Research Article
Simulating Pattern-Process Relationships to Validate Landscape Genetic Models

A. J. Shirk, 1 S. A. Cushman, 2 and E. L. Landguth 3

1 Climate Impacts Group, University of Washington, Box 355672, Seattle, WA 98195, USA
2 Rocky Mountain Research Station, United States Forest Service, Missoula, MT 59808, USA
3 Division of Biological Sciences, University of Montana, Missoula, MT 59812, USA

Correspondence should be addressed to A. J. Shirk, ashirk@u.washington.edu

Received 28 September 2011; Accepted 6 January 2012

Academic Editor: Daniel Rubenstein

Copyright © 2012 A. J. Shirk et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Landscapes may resist gene flow and thereby give rise to a pattern of genetic isolation within a population. The mechanism by which a landscape resists gene flow can be inferred by evaluating the relationship between landscape models and an observed pattern of genetic isolation. This approach risks false inferences because researchers can never feasibly test all plausible alternative hypotheses. In this paper, rather than infer the process of gene flow from an observed genetic pattern, we simulate gene flow and determine if the simulated genetic pattern is related to the observed empirical genetic pattern. This is a form of inverse modeling and can be used to independently validate a landscape genetic model. In this study, we used this approach to validate a model of landscape resistance based on elevation, landcover, and roads that was previously related to genetic isolation among mountain goats (Oreamnos americanus) inhabiting the Cascade Range, Washington (USA). The strong relationship between the empirical and simulated patterns of genetic isolation we observed provides independent validation of the resistance model and demonstrates the utility of this approach in supporting landscape genetic inferences.

1. Introduction

Landscapes influence mating probabilities within a population by resisting the movement of breeding individuals or their propagules [1]. This influence can have profound effects on population viability. In particular, permeable landscapes favor the formation of large breeding pools, increasing the local effective population size, potentially boosting demographic rates, improving mate choice, and increasing allelic diversity [2]. Conversely, in vast, highly resistive, or fragmented landscapes, smaller, more isolated breeding pools limit the local effective population size, potentially diminishing demographic rates, reducing mate choice, and preventing the influx of new genetic variation [3, 4]. Such populations face greater risk of decline and extirpation due to a combination of demographic and genetic factors, including potential Allee effects [5], stochastic environmental fluctuations [6], random fixation of deleterious recessive alleles [7], inbreeding depression [8–12], and in the long-term, lack of adaptive potential in a changing environment [13].

These demographic and genetic threats to population persistence suggest the most viable populations, all other factors being equal, would exist in landscapes that permit all individuals to freely interbreed. Panmictic populations are probably rare in the wild, however [14, 15]. More commonly, populations exhibit substructure arising from one of several mechanisms that cause genetic isolation. Isolation by barrier (IBB), for example, occurs when strong migration barriers (e.g., major roads) sharply reduce gene flow, generally over a relatively small geographic extent [16]. Alternatively, genetic isolation can accrue as a function of distance (isolation by distance or IBD), particularly for species with migration rates relatively unaffected by landscape heterogeneity [17]. For species whose dispersal is sensitive to landscape heterogeneity (owing to environmental gradients such as elevation, landcover, climate, slope), isolation by landscape resistance (IBR) may be observed [18]. Each of these modes of genetic isolation has the potential to reduce population viability due to demographic and genetic factors, particularly when barrier effects are strong, the distance between individuals...
or subpopulations is large relative to dispersal capability, or landscape resistance is high.

For populations threatened by connectivity issues, the particular mechanism of genetic isolation responsible for the reduced connectivity bears strongly on the appropriate conservation responses. Mitigation efforts for populations isolated by barriers, for example, might focus on barrier crossing structures [19]. For populations isolated by distance, the focus might be on establishing stepping stone populations to reduce distances between breeding individuals [20]. Among populations isolated by landscape resistance gradients, it may be possible to reduce the resistance of some landscape features or link habitat areas by establishing or protecting low resistance movement corridors [21, 22].

Despite the importance of understanding the mechanism of genetic isolation to the management of populations threatened by poor connectivity, few studies rigorously evaluate the relative support for a range of plausible hypotheses. Observing a correlation between a landscape model and a genetic pattern and inferring causation is dangerous because of the potential for spurious correlations between the model under evaluation and an untested alternative but causal model [23]. To reduce the risk of false inferences, some recent landscape genetics studies [24, 25] have used partial Mantel’s tests [26] to rigorously evaluate the relative support for a range of hypotheses related to IBD, IBB, and IBR. While this approach decreases the potential for false inference, in reality, the full range of possibility can never be tested. As such, the risk of false inference is always present.

In contrast to approaches that correlate landscape genetic models with patterns of genetic isolation, population genetic simulations allow the researcher to specify the landscape and demographic factors that govern gene flow and then determine if the simulated pattern of genetic isolation that arises matches the empirical pattern of isolation (as measured by inter-individual genetic distances). Simulations are powerful tools in landscape genetics [27–29], yet to-date they have not been used to in conjunction with empirical landscape genetic models as a means to reduce the potential for false inferences. Though matching simulated genetic patterns to empirical patterns is also subject to false inferences, error in simulations is largely independent of error in the empirical analysis approach. Thus, analysis using both methods on the same data would represent a more robust means of landscape genetic inference.

As an example of this concept, consider a mountainous landscape with major highways in some, but not all, low elevation valleys. The primary determinant of genetic isolation in this landscape for the species under study is related to the barrier effect of highways, not elevation per se. If one models resistance in this landscape based only on elevation, the conclusion would likely be that low elevation is a barrier to gene flow which fragments the study area into genetically isolated subpopulations. This false inference would be made because the model is spuriously correlated to the true driver of genetic isolation. If one were to simulate gene flow on this landscape based on assigning high resistance to low elevation, the resulting pattern of genetic isolation would likely be a poor fit to the empirical pattern, because in reality, gene flow would cross low elevation valleys without roads and greatly reduce genetic differentiation across the study area, depending on the configuration of the landscape. In this way, simulation may be used to validate empirical landscape genetic analysis.

In this study, we coupled the empirical correlative approach that predominates in the field of landscape genetics today with landscape genetic simulations. Previously, we identified a spatial model of landscape resistance (X) among mountain goats in the Cascade Range, Washington, USA, that was most related to the observed pattern of genetic isolation (Y) relative to a range of alternative landscape models [25]. Here, we test the validity of this model by simulating the process of gene flow on landscape X and determine if the observed empirical pattern of genetic isolation Y matches the simulated pattern of genetic isolation Y’. If the landscape resistance model is robust, simulated gene flow on the landscape should result in a genetic pattern similar to the real population, and using partial regression methods to account for the effect of alternative landscape models should not significantly decrease that relationship. The added rigor of this approach reduces the potential for false inferences and thereby affords greater confidence in the accuracy of landscape models. This approach is particularly valuable when applied to landscape models that will be used to inform conservation efforts, where managers can ill afford to waste limited resources and the consequences of management action based on false inferences may actually reduce population viability.

2. Methods

2.1. Study Area. The study area included approximately 36,500 km² of the Cascade Range of Washington State, USA, extending along a north-south axis 315 km from Mount Adams in the south to the Canadian border in the north (Figure 1). The landscape is mostly mountainous and covered with montane forests, except at high elevations, where subalpine parkland, rocky alpine summits, and glaciers predominate. Elevation varies from near-sea level to almost 4,400 m. A major transportation corridor, Interstate 90 (I90), crosses the study area on an east-west axis roughly in the center of the range. In addition, three state highways run east-west across the range (two north and one south of I90) and numerous other highways and secondary roads intrude towards the crest from lower elevation valleys.

2.2. Landscape Model (X). We previously identified a spatial model of landscape resistance (i.e., a resistance surface; [30]) based on suboptimal elevation, landcover, and the presence of roads as the most related to genetic isolation among mountain goats within the study area [25]. This isolation by resistance (IBR) model is represented as a raster where each grid cell is assigned a resistance to gene flow. We used the ArcGIS COSTDISTANCE function to calculate a matrix of the pairwise landscape distances (accumulated least-cost-path) between all simulated mountain goat locations based on the resistance values of this model. These pairwise
landscape distances were then used to specify dispersal and mating probabilities in the simulations described below.

In addition to the IBR model, we evaluated the relative support for two alternative landscape models. We formed an isolation by distance model (IBD) based on the pairwise Euclidean distance between all simulated individuals. We also formed an isolation by barrier model based on the assumption that I90 was a complete barrier to gene flow in an otherwise panmictic population. Thus, in the barrier model, pairwise landscape distances between individuals on the same side of I90 were set to zero and landscape distances spanning I90 were set to infinity.

2.3. Observed Genetic Distance ($Y$). We previously collected 147 genetic samples from mountain goats throughout the study area and genotyped them at 18 microsatellite loci [25]. We excluded 14 of these genotypes from further analysis based on their apparent admixture with nonnative mountain goats translocated to the Cascade Range in the 1980s, as previously reported [25]. From these genotypes, we calculated a matrix of pairwise genetic distances between all individuals based on the distance along the first principal components axis [31, 32] using R 2.10 (R Development Core Team 2009).

2.4. Simulated Genetic Distance ($Y'$). We used CDPOP version 0.84 [33] to simulate 1500 generations (~ six years per generation) of mountain goat gene flow on the three landscape models described above (IBB, IBD, and IBR). The first 1000 generations represent the latter part of the Holocene
prior to the wide-scale anthropogenic landscape changes of the modern era. This lengthy "burn-in" period allowed for the simulated population to reach genetic equilibrium prior to modern day landscape changes. After generation 1000, the simulations represent the modern time period leading up to the present day at generation 1020. We also simulated future generations, up to 1500.

CDPOP is an individual-based spatially explicit population genetic simulator. The program tracks alleles across individuals over time based on dispersal and mating. Dispersal and mating probabilities in CDPOP are governed by the pairwise landscape distances (least-cost-path) between individuals given a landscape model resistance surface ($X$). We performed 100 Monte Carlo iterations per simulation. Each simulation tracked 18 loci with 10 alleles per locus randomly assigned to the first generation (the same number of loci used in [25]). We set the mutation rate to 0.0005 based on a k-allele model. We specified mating to be sexual with nonoverlapping generations, diploid individuals, and an equal sex ratio at birth. For each simulation, we used a constant population size of 2392 mountain goats (the current estimated population size, Washington Department of Fish and Wildlife, unpublished data) inhabiting the study area. Initial conditions included random starting allele frequencies and 10 alleles per locus. Starting locations were based on the sampling locations of the 135 mountain goats genotyped in Shirk et al. [25] or, for the remaining 2257 goats, locations were randomly assigned within blocks of known habitat, each of which was associated with a current population estimate (based on WDFW survey data, unpublished).

In the first 1000 generations of the IBB simulation, mating was random (i.e., a panmictic population). For the last 500 generations, we stipulated that 190 fragmented the population into north and south Cascade subpopulations (NOCA and SOCA, resp., Figure 1), with random mating within each subpopulation but no mating between subpopulations. This definition of subpopulations was based on the previously reported identification of two genetic clusters [25] using the STRUCTURE algorithm [34]. Further subdivision of the north and south Cascade subpopulations was not supported by STRUCTURE [25].

We simulated IBD by relating mating probability to the inverse square of the Euclidean distance between individuals. This mating probability was constant during all 1500 generations of the simulation.

We simulated IBR by relating mating probability to the inverse square of the accumulated-least cost path distance between individuals based on the resistance surface identified in Shirk et al. [25] as most related to the population’s genetic isolation. During the first 1000 generations, we modified this resistance surface by removing all additional resistance due to anthropogenic landscape features, including roads, urban areas, agriculture, and large reservoirs. During the last 500 generations, we used the original modern era resistance surface.

For each simulated generation, we calculated the PCA-based genetic distance ($Y$; as described above) between the 135 simulated mountain goats whose locations matched the sampled goats from the actual population.

2.5. Causal Modeling. We employed a simple Mantel test [35] within the R package "ecodist" [36] to ascertain the convergence between simulated genetic distance ($Y'$) and the actual genetic distance ($Y$) for each Monte Carlo iteration across 1500 generations. To potentially identify a causal model among competing alternative models, we also employed partial Mantel’s tests [26] as implemented in the R package “ecodist.” Partial Mantel’s tests control for the variance explained by alternative landscape hypotheses when evaluating the relationship between $Y$ and $Y'$. If little or no relationship exists between $Y$ and $Y'$ after partialling out a competing alternative hypothesis, the model can be rejected as spurious. Conversely, a causal model would meet the expectation of a strong relationship between $Y$ and $Y'$, even after partialling out the effect of alternative hypotheses. Partial Mantel tests have been shown to offer high power with low type I and type II error rates in identifying causal spatial genetic relationships and rejecting spurious models, even when the degree of correlation between models is high [23].

3. Results

3.1. Simulations. All simulations began the first generation with random starting allele frequencies. At time zero, there was no significant simple Mantel’s correlation between simulated genetic distance ($Y'$) and the current population’s genetic distance ($Y$; Figure 2) for any of the three landscape hypotheses. For the isolation by resistance simulation, the simple correlation between $Y$ and $Y'$ rapidly increased to 0.6 over the first 100 generations then stayed nearly constant until the shift to the modern landscape at $t = 1000$, where the correlation increased to nearly 0.7 over a period of 20 generations leading up to the simulated present-day population (Figure 2(a)). Partialling out the variance explained by the IBB hypothesis did not reduce the strength of the IBR hypothesis correlation. Partialling out the variance explained by the IBD hypothesis reduced the IBR hypothesis correlation by approximately 0.05 or 0.10 for the modern or historic portions of the simulation, respectively, but the overall relationship between $Y$ and $Y'$ remained high. The IBD landscape simulated genetic distance showed a weaker simple Mantel correlation with the modern population’s genetic distance, and this correlation decreased over time. Partialling out the variance explained by the IBR hypothesis reduced the correlation significantly for all generations (Figure 2(b)). Simulating gene flow on the IBB landscape resulted in a pattern of genetic isolation that was not significantly related to the observed empirical pattern at any generation (Figure 2(c)).

3.2. Causal Modeling. Only the IBR hypothesis met the statistical expectations of a causal model (i.e., partialing out the variance explained by competing alternative hypotheses did not greatly diminish the relationship between $Y$ and $Y'$, Figure 2).

4. Discussion

A primary goal of landscape genetic analysis is to infer causal relationships between the process of gene flow occurring over
Figure 2: Correlation between observed empirical genetic distance ($Y$) and simulated genetic distance ($Y'$). (a), the simple Mantel correlation between $Y$ and $Y'$ for the IBR landscape is depicted across 1500 simulated generations (black line). Partial Mantel correlations controlling for the variance explained by the IBB hypothesis (light gray line) and IBD hypothesis (dark gray line) are also shown. (b), the simple Mantel correlation between $Y$ and $Y'$ for the IBD landscape is depicted (dark gray line), with the partial correlations controlling for the variance explained by the IBR (black line) and IBB (light gray line) also shown. (c), the simple Mantel correlation between $Y$ and $Y'$ for the IBB landscape is depicted (light gray line), with the partial correlations controlling for the variance explained by the IBR (black line) and IBB (dark gray line) also shown. In all panels, the gray vertical bar indicates the modern day (generation 1020), and the dotted lines indicate the standard deviation of the Mantel correlations across 10 Monte Carlo iterations.
a landscape and the subsequent patterns of genetic variation that arise over time [37]. As increasingly altered and fragmented landscapes reduce gene flow for many species [38], the need for robust landscape genetic inference to help inform connectivity conservation is high [39]. Indeed, a spatially explicit understanding of landscape effects on gene flow could guide such activities as the identification and establishment of migration corridors [40], construction of highway crossing structures, reserve network planning [41], and predicting the ability of species to shift ranges in response to climate change [42].

The potential for this emerging field to inform connectivity conservation is, of course, dependent upon the accuracy of the inference relating landscapes to gene flow. The vast majority of published spatially explicit landscape models are identified due to their correlation with an observed pattern of genetic isolation. Correlations alone, however, are not definitive proof of causation [43]. The variables comprising a landscape model may simply be spuriously related to the true driver(s) of genetic isolation. If one fails to evaluate the causal mechanism as a model variable, and instead focuses on a spurious but correlated variable, one risks misinforming conservation efforts. Importantly, the consequences of false inference could be highly detrimental to effective conservation efforts.

As a simple hypothetical example, consider the earlier example where major highways are a barrier to gene flow, but because highways are only in low elevation valleys, there is a potential for false inferences if one focuses on elevation as the driver of genetic isolation. The conservation implication of this false inference would be to protect or create a continuous network of high elevation linkages. In reality, a strategy focused on constructing a few key highway crossing structures would likely be much more effective.

This example highlights the importance of testing a full range of plausible causal models (e.g., [44]). Because of correlations between alternative landscape models, partial regression approaches allow for causal model inferences among competing but correlated landscape hypotheses and thereby afford greater rigor (e.g., [24, 25]). Validation of models with new data also adds to the rigor and confidence in causal models [45]. Though the risk of false inference can be reduced by these measures, it would be desirable to have an independent test to validate these inferences [23, 46].

Returning to our example, consider how the model based erroneously on the effect of elevation might perform if it were used as the basis to simulate gene flow. Low elevation valleys without major roads would have high resistance in this model, but in reality would offer low resistance because roads are the major driver of genetic isolation, not elevation. Simulating gene flow on this poor landscape model would result in greater simulated genetic isolation across roadless valleys than exists in reality. Consequently, the simulated genetic distance would likely be a poor fit to the actual genetic distance among individuals in the population. This example highlights how error in the empirical and simulation approaches is independent. Error arising from a spuriously correlated model is not likely to perform well in a simulation, where error arises from stochastic population genetic processes like drift, mutation, and migration.

In this analysis, when we simulated gene flow on a resistance model identified as causal in the empirical model selection approach of Shirk et al. [25], the simulated pattern of genetic isolation was strikingly related to the current observed pattern. Importantly, simulations of alternative models based on isolation by barrier or distance did not meet the expectations of a causal model. This independent test strongly affirms the conclusions of Shirk et al. [25]. Namely, that for mountain goats in the Cascade Range, Washington (USA), a resistance model in which gradients of elevation, landcover, and major highways govern gene flow. The strong relationship between this model and the current pattern of genetic isolation using both empirical and simulation approaches represents a robust spatial understanding of the process of gene flow on this landscape. Agencies charged with improving habitat connectivity for this species within the study area may use this model with greater confidence that it will accurately inform ongoing conservation efforts.

While the pattern of genetic isolation in the simulated present-day generation closely matched the current population’s genetic isolation, it is interesting to note that the simulated historic population was also highly related to the current population. Some studies note contemporary landscapes have a strong effect on population structure [47, 48]. Others have shown historic landscapes have a greater effect [49]. In the Cascade Range, relatively little landscape change has occurred over much of the mountain goat’s high elevation habitat. The main exception has been the construction of several major highways. The similarity between the modern and historic landscapes likely explains the strong relationship between the simulated historic population and the modern population’s genetic distance.

If a particular model is highly supported by this combination of empirical and simulation analysis, researchers may also have greater confidence projecting the simulation forward into future generations. Here, we simulated mountain goat population genetics several hundred generations into the future. That the genetic isolation of simulated future generations is highly related to the current population suggests much of the effect of modern landscape change is already reflected in the population’s genetic structure. Though not the focus of this study, these forward simulations and others based on alternative landscapes will be used to estimate how landscape change may shape this population’s genetic diversity over time.

There are two important limitations to this approach that should be noted. First, there are important assumptions inherent in any simulation software. In CDPOP, simplifications such as nonoverlapping generations and a constant population size over time do not accurately reflect mountain goat population demographics within the study area. Yet despite these simplistic assumptions, the simulation was able to reproduce a population genetic structure remarkably similar to the actual population. Ongoing development of CDPOP will add greater flexibility to model dynamic populations with overlapping generations. Such advances should
increase the power of simulations to support landscape genetic inferences [28].

Another limitation of this study was the simplistic treatment of landscape change from the historic to the modern period. We simulated mountain goat population genetics within the study area in two steps—one based on a modern landscape model, and another based on removing the effect of roads, urbanization, and other anthropogenic factors from that model. In reality, these landscape changes took place over about 20 generations. A detailed spatial representation of landscape change over time is easily incorporated into simulations, but in practice, limited availability of spatial data that captures landscape change makes this task difficult. Despite the simplistic treatment of this issue in our simulations, the strength of the relationship between the simulated and modern population suggests this limitation did not significantly hinder our ability to closely reproduce the current population’s pattern of genetic isolation.

5. Conclusion

We believe that combining empirical model selection with simulation analysis greatly strengthens landscape genetic inference. Empirical analysis can evaluate the relative support among a large suite of plausible alternative models, and decisively reject those that are inconsistent with the pattern in the data. This results in identification of an inferred process governing gene flow. It is then possible to simulate gene flow under the inferred process and produce a pattern of population structure and differentiation. If this deductive analysis produces population genetic structure that is highly related to that observed in the empirical population, it provides strong support for the identified process.

Acknowledgment

The authors thank Cliff Rice, Washington Department of Fish and Wildlife, for providing spatial data regarding mountain goat habitat and population size within the study area.

References

8 International Journal of Ecology


