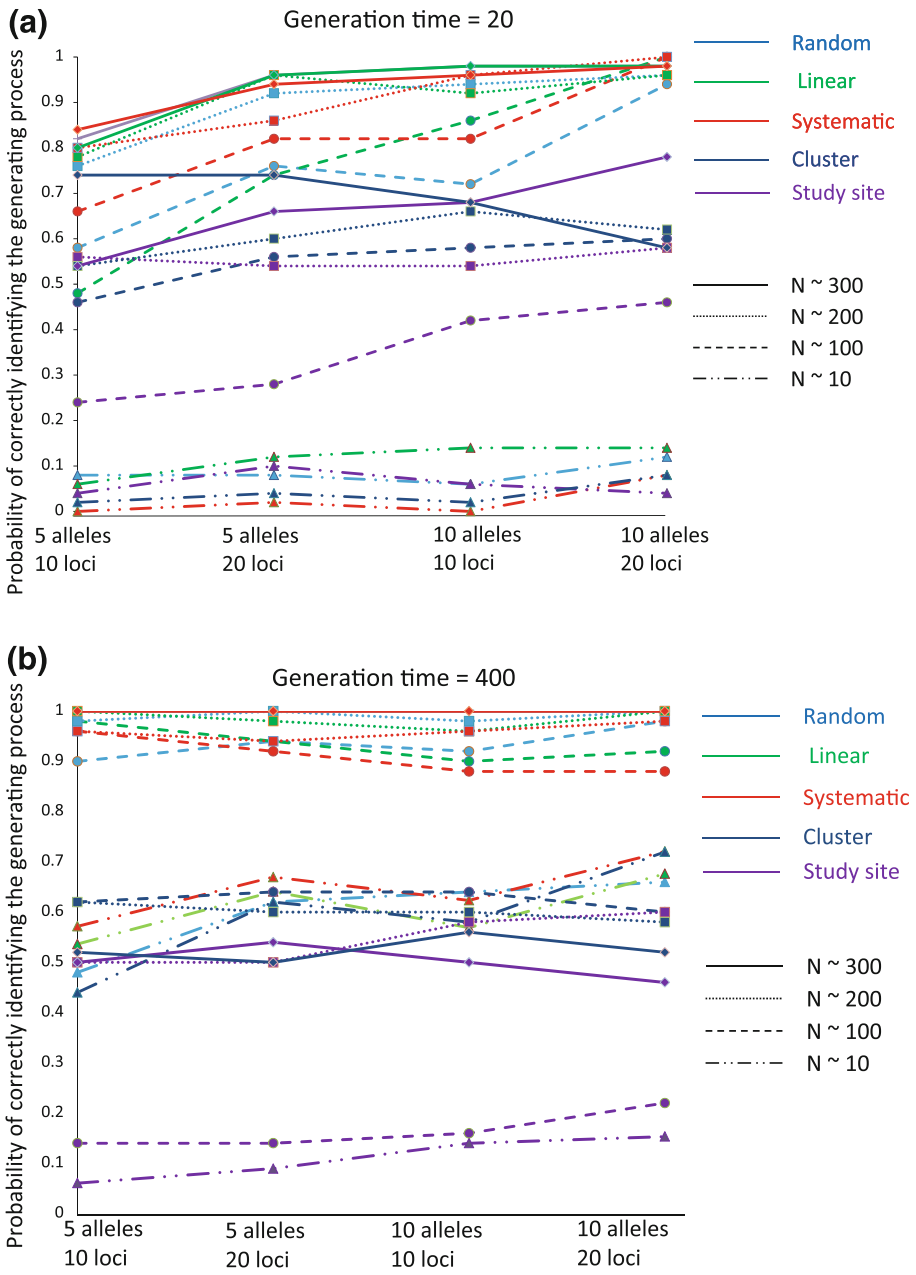

Fig. 2 Probability of successfully identifying the generating process for all combinations of the number of alleles, number of loci used, sample size, generational time, and sampling design. **a** shows the relationships at disequilibrium (generation 20), and **b** shows the relationships at equilibrium (generation 400). The sampling designs are represented by *different colors* (random is *light blue*, linear is *green*, systematic is *red*, cluster is *dark blue*, and study site is *purple*). The different sample sizes are represented with different line styles. At generation 400 when the sample size was ~300, the random, linear, and systematic sampling designs all had a probability of 1 with all combinations of alleles per locus and number of loci used. (Color figure online)

increased the number of alleles, loci, sample size, and generation time, the probability of successfully identifying the underlying generating process also increased (Fig. 2). The positive relationship between our covariates and the probability of success was present for both the non-equilibrium (generation 20) and equilibrium (generation 400) samples. However, the pattern of increasing probability of success with increasing values in the covariates was not present in the cluster and the study site sampling designs (Fig. 2). The 95th percentile for the probability of success distribution in the cluster and study site sampling designs did not cross 0.95 under any combination of alleles, loci, nor sample size at either generation. These results demonstrated an overall much lower probability of success compared to the random, linear, and systematic sampling designs. Furthermore, the random, linear, and systematic sampling designs had a higher probability of success in the equilibrium population samples (generation 400), than the non-equilibrium (generation 20). The relationship was reversed for the study site sampling design in which the probability of success was lower in the equilibrium population samples. The pattern for the cluster sampling design was different at different samples sizes. In lower sample sizes, the cluster design had a higher probability of success at generation 400 but with higher sample sizes the probability of success was greater at generation 20.

The ANCOVA analysis supported the conclusions based on the distributions presented by Fig. 3 There were significant differences between the slopes of the regression fit to each sampling scheme, suggesting the difference among groups—adjusted for the covariates—are unlikely to have occurred by chance. This was true for analysis at generation 20 ($F_{13,\ 79} = 90.13$, $P < 0.001$, $r^2 = 0.95$) and generation 400 ($F_{13,\ 79} = 71.65$, $P < 0.001$, $r^2 = 0.95$). The predicted values from the ANCOVA suggested the study site and cluster sampling methods performed poorly and the other sampling designs were essentially equal at generations 20 and 400 (Fig. 3a–b). Post hoc pairwise Fisher–Hayter comparisons found significant ($P < 0.05$) differences between the study site sampling design and all others at generation 20, with the exception of the cluster sampling design (i.e., no significant difference between cluster and study site). The cluster sampling design was also significantly different from all methods except the study site design. Pairwise comparisons among the random, linear, and systematic designs did not differ statistically ($P > 0.05$). The random design had a mean probability of success of 0.67 (median = 0.79). The linear design had a mean of 0.68 and median = 0.83. The systematic design had a mean (0.67) and median (0.83). Thus, the random, linear, and systematic designs were equal in their capacity to correctly identify the underlying processes at generation 20 and the study site and cluster designs were less accurate, and not different from each other in post hoc statistical comparisons. The results of the post hoc pairwise Fisher–Hayter tests were similar at generation 400; however, the cluster sampling design performed better at generation 400 compared to the study site design. All post hoc pairwise Fisher–Hayter comparisons at generation 400 were significantly different from the study site design ($P < 0.05$), including the cluster design. The cluster design performed better than the study site design; however, the cluster design was significantly different from the other 3 designs. There was no statistically significant difference between combinations of linear, systematic, and random ($P > 0.05$). Again, the random and linear sampling designs had the highest mean and median probability of success (random $\bar{x} = 0.88$, median = 0.98; linear $\bar{x} = 0.90$, median = 0.97; systematic $\bar{x} = 0.84$, median = 0.95; study site $\bar{x} = 0.32$, median = 0.34; clustered mean = 0.59, median = 0.60).

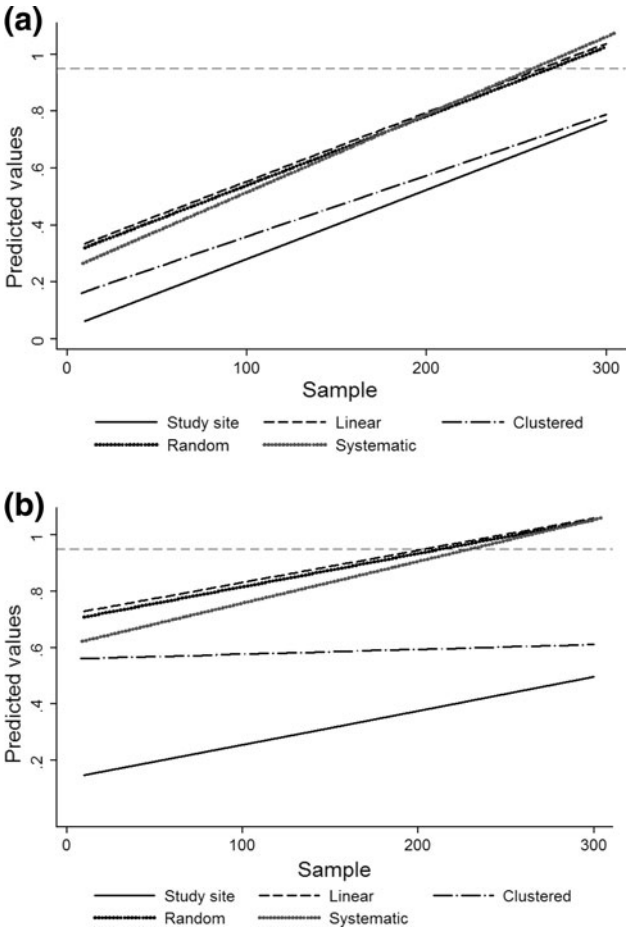

Fig. 3 Model predicted values across sample size for each of the three sampling designs for **a** generation 20 and **b** generation 400. The predicted values were generated using a general linear model ANCOVA with probability of success as the response variable, sampling design as the grouping variable, and alleles, loci, and sample size entered as covariates. The *horizontal dashed line* indicates a value of 0.95 in the response variable

Discussion

Our study examined the impact of five different sampling regimes. The random design was a true spatial random sample drawn from a continuously distributed underlying population, and represented an idealistic sampling approach that was predicted to perform very well in terms of identifying the underlying processes. Generally, the random, linear, and systematic sampling regimes behaved as expected in terms of our covariates (Figs. 2, 3) for this resistance landscape and simulated population. Those three sampling regimes demonstrated higher probabilities of correctly detecting the generating process with increasing levels of polymorphisms, numbers of loci, sample sizes, and equilibrium simulation conditions (generation 400). These patterns are consistent with those presented in other simulations studies (Murphy et al. 2008; Landguth et al. 2011). Fortunately, two of the more realistic sampling regimes (linear and systematic) performed nearly as well as the more idealistic random sampling design described in Landguth et al. (2011). However, the cluster and single study site sampling regimes did not perform well and were unable to identify the correct generating process with a probability of success >95 % (Figs. 2, 3). Contrary to the

results from the other sampling regimes, at higher sample sizes the cluster and study site sampling regimes performed worse at equilibrium conditions (generation 400) than at non-equilibrium conditions (generation 20).

The cluster sampling regime is one of the more common sampling approaches in genetics and therefore, the overall poor performance of this approach is of particular interest. The cluster sampling regime (by definition) groups individuals that are close together and likely to be closely related. Thus, sampling clustered portions of the individuals (specifically those close together and highly related) captures only a small subset of the resistance surface and thereby restricts the opportunity to detect responses to widely different landscape features from across the study area. Our clustering design in this example divided the study area into four quadrants, with one division (north/south) roughly following the valley and major roadway. This division was deliberate, as many authors have emphasized the need to incorporate a priori hypotheses into study designs (Balkenhol et al. 2009; Anderson et al. 2010). Even worse, is the single study site design, in which a group of samples are all collected from one area and used to make inference to a much larger landscape. There are many parallels between habitat selection studies that assess habitat connectivity and landscape genetic studies that assess functional connectivity. Multiple habitat selection studies have shown that the application of habitat selection models to novel areas outside of the study sites used to develop the models should be done with extreme caution (Miller et al. 2004; Coe et al. 2011). Good model performance in novel areas requires similar composition of resources to the single study site, or sufficient generality in model form. The spatial interpolation and extrapolation of habitat selection models is most accurate when the variation and availability of habitat types is approximately the same in the novel areas (Mladenoff et al. 1999; Aarts et al. 2008). Our results suggest that similar cautions are prudent when extrapolating landscape genetics studies outside of the study sites in which they were developed. Similar to the cluster design, the single study site design only captured a small amount of the variation in the underlying resistance surface, and therefore, had limited capacity to detect the underlying landscape components influencing the functional connectivity.

Several authors have emphasized the need to carefully consider the spatial scale at which life processes take place (i.e., home range or dispersal) and advocate that sampling regimes and analysis metrics should be dictated explicitly by the ecological characteristics of the species being studied (Balkenhol et al. 2009; Schwartz and McKelvey 2009; Anderson et al. 2010; Cushman and Landguth 2010; Jaquiéry et al. 2011). This simulation study used a resistance surface developed for black bears in North America, and individuals were simulated using black bear life history characteristics (i.e., continuously distributed throughout suitable habitat, mating and dispersal emulating black bear biology with moderate species-specific dispersal strategies relative to their distribution). Thus, sampling only one (single study site) or a few (cluster) regions of the study area and then making inferences about genetic responses to landscape processes across the entire study area and varying levels of landscape resistance is clearly inadvisable in this example. This underscores the importance of carefully considering the life history of the organism and designing sampling regimes appropriately to avoid flawed inferences. Cluster sampling is likely more appropriate for organisms that are patchily distributed such as amphibians, alpine species, and some plants (Jacquemyn et al. 2006; Fedy et al. 2008; Murphy et al. 2010), than for species whose distribution is more continuous.

The random sampling design represented an idealistic sampling approach, that would be difficult, if not impossible, to replicate in a field study. The systematic sampling regime performed nearly as well as the random, particularly at higher sample sizes. However, the linear sampling approach, which has been used in many field studies, performed equally as well in our study as the random design. The excellent performance of the linear sampling design in our study is likely a function of the road network being distributed over the entire study area, and thus resulting in a sample that well represents the underlying resistance surface. Our resistance surface was developed by Cushman et al. (2006) and has minimum resistance at medium elevation (classification of low elevation of 7, medium elevation of 1, and high elevation of 10), in forested areas (classification of 1 for forested areas and 10 in non-forested areas), and away from roads (particularly paved highways, in the classification of roads giving 0 resistance to non-roads, 5 to minor roads, and 50 for interstates). That is, the resistance surface results in a moderate cost to individuals that cross minor roads and a high cost to cross major highways (i.e., a 50 classification compared to the other data layers, which had a high classification of 10). By sampling on both sides of minor and major roads using the linear sampling regime, we were likely to capture the extreme genetic distances across this particular boundary (similar to sampling on both sides of a complete barrier (Landguth et al. 2010)). Thus, it was likely the focus on sampling on both sides of a barrier that increased our probability of correctly identifying the underlying resistance surface (nearly the highest classification in the Cushman et al. (2006) surface). Additionally, in this simulation scenario the coverage of roads was quite extensive (Fig. 1c). If this simulation study were carried out in a landscape with fewer roads the linear sampling design may not have performed as well as the random and could potentially be outperformed by the systematic

sampling design. This again underscores the importance of carefully considering both the spatial and temporal patterns of not only the species in question but also the ecological and anthropogenic forces that act upon that species when designing a landscape genetic study (Anderson et al. 2010; Segelbacher et al. 2010; Cushman and Landguth 2010).

Landscape genetics is a relatively new field of study and it is still unclear how sampling strategies affect our inferences and conclusions (Balkenhol et al. 2009; Segelbacher et al. 2010; Epperson et al. 2010; Balkenhol and Landguth 2011). This study establishes the foundation for future research on the influence of sampling strategies on our conclusions in field-based landscape genetics studies. For example, this study investigated only one model of landscape resistance. While this is a logical first step in exploring the behavior of different sampling regimes, it would be valuable to extend this analysis to a wide range of alternative landscapes that vary in regards to landscape composition, complexity, and strength (i.e., as our linear sampling results suggest the importance of sampling across different gradients in resistance). Several authors have emphasized the importance of replicating analyses across multiple study areas to provide a more generalized view of the relative influence of landscape structure on gene flow (Segelbacher et al. 2010; Short Bull et al. 2010) and this is true for simulation studies as well. Additionally, this study used Mantel and partial Mantel testing, a method whose use is not above criticism. Raufaste and Rousset (2001) suggested that partial Mantel tests were "inadequate" because the permutation process was erroneous and the associated *P* value did not represent type I error. Legendre and Fortin (2010) have shown that the power to detect a spatial relationship when one is present in the data can be lower using Mantel tests than alternative methods that are not based on pair-wise distances. However, Legendre and Fortin (2010) recognized that Mantel tests are appropriate when testing hypotheses that can only be formulated using pair-wise distances. Guillot and Rousset (2012) disagree with Legendre and Fortin (2010), however, arguing that their simulations did not include data with autocorrelation, therefore not speaking to the criticisms of partial Mantel tests. It is not our intension here to assess the performance of the partial Mantel test in relation to other analytical techniques, but rather to investigate the effects of sampling design when using Mantel tests as they are commonly used in landscape genetics (Storfer et al. 2010). Future research focused on exploring the power of Mantel tests compared to other distance-based methods, such as distance based redundancy analyses should help resolve this issue. Finally, this simulation emulates a relatively mobile animal population (black bear) that is continuously distributed across the landscape. The results of this study may not apply to species with non-continuous distributions, those with low levels of mobility or dispersal, or those with less generalizable life characteristics (e.g., species with highly skewed mating systems, species where the behavior of one gender is radically different from the other).

Conclusions

We demonstrated that sampling design is an important factor that can influence our ability to correctly infer the impact of landscape pattern on gene flow. While previous research (Landguth et al. 2011) described the relationship among some elements of sample design (sample size, level of polymorphism of markers, and number of markers) questions about the applicability of that research to empirical study designs where a random sampling design may be difficult to achieve, remained. This study showed that some more realistic, and logistically achievable, sampling regimes (linear and systematic) performed nearly as well as the random sampling design implemented in Landguth et al. (2011) and at realistic numbers of individuals sampled. For a simulated continuously distributed species, we found that random, linear, and systematic sampling regimes performed well with high sample sizes (>200) and higher levels of polymorphism (10 alleles per locus) and number of molecular markers (20). While this number of loci and level of polymorphism have been difficult to achieve in the past, next generation sequencing methods can now resolve those issues (Castoe et al. 2012). The cluster and single study site sampling regimes were not able to correctly identify the generating process using the Mantel approach and thus, are not advisable strategies for scenarios similar to our simulations. Our research emphasizes the importance of sampling data at ecologically appropriate spatial and temporal scales, with careful consideration of high resistance landscape components that are likely to influence the species genetic structure. Additionally, simulating sampling designs a priori could help guide field data collection efforts.

Acknowledgments The use of any trade, product, or firm names is for descriptive purposes only and does not imply endorsement by the U.S. Government. We thank C. Funk and three anonymous reviewers for helpful comments on this manuscript.

References

Aarts G, MacKenzie M, McConnell B, Fedak M, Matthiopoulos J (2008) Estimating space-use and habitat preference from wildlife telemetry data. Ecography 31:140–160

Agee JK, Johnson DR (1987) Ecosystem management for parks and wilderness. University of Washington Press, Seattle

Anderson CD, Epperson BK, Fortin M-J, Holdregger R, James PM, Rosenberg M, Scriber KT, Spear ST (2010) Considering spatial

and temporal scale in landscape-genetic studies of gene flow. Mol Ecol 19:3565–3575
Balkenhol N, Landguth EL (2011) Simulation modeling in landscape genetics: on the need to go further. Mol Ecol 20:667–670
Balkenhol N, Gugerli F, Cushman SA, Waits LP, Coulon A, Arntzen JW, Holderegger R, Wagner HH (2009) Identifying future research needs in landscape genetics: where to from here? Landsc Ecol 24:455–463
Barba MD, Waits LP, Genovesi P, Randi E, Chirichella R, Cetto E (2010) Comparing opportunistic and systematic sampling methods for non-invasive genetic monitoring of a small translocated brown bear population. J Appl Ecol 47:172–181
Bowcock AM, Ruiz-Linares A, Tomfohrde J, Minch E, Kidd JR, Cavalli-Sforza LL (1994) High resolution of human evolutionary trees with polymorphic microsatellites. Nature 368:455–457
Bruggeman DJ, Wiegand T, Fernandez N (2010) The relative effects of habitat loss and fragmentation on population genetic variation in the red-cockaded woodpecker (Picoides borealis). Mol Ecol 19:3679–3691
Castoe TA, Poole AW, de Koning APJ, Jones KL, Tomback DF, Oyler-McCance SJ, Fike JA, Lance SL, Streicher JW, Smith EN, Pollock DD (2012) Rapid microsatellite identification from Illumina paired-end genomic sequencing in two birds and a snake. PLoS ONE 7(2):e30953. doi:10.1371/journal.pone.0030953
Cegelski CC, Waits LP, Anderson NJ, Flagstad O, Strobeck C, Kyle CJ (2006) Genetic diversity and population structure of wolverine (Gulo gulo) populations at the southern edge of their current distribution in North America with implications for genetic viability. Conserv Genet 7:197–211
Chapin S, Kofinas GP, Folke C (2009) Principles of ecosystem stewardship: resilience-based natural resource management in a changing world. Springer Science and Business Media, New York
Coe PK, Johnson BK, Wisdom MJ, Cook JG, Vavra M, Nielson RM (2011) Validation of elk resource selection models with spatially independent data. J Wildl Manage 75:159–170
Coulon A, Cosson JF, Angibault JM, Cargnelutti B, Galan M, Morellet N, Petit E, Aulagnier S, Hewison AJM (2004) Landscape connectivity influences gene flow in a roe deer population inhabiting a fragmented landscape: an individual-based approach. Mol Ecol 13:2841–2850
Cushman SA, Landguth EL (2010) Spurious correlations and inference in landscape genetics. Mol Ecol 19:3592–3602
Cushman SA, McKelvey KS, Hayden J, Schwartz MK (2006) Gene-flow in complex landscapes: testing multiple models with causal modeling. Am Natur 168:486–499
Dupanloup I, Schneider S, Excoffier L (2002) A simulated annealing approach to define the genetic structure of populations. Mol Ecol 11:2571–2581
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
Eckert AJ, Bower AD, González-Martínez SC, Wegrzyn JL, Coop G, Neale DB (2010) Back to nature: ecological genomics of loblolly pine (Pinus taeda, Pinaceae). Mol Ecol 19:3789–3805
Epperson BK, McRae BH, Scribner K, Cushman SA, Rosenberg MS, Fortin M-J, James PA, Murphy M, Manel S, Legendre P, Dale MRT (2010) Utility of computer simulations in landscape genetics. Mol Ecol 19:3549–3564
ESRI (1999–2008) Environmental System Research Institute, Redlands
Fedy BC, Martin K, Ritland C, Young J (2008) Genetic and ecological data provide incongruent interpretations of population structure and dispersal in naturally subdivided populations of white-tailed ptarmigan (Lagopus leucura). Mol Ecol 17:1905–1917
François O, Ancelet S, Guillot G (2006) Bayesian clustering using hidden markov random fields in spatial population genetics. Genetics 174:805–816
Freedman AH, Thomassen HA, Buermann W, Smith TB (2010) Genomic signals of diversification along ecological gradients in a tropical lizard. Mol Ecol 19:3773–3788
Galindo HM, Pfeiffer-Herbert AS, McManus MA, Chao Y, Chai F, Palumbi SR (2010) Seascape genetics along a steep cline: using genetic patterns to test predictions of marine larval dispersal. Mol Ecol 19:3692–3707
Gamache I, Jaramillo-Correa JP, Payette S, Bousquet J (2003) Diverging patterns of mitochondrial and nuclear DNA diversity in subartic black spruce: imprint of a founder effect associated with postglacial colonization. Mol Ecol 12:891–901
Gauffre B, Estoup A, Bretagnolle V, Cosson JF (2008) Spatial genetic structure of a small rodent in a heterogenous landscape. Mol Ecol 17:4619–4629
Goslee SC, Urban DL (2007) The ecodist package for dissimilarity-based analysis of ecological data. J Stat Software 22:1–19
Guillot G, Rousset F (2012) On the use of simple and partial Mantel tests in presence of spatial autocorrelation. arXiv:1112.0651v1 [q-bio.PE]
Holderegger R, Wagner HH (2008) Landscape genetics. Bioscience 58:199–207
Jacquemyn H, Brys R, Honnay O, Hermy M, Roldán-Ruiz I (2006) Sexual reproduction, clonal diversity and genetic differentiation in patchily distributed populations of the temperate forest herb Paris quadrifolia (Trilliacaea). Oecologia 147:434–444
Jaquie′ry J, Broquet T, Hirzel AH, Yearsley J, Perrin N (2011) Inferring landscape effects on dispersal from genetic distances: how far can we go? Mol Ecol 20:692–705
Kendall KC, Stetz JB, Roon DA, Waits LP, Boulanger JB, Paetkau D (2008) Grizzly bear density in Glacier National Park, Montana. J Wildl Manage 72:1693–1705
Knowles LL, Alvarado-Serrano DF (2010) Exploring the population genetic consequences of the colonization process with spatio-temporally explicit models: insights from coupled ecological, demographic and genetic models in montane grasshoppers. Mol Ecol 19:3727–3745
Landguth EL, Cushman SA (2010) CDPOP: a spatially explicit cost distance population genetics program. Mol Ecol Res 10:156–161
Landguth EL, Cushman SA, Schwartz MK, McKelvey KS, Murphy M, Luikart G (2010) Quantifying the lag time to detect barriers in landscape genetics. Mol Ecol 19:4179–4191
Landguth EL, Fedy BC, Oyler-McCance SJ, Garey AL, Emel SL, Mumma M, Wagner HH, Fortin M-J, Cushman SA (2011) Effects of sample size, number of markers, and allelic richness on the detection of spatial genetic pattern. Mol Ecol Res. doi: 10.1111/j.1755-0998.2011.03077.x
Legendre P, Fortin M-J (2010) Comparison of the Mantel test and alternative approaches for detecting complex multivariate relationships in the spatial analysis of genetic data. Mol Ecol Res 10:831–844
Legendre P, Legendre L (1998) Numerical ecology, 2nd English edn. Elsevier, Amsterdam
Manel S, Schwartz MK, Luikart G, Taberlet P (2003) Landscape genetics: combining landscape ecology and population genetics. Trends Ecol Evol 18:189–197
Manel S, Poncet BN, Legendre P, Gugerli F, Holderegger R (2010) Common factors drive adaptive genetic variation at different spatial scales in Arabis alpina. Mol Ecol 19:3824–3835
Mantel N (1967) The detection of disease clustering and a generalized regression approach. Cancer Res 27:209–220
Martinez JG, Carranza J, Fernandez-Garcia JL, Sanchez-Prieto CB (2002) Genetic variation of red deer populations under hunting exploitation in southwestern Spain. J Wildl Manage 66:1273–1282

McRae BH, Beier P (2007) Circuit theory predicts gene flow in plant and animal populations. Proc Natl Acad Sci USA 104:19885–19890

Miller JR, Turner MG, Smithwick AH, Dent CL, Stanley EH (2004) Spatial extrapolation: the science of predicting ecological patterns and processes. Bioscience 4:310–320

Mladenoff DJ, Sickley TA, Wydeven AP (1999) Predicting gray wolf landscape recolonization: logistic regression models vs. new field data. Ecol Appl 9:37–44

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

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

Oyler-McCance SJ, Taylor SE, Quinn TW (2005) A multilocus genetic survey of greater sage-grouse across their range. Mol Ecol 14:1293–1310

Pernetta AP, Allen JA, Beebee TJC, Reading CH (2011) Fine-scale population genetic structure and sex-biased dispersal in the smooth snake (Coronella austriaca) in southern England. Heredity 107:231–238

Pritchard JK, Stephens M, Donnelly PJ (2000) Inference of population structure using multilocus genotype data. Genetics 155:945–959

Raufaste N, Rousset F (2001) Are partial mantel tests adequate? Evolution 55:1703–1705

R Development Core Team (2009) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org

Schwartz MK, McKelvey KS (2009) Why sampling scheme matters: the effect of sampling scheme on landscape genetic results. Conserv Genet 10:441–452

Schwartz MK, Copeland JP, Anderson NJ, Squires JR, Inman RM, McKelvey KS, Pilgrim KL, Waits LP, Cushman SA (2009) Wolverine gene flow across a narrow climatic niche. Ecology 90:3222–3232

Segelbacher G, Cushman SA, Epperson BK, Fortin M-J, Francois O, Hardy OJ, Holdregger R, Taberlet P, Waits LP, Manel S (2010) Applications of landscape genetics in conservation biology: concepts and challenges. Conserv Genet 11:375–385

Selkoe KA, Watson JR, White C, Horin TB, Iacchei M, Mitarai S, Siegel DA, Gaines SD, Toonen RJ (2010) Taking the chaos out of genetic patchiness: seascape genetics reveals ecological and oceanographic drivers of genetic patterns in three temperate reef species. Mol Ecol 19:3708–3726

Shirk AJ, Wallin DO, Cushman SA, Rice CG, Warheit KI (2010) Inferring landscape effects on gene flow: a new multi-scale model selection framework. Mol Ecol 19:3603–3619

Short Bull RA, Cushman SA, Mace R, Chilton T, Kendall KC, Landguth EL, Schwartz MK, McKelvey K, Allendorf F, Luikart G (2010) Why replication is important in landscape genetics: American black bears in the Rocky Mountains. Mol Ecol 20:1092–1107

Sork V, Waits L (2010) Contributions of landscape genetics—approaches, insights, and future potential. Mol Ecol 17:3489–3495

Spear S, Balkenhol N, Fortin M-J, McRae B, Scribner K (2010) Use of resistance surfaces for landscape genetic studies: considerations for parameterization and analysis. Mol Ecol 19:3576–3591

Storfer A, Murphy M, Evans J, Goldberg C, Robinson S, Spear S, Dezzani R, Delmelle E, Vierling L, Waits L (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

Thomassen HA, Cheviron ZA, Freedman AH, Harrigan RJ, Wayne RK, Smith TB (2010) Spatial modelling and landscape-level approaches for visualizing intra-specific variation. Mol Ecol 19:3532–3548

Trombulak S, Baldwin R (2010) Landscape-scale conservation planning. Springer, New York