MOLECULAR ECOLOGY RESOURCES
Molecular Ecology Resources (2012) 12, 276–284 doi: 10.1111/j.1755-0998.2011.03077.x

Effects of sample size, number of markers, and allelic richness on the detection of spatial genetic pattern

ERIN L. LANDGUTH,* BRADLEY C. FEDY,†,‡ SARA J. OYLER-McCANCE,‡ ANDREW L. GAREY,§ SARAH L. EMEL,¶ MATTHEW MUMMA,** HELENE H. WAGNER,†† MARIE-JOSÉE FORTIN†† and SAMUEL A. CUSHMAN‡‡

*Division of Biological Sciences, University of Montana, Missoula, MT 59812, USA, †Natural Resource Ecology Laboratory, Colorado State University, Fort Collins, CO 80526, USA, ‡U.S. Geological Survey, Fort Collins Science Center, Fort Collins, CO 80526, USA, §VCU Rice Center, Virginia Commonwealth University, Richmond, VA 23284, USA, ¶School of Biological Sciences, Washington State University, Pullman, WA 99164, USA, **Department of Fish and Wildlife Resources, University of Idaho, Moscow, ID 83844, USA, ††Department of Ecology and Evolutionary Biology, University of Toronto, Toronto, ON, M5S 3G5 Canada, ‡‡USDA Forest Service, Rocky Mountain Research Station, Flagstaff, AZ 86001, USA*

Abstract

The influence of study design on the ability to detect the effects of landscape pattern on gene flow is one of the most pressing methodological gaps in landscape genetic research. To investigate the effect of study design on landscape genetics inference, we used a spatially-explicit, individual-based program to simulate gene flow in a spatially continuous population inhabiting a landscape with gradual spatial changes in resistance to movement. We simulated a wide range of combinations of number of loci, number of alleles per locus and number of individuals sampled from the population. We assessed how these three aspects of study design influenced the statistical power to successfully identify the generating process among competing hypotheses of isolation-by-distance, isolation-by-barrier, and isolation-by-landscape resistance using a causal modelling approach with partial Mantel tests. We modelled the statistical power to identify the generating process as a response surface for equilibrium and non-equilibrium conditions after introduction of isolation-by-landscape resistance. All three variables (loci, alleles and sampled individuals) affect the power of causal modelling, but to different degrees. Stronger partial Mantel *r* correlations between landscape distances and genetic distances were found when more loci were used and when loci were more variable, which makes comparisons of effect size between studies difficult. Number of individuals did not affect the accuracy through mean equilibrium partial Mantel *r*, but larger samples decreased the uncertainty (increasing the precision) of equilibrium partial Mantel *r* estimates. We conclude that amplifying more (and more variable) loci is likely to increase the power of landscape genetic inferences more than increasing number of individuals.

Keywords: causal modelling, CDPOP, isolation-by-barrier, isolation-by-distance, isolation-by-landscape resistance, partial Mantel test, sampling, simulation modelling

Received 25 June 2011; revision received 9 September 2011; accepted 9 September 2011

Introduction

The sensitivity of landscape genetic analyses to variation in the sampling scheme used to obtain the genetic data remains largely unexplored (Balkenhol *et al.* 2009; Segelbacher *et al.* 2010). Schwartz & McKelvey (2009) demonstrated that clustered spatial sampling of individuals in a population in which gene flow is governed by isolation-by-distance (IBD) can lead to incorrect identification of non-existent, discrete populations. While the effects of the spatial sampling design on landscape genetic inference has received some attention, much less is known about the effects of the study design in terms of the number of sampled individuals, number of loci analysed per individual (but see Murphy *et al.* 2008), and number of alleles per locus on the ability to correctly and reliably identify the generating process. 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 & Landguth 2011).

Individual-based landscape genetic analytical methods excel at quantifying correlations between current

Correspondence: Erin L. Landguth, Fax: +1 406 243 4184; E-mail: erin.landguth@mso.umt.edu

© 2011 Blackwell Publishing Ltd

landscape structure and current spatial genetic structure. However, rates and patterns of landscape change may affect the emergence, change and loss of genetic structure (Ezard & Travis 2006; Murphy *et al.* 2008; Landguth *et al.* 2010). Given that species conservation and management is primarily concerned with recent or predicted future changes, if landscape genetics is to be used in these contexts then there is an urgent need to rigorously assess both the effects that legacies of past landscape change have on observed genetic patterns and the speed at which these genetic patterns change in response to alterations to existing landscapes. Simulation modelling approaches are particularly valuable to explore the effects of sampling on the detectability of landscape genetic pattern–process relationships, as they allow explicit control over the generating process of gene flow and the sampling design (Balkenhol *et al.* 2009; Epperson *et al.* 2010; Balkenhol & Landguth 2011). For a systematic assessment of the effects of study design on landscape genetic inference, it is important that the generating process is known without error.

A common analytical approach to associate landscape patterns with gene flow processes uses pair-wise calculation of cost distances (e.g. Coulon *et al.* 2004; Cushman *et al.* 2006; McRae 2006; Spear *et al.* 2010). Pair-wise cost distances among many sampling locations are calculated from a landscape resistance model and correlated with pair-wise genetic distances among the same sampling locations, using statistical methods such as Mantel (1967) and partial Mantel tests (Smouse *et al.* 1986). These tests allow direct association of landscape pattern (expressed in cost distances among individuals) and population genetic process (expressed in terms of pair-wise genetic differentiation among spatially referenced genotypes; e.g. Cushman & Landguth 2010). Legendre & 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 & Fortin (2010) recognized that Mantel tests are appropriate when testing hypotheses that can only be formulated using pair-wise distances.

Causal modelling (Legendre & Legendre 1998) has recently been applied to model selection in landscape genetics (Cushman *et al.* 2006). The method tests alternative hypotheses by conducting a series of diagnostic Mantel and partial Mantel tests. Despite concerns regarding low power of Mantel and partial Mantel tests, recent simulation analyses have shown that individual-based analysis of pair-wise relationships between genetic distances and cost distances in complex landscapes using a causal modelling framework have high power to identify the correct landscape resistance hypothesis and reject incorrect, but correlated, alternative hypotheses. For example, Cushman *et al.* (2006) demonstrated the ability of a causal modelling framework to identify a peak of support for the most supported out of 110 alternative models for American black bear (*Ursus americanus*) gene flow in northern Idaho, USA. More recently, Cushman & Landguth (2010) used simulation modelling to demonstrate that causal modelling with partial Mantel tests has a high power to identify the generating process of genetic differentiation in a spatially continuous population and reject incorrect but often highly correlated alternative hypotheses. This simulation study, which assumed a large number of highly variable loci amplified for a large sample of individuals, demonstrated that causal modelling is effective in identifying the correct model out of a large set of models comprised of a full factorial of a number of variables potentially associated with gene flow.

The objective of this paper is to investigate how the power of causal modelling with partial Mantel tests depends on number of individuals, number of loci, and number of alleles per locus. In this study, we use a spatially-explicit, individual-based, simulation program to generate spatial patterns of genetic similarity among organisms across a spatially heterogeneous landscape, which we evaluated at two time steps of 20 and 400 non-overlapping generations (Landguth & Cushman 2010). We then assess how variation in the number of sampled individuals, number of loci, and number of alleles per locus affects the strength of observed landscape genetic pattern–process relationships. Finally, we model the statistical power of causal modelling with partial Mantel tests to identify the correct model of isolation-by-landscape resistance (IBR) from among competing models of IBD and isolation-by-barrier (IBB) as a function of number of individuals, number of alleles per locus and number of loci used.

Models and methods

Simulation program

We used a spatially-explicit, landscape genetic program (CDPOP version 0.85, Landguth & Cushman 2010) to simulate genetic exchange across 500 non-overlapping generations among 1000 individuals as functions of individual-based movement through mating and dispersal on a given landscape. The landscape pattern is represented through a resistance surface with grid cell values representing costs of movement through the landscape. In CDPOP, mating and dispersal are modelled 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

© 2011 Blackwell Publishing Ltd

dispersal) cost functions are scaled to a user-specified maximum dispersal distance.

Our goal was to assess the sensitivity of landscape genetic inference to number of individuals and number and variability of molecular markers, everything else being equal for maximum comparability, rather than assessing sensitivity to characteristics of the landscape. Therefore, we used the same landscape resistance model for all simulations. We adopted an empirically tested model of landscape resistance to black bear movement in Northern Idaho, USA, from Cushman *et al.* (2006) to ensure that the simulated scenario mimicked a realistic system (Fig. 1; extent of approximately 3000 km^2 with resistance values ranging from 1 to 62 in 90-m grid cells). On this landscape resistance model, we initialized the 1000 individual locations in the Universal Transverse Mercator (UTM) coordinate system by populating grid cell values in a hexagonal pattern at 1.6-km spacing unless the cell value was >6 (Fig. 1). This was carried out to place individuals in habitat that was relatively suitable for the species, given this landscape resistance hypothesis. We simulated movement (mating and dispersal) 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 low dispersing organism and corresponding to the range of positive spatial autocorrelation of genetic 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 ≤39 200 cost-units apart with probability of mating or dispersal distance within that limit specified by an inverse-square probability function (Landguth & Cushman 2010). See Fig. 1 for an example individual that has the corresponding radius of 39 200 cost-units.

We initialized genotypes for the 1000 individuals by randomly assigning allelic states across an initial population with a random sex assignment that contained 25 loci (which were subsequently sub-sampled) with the *k*-allele mutation rate set to 0.0. CDPOP simulates spatially referenced genotypes for all individuals at each generation with independent assortment and no linkage disequilibrium in Mendelian inheritance. We set mating parameters 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 1000 at every generation, and the remaining offspring were discarded once all the 1000 locations were occupied by a dispersing individual. This is equivalent to forcing

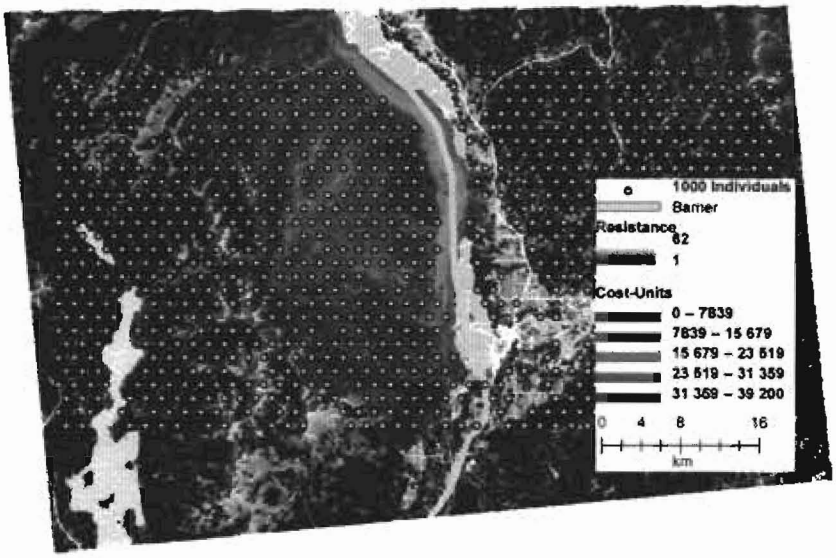

Fig. 1 Simulated population, landscape and maximum movement distance. 1000 individuals (yellow points) located on the resistance surface from Cushman *et al.* (2006) with extent of approximately 3000 km^2 and resistance values ranging from 1 to 62 in 90-m grid cells. An example individual is selected (blue point) to show the corresponding 39 200 cost-unit radius. This maximum cost distance value constrains all mate choices and dispersal distances to be ≤39 200 cost-units apart with probability of mating or dispersal distance within that limit specified by an inverse-square probability function. In addition, the yellow line represents the barrier delineation used for the isolation-by-barrier distance calculation.

© 2011 Blackwell Publishing Ltd

emigrants out of the study area once all available home ranges are occupied (Landguth & Cushman 2010).

Simulation scenarios

We ran 50 replicate simulations for each level of alleles per locus (3, 5, 10, 15, 20 and 25 maximum alleles at the beginning of each simulation run, thus simulating a panmictic initial population with maximum allelic diversity), with 25 loci and 1000 individuals. Simulations were performed under IBR for a period of 500 non-overlapping generations. Once simulation runs were completed, we randomly sequentially sub-sampled loci and individuals according to a complete three-way factorial design with factors (i) initial number of alleles per locus, with six levels (see above); (ii) number of loci, with five levels (5, 10, 15, 20, 25 loci); and (iii) number of individuals, with 23 levels (1000, 900, ..., 500, 450, ..., 100, 90, ..., 10 individuals). Each simulation was evaluated at two time steps (20 and 400 generations), resulting in 13 800 data sets that will be referred to as scenarios. 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, partial Mantel r equilibrium was achieved in all simulations and the relationships between genetic and the association between landscape pattern and genetic structure had stabilized.

Statistical analysis of simulation results

Inter-individual genetic distance was calculated as the proportion of shared alleles (Bowcock *et al.* 1994), and landscape-cost distance model (i.e. IBR distance; e.g. Cushman *et al.* 2006; Spear *et al.* 2010) 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 version 9.2 (ESRI 2009). Euclidean distance (IBD distance) was calculated from the UTM coordinates between all pairs of individuals. The barrier-cost distance (IBB distance) was represented as a model matrix similarly to the Legendre & Legendre (1998), with pair-wise distance equal to 1 for two individuals from opposite sides of a complete barrier separating approximately half of the 1000 individuals, and pair-wise distance equal to 0 for two individuals from the same side of the barrier (panmixia; see Fig. 1 for the barrier delineation).

For each scenario, we performed a partial Mantel test to correlate genetic distance to IBR distance accounting for IBD distance using the library 'ecodist' version 1.1.3 (Goslee & Urban 2007) in the statistical software package R (R Development Core Team 2009). Because of 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 modelling, which involves a series of diagnostic Mantel and partial Mantel tests (e.g. Cushman *et al.* 2006). These 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 1999 permutations, corresponding to a 0.005 precision for the cutoff value, $\alpha = 0.05$.

Successful identification of the underlying generating process, say IBR for example, 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 modelled the probability of successfully identifying the correct landscape resistance model for varying combinations of three covariates (number of individuals, number of loci, and number of alleles per locus) separately for the two time steps of 20 and 400 generations. The modelled response variable was the probability of correctly identifying the generating process (IBR) as described earlier, and this probability was skewed towards success (i.e., probability = 1). Therefore, we used a hurdle model approach (Martin *et al.* 2005; ver Hoef & Jansen 2007), which is a two-step process. First, we modelled the response variable with a logistic model using a binary response variable where 1 equals successful identification of the generating process and 0 equals unsuccessful process identification. For the predicted probability of success, we used a cut-off of 0.95, so that parameter combinations with predicted success probabilities ≥ 0.95 received a predicted response value of 1, those with predicted probabilities of <0.95 a value of 0. In order to allow for nonlinear response (such as levelling off after an initial increase with increasing numbers of samples, loci, or alleles per locus) of probability of success to each of the three covariates, we included both linear and quadratic terms. We ran all possible model combinations, and the top model for each time step was selected by minimizing Akaike's Information Criterion (AIC).

The second step of the hurdle modelling approach involved modelling the observations with a predicted probability <0.95 using linear regression. The probabilities

© 2011 Blackwell Publishing Ltd

were not subjected to arcsine transformation (Warton & Hui 2011). We analysed the same covariates as in the logistic (binary) component of the analysis. Again, all possible model combinations were run, and the top model for each time step was selected based on AIC.

Results

Mean equilibrium partial Mantel *r* was low when number of loci and alleles were both small and rose convexly with increases in alleles and loci (Fig. 2a,b). This indicated a rapid nonlinear increase in equilibrium partial Mantel *r* with increases in alleles and loci. The equilibrium partial Mantel *r* was <0.4 when five loci with three alleles per locus were analysed, and rose to over 0.70 when 25 loci with 25 alleles each were analysed. In contrast to the relatively large effects of number of loci and number of alleles per locus, the number of individuals sampled from the population had little effect.

However, a comparison between the overall partial Mantel *r* values from a sample of 1000 individuals (Fig. 2a) to a sample of 10 individuals (Fig. 2b) revealed a smoother convex pattern with higher number of individuals (i.e. higher precision). In fact, strong effects are seen with number of alleles per locus, number of loci analysed per individual, and number of individuals on the standard deviation of equilibrium partial Mantel *r* (Fig. 2c,d). Variability in equilibrium partial Mantel *r* among simulation runs decreased substantially with increases in either number of alleles per locus, number of loci per individual, or number of individuals sampled from the population.

Using the causal modelling framework, we found that number of individuals, number of loci per individual, and number of alleles per locus all contributed to the probability of success in identifying the generating process IBR, but to different degrees. In models for both generation 20 and 400, the most supported logistic and linear models based on AIC included all three predictors (Table 1). The inclusion of the quadratic term for each variable in the top models suggested the positive influence of each covariate reaches an asymptote across the range of values we examined. For ease of interpretation, we did not standardize covariates (e.g. z-scores) prior to analysis. The coefficient estimates (Table 1) are directly influenced by the scale of the predictor variable. Therefore, the small beta estimates for the number of individual variable should be interpreted with caution. Increasing number of individuals had an influence on the predicted probability surface at generation 20 (Fig. 3a–c). The number of alleles per locus and the number of loci had very similar scales, allowing for direct comparison of the beta values. Number of alleles per locus had a slightly larger impact on probability of success at a given number of individuals than did the number of loci used. The probability of a correct causal model result increased

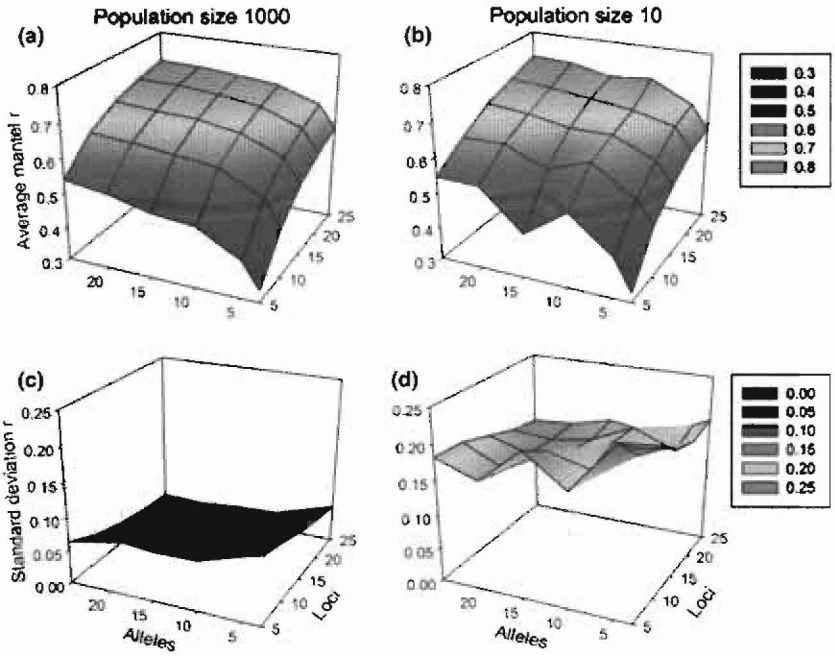

Fig. 2 Surfaces showing equilibrium partial Mantel *r* values at generation 400 as a function of number of loci and number of alleles for: 1000 individuals (a), 10 individuals (b). Surfaces for: standard deviation of equilibrium partial Mantel *r* for 1000 individuals (c), and 10 individuals (d).

© 2011 Blackwell Publishing Ltd

Table 1 The top logistic and linear models (chosen using AIC) for probability of success within the casual modelling framework

Logistic models								
Gen	L	L^2	S	S^2	A	A^2	Int	AIC
20	0.808	−0.018	0.031	−0.000	0.963	−0.018	−22.332	308.059
400	0.194	−0.004	0.036	−0.000	0.368	−0.009	−7.548	373.867
Linear models								
Gen	L	L^2	S	S^2	A	A^2	Int	AIC
20	0.033	−0.001	0.003	−0.000	0.086	−0.002	−0.287	−156.441
400	0.024	−0.001	0.003	−0.000	0.013	−0.000	0.781	−277.686

Gen, generation; L, number of loci; S, number of individuals sampled; A, number of starting alleles; Int, intercept; AIC, Akaike's Information Criterion.

dramatically with increasing number of individuals, number of alleles and number of loci (Fig. 3). At generation 400, the probability of success was high for all loci by allele combinations with number of individuals higher than 100.

Discussion

Much work in landscape genetics has focused on testing hypothesized relationships between landscape pattern and genetic differentiation in spatially structured populations (Coulon *et al.* 2006; Cushman *et al.* 2006; McRae 2006; Schwartz *et al.* 2009; Shirk *et al.* 2010). This analysis formally evaluates the effects of number of loci, number of alleles per locus and number of sampled individuals on landscape genetic inference using a causal modelling framework. By using an individual-based simulation program specifically designed to simulate gene-flow processes across resistance landscapes (Landguth & Cushman 2010), we were able to control the generating process and conduct a simulation experiment in which number of loci, alleles per locus and number of individuals sampled were systematically varied. This factorial simulation provided a robust means to comprehensively evaluate the interactive effects of different numbers of individuals, loci and alleles on the resulting relationships between landscape pattern and gene flow.

We evaluated both the equilibrium value of the partial Mantel correlation between genetic distance and landscape-cost distance, and the ability of causal modelling with partial Mantel tests to correctly identify IBR as the generating process. The equilibrium partial Mantel correlation between genetic and cost distances is a convex function of number of loci, number of alleles, and number of individuals sampled from the population. The magnitude of correlations between landscape pattern and genetic structure of a population of organisms is highly dependent on the number of loci analysed per individual, and the number of alleles at each of those loci. The number of individuals used in the analysis has little effect on the equilibrium value of this correlation. However, the variability of the correlation among pairs of individuals decreases substantially with increasing number of individuals, as well as increasing loci and alleles. Thus, stronger relationships will be found when larger numbers of loci are used, and when loci are more variable. This implies that two otherwise identical studies with differing sampling intensities are likely to find landscape patterns with vastly different Mantel r values and thus effect sizes, reconfirming that comparisons between studies with different number of loci or alleles per locus is problematic.

Under the simulation conditions used here, the power of a causal modelling approach to correctly identify the generating process is nearly perfect under equilibrium genetic conditions. Under nonequilibrium conditions, the power to correctly identify the generating process increases substantially with the number of loci, alleles per locus, and individuals analysed. The number of loci used in landscape genetic studies varies but is typically at least 10, and is likely to increase in the near future with improving genomic methods to isolate microsatellite loci. The number of alleles per locus is also quite variable (e.g. 3–25 alleles per locus), and unfortunately beyond the control of the researcher. Examining the probability of correctly identifying the generating process at 15 loci, with 10 alleles per locus (a realistic scenario) reveals that our ability to correctly identify the generating process is low (<0.20) with a 100 individuals. As the number of individuals increases to 300, the probability increases to ~0.7 and is essentially 1.0 with 500 individuals. Interestingly, if 25 loci were available with 25 alleles per locus, even a small number of individuals (100) could be used

© 2011 Blackwell Publishing Ltd

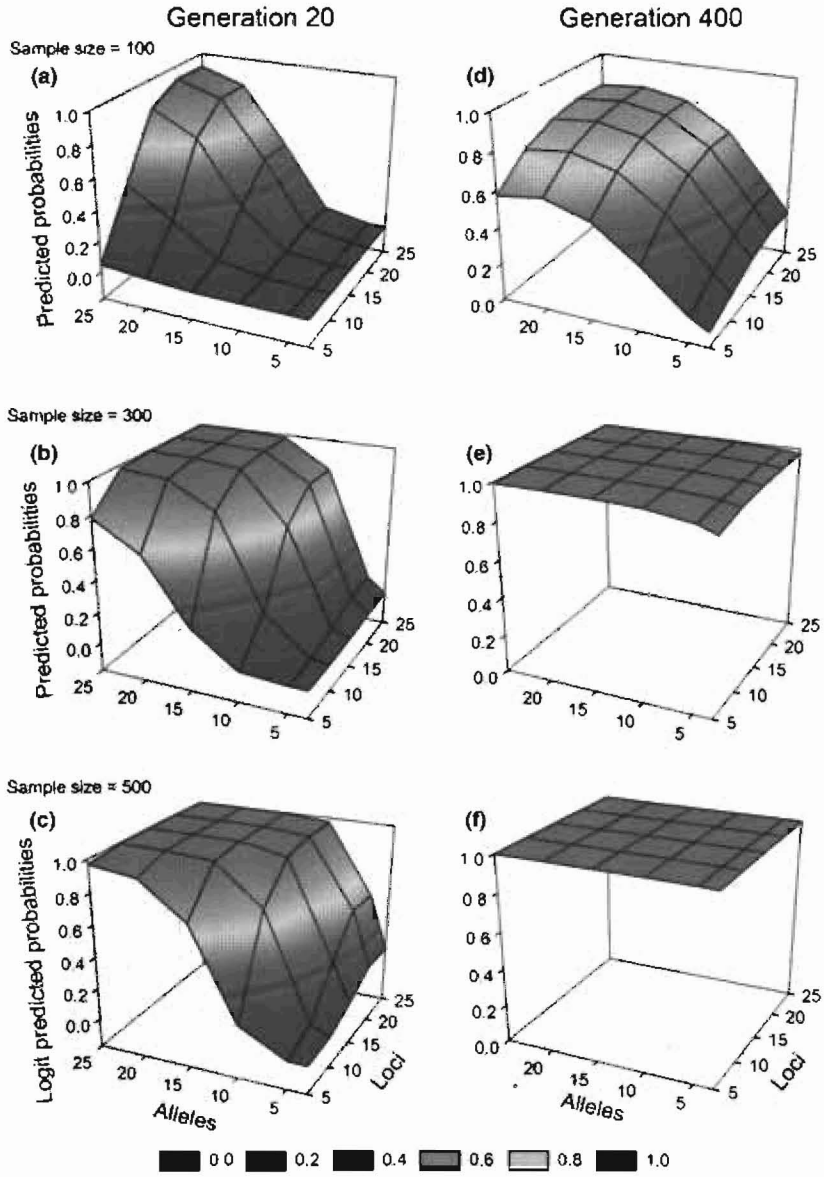

Fig. 3 Relationship between number of alleles, number of loci and the probability of success using the top logistic models for three number of individuals: 100 at generation 20 (a), 100 at generation 400 (b), 300 at generation 20 (c), 300 at generation 400 (d), 500 at generation 20 (e), and 500 at generation 400 (f).

to correctly identify a generating landscape pattern. To date, polymorphism in microsatellites used in landscape genetic studies are not that high, so that scenario may not be realistic at this point in time. Given these effects, there may be some trade-off between the number of loci to be analysed and the number of individuals to be sampled and researchers may obtain better results from increasing the number of loci analysed rather than the number or individuals sampled, provided that loci are highly polymorphic (also see important caveats below). Importantly, number of individuals below 500 with low to moderate levels of polymorphism of alleles (5–10) can fail to identify the generating process even with large numbers of loci.

In evaluating both the power of the causal modelling approach to correctly identify the generating process and the influence on the equilibrium mean Mantel r value, we found that the number of loci analysed and number of alleles per locus were more influential than number of individuals sampled. However, increasing the number of individuals did have an impact on the standard deviation of the equilibrium partial Mantel r (Fig. 2c,d) and increased the probability of success using the causal modelling approach (Table 1, Fig. 3). Thus, while

© 2011 Blackwell Publishing Ltd

increasing the number of individuals sampled did not increase mean equilibrium partial Mantel r (accuracy), it did decrease the uncertainty (increasing the precision) of equilibrium partial Mantel r estimates and increase the power of the causal modelling approach to correctly identify the generating process.

Our results support the conclusions of Cushman & Landguth (2010) who showed that causal modelling using partial Mantel tests has a high power to identify the underlying process of genetic differentiation while rejecting correlated alternative hypotheses. By evaluating the performance of the causal modelling approach in a fully factorial design covering a broad range of numbers of alleles per locus, number of loci, and number of sampled individuals, we demonstrated that when a population has reached a genetic drift equilibrium, causal modelling has nearly perfect ability to correctly identify the driving process regardless of the number of loci or alleles included in analysis. Our simulations also demonstrated that even prior to establishment of equilibrium, causal modelling can be effective, but its power in such cases depends strongly on highly polymorphic loci and reasonably large sample sizes of individuals. This study suggests that for species with modest generation times (e.g., >5 years) and naturally low dispersal, our ability to accurately and reliably detect a signal generated by recently emerged landscape-level barriers to dispersal (e.g. roads or fragmentation of formerly continuous habitat) is really very limited. This is in line with Landguth *et al.* (2010) who showed that genetic studies of species with low dispersal might not detect effects of landscape barriers for many generations, even if the landscape change has resulted in complete isolation of subgroups of a previously connected population.

A few caveats are worth mentioning, which should lead to future work. First, all of these studies relied on the use of Mantel and partial Mantel testing and future work should explore the power of this test in comparison with other distance-based methods, such as distance-based redundancy analysis (see Legendre & Fortin 2010). Second, all of these studies used one model of landscape resistance. It would be valuable to extend this analysis to a wide range of alternative landscapes expressing a range of composition and structure and explore how landscape patterns interact with sampling intensity. Third, it is important to note that this study approach was designed for systems of individuals (i.e. more realistically represents an animal system) and that the consequences for population-based designs may be different. Finally, our study used a spatial random sample drawn from a continuously distributed underlying population. In reality, such a population rarely (if ever) exists and such a sampling design may be nearly impossible to attain or may not be appropriate for a species whose distribution is particularly patchy. The effects of spatial sampling scheme were not the focus of this study, and we warn that effect of differing sampling strategies may have an important impact on the ability to determine the generating process. Future research should more fully address the effects of spatial sampling scheme on landscape genetic inferences. Schwartz & McKelvey (2009) demonstrated that clustered sampling can lead to erroneous conclusions about the number of populations when using Bayesian clustering approaches to define populations when IBD is a driving force. It would be interesting to evaluate how a wide range of spatial sampling strategies affect the performance of individual-based, cost-distance landscape genetic modelling approaches, such as those used in this paper.

Conclusions

Taken together, these results have several implications. First, the magnitude of partial Mantel correlations is highly sensitive to the number of alleles per locus and the number of loci analysed, reconfirming caution when comparing results from different studies. Second, our results show that all three variables (loci, alleles, and sampled individuals) affect the power of causal modelling, but to different degrees. In empirical landscape genetic research, there is often a trade-off in cost between sampling more individuals or amplifying more loci per individual. Our results suggest that when trade-offs exist between number of individuals and number of loci (assuming that the loci involved are highly polymorphic), researchers in general may gain a greater benefit in terms of the power to identify the generating process by investing in relatively more (and more polymorphic) markers for relatively fewer individuals. This result is especially relevant for the detection of effects on population connectivity from recent landscape change and in the context of conservation genetics for species of concern, which tend to be rare, cryptic, or difficult to capture, such that it is often impossible to obtain large samples of individuals.

Acknowledgements

This work resulted from a Distributed Graduate Seminar (Developing Best Practices for Testing Landscape Effects on Gene Flow), conducted through the National Center for Ecological Analysis and Synthesis, a centre funded by the National Science Foundation grant #EF-0553768, the University of California, Santa Barbara, and the State of California. E.L.L. was supported in part by funds provided by the Rocky Mountain Research Station, Forest Service, U.S. Department of Agriculture and by the National Science Foundation grant #DGE-0504628. 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 L. Waits and M. Murphy for helpful comments on this manuscript.

© 2011 Blackwell Publishing Ltd

References

Balkenhol N, Landguth EL (2011) Simulation modeling in landscape genetics: on the need to go further. *Molecular Ecology*, **20**, 667–670.

Balkenhol N, Gugerli F, Cushman SA *et al.* (2009) Identifying future research needs in landscape genetics: where to from here? *Landscape Ecology*, **24**, 455–463.

Bowcock AM, Ruiz-Linares A, Tomfohrde J *et al.* (1994) High resolution of human evolutionary trees with polymorphic microsatellites. *Nature*, **368**, 455–457.

Coulon A, Cosson JF, Angibault JM *et al.* (2004) Landscape connectivity influences gene flow in a roe deer population inhabiting a fragmented landscape: an individual-based approach. *Molecular Ecology*, **13**, 2841–2850.

Coulon A, Guillot G, Cosson JF *et al.* (2006) Genetic structure is influenced by landscape features: empirical evidence from a roe deer population. *Molecular Ecology*, **15**, 1669–1679.

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 models with causal modeling. *The American Naturalist*, **168**, 486–499.

Epperson BK, McRae BH, Scribner K *et al.* (2010) Utility of computer simulations in landscape genetics. *Molecular Ecology*, **19**, 3549–3564.

ESRI (Environmental Systems Resource Institute) (2009) *ArcMap 9.2*. ESRI, Redlands, California.

Ezard THP, Travis JMJ (2006) The impact of habitat loss and fragmentation on genetic drift and fixation time. *Oikos*, **114**, 367–375.

Goslee SC, Urban DL (2007) The ecodist package for dissimilarity-based analysis of ecological data. *Journal of Statistical Software*, **22**, 1–19.

ver Hoef JM, Jansen JK (2007) Space-time zero-inflated count models of Harbor seal. *Environmetrics*, **18**, 697–712.

Landguth EL, Cushman SA (2010) CDPOP: a spatially explicit cost distance population genetics program. *Molecular Ecology Resources*, **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. *Molecular Ecology*, **19**, 4179–4191.

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. *Molecular Ecology Resources*, **10**, 831–844.

Legendre P, Legendre L (1998) *Numerical Ecology*, 2nd English edn. Elsevier, Amsterdam.

Mantel N (1967) The detection of disease clustering and a generalized regression approach. *Cancer Research*, **27**, 209–220.

Martin TG, Wintel BA, Thodes JR *et al.* (2005) Zero tolerance ecology: improving ecological inference by modeling the source of zero observations. *Ecology Letters*, **8**, 1235–1246.

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

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.

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. *Conservation Genetics*, **10**, 441–452.

Schwartz MK, Copeland JP, Anderson NJ *et al.* (2009) Wolverine gene flow across a narrow climatic niche. *Ecology*, **90**, 3222–3232.

Segelbacher G, Cushman SA, Epperson BK *et al.* (2010) Applications of landscape genetics in conservation biology: concepts and challenges. *Conservation Genetics*, **11**, 375–385.

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

Smouse PE *et al.* (1986) Multiple regression and correlation extensions of the Mantel test of matrix correspondence. *Systematic Zoology*, **35**, 627–632.

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. *Molecular Ecology*, **19**, 3576–3591.

Warton DI, Hui FKC (2011) The arcsine is asinine: the analysis of proportions in ecology. *Ecology*, **92**, 3–10.

© 2011 Blackwell Publishing Ltd