MOLECULAR ECOLOGY

Molecular Ecology (2010) **19**, 3592–3602 doi: 10.1111/j.1365-294X.2010.04656.x

Spurious correlations and inference in landscape genetics

SAMUEL A. CUSHMAN* and ERIN L. LANDGUTH†

**USDA Forest Service, Rocky Mountain Research Station, 800 E Beckwith, Missoula, MT 59801, USA, †Individualized Interdisciplinary Graduate Program, University of Montana, 800 E Beckwith, Missoula, MT 59801, USA*

Abstract

Reliable interpretation of landscape genetic analyses depends on statistical methods that have high power to identify the correct process driving gene flow while rejecting incorrect alternative hypotheses. Little is known about statistical power and inference in individual-based landscape genetics. Our objective was to evaluate the power of causal-modelling with partial Mantel tests in individual-based landscape genetic analysis. We used a spatially explicit simulation model to generate genetic data across a spatially distributed population as functions of several alternative gene flow processes. This allowed us to stipulate the actual process that is in action, enabling formal evaluation of the strength of spurious correlations with incorrect models. We evaluated the degree to which naïve correlational approaches can lead to incorrect attribution of the driver of observed genetic structure. Second, we evaluated the power of causal modelling with partial Mantel tests on resistance gradients to correctly identify the explanatory model and reject incorrect alternative models. Third, we evaluated how rapidly after the landscape genetic process is initiated that we are able to reliably detect the effect of the correct model and reject the incorrect models. Our analyses suggest that simple correlational analyses between genetic data and proposed explanatory models produce strong spurious correlations, which lead to incorrect inferences. We found that causal modelling was extremely effective at rejecting incorrect explanations and correctly identifying the true causal process. We propose a generalized framework for landscape genetics based on analysis of the spatial genetic relationships among individual organisms relative to alternative hypotheses that define functional relationships between landscape features and spatial population processes.

Keywords: causal modelling, CDPOP, landscape genetics, landscape resistance, partial Mantel test, simulation modelling, spurious correlation

Received 16 September 2009; revision received 28 January 2010; accepted 10 February 2010

Introduction

Recent advances in landscape genetics have greatly facilitated developing rigorous, species-specific, and multivariate characterizations of habitat connectivity for animal species (Manel *et al.* 2003; Cushman 2006; Holderegger & Wagner 2006; Storfer *et al.* 2007; Segelbacher *et al.* 2010). Landscape genetics explicitly quantifies the effects of landscape composition, configuration and matrix quality on spatial patterns in neutral and adaptive genetic variation and underlying microevolutionary processes (Manel *et al.* 2003; Holderegger & Wagner 2006, 2008; Storfer *et al.* 2007). Balkenhol *et al.* (2009a) argue that while a very exciting emerging field, landscape genetics is currently limited by several major theoretical and methodological challenges. For example, they note that while many statistical methods have been proposed for analysing the spatial distribution of genetic variation, and for linking observed genetic patterns to landscape characteristics, it is still unclear under which conditions the various methods produce accurate, valid and repeatable results in a landscape genetics context.

Correspondence: Samuel A. Cushman, Fax: +1 406 543 2663; E-mail: scushman@fs.fed.us

There have been remarkably few quantitative comparisons among alternative statistical methods of inferring pattern-process relationships in landscape genetics. Perhaps the most comprehensive such analysis to date is Balkenhol *et al.* (2009b) in which the authors used a simulation model to produce spatial genetic patterns with known relationships to landscape features and evaluated Type I error rates and power of several alternative statistical methods to correctly identify the driving process. That analysis differs from this in that it used a population-based spatial migration model with migration probabilities depending on cost distance rather than an individual-based cost-distance approach. In addition, while it evaluated the power of Mantel tests, it did not evaluate the power of causal modelling using resemblance matrices, which is the focus of this paper.

Balkenhol *et al.* (2009a) note that most past landscape genetic studies have assumed certain patterns of population structure a priori and limited analysis to relatively simple null-hypothesis testing, such as testing for the presence of a barrier, rather than comparing the evidence for competing hypotheses involving more complex landscape effects. They argue that this may lead to important misinterpretations. For example, many population and landscape genetic studies have used *F*-statistics (Wright 1943) or assignment tests (Pritchard *et al.* 2000; Corander *et al.* 2003; Francois *et al.* 2006) to relate genetic differences among predefined subpopulations, propose interpopulation distance relationships (Witherspoon *et al.* 2007), identify putative movement barriers (Manni *et al.* 2004; Funk *et al.* 2005) or correlations with landscape features (Vitalis & Couvet 2001; Spear *et al.* 2005). Often once discrete subpopulations have been identified, post hoc analyses are performed, correlating observed genetic patterns with interpopulation distance or putative movement barriers (e.g. Proctor *et al.* 2005). Such ad hoc or post-hoc ''hypothesis'' testing runs serious risks of producing erroneous conclusions (Holderegger & Wagner 2008). This is particularly true given that populations often have substantial internal structure (Wright 1943; Gompper *et al.* 1998; Van Horn *et al.* 2004).

Assuming the existence of discrete subpopulation structure, followed by application of methods designed to detect such structure, followed by post-hoc analysis to identify potential causes for the inferred population structure is a perilous path of inference. The implicit assumptions underlying an ecological analysis in large part drive the questions asked and the methods used, which in turn constrain the results that will be obtained (Cushman & Huettmann 2009). Methods to delineate discrete populations are known to identify boundaries even in continuously structured populations (Schwartz & McKelvey 2009). Thus, beginning with an assumption of discrete populations and then utilizing methods designed to detect them runs a large risk of producing erroneous inferences about genetic structure. The subsequent post-hoc analysis of putative subpopulations with respect to selected landscape features adds a second level of inferential peril. Such analyses begin with an idea of population substructure and then seek correlations with landscape features that are coincident; when such correlations are found the researchers typically assume that these coincident features are the cause of the inferred substructure. Observing a coincidence between a landscape feature and a putative subpopulation boundary does not confirm the role of that feature in creating the putative population boundary. This is an example of the logical deductive fallacy of affirming the consequent (Cushman & Huettmann 2009).

Simulation models have a particularly important role to play in investigating the ability of alternative statistical methods to correctly attribute the causes of observed genetic substructure. They provide control over simulated process such that there is no ambiguity as to causes of observed genetic patterns (Epperson *et al.* 2010). In this paper we used an individual-based, spatially-explicit simulation model to generate genetic data across a spatially distributed population as functions of several alternative gene flow processes. First, we evaluate the degree to which naïve correlational approaches can lead to incorrect attribution of the driver of observed genetic structure. Second, we assess the power of causal modelling with partial Mantel tests on resistance gradients to identify a correct explanatory model and reject incorrect alternative models. Third, we quantify how rapidly after the gene flow process is initiated that we are able to reliably detect the effect of the correct model and reject the incorrect models. Finally, we propose a generalized framework for landscape genetic inference based on analysis of the spatial genetic relationships among individual organisms across a range of alternative hypotheses that define functional relationships between multiple landscape features and spatial population processes.

Methods

Simulation scenarios

We use the CDPOP individual-based landscape genetics model (Landguth & Cushman 2010) to conduct all simulations. CDPOP models genetic exchange for a given resistance surface and $n_{x,y}$ located individuals as functions of individual-based movement through mating and dispersal, vital dynamics, and mutation. The model represents landscape structure flexibly as

resistance surfaces whose value represents the step-wise cost of crossing each location. Mating and dispersal are modelled as probabilistic functions of cumulative cost across these resistance surfaces. The model provides a framework for provides a framework for simulating the emergence of spatial genetic structure in populations resulting from specified landscape resistance processes governing organism movement behaviour.

In this analysis, we simulate genetic exchange without mutation in a population of 1248 individuals under three landscape resistance scenarios: (a) isolation by Euclidean distance (Fig. 1a); (b) isolation by barrier (Fig. 1b) and (c) and isolation by landscape resistance (Fig. 1c). In the Euclidean case all resistance values are set to 1, making all mating and dispersal a function of Euclidean distance alone. In the barrier case, there is resistance of 1 resulting in a Euclidean distance function in all cells except for the barrier, which has a value much greater than 1 to stipulate an absolute barrier to dispersal or mate choice dividing the landscape. In the landscape case, resistance is a continuous function of multiple environmental attributes. For this paper, we use the landscape resistance surface identified by Cushman *et al.* (2006) as the most supported out of 110 alternative models for black bear (*Urusus americanus*) gene flow in northern Idaho, USA. For each resistance surface and the 1248 individuals, cost distance matrices were calculated as the cumulative cost associated with traversing the least cost path from each individual's location to every other individual's location using ArcGIS COSTDISTANCE function (ESRI 1999–2008). We initialized the locations for 1248 individuals on each of these cost surfaces by populating the landscape resistance surface (Fig. 1c) with locations on a grid with a 1.6 km spacing that fell on pixels with landscape resistance value of less than 6. This was done to place individuals in habitat that was relatively suitable for the species given this landscape resistance hypothesis. The same 1248 points were used in all simulations for all three resistance processes. We initialized genotypes for these 1248 individuals by randomly assigning allelic states across 10 loci each with 10 alleles.

We simulated change in individual genotypes across this grid of 1248 individuals for each of the three resistance processes following random assortment in Mendalian inheritance. For each scenario we produced 100 replicate simulations to quantify the variability in spatial genetic structure produced. We ran all simulations for 1000 non-overlapping generations. We simulated mating with male replacement and without female replacement at each generation. Each mate pair was specified to have a number of offspring following a Poisson process with mean of 4, to guarantee a positive lambda value to ensure that all locations were filled at each time step and to avoid immigrants. We maintained a constant population of 1248 by discarding the remaining offspring once all the 1248 grid locations were occupied. This is equivalent to forcing emigration out of the study area once all available home ranges are occupied (Landguth & Cushman 2010).

We used an inverse-square probability function to select mates and offspring dispersal destinations as

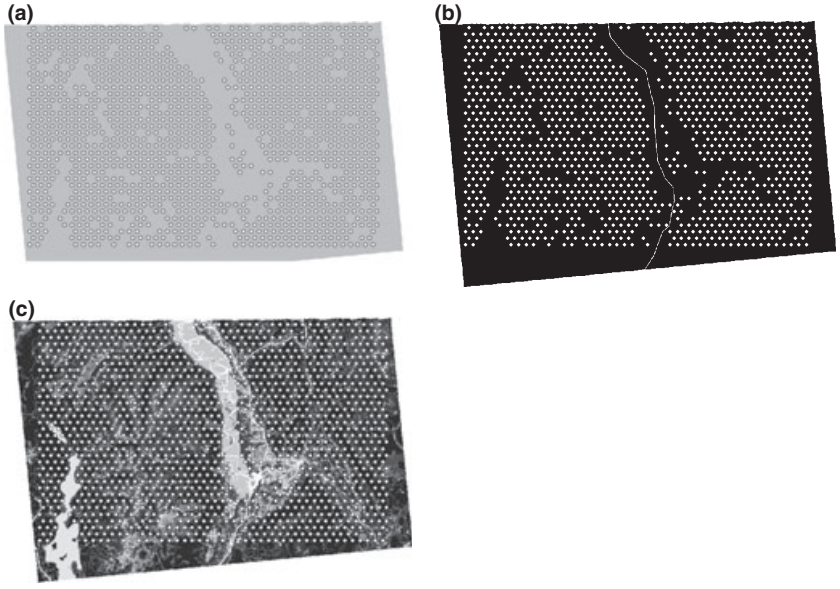

Fig. 1 The three resistance models used in this analysis: (a) isolation by Euclidean distance; (b) isolation by barrier and (c) isolation by landscape cost distance. White points indicate locations of the 1248 individual organisms forming the simulated population.

functions of cost distance, with a maximum movement threshold set at 39 200 cost units, 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 results in all mate choices and dispersal distances to be less than or equal to 39 200 cost-units apart, with probability of mate distance or dispersal distance within that limit specified by an inverse square function. This is reasonable given that spatial processes on a two-dimensional surface usually are governed by an inverse square relationship, and probability of moving to a particular location on a plane is inversely related to the square of the distance from the origin to the destination.

Causal modelling with partial Mantel tests

At every 10-year time step for each of the 100 Monte Carlo runs and each of the three resistance processes, we used the Bray–Curtis percentage dissimilarity measure to produce genetic distance matrices for all pairs of individuals. Then we calculated three Mantel tests (Mantel 1967) and six partial Mantel tests to assess the degree of association between each genetic distance matrix and the three cost distance matrices computed from each resistance surface and the 1248 individuals (Table 1). These tests were selected to allow us to implement the causal modelling framework used by Cushman *et al.* (2006).

Causal modelling with distance matrices using partial Mantel tests provides expected outcomes in terms of significance and non-significance of a series of tests that can be used to reject explanations that are not consistent with the expectations of the causal model. For example, Fig. 2 shows the three causal models corresponding to each of the three resistance processes of isolation by distance, isolation by barrier and isolation by landscape resistance, and lists the expected outcomes of Mantel and partial Mantel tests for each process. The isolation by distance hypothesis proposes that gene flow is a function of Euclidean distance among individuals and is not independently related to barriers or landscape structure. In contrast, the barrier model proposes that gene flow is a function of Euclidean distance on either side of an absolute barrier and is not driven by global isolation by distance or differential landscape resistance. The landscape resistance model proposes that gene flow is a function of landscape resistance as specified in Fig. 1c and is not driven by isolation by Euclidean distance or any barrier features in the landscape. We preformed the simple Mantel and partial Mantel tests listed in Table 1 using the ECODIST package in *r*, with 1999 permutations to calculate statistical significance.

Table 1 List of Mantel and partial Mantel tests calculated between genetic distance and cost distance

Independent variable	Variable partialled out	Acronym
Landscape	None	L
Distance	None	D
Barrier	None	B
Barrier	Landscape	B|L
Barrier	Distance	B|D
Landscape	Barrier	L|B
Landscape	Distance	L|D
Distance	Barrier	D|B
Distance	Landscape	D|L

Results

Simple Mantel correlations

CDPOP simulation of isolation by Euclidean distance produced significant correlation between barrier model cost distance and genetic distance within 10 generations (Fig. 3g). The correlation rose rapidly to asymptote at approximately $r = 0.4$ in 120 generations. Similarly, implementing the barrier process in CDPOP produced significant correlation between barrier model cost distance and genetic distance within 10 generations (Fig. 3d), which increased rapidly to an asymptote at approximately 300 generations. Simulating isolation by landscape resistance resulted in strong correlation between landscape cost distance and genetic distance, with significant correlation apparent at the first reported time step of 10 generations after the beginning of the simulation (Fig. 3d). The correlation between landscape cost distance and genetic distance increased rapidly to an asymptote at approximately 300 generations. This shows that the spatial process of isolation by landscape resistance results in detectable spatial genetic patterns almost immediately, but takes several hundred generations to equilibrate at this population size and dispersal distance.

Spurious correlations

The simple Mantel tests between the cost-distances corresponding to the incorrect alternative hypotheses are nearly identical to those associated with the simulated actual process in all three cases (Fig. 3b,c,e,f,h,i). These are spurious correlations, as we stipulated the driving process and therefore know there is no independent effect of these alternative hypotheses. In all three cases, this spurious effect is detected almost immediately and is very strong. In the case where we stipulated an isolation by distance process, the simple Mantel tests between the cost-distances corresponding

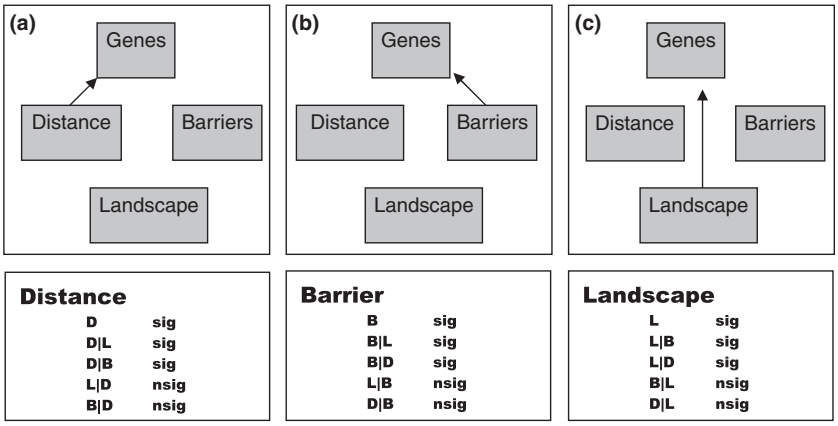

Fig. 2 Schematic showing the three causal models evaluated in this analysis. (a) Genetic structure is a function of Euclidean distance among individuals with no independent relationship with barriers or landscape resistance; (b) genetic structure is a function of barriers with no independent relationship with Euclidean distance or landscape resistance and (c) genetic structure is a function of landscape resistance with no independent relationship with Euclidean distance or barriers. Below each causal model are lists of the diagnostic Mantel tests used to evaluate support and expected significance outcomes if that causal model were correct. For example, in the Distance causal model: D—simple Mantel test between genetic distance and Euclidean distance, D|L—partial Mantel test between genetic distance and Euclidean distance, partialling out landscape cost distance, D|B—partial Mantel test between genetic distance and Euclidean distance, partialling out barrier distance, L|D—partial Mantel test between genetic distance and landscape cost distance, partialling out Euclidean distance, B|D—partial Mantel test between genetic distance and barrier distance, partialling out Euclidean distance.

to the landscape and barrier hypotheses produce very highly significant spurious correlations which are nearly identical to those produced by the correct distance model (Fig. 3). Similarly, in the case where we stipulated an isolation by barrier process, Mantel *r* values for the landscape hypothesis are nearly identical to those produced by the correct barrier resistance process, and those for the distance hypothesis only slightly lower. Likewise, in the simulation of isolation by landscape resistance, Mantel *r* values for the barrier hypothesis are nearly identical to those produced by the correct landscape resistance process, and those for the distance hypothesis only slightly lower. These spurious effects are detected almost immediately and reach an asymptotic effects size equivalent to that of the correct stipulated process.

Power of causal modelling

In all three simulations stipulating isolation by distance, barrier and landscape resistance, causal modelling consistently rejected the incorrect alternative hypotheses and supported the correct causal process (Figs 4–6). Causal modelling had equivalent power for each of the three stipulated processes. Therefore, we only report power results for the case where landscape resistance was the simulated driving process. At 10 generations causal modelling had power of 0.82 to correctly reject barrier and distance models and support the landscape model (Fig. 7). Power rose to 86% at 30 generations, 90% at 40, and 96% of the time causal modelling correctly rejected the incorrect models and affirmed the correct landscape hypothesis after 50 generations of simulation time. Beyond 50 generations causal modelling had apparent power of that varied between 96% and 100%, due to the stochastic nature of the simulation.

Discussion

Spurious correlations in landscape genetics

Few studies have explored the performance of different statistical approaches to evaluate pattern-process relationships in landscape genetics. Balkenhol *et al.* (2009a,b) provides the most comprehensive analysis to date of performance of a range of statistical tools in terms of Type I error rates. They report Mantel and partial Mantel tests have high Type I error rates in their simulations. We confirm the high risk of Type I error when testing pattern-process relationships with simple Mantel tests. However, we do not believe this is a problem with the Mantel test per se, as much as a fundamental attribute of spatially correlated alternative causal hypotheses in landscape genetics (e.g. Murphy *et al.* 2008). Our analysis demonstrates that most alternative hypotheses of factors driving gene flow across landscapes will produce correlated expectations. For example, in our simulation isolation by Euclidean distance, a movement barrier and a complex landscape

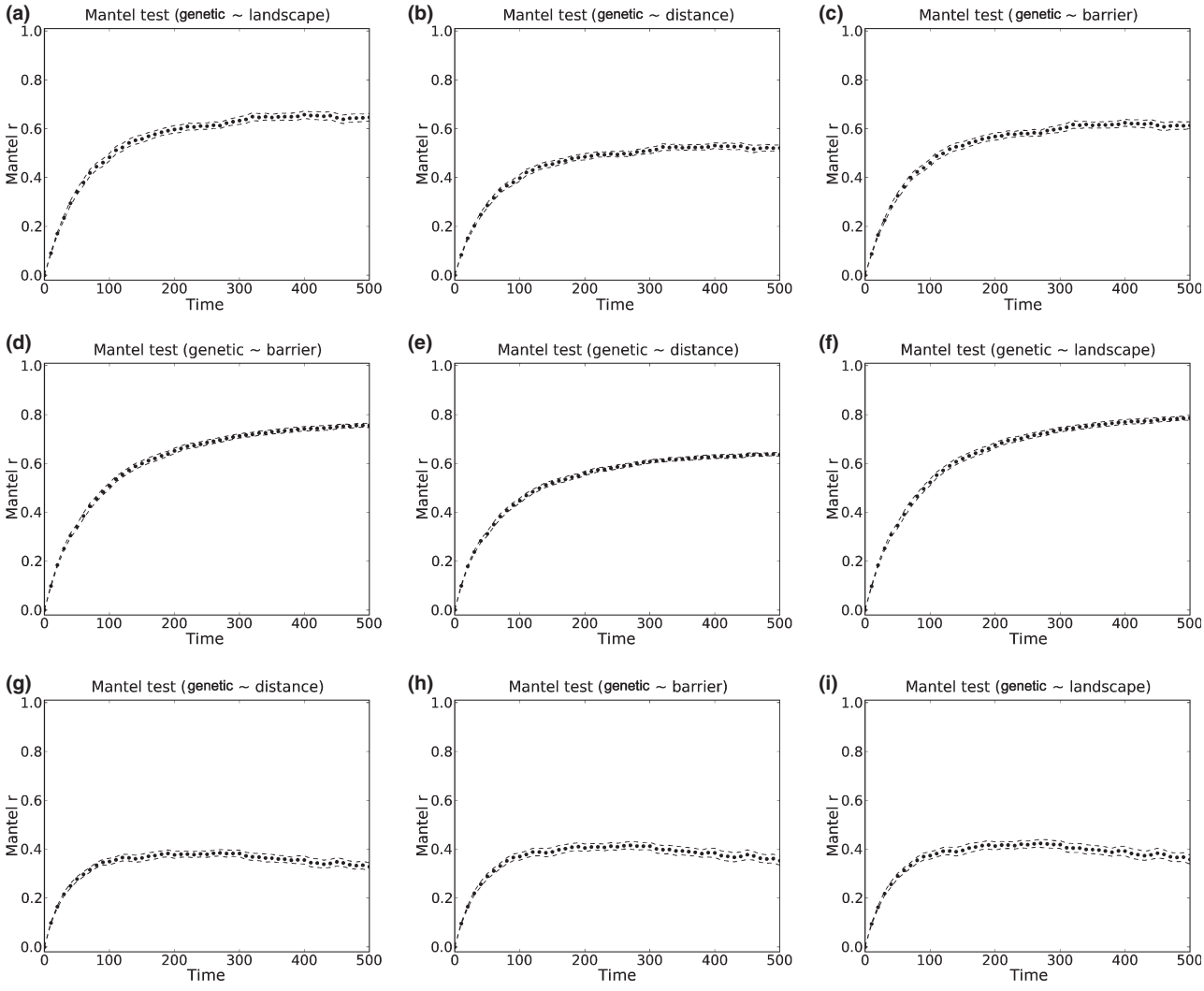

Fig. 3 Plot of simple Mantel r values across the first 500 generations of simulation time showing the correlation between genetic distances and the correct stipulated process (3a,d,g) and spurious correlations between genetic distance and incorrect alternative hypotheses (3b,c,e,f,h,i). Error bars show 95% confidence interval across 100 Monte Carlo simulation runs. The panels show correlation between genetic distance and (3a) Euclidean distance when Euclidean distance is the stipulated process, (3d) the barrier model when it is the stipulated process, (3g) the landscape model when it is the stipulated process. The remaining panels are spurious correlations between genetic distance and (3b) the distance model and (3c) the landscape resistance model when barrier is the stipulated process; between genetic distance and (3e) the distance model and (3f) the landscape model when the barrier model is the stipulated process; between genetic distance and (3h) the barrier model and (3i) the landscape model when isolation by distance is the stipulated gene flow process.

resistance hypothesis produce highly correlated cost distance expectations, although they are conceptually very different hypotheses. Similarly, using a simulation approach Murphy *et al.* (2008) showed that spurious isolation by distance signals were a common byproduct of a variety of spatial gene flow processes. Finding strong spurious relationships in both simulated and empirical data sets emphasizes the need for carefully constructed a priori hypothesis statements and powerful statistical methods to evaluate support for competing alternative hypotheses. This highlights the need to utilize more sophisticated and robust approaches in landscape genetic inference.

Dangers of naïve confirmatory analysis

Landscape genetics has focused on describing and mapping population units (Pritchard *et al.* 2000; Corander *et al.* 2003; Francois *et al.* 2006), measuring gene flow (Balkenhol *et al.* 2009a) and modelling factors that predict rates and patterns of gene flow within and between populations (Coulon *et al.* 2004; Cushman *et al.* 2006;

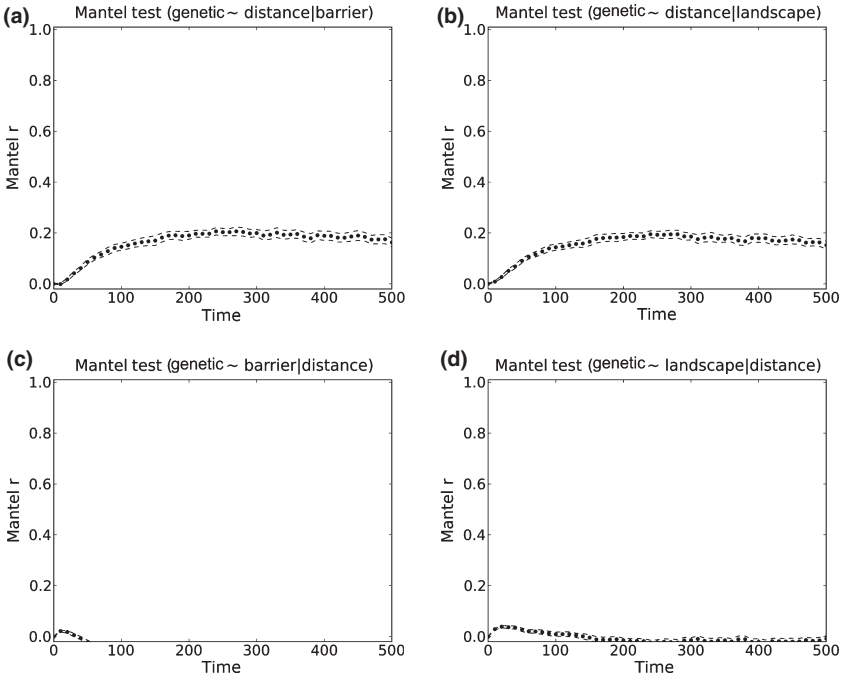

Fig. 4 Plot of diagnostic partial Mantel tests over the first 500 generations of simulation time for the case when the stipulated process simulated in CDPOP is isolation by Euclidean distance, with no independent effect of barriers or landscape resistance. (a) partial Mantel test between genetic distance and Euclidean distance, partialling out the effect of barriers; (b) partial Mantel test between genetic distance and Euclidean distance, partialling out the effect of landscape resistance; (c) partial Mantel test between genetic distance and barrier distance, removing the effect of Euclidean cost distance and (d) partial Mantel test between genetic distance and landscape cost distance, removing the effect of Euclidean distance. These four tests all support the correct causal model and refute the incorrect alternative models, with significant Mantel correlations appearing within 10 generations of initiation of the causal process.

McRae & Beier 2007; Segelbacker *et al.* 2010). In many cases researchers have proposed a model, sought evidence consistent with that model, and finding such evidence, concluded that the proposed model of population structure was correct. However, it is very common that multiple alternative landscape genetic hypotheses are highly correlated (Murphy *et al.* 2008; fig. 3 in this paper). This leads to very high risk of Type I error in applying simple correlational approaches seeking to confirm the expectation of a hypothesis. Our results confirm the high risk of incorrect inference using simple Mantel tests due to high Type I error rates due to intercorrelation of alternative hypotheses. Concluding a hypothesis is supported using naïve correlations is an example of the logical deductive fallacy of affirming the consequent (Cushman & Huettmann 2009), in which a result consistent with a hypothesis is mistaken for demonstration of the truth of that hypothesis.

Causal modelling and type I error

Cushman *et al.* (2006) proposed a causal modelling approach to lessen the risk of attributing incorrect causal mechanisms for observed spatial genetic patterns due to spurious correlations. Using multi-model inference on an empirical data set, Cushman *et al.* (2006) demonstrated that had they tested the single hypothesis of isolation by barriers against a null hypothesis of global panmixia they would have incorrectly affirmed the barrier hypothesis. Likewise, had they evaluated the isolation by barrier and isolation by distance model they would have incorrectly affirmed the isolation by distance model, as it was more supported than the barrier model. This again would be an inferential error, as the distance model was not among the most supported models.

The analyses presented in this paper were designed to evaluate the severity of spurious correlations in landscape genetic studies of gene flow, and to assess the efficacy of the causal modelling framework using partial Mantel tests to reject incorrect hypotheses and support the correct causal process. The individual-based landscape-genetic simulation model we use provides control over the implemented process. The model allowed us stipulate the actual process that is in action. This, in turn, allowed us to formally evaluate the strength of spurious correlations with incorrect models and to evaluate the rigor of the causal modelling approach.

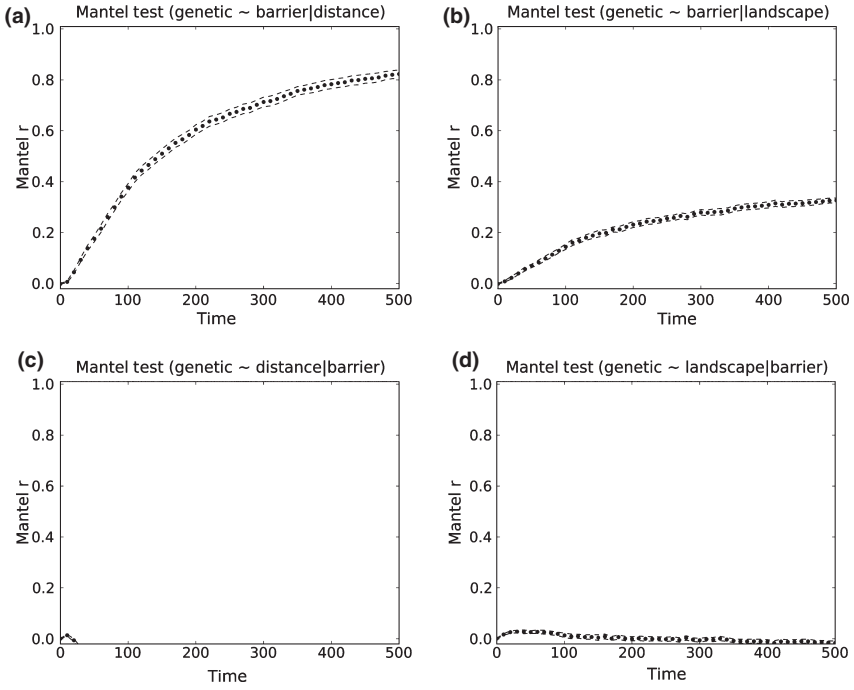

Fig. 5 Plot of diagnostic partial Mantel tests over the first 500 generations of simulation time for the case when the stipulated process simulated in CDPOP is isolation by the barrier model, with no independent effect of Euclidean distance or landscape resistance. (a) partial Mantel test between genetic distance and barrier distance, partialling out the effect of Euclidean distance; (b) partial Mantel test between genetic distance and barrier distance, partialling out the effect of landscape resistance; (c) partial Mantel test between genetic distance and Euclidean distance, removing the effect of barrier model cost distance and (d) partial Mantel test between genetic distance and landscape cost distance, removing the effect of barrier model cost distance. These four tests all support the correct causal model and refute the incorrect alternative models, with significant Mantel correlations appearing within 10 generations of initiation of the causal process.

The results clearly show that there is high risk of incorrect attribution of causality in the absence of formal evaluation of support for competing alternative hypotheses in landscape genetics. In this analysis, each simulated process produced genetic structure that was correlated with the two incorrect hypotheses nearly as strongly as with the correct causal process. Without formal causal modelling to partial out the effects of each alternative model it would have been impossible to identify which model was correct. Our results demonstrate that partial Mantel tests in a causal modelling framework do not suffer from high Type I error rates, in marked contrast to simple Mantel tests. Specifically, the spurious relationships with incorrect alternative hypotheses were removed effectively by the partial Mantel tests implemented in the causal modelling framework. This suggests that contrary to previous findings (Balkenhol *et al.* 2009b), partial Mantel do not appear to suffer from elevated Type I error rates, at least when implemented in a causal modelling framework.

In most published spatial genetics work, isolation by barriers into discrete panmictic subpopulations has been proposed as the expectation and methods designed to detect such populations are employed. However, given the strength and universality of spurious correlations between genetic structure and barrier, distance and landscape models, this simplistic approach to hypothesis confirmation produces equivocal inferences and may frequently lead to affirmation of incorrect explanatory models.

Type II error, power and causal modelling

Our results also demonstrate that causal modelling with partial Mantel tests appears to be very robust in its ability to reject incorrect causal mechanisms and correctly identify the driving process that is responsible for the observed pattern of genetic variation across the landscape. In our analysis, causal modelling had very high power to correctly identify the driving cause and reject the spurious effects of correlated alternative hypotheses. Importantly, causal modelling had a power of 86% to correctly identify a causal process and reject incorrect alternatives within 20 generations after the initiation of the process, and nearly perfect power after

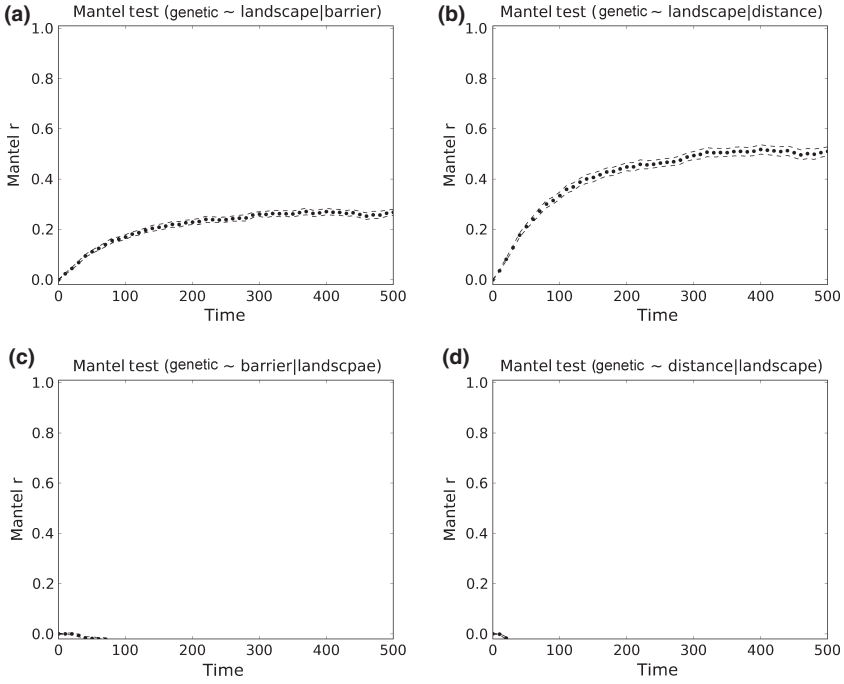

Fig. 6 Plot of diagnostic partial Mantel tests over the first 500 generations of simulation time for the case when the stipulated process simulated in CDPOP is isolation by landscape resistance, with no independent effect of Euclidean distance or barrier. (a) partial Mantel test between genetic distance and landscape costdistance, partialling out the effect of barrier model distance; (b) partial Mantel test between genetic distance and landscape cost distance, partialling out the effect of barrier model resistance; (c) partial Mantel test between genetic distance and Euclidean distance, removing the effect of landscape cost distance and (d) partial Mantel test between genetic distance and Euclidean distance, removing the effect of landscape cost distance. These four tests all support the correct causal model and refute the incorrect alternative models, with significant Mantel correlations appearing within 10 generations of initiation of the causal process.

30 generations. The CDPOP model implements stochastic, individual-based mating and dispersal as functions of the stipulated landscape resistance hypotheses. Observing near perfect performance of the causal modelling approach across 100 runs of a stochastic model we believe is strong evidence for its effectiveness. Importantly, the performance of the causal modelling approach was equally high when the stipulated process was isolation by barriers, isolation by distance or isolation by landscape resistance. This suggests it may provide a generalized framework for evaluating support among alternative hypotheses regarding the causes driving patterns of gene flow.

Landscape resistance, multi-model inference and causal modelling: a generalized framework for landscape genetics

We believe that the combination of landscape resistance models as predictors of individual-level genetic variation and causal modelling provides a robust framework for landscape genetic inference. Specifically, representing the processes that drive population structure as continuously resistant landscapes and evaluating the support for alternative landscape resistance models provides a common framework for judging relative support among alternative models. Classic models of isolation by distance, and isolation by barriers into discrete panmictic populations can be specified as particular landscape resistance surfaces within this framework. The extreme null model of global panmixia can also be specified as a special landscape resistance model (Landguth & Cushman 2010). This allows a formal evaluation of the degree of support for these classic explanations relative to alternative models in which differential resistance across complex landscapes drives gene flow. Indeed, our results suggest that putting these alternative explanations into a common analysis in which the degree of support for each can be evaluated relative to the others is essential. Without such formal evaluation, the severe risk of spurious correlations will render most analyses equivocal. However, an individual-level, resistance-process based framework allows the expression of all alternative models of the factors driving population structure in comparable forms that can be tested in a common framework, such as causal modelling.

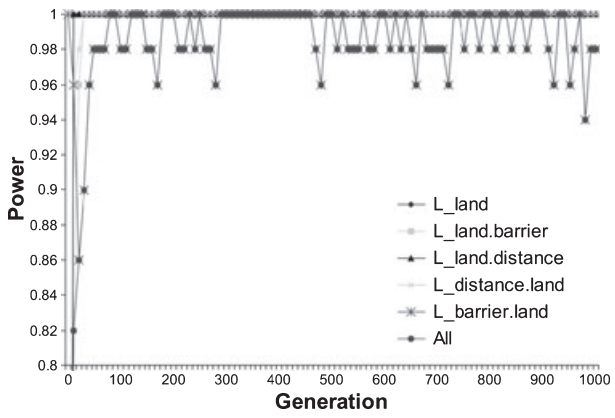

Fig. 7 Plot of power to affirm the correct causal model and refute the incorrect alternative causal models across generations since inception of the causal process. All: power for correct result from all five diagnostic Mantel tests; L_land: power for correct result for simple Mantel test between genetic distance and landscape cost distance; L_land.barrier: power for correct result for partial Mantel test between genetic distance and landscape cost distance partilling out barrier; L_land.distance: power for correct result for partial Mantel test between genetic distance and landscape cost distance, partialling out Euclidean distance; L_distance.land: power for correct result for partial Mantel test between genetic distance and Euclidean distance, partialling out landscape cost distance; L_barrier.land: power for correct result for partial Mantel test between genetic distance and barrier distance, partialling out landscape cost distance.

Combining empirical analysis and simulation in landscape genetics

In landscape genetics, the integration of empirical and simulation analyses provides a tremendous opportunity to increase understanding of spatial population processes (Cushman 2006; Balkenhol *et al.* 2009a) and greatly lessen risk of inferential errors regarding pattern-process relationships. Specifically, empirical analyses, such as Cushman *et al.* (2006), provide inductive inference of process from pattern in the data by ranking support for multiple alternative hypotheses. There is still substantial risk of inferential error through this inductive process; the true driving process may not be among those evaluated. Simulation modelling to recreate the genotypes of individuals in continuous populations that would be expected given the inferred process provides the reverse, deductive inference (Landguth & Cushman 2010). There is still a risk of inferential error through this deduction; there may be multiple processes that could create the same pattern in the data. However, by linking empirical analysis and simulation modelling we can obtain a higher level of confidence in our results (Cushman 2006). Specifically, seeing that analysis of a given empirical data set provides strongest support for a particular process, and then observing that the process is able to recreate the empirical genetic pattern independently based on theoretical processes implemented in a simulation model provides a strong test of the inferred pattern-process relationship. Our analysis provides one example of how this can work, with Cushman *et al.* (2006) inferring a causal process from empirical data, and the simulation results presented here confirming that that process creates a spatial genetic pattern consistent with empirical causal modelling results.

Scope and limitations

Here we have shown that the causal modelling framework effectively identifies a correct hypothesis while rejecting incorrect explanations in individual-based landscape genetic analysis. However, the challenge of proposing the correct model is enormous. In this paper we had the luxury of being able to stipulate the correct model and simulate the genetic structure that would result from it. In empirical analysis this is never the case. In empirical analysis we don't know the true causal process, and must propose plausible alternatives to test. However, causal modelling can be linked with a multi-model approach to more rigorously implicate causality and reduce the risk of affirming the consequent in landscape genetics (e.g. Cushman *et al.* 2006). Often this is best done by evaluating support among multiple competing landscape resistance models within the causal modelling framework (e.g. Cushman *et al.* 2006).

This analysis evaluated power to discriminate among isolation by distance, barrier and landscape resistance alternative hypotheses. Another important question not evaluated in this paper is the success of the causal modelling approach to correctly discriminate among correlated alternative landscape resistance hypotheses.

Our analysis is based on an individual-based simulation model. We think this is a powerful framework to evaluate this question, given it allows the stipulation of a driving process and explicit control over how it is enacted. However, as in the case of any model, our simulation is a simplification. Further research is needed to evaluate how additional factors not addressed here may influence the performance of the statistical approach implemented here. For example, our simulation used a fixed population size across generations. It will be informative to evaluate how landscape genetic simulation results are influenced by dynamic population size. Similarly, our simulation is a grid-based model, and not a fully spatially-dynamic agent based model. This means that individuals can only occupy a finite number of discrete locations in the landscape, rather than having full flexibility to move to any location as a stochastic movement process. This simplification is necessary to achieve feasible processing times.

References

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

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

Corander J, Waldmann P, Sillanpaa MJ (2003) Baysian analysis of genetic differentiation between populations. *Genetics*, **166**, 367–374.

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.

Cushman SA (2006) Effects of habitat loss and fragmentation on amphibians: a review and prospectus. *Biological Conservation*, **128**, 231–240.

Cushman SA, Huettmann FF (2009) Ecological knowledge, theory and information in space and time. In: *Spatial Complexity, Ecoinformatics and the Conservation of Animal Populations* (eds Cushman SA, Huettman FF), pp. 3–18, Springer, Tokyo.

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

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

ESRI 1999–2008. ArcGIS 9.3. Redlands, CA.

Francois O, Ancelet S, Guillot G (2006) Bayesian clustering using Hidden Markov Random Fields in spatial genetics. *Genetics*, **174**, 805–816.

Funk CW, Blouin MS, Corn PS *et al.* (2005) Population structure of Columbia spotted frogs (*Rana luteivetris*) is strongly affected by the landscape. *Molecular Ecology*, **14**, 483–496.

Gompper M, Gittleman JL, Wayne RK (1998) Dispersal, philopatry, and genetic relatedness in a social carnivore: comparing males and females. *Molecular Ecology*, **7**, 157–163.

Holderegger R, Wagner HH (2006) A brief guide to landscape genetics. *Landscape Ecology*, **21**, 793–796.

Holderegger R, Wagner HH (2008) Landscape genetics. *BioScience*, **58**, 199–207.

Landguth EL, Cushman SA (2009) CDPOP: an individual-based, cost-distance spatial population genetics model. *Molecular Ecology Resources*, **10**, 156–161.

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

Manni F, Guerard E, Heyer E (2004) Geographic patterns of (genetic, morphologic, linguistic) variation: how barriers can be detected by using Monmonier's algorithm. *Human Biology*, **76**, 173–190.

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

McRae BH, Beier P (2007) Circuit theory predicts gene flow in plant and animal populations. *Proceedings of the National Academy of Sciences, USA*, **104**, 19885–19890.

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

Pritchard JK, Stephens P, Donnelly P (2000) Inference of population genetic structure using multilocus genotype data. *Genetics*, **155**, 945–959.

Proctor MF, McLellan BN, Strobeck C, Barclay RMR (2005) Genetic analysis reveals demographic fragmentation of grizzly bears yielding vulnerability by small populations. *Proceedings of the Royal Society Series B*, **272**, 2409–2416.

Schwartz MK, McKelvey KS (2009) Why sampling scheme matters: the effect of sampling scheme on landscape genetic results. *Conservation Genetics*, doi: 10.1007/s10592-10008-19622-10591.

Segelbacher G, Cushman SA, Epperson BK, Fortin M-J, Francois O, Hardy DJ, Holderegger R, Taberlet P, Waits LP, Manel S (2010) Applications of landscape genetics in conservation biology: concepts and challenges. *Conservation Genetics*, doi:10.1007/s10592-009-0044-5.

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

Spear JR, Walker JJ, McCollom TM, Pace NR (2005) Hydrogen bioenergetics in the Yellowstone geothermal ecosystem. *Proceedings of the National Academy of Sciences, USA*, **102**, 2555–2560.

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

Van Horn RC, Engh AL, Scribner KT, Funk SM, Holekamp KE (2004) Behavioral structuring of relatedness in the spotted hyena (*Crocuta crocuta*) suggests direct fitness benefits of clan-level cooperation. *Molecular Ecology*, **13**, 449–458.

Vitalis R, Couvet D (2001) Estimation of effective population size and migration rate from one- and two-locus identity measures. *Genetics*, **157**, 911–925.

Witherspoon DJ, Wooding S, Rogers AR, Marchani EE, Watkins WS, Batzer MA, Jorde LB (2007) Genetic similarities within and between human populations. *Genetics*, **176**, 351–359.

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