MOLECULAR ECOLOGY
RESOURCES
Molecular Ecology Resources (2011) **11**, 922–934 doi: 10.1111/j.1755-0998.2011.03035.x

sGD: software for estimating spatially explicit indices of genetic diversity

A.J. SHIRK* and S.A. CUSHMAN†
Climate Impacts Group, Joint Institute for the Study of the Atmosphere (JISAO), University of Washington, Box 355672, Seattle, Washington, 98195-5762 USA, †USDA Forest Service, Rocky Mountain Research Station, 800 E. Beckwith, Missoula, Montana 59801 USA

Abstract

Anthropogenic landscape changes have greatly reduced the population size, range and migration rates of many terrestrial species. The small local effective population size of remnant populations favours loss of genetic diversity leading to reduced fitness and adaptive potential, and thus ultimately greater extinction risk. Accurately quantifying genetic diversity is therefore crucial to assessing the viability of small populations. Diversity indices are typically calculated from the multilocus genotypes of all individuals sampled within discretely defined habitat patches or larger regional extents. Importantly, discrete population approaches do not capture the clinal nature of populations genetically isolated by distance or landscape resistance. Here, we introduce spatial Genetic Diversity (sGD), a new spatially explicit tool to estimate genetic diversity based on grouping individuals into potentially overlapping genetic neighbourhoods that match the population structure, whether discrete or clinal. We compared the estimates and patterns of genetic diversity using patch or regional sampling and sGD on both simulated and empirical populations. When the population did not meet the assumptions of an island model, we found that patch and regional sampling generally overestimated local heterozygosity, inbreeding and allelic diversity. Moreover, sGD revealed fine-scale spatial heterogeneity in genetic diversity that was not evident with patch or regional sampling. These advantages should provide a more robust means to evaluate the potential for genetic factors to influence the viability of clinal populations and guide appropriate conservation plans.

Keywords: clinal population, genetic diversity, mountain goat, sampling method, simulation

Received 10 June 2010; revision received 30 April 2011; accepted 11 May 2011

Introduction

Anthropogenic landscape modifications and climate change are profoundly changing the size and distribution of many animal populations (Harte *et al.* 2004; Fischer & Lindenmayer 2007). Although the ranges of some species are expanding, most are contracting owing to a combination of factors including habitat loss, habitat fragmentation and harvest (Rodríguez 2002; Laliberte & Ripple 2004; Wiegand *et al.* 2005). Small remnant populations inhabiting fragmented landscapes are susceptible to extirpation by genetic processes (Lacy 1997; Gaggiotti 2003; Keyghobadi 2007).

Accurate estimates of population genetic diversity are crucial to evaluating the potential for genetic processes to affect population viability, monitoring at-risk populations and designing conservation plans to improve genetic diversity (Schwartz *et al.* 2007). Population genetic diversity indices are calculated from the genotypes of a group of sampled individuals, typically at neutral genetic marker loci. Generally, individuals are grouped either by the habitat patch in which they were sampled (e.g. Baker *et al.* 2008) or from a larger regional extent (e.g. Ernest *et al.* 2003). Regions are typically delineated by expert knowledge, arbitrary political or geographical boundaries, or genetically defined based on assignment tests such as STRUCTURE (Pritchard *et al.* 2000).

Both patch and regional clustering of samples make strong assumptions regarding the population's genetic structure and the resulting pattern of genetic diversity across a landscape. Specifically, these approaches implicitly assume that each discretely bounded sampling unit represents a single panmictic group of individuals isolated from other patches or regions. These assumptions reflect an island model of population structure (Wright 1931) where barriers subdivide the population

Correspondence: Andrew Shirk, Fax: 360 753 7737;
E-mail: ashirk@u.washington.edu

into discrete subpopulations (here we refer to this as isolation by barrier, or IBB). Importantly, the assumption of panmixia within each subpopulation would be expected to produce an internally uniform pattern of genetic diversity (i.e. no spatial variation in diversity indices within each patch or region).

An island model reflecting IBB, however, is not the only way in which a population may be structured. Many populations exhibit genetic isolation by distance (isolation by distance or IBD; Wright 1943). IBD produces a clinal population structure where genetic distance between individuals or subpopulations is a function of the Euclidean distance separating them. Other populations have been linked to genetic isolation arising from the variable resistance of the landscape to gene flow (isolation by resistance or IBR; Cushman *et al.* 2006; McRae 2006). IBR produces a clinal population structure where genetic distance between individuals is a function of cost-weighted distance (i.e. distance weighted to account for the effect of landscape resistance) rather than Euclidean distance. For populations structured by IBD or IBR, genetic diversity indices may vary greatly through space as a function of the local number of breeding individuals and their relative isolation by gradients of distance or landscape resistance (Wright 1946; Chambers 1995; Amos & Harwood 1998).

Although many populations are structured by IBD or IBR, patch and regional sampling (with their attendant island model assumptions) are the only approaches commonly used to estimate genetic diversity. The use of these approaches on a clinal population poses a challenge in that discrete boundaries do not capture the gradients of relatedness evident in clinal populations (Schwartz & McKelvey 2009). Indeed, for many clinal populations, there are often no obvious regions or habitat patches that bound groups of individuals in a biologically meaningful way. Decisions on how to group individuals for genetic diversity estimates of clinal populations can therefore be highly subjective.

Population genetic theory would predict that arbitrarily imposing discrete boundaries on a clinal population may lead to error in diversity estimates and mask the spatially complex patterns of genetic diversity expected of a clinal population (Chambers 1995; Hartl *et al.* 1997). For example, if a clinal population is distributed in clusters, it may be possible to delineate patch extents around each cluster. However, this approach does not account for the potential of proximate clusters to be part of the same breeding pool. Thus, patch sampling of a clustered but clinal population may result in multiple estimates of diversity (one per patch) for the same breeding pool, with each estimate based on a smaller sample size and therefore more prone to error. Conversely, use of a regional sampling approach on a clinal population risks grouping individuals from different local breeding pools into the same regional extent, thus violating the assumption of panmixia. In this event, genetic diversity may vary in space within the region, but the single regional estimate masks that variation. Moreover, a regional estimate that groups individuals from different breeding pools would likely be biased owing to the Wahlund effect (Wahlund 1928), which inflates heterozygosity and inbreeding indices.

The potential for patch or regional sampling scheme to a clinal population to increase error and mask spatial complexity in genetic diversity patterns has important implications for conservation and management. If a population does not meet the underlying assumptions of these approaches, areas where genetic diversity is low might go undetected and the threat genetic factors may pose to population viability may be understated. Although new methods have been developed to characterize the genetic structure of clinal populations (e.g. Miller 2005), to our knowledge, no method that fully captures the complex patterns of genetic diversity expected of a clinal population has been published.

To meet this need, we propose an alternative approach that avoids partitioning genetic data into discrete patch or regional boundaries. Instead, we suggest that a more biologically relevant grouping of individuals would naturally be that of the genetic neighbourhood. The idea of a genetic neighbourhood for populations isolated by distance was first introduced by Wright (1946). Recently, more general approaches to estimating genetic neighbourhood distances have been developed that are appropriate for any mode of genetic isolation. One common approach is to infer genetic neighbourhoods from correlograms depicting the autocorrelation of genotypes across a range of distance classes (Waser & Elliott 1991; Campbell & Dooley 1992; White & Svendsen 1992). This approach has the advantage of not requiring additional data beyond the genotypes already obtained to estimate genetic diversity.

In this study, we describe a way to use the concept of genetic neighbourhoods to capture spatially complex patterns of diversity expected of clinal populations while avoiding error introduced by patch or regional sampling. Specifically, with an estimate of genetic neighbourhood distance and a spatial landscape model capturing the mechanism of genetic isolation in the population, we use cost-weighted distance methods to identify the spatial extent of the genetic neighbourhood surrounding each sampled individual. Diversity indices can then be estimated for each neighbourhood based on all individuals within the neighbourhood extent. The result is a point estimate of diversity for every sampling location based on a genetic neighbourhood surrounding that point. Importantly, an individual may belong to multiple neigh-

bourhoods resulting in shifting membership across the study area, just as local breeding pools in a clinal population shift in overlapping groups across space. Tying membership to a spatial model of gene flow limits the extent of the neighbourhood to reflect local breeding pools that are not strongly isolated by barriers, distance or landscape resistance.

To make this approach broadly available, we developed a software tool, spatial Genetic Diversity (sGD), that estimates five indices of genetic diversity (observed heterozygosity, expected heterozygosity, inbreeding coefficient, allelic diversity and mean number of alleles per locus) as well as sample size in a spatially explicit manner based on clustering individuals by genetic neighbourhood. sGD is designed for use with codominant neutral genetic markers and diploid species, particularly those forming clinal populations. This tool is freely available as a Python script or as an ArcGIS toolbox (ESRI). We demonstrate its application on three simulated populations inhabiting artificial landscapes where genetic isolation is a function of barriers, distance or landscape resistance to gene flow. We also demonstrate sGD on empirical data from mountain goats in the Cascade Range, Washington, USA. Finally, we compare the inferences using sGD to inferences based on patch and regional sampling approaches where possible using the same simulated and empirical data.

Methods

Overview of comparison between patch sampling, regional sampling and sGD

We evaluated the performance of patch sampling, regional sampling and sGD on three simulated populations isolated by barrier (IBB), distance (IBD) or landscape resistance (IBR), and one empirical population (mountain goats in the Washington Cascade Range, USA) isolated by landscape resistance. Two alternative methods and four populations potentially yield a total of eight comparisons with sGD; however, three of these comparisons were not appropriate. Specifically, there were no appropriate patch boundaries in the IBB or IBD simulations, and there were not sufficient numbers of mountain goats sampled per patch in the Washington landscape to reliably estimate diversity indices. Eliminating these, we focused on the following comparisons (i) sGD and regional sampling on a simulated population exhibiting IBB, (ii) sGD and regional sampling on a simulated population exhibiting IBD, (iii) sGD and regional sampling on a simulated population exhibiting IBR, (iv) sGD and patch sampling on a simulated population exhibiting IBR and (v) sGD and regional sampling on a real mountain goat population exhibiting IBR.

Simulated landscapes and populations

We generated three hypothetical landscapes for use in population genetic simulations designed to evaluate the performance of sGD across a range of possible modes of genetic isolation. Each landscape was a square measuring 25.6 km per side at a resolution of 100 m per grid cell (i.e. 256 cells × 256 cells) and was inhabited by 576 simulated individuals.

The first landscape was designed to reflect genetic isolation by barrier (IBB). The population in this landscape was uniformly distributed in a regular grid pattern but separated into two subpopulations by a barrier (one grid cell in width) running north–south through the centre of the landscape (Fig. 1, top row, left panel). We generated a raster GIS model representing this landscape's resistance to gene flow (i.e. a resistance surface; Spear *et al*. 2010) by assigning a high resistance of 100 to the barrier feature grid cells (100 * 100 m yields a cost-weighted distance of 10 km to cross the barrier). Assigning a resistance of zero to grid cells on either side of the barrier resulted in no resistance (0 * 100 m per cell yields no resistance to gene flow) within each subpopulation, thus making each subpopulation internally panmictic in the simulation.

The second landscape was designed to reflect genetic isolation by distance (IBD). The population in this landscape was also uniformly distributed in a regular grid pattern (as in the IBB simulation). We generated a resistance surface representing IBD by assigning all grid cells a resistance of one (Fig. 1, middle row, left panel). Thus, the cost-weighted distance of crossing a single grid cell (1 * 100 m) was equal to the Euclidean distance across the grid cell.

The third landscape was designed to reflect genetic isolation by resistance (IBR). We generated this landscape with QRULE (Gardner & Urban 2007) such that each grid cell was assigned a resistance value ranging from 1 to 10. The lowest-resistance grid cells were clustered such that they had a contagion of 0.1 and represented 10% of the landscape (Fig. 1, bottom row, left panel). A population of 576 individuals was randomly distributed in an irregular pattern related to a probability of 1/resistance. Thus, most individuals were located in the least resistive grid cells.

To compare a regional sampling approach to sGD, we generated regional extents corresponding to two subpopulations divided by the barrier feature in the IBB simulation (Fig. 1, top row, right panel), one panmictic population in the IBD simulation (Fig. 1, middle row, right panel) and two subpopulations divided by a north–south expanse of high resistance in the IBR simulation (Fig. 1, bottom row, right panel). We also sought to compare a patch sampling approach to sGD by generating four patch boundaries surrounding the largest clusters of

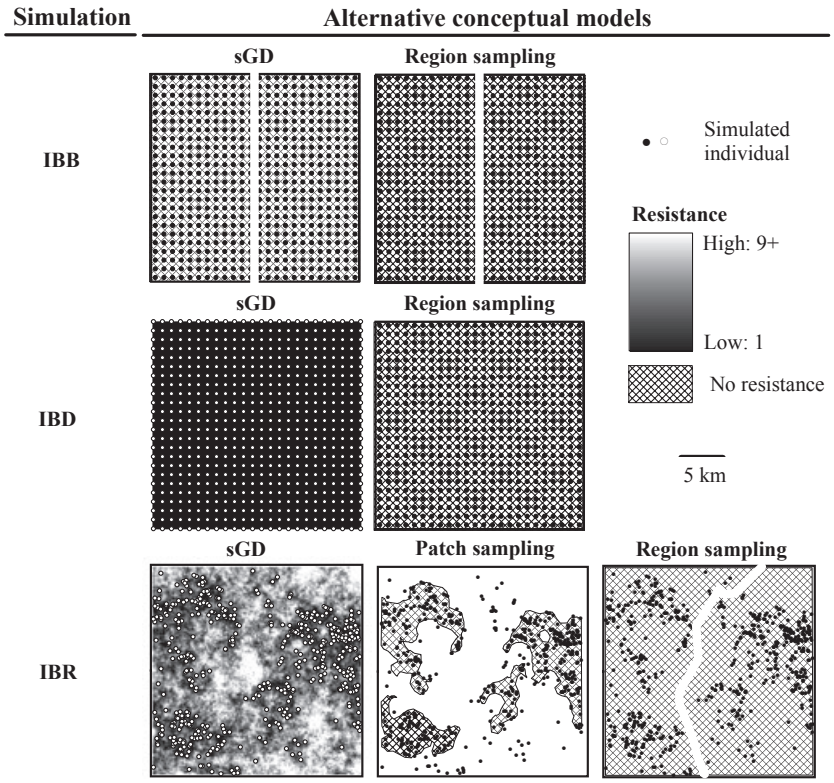

Fig. 1 Alternative conceptual models for simulated populations. In the IBB simulation, regional sampling assumes two panmictic subpopulations separated by a barrier (top row). In the IBD simulation (middle row), regional sampling assumes a single panmictic population. In the IBR simulation (bottom row), patch and regional sampling assume that barriers separate patches or regions that are otherwise panmictic (no resistance within each discrete patch or region). Conversely, sGD genetic neighbourhoods are based on the same spatial model that drives gene flow in the simulation. In the IBB simulation, the spatial model is identical to the assumptions of regional sampling; however, the spatial models depart from the assumptions of regional and patch sampling in the IBD and IBR simulations.

individuals in the IBR simulation (Fig. 1, bottom row, middle panel).

Empirical landscape and population

We also evaluated the performance of sGD using empirical data. The empirical study area included approximately 8000 km^2 of the southern Cascade Range, Washington, USA. The majority of the area is encompassed by the Snoqualmie National Forest, Gifford Pinchot National Forest and Mount Rainier National Park. The landscape is mountainous and covered with montane forests, except at elevations above about 1400 m, where subalpine parkland, rocky alpine summits and glaciers predominate. Elevation varies from near sea level to almost 4400 m. Two highways with average annual daily traffic volumes of between 1200 and 1700 vehicles (much of the volume occurs during daylight hours in the summer months) run east–west across the centre of the study area. In addition, numerous low-use unpaved forest roads exist on federal and private timber lands. No large towns or cities exist within the study area. A previous study (Shirk *et al.* 2010) identified a model of landscape resistance as causal relative to alternative resistance models or null models of isolation by distance or barrier. We used this resistance model to define genetic neighbourhoods in sGD as described below.

The south Cascade landscape is inhabited by approximately 1200 mountain goats patchily distributed in areas where suitable habitat exists (C. Rice, personal communication). In a previous study (Shirk *et al.* 2010), we genotyped 71 mountain goats from this region at 18 microsatellite loci (approximately 6% of the population). All loci were in Hardy–Weinberg and linkage equilibrium after Bonferroni correction for multiple tests.

Pairwise landscape distances between individuals

Calculating a matrix of pairwise landscape distances (i.e. cost-weighted distance; Adriaensen *et al.* 2003) between all simulated individuals or sampled mountain goats was a prerequisite to estimating the genetic neighbourhood size and also an input in the genetic simulations. We calculated least-cost distance matrices (i.e. the pairwise least-cost distance between all individuals) for each of the three simulated landscape resistance models and the empirical mountain goat resistance model using the Landscape Genetics ArcToolbox (downloaded from http://purl.org/NET/python_land_gen/arcgis_toolbox).

Population genetic simulations

We used the population genetic simulation CDPOP (Landguth & Cushman 2010) to test the performance of

sGD in each of the three hypothetic landscapes described previously. 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 (i.e. cost-weighted distance) between individuals given a landscape model resistance surface.

We ran each simulation for 500 generations, iterated for a total of 100 Monte Carlo runs. Parameters for the simulated population were designed to simulate a small vagile animal with limited dispersal capability relative to the size of the landscape, thus allowing for genetic differentiation and spatial variation in genetic diversity to occur over time. We specified reproduction to be sexual with nonoverlapping generations and a mating probability related to the inverse square of the effective landscape distance (based on IBB, IBD or IBR) as specified in pairwise distance matrices. We limited mating and dispersal to occur at a maximum cost-weighted distance of 2 km in the IBB and IBD simulations and 5 km in the IBR simulation. We also specified the age of first reproduction to be 1 year, a 1:1 sex ratio at birth, random starting allele frequencies, 12 loci, 10 alleles per locus, a mutation rate = 0.005 per generation and a k-allele model of mutation.

Pairwise genetic distance between individuals

Calculating a matrix of pairwise genetic distances between all simulated individuals or sampled mountain goats was a prerequisite to estimating genetic neighbourhood size. For each simulated population and the empirical mountain goat population, we generated a table with rows representing sampled individuals and columns representing each allele present in the population. Next, we tallied the number of each allele observed for a given individual (0, 1, or 2). We used the ecodist package in R 2.10 (R Development Core Team) to calculate Bray–Curtis dissimilarity from this table of allele use and thereby generate a matrix of pairwise genetic distance between individuals. This index ranges from 0 (no alleles in common) to 1 (all alleles are identical). We chose this metric of genetic distance as it makes no population genetic assumptions nor does it assume a particular mechanism of mutation. An R script for generating this genetic distance matrix is included with sGD.

Genetic neighbourhood size

We used the ecodist package in R 2.10 (Goslee & Urban 2007) to estimate genetic neighbourhood diameter for each simulated and empirical population based on the cost-weighted distance at which pairwise genetic distance is no longer positively correlated. We identified this point using a Mantel correlogram characterizing the correlation of pairwise genetic distances among individuals (based on Bray–Curtis dissimilarity as described previously) across multiple ranges of pairwise cost-weighted distances (based on the landscape resistance models as described previously). The largest distance class retaining a positive Mantel correlation was chosen as the genetic neighbourhood diameter. An R script for generating this correlogram is included with sGD.

sGD estimates of genetic diversity indices

spatial Genetic Diversity is a software tool designed to estimate spatially explicit indices of genetic diversity from microsatellite genotypes. sGD was written in Python 2.5 (available from http://www.python.org). It is available in two versions: one configured to be used with a Python interface and the other configured as a toolbox for use in ArcGIS (ESRI). Both versions use the same inputs; however, the ArcGIS version produces a shapefile in addition to the comma-delimited table output. The Python and ArcGIS versions of sGD as well as a user manual, sample input files and R 2.10 code to calculate pairwise genetic distance matrices and produce Mantel correlograms can be downloaded at https://sites.google.com/a/uw.edu/shirk/.

spatial Genetic Diversity inputs include a microsatellite genotype table, a matrix of pairwise landscape distances between sampled individuals (based on the landscape resistance models as described earlier), a minimum sample size per genetic neighbourhood (in this study, we used a minimum of 10) and an estimate of the genetic neighbourhood diameter. From these inputs, sGD identifies the membership of the genetic neighbourhood surrounding each sampled location. For each location, the individual(s) sampled at that location is, by definition, part of the neighbourhood. In addition, any nearby individuals within the neighbourhood radius (specified in cost-weighted distance and read from the input matrix of pairwise cost-weighted distances) are also included in the neighbourhood. We use the radius rather than diameter as a threshold to ensure that all individuals within the neighbourhood are located no further than a full neighbourhood diameter apart.

Once neighbourhoods have been defined for each sampling location, sGD determines which neighbourhoods meet the user-specified minimum sample size requirement. For those neighbourhoods meeting this requirement, sGD calculates five genetic diversity indices based on the codominant marker genotypes of all individuals in the neighbourhood. Specifically, sGD estimates observed heterozygosity, H_O, based on the equation:

$$H_O = 1 - \sum_k \sum_i P_{kii}/np$$

where P_{kii} is the frequency of genotype A_iA_i in sample k and np is the number of samples. sGD estimates Nei's gene diversity (Nei 1973), H_S, based on the equation:

$$H_S = \frac{\tilde{n}}{\tilde{n}-1}\left[1 - \sum_i \overline{p_i^2} - H_O/2\tilde{n}\right]$$

where

$$\overline{p_i^2} = \sum_k p_{ki}^2/np \quad \text{and} \quad \tilde{n} = \frac{np}{\sum_k 1/n_k}$$

spatial Genetic Diversity estimates Wright's inbreeding coefficient (Wright 1922), F_{IS}, based on the equation:

$$F_{IS} = 1 - \left(\frac{H_O}{H_S}\right)$$

spatial Genetic Diversity estimates allelic richness (Mousadik & Petit 1996), A_r, based on the equation:

$$A_r = \sum_{i=1}^{n_i}\left[1 - \frac{\binom{2N - N_i}{2n}}{\binom{2N}{2n}}\right]$$

where n is the smallest number of individuals genotyped per locus, N is the total number of individuals genotyped in a particular locus and N_i is the number of alleles of type i observed among N individuals. Finally, sGD estimates the mean number of alleles per locus, A, based on the equation:

$$A = (1/k)\sum_{i=1}^{k} n_i$$

where n is the number of observations of allele i at the kth locus. Note that A is highly sensitive to sample size. All other indices calculated by sGD are either inherently not sensitive to sample size (H_O, F_{IS}) or corrected for sample size (H_S, A_r). The neighbourhood sample size and five estimates of diversity are written to a comma-delimited text file (.csv) in both versions and also to an ESRI shapefile in the ArcGIS version of sGD.

Patch and regional sampling estimates of genetic diversity

As a comparison to the neighbourhood grouping of sGD, we also calculated the same diversity indices with FSTAT 2.2 (Goudet 1995) by grouping individuals by patch or region for the IBR simulation and by region for the IBB simulation. For the patch grouping, four patches were defined and diversity indices were calculated for all individuals in the simulation falling within the patch after 500 generations. For the regional grouping, two regions were defined based on the barrier feature of the IBB simulation and a high resistance zone in the landscape of the IBR simulation. Diversity indices were calculated for all individuals in the simulation falling within each region after 500 generations.

Validation of sGD diversity estimates

We estimated each of the sGD diversity indices using FSTAT 2.2 for 100 genetic neighbourhoods subsampled from the IBB, IBD, IBR and south Cascade landscapes. In each case, the FSTAT and sGD neighbourhood estimates were identical (data not shown).

Results

sGD vs. regional sampling in the IBB simulation

In the IBB simulation, we based regional sampling on two regional extents (separated by a strong barrier) each with a population of 288 individuals (Fig. 1). Within each region, diversity indices calculated in FSTAT 2.2 were identical to the point estimates of diversity calculated by sGD (Table 1).

In the IBB simulation, the autocorrelation of pairwise genetic distances reached zero at any cost-weighted distance greater than zero (Fig. 2). Thus, each neighbourhood in sGD contained all individuals (288) on the same side of the barrier, but none from the opposite side of the barrier, as was the case with regional sampling (Fig. 3). The spatial representation of diversity in sGD using both methods was also identical (compare regional sampling diversity estimates of IBB simulation in Table 1 with sGD estimates in Fig. 4, first column).

sGD vs. regional sampling in the IBD simulation

In the IBD simulation, regional sampling was based on one regional extent (the entire landscape), with a population of 576 individuals (Fig. 1, Table 1). The autocorrelation of pairwise genetic distance in the IBD landscape reached zero at a cost-weighted distance of 3.1 km (Fig. 2). Genetic neighbourhoods in the IBD landscape were circular (Fig. 3). The average genetic neighbourhood population size was 371, significantly lower ($P < 0.05$, z-test) than the total population size of the region (Table 1). sGD estimates of heterozygosity, gene diversity and allelic diversity were all significantly lower ($P < 0.05$, z-test) than the regional estimates (Table 1). sGD neighbourhood population sizes were higher in the centre of the landscape than near the periphery (Fig. 4, second column). Based on sGD, gene and allelic diversity were also higher in the centre of the landscape than the

Table 1 Sample size (N) and genetic diversity indices are given for three simulated and one empirical landscape. Diversity indices from simulations represent the mean value of 100 iterations. Diversity indices for each landscape model are summaries for each patch or regional extent shown in Fig. 4 and were estimated using one of three sampling methods, including patch sampling, regional sampling or sGD. sGD estimates represent the mean (SD) estimate among all neighbourhoods within the patch or region. Diversity indices include observed heterozygosity (H_O), Nei's gene diversity (H_S), inbreeding coefficient (F_{IS}), allelic richness (A_r) and mean number of alleles per locus (A). sGD estimates that are significantly different ($P < 0.05$, Z-test) than the patch or regional estimate are in bold

Model	Method	Extent	N	H_O	H_S	F_{IS}	A_r	A
IBB	Region	Region1	288	0.41	0.41	0.01	3.66	3.66
IBB	sGD	Region1	288 (0)	0.41 (0)	0.41 (0)	0.01 (0)	3.66 (0)	3.66 (0)
IBB	Region	Region2	288	0.41	0.41	0.01	3.66	3.66
IBB	sGD	Region2	288 (0)	0.41 (0)	0.41 (0)	0.01 (0)	3.66 (0)	3.66 (0)
IBD	Region	Region1	576	0.61	0.64	0.04	6.25	6.25
IBD	sGD	Region1	**371** (88)	**0.58** (0.01)	**0.61** (0.01)	0.04 (0.01)	**5.46** (0.09)	**5.71** (0.20)
IBR	Patch	Patch1	134	0.32	0.39	0.16	3.07	3.75
IBR	sGD	Patch1	125 (16)	**0.22** (0.01)	**0.31** (0.05)	**0.29** (0.07)	**2.15** (0.15)	**2.96** (0.25)
IBR	Patch	Patch2	205	0.20	0.26	0.23	2.76	3.58
IBR	sGD	Patch2	233 (23)	**0.23** (0.01)	**0.28** (0.01)	**0.18** (0.02)	**2.04** (0.03)	**3.18** (0.13)
IBR	Patch	Patch3	37	0.29	0.28	−0.02	2.42	2.42
IBR	sGD	Patch3	**106** (57)	**0.24** (0.01)	**0.30** (0.01)	**0.22** (0.02)	**2.15** (0.04)	**3.16** (0.13)
IBR	Patch	Patch4	96	0.19	0.27	0.27	2.71	3.17
IBR	sGD	Patch4	**119** (8)	**0.23** (0.01)	0.30 (0.02)	0.23 (0.06)	**2.09** (0.09)	2.97 (0.23)
IBR	Region	Region1	300	0.22	0.30	0.24	4.59	4.67
IBR	sGD	Region1	196 (72)	0.23 (0.01)	0.28 (0.03)	0.19 (0.05)	**2.07** (0.12)	**3.16** (0.25)
IBR	Region	Region2	276	0.29	0.51	0.42	4.58	4.58
IBR	sGD	Region2	**117** (24)	**0.22** (0.01)	**0.30** (0.05)	**0.26** (0.08)	**2.12** (0.15)	**2.94** (0.27)
WA	Region	Region1	71	0.36	0.40	0.11	4.30	4.33
WA	sGD	Region1	**27** (6)	0.35 (0.03)	0.38 (0.02)	0.09 (0.03)	**2.90** (0.05)	**3.31** (0.25)

periphery. Conversely, inbreeding was lower and observed heterozygosity was higher near the periphery of the landscape than the centre.

sGD vs. patch sampling in the IBR simulation

In the IBR simulation, regional sampling was based on four patch extents varying in population size from 37 to 205 individuals (Fig. 1, Table 1). The autocorrelation of pairwise genetic distances in the IBR landscape reached zero at a cost-weighted distance of 7.9 km (Fig. 2). Genetic neighbourhoods were irregularly shaped owing to the variable resistance of the IBR landscape (Fig. 3). The average genetic neighbourhood population size surrounding individuals within each patch differed significantly ($P < 0.05$, z-test) from the total population size for patch 3 (Table 1). There were also significant differences ($P < 0.05$, z-test) in heterozygosity, allelic diversity and inbreeding between the patch-level estimates and the average sGD estimates of individuals within each patch (Table 1). Based on sGD estimates, the pattern of genetic diversity in the IBR landscape was spatially complex, with generally higher heterozygosity and allelic diversity and lower inbreeding in the centre of the landscape (Fig. 4, third column). This heterogeneity was not evident in the patch-level depiction (Fig. 4, fourth column).

sGD vs. regional sampling in the IBR simulation

In the IBR simulation, regional sampling was based on two regional extents with 300 and 276 individuals, respectively (Fig. 1, Table 1). The average genetic neighbourhood population size surrounding individuals within region 2 was significantly lower ($P < 0.05$, z-test) than the total population size of the region (Table 1). The regional estimates of heterozygosity, allelic diversity and inbreeding were generally significantly higher ($P < 0.05$, z-test) than the sGD estimates from neighbourhoods surrounding individuals within each region (Table 1). As was the case with patch sampling, the spatial heterogeneity in diversity indices detected by sGD was not evident with regional sampling (Fig. 4, fifth column).

sGD vs. regional sampling in the south Cascade Range

In the empirical mountain goat analysis, the regional sampling approach was based on a single region (based on the STRUCTURE analysis from Shirk *et al.* 2010) with a sample size of 71 individuals. Diversity indices based on this approach estimated in FSTAT were intermediate between the highest and lowest values calculated with sGD (Table 1, Fig. 5). In sGD, autocorrelation of pairwise genetic distances reached zero at a cost-weighted distance

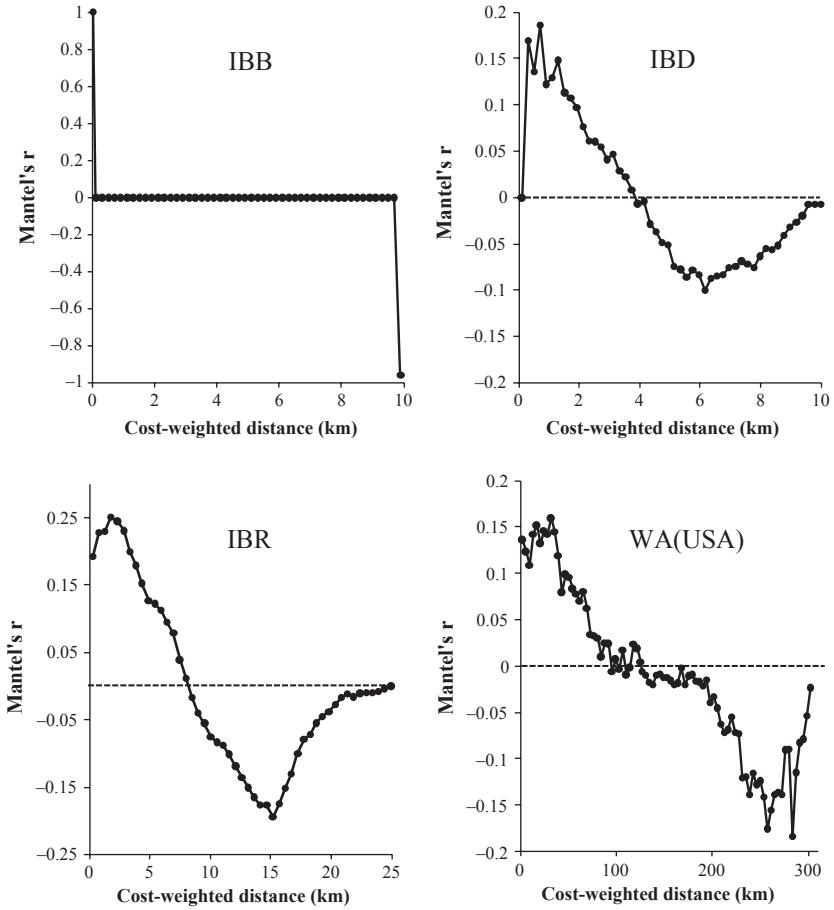

Fig. 2 Correlograms depicting the spatial autocorrelation in Bray–Curtis genetic distance among individuals across a range of cost-weighted distance classes for the IBB, IBD, IBR and empirical mountain goat population. We defined the diameter of a genetic neighbourhood as the largest significant distance class with a positive correlation.

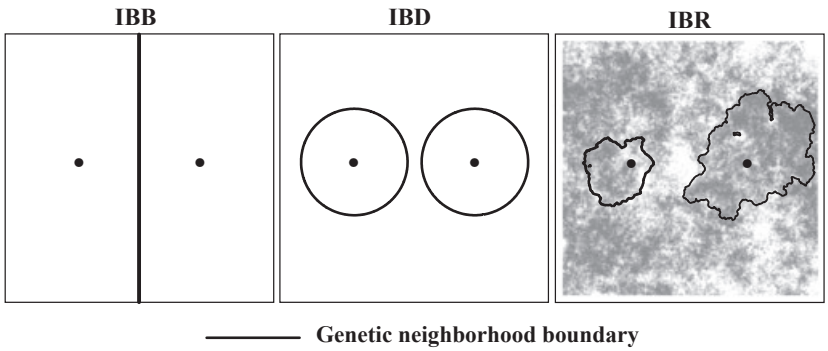

Fig. 3 An example of the sGD genetic neighbourhoods surrounding two simulated individuals in the IBB, IBD and IBR simulations. The spatial extent of the neighbourhoods depended on the resistance in the landscape and the neighbourhood distance inferred from spatial autocorrelation correlograms. In the IBB landscape, the neighbourhood extent was determined by the barrier feature. In the IBD landscape, the neighbourhood extent was determined by the neighbourhood distance. In the IBR landscape, the neighbourhood extent was determined by the neighbourhood distance modified by the effect of landscape resistance.

of 91.3 km (Fig. 2). The average sGD genetic neighbourhood population size based on this distance was 27, significantly lower than the regional population size. Based on sGD, we observed a trend towards generally lower observed heterozygosity and gene diversity as well as higher inbreeding from north to south across the study area (Fig. 5). The average number of alleles increased from north to south, but when corrected by rarefaction, allelic richness showed no strong trend, although there was variation across the study area (Fig. 5).

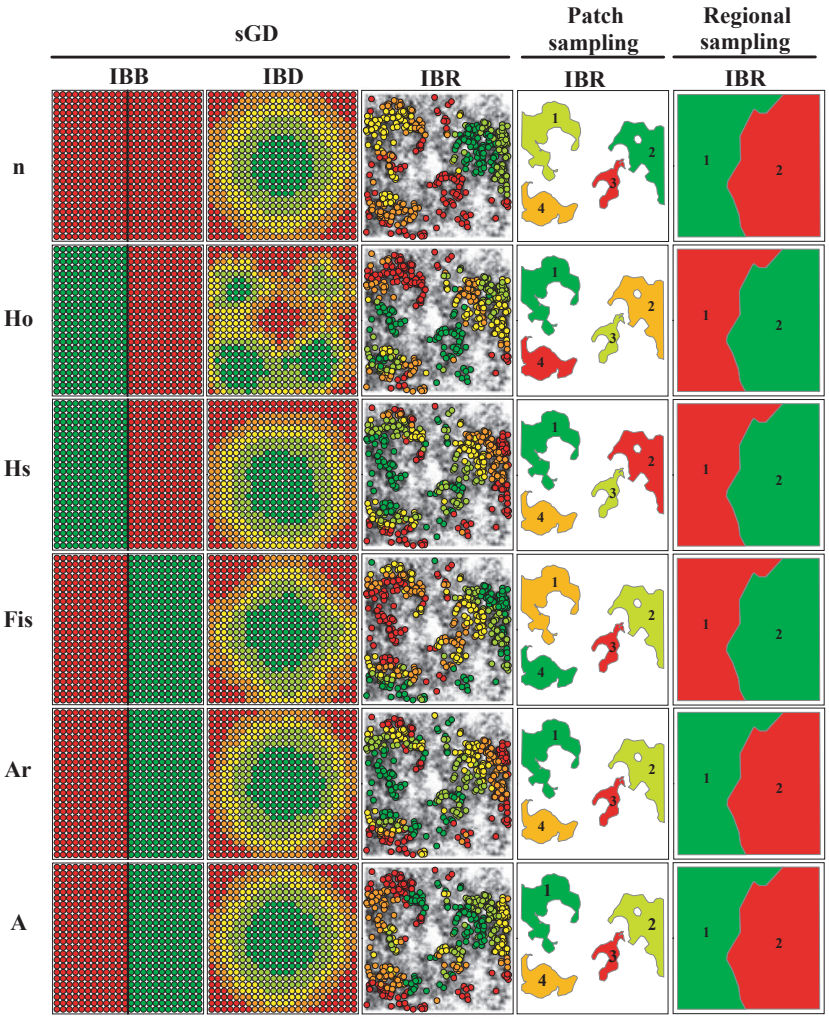

Fig. 4 Spatial patterns of sample size (*N*) and genetic diversity calculated by sGD (on the IBB, IBD and IBR landscapes), patch sampling (on the IBR landscape) or regional sampling (on the IBD and IBR landscape). The colour indicates the percentile of the index based on the range of values in each landscape (red 0–20%, yellow 20–40%, orange 40–60%, light green 60–80%, dark green 80–100%). Diversity indices include observed heterozygosity (H_o), Nei's gene diversity (H_s), inbreeding coefficient (F_{IS}), allelic richness (A_r) and mean number of alleles per locus (A). The actual values of these indices are provided in Table 1.

Neighborhood size and genetic diversity estimates

Genetic diversity indices were generally not strongly correlated with neighborhood population size (*N*) in the IBD and IBR simulations or the empirical mountain goat population (Supplemental Table 1). This suggests other factors like landscape effects on gene flow influenced genetic diversity in these clinal populations. An exception was the mean number of alleles, which was expected to be strongly related to sample size. The sGD estimate of allelic richness, which is adjusted for sample size, was not strongly correlated with *N* (Supplemental Table 1).

Discussion

In this study, we compared the inferences regarding population genetic diversity based on three different sampling approaches. We found that the way in which sampled individuals were grouped influences not only the estimate of genetic diversity but also the grain of its spatial representation across the landscape. Specifically, our results demonstrate that grouping individuals into units that do not match the population's genetic structure can lead to error in diversity estimates and mask spatial heterogeneity.

The mismatch between sampling unit and population structure was particularly apparent when we used a regional sampling approach in simulations (IBR and IBD) and the empirical mountain goat population. In each of these cases, regional estimates were based on a population size significantly larger than the average genetic neighbourhood population size within the region. As a result of grouping individuals from distinct neighbourhoods, regional sampling inflated estimates of heterozygosity and inbreeding relative to the sGD estimates, as expected owing to the Wahlund effect (Wahlund 1928). The discrete regional boundaries also failed to account for the shifting neighbourhood membership over space inherent in a clinal population. As a result, the regional sampling approach yielded a single estimate that masked the substantial spatial variance in diversity

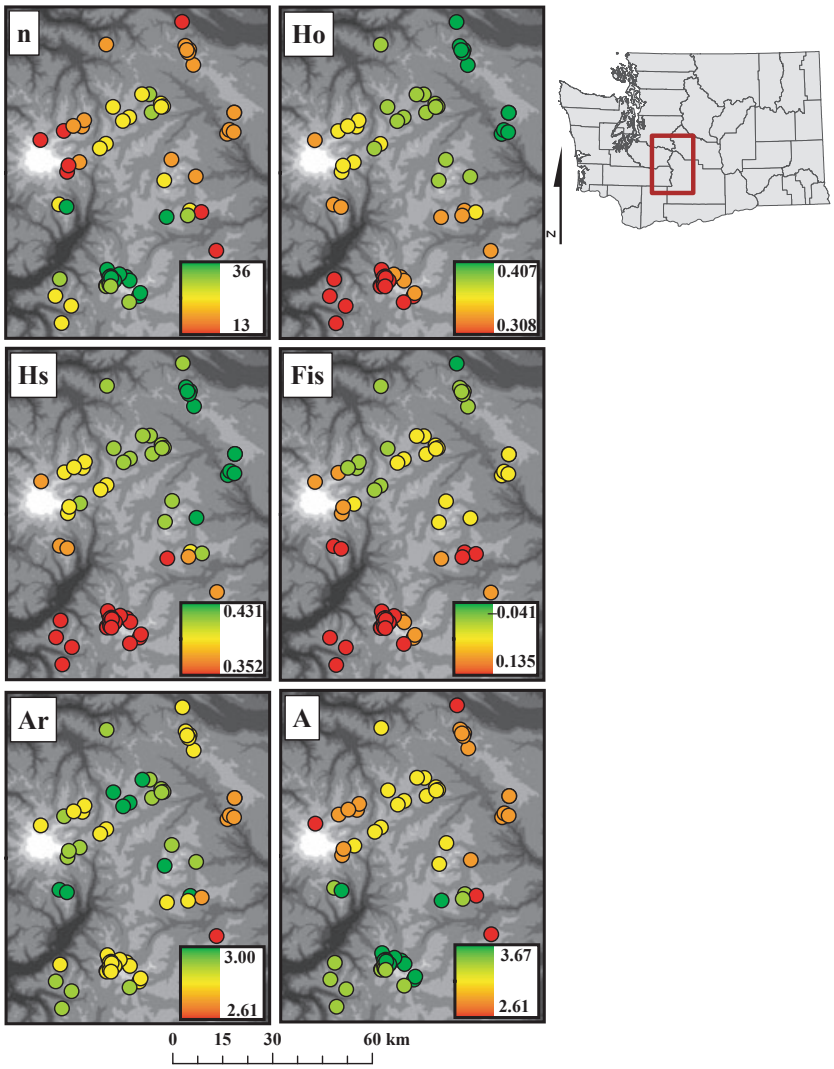

Fig. 5 Spatial patterns of sample size (N) and genetic diversity calculated by sGD for mountain goats in the south Cascade Range, Washington, USA. Diversity indices include observed heterozygosity (H_o), Nei's gene diversity (H_s), inbreeding coefficient (F_{IS}), allelic richness (A_r) and mean number of alleles per locus (A).

we observed in the IBD, IBR and south Cascade landscapes with sGD.

Rather than overstating neighbourhood membership as was the case with regional sampling, patch sampling in the IBR simulation generally understated neighbourhood size. This was reflected in the mean sGD neighbourhood size surrounding individuals sampled within each patch, which was greater than the total patch population size in three of the four patches. This mismatch highlights how a discrete patch boundary can fail to account for gene flow linking multiple patches together into a single neighbourhood. Artificially partitioning the sample data from a genetic neighbourhood into multiple patch estimates reduces sample size and increases error. Patch sampling also assumes a uniform pattern of genetic diversity within each patch, and this was clearly not supported when compared with the sGD estimates of diversity in the IBR simulation. Together, these results suggest that the most accurate indices of genetic diversity occur when sampled individuals are grouped based on their genetic unit of organization.

spatial Genetic Diversity offers an alternative approach that groups individuals by genetic neighbourhood with the expectation that it would more accurately reflect the complex patterns inherent in clinal populations. The extent of a neighbourhood in sGD depends on the landscape model's resistance to gene flow and the distance defining the neighbourhood's maximum limit. The IBB, IBD and IBR simulations demonstrate how neighbourhood extents in sGD reflect these dynamics. For example, in the IBB simulation, sGD neighbourhoods were comprised of all individuals sampled within each discrete subpopulation, as expected when strong barriers form isolated but internally panmictic subpopulations. In the IBD simulation, sGD neighbourhoods formed a circle about the sampled point, as expected when a population

is genetically isolated by distance (Wright 1943, 1946). In the IBR simulation, sGD neighbourhood size and shape varied owing to complex patterns of resistance in the landscape, as expected when a population is isolated by landscape resistance (Chambers 1995). These examples demonstrate sGD's capacity to match the sampling unit to the genetic structure of the population, thereby providing an ecologically relevant and highly flexible basis for grouping individuals.

spatial Genetic Diversity's use of potentially overlapping neighbourhoods centred about each sampling location changes the unit of observation to the point rather than the patch or region. This increased resolution offers a much greater spatial understanding of genetic diversity patterns in clinal populations than are possible with patch or regional sampling. This was evident in the IBR simulation, where sGD detected generally higher levels of heterozyogosity and allelic diversity with lower inbreeding in the centre of the landscape. The complex pattern reflects the varying local population size and the heterogeneous landscape's resistance to gene flow. Although mating and dispersal were limited to a cost-weighted distance of 5 km in this simulation, the neighbourhood size extended to 7.9 km. This reflects the ability of gene flow over multiple generations to maintain positive autocorrelation in allele frequencies beyond the limit of any single dispersal event.

sGD also provides a means to estimate diversity in continuous populations isolated by distance. Imposing patch or regional boundaries for the purposes of estimating genetic diversity indices in continuous populations would be highly subjective (Schwartz & McKelvey 2009). sGD, however, was able to characterize changes in diversity across the simulated IBD landscape as genetic neighbourhood membership changed over space. Although mating and dispersal were limited to a cost-weighted distance of 2 km in this simulation, the neighbourhood size extended to 3.1 km, reflecting the influence of gene flow in defining neighbourhoods, as in the IBR simulation. The larger neighbourhood population size in the centre of the IBD landscape relative to the edges resulted in greater retention of allelic diversity and expected heterozygosity over time. Surprisingly, observed heterozygosity was lower and inbreeding was higher in the centre than near the periphery. This result reflects a limitation of the CDPOP algorithm based on its requirement that all females must mate every generation. Females along the periphery of the landscape did not have as many available males nearby to breed with compared with the centre of the landscape; thus, the average distance males moved to mate with females in corner areas was higher than in the centre. This greater distance meant breeding pairs near the periphery were less related than those near the centre and thereby increased observed heterozygosity and lowered inbreeding. In natural populations, there is no requirement for all females to breed; hence, this artefact of the simulation would not occur. Future versions of CDPOP will allow for more dynamic populations not constrained in this way (E. Landguth, personal communication).

In addition to characterizing complex diversity patterns in clinal populations, sGD was also capable of accurately capturing the degree and pattern of genetic diversity in a discretely structured population. In the IBB simulation, both the estimate and the spatial representation of diversity were identical in sGD and the regional sampling approach. Thus, sGD is a flexible tool capable of estimating indices and patterns of genetic diversity in any population, regardless of the mechanism of genetic isolation.

While population genetic simulations allowed us to compare sGD with regional and patch sampling approaches in a controlled environment, the mountain goat analysis offered an extension of this comparison to empirical data. This population was previously shown to be clinally structured based on landscape resistance arising from suboptimal landcover, elevation and roads (Shirk *et al.* 2010). Imposing a patch or regional sampling scheme on this population ignores its genetic organization and adds the assumptions of an island population model to a case where they are not supported (Shirk *et al.* 2010). In addition, although mountain goats are patchily distributed in the study area, patch sampling of this population would be impractical owing to the difficulty of sampling in rugged mountain goat habitat and the generally low population density, both of which preclude obtaining sufficient sample size per patch to accurately estimate diversity indices. Imposing a regional structure would also be difficult, as no obvious barriers exist in this landscape that would form discrete subpopulations. STRUCTURE assignment tests (Pritchard *et al.* 2000) previously revealed a single south Cascade population, with further subdivision not supported (Shirk *et al.* 2010). We considered this to be strong evidence for a single population and therefore estimated diversity based on all sampled individuals in the south Cascades as a regional estimate comparison to sGD.

Diversity indices calculated for south Cascade mountain goats based on a single region masked strong gradients of diversity that were apparent with sGD and also appeared biased by the Wahlund effect (heterozygosity and inbreeding were higher in the regional sample compared with the neighbourhood average). The pattern detected by sGD reflected generally declining heterozygosity and increased inbreeding towards the southern extreme of the population. Interestingly, this matches the

expectation of declining diversity in peripheral populations (Eckert *et al.* 2008). Indeed, the south Cascade Range represents the southern extreme of the coastal North American mountain goat range. Importantly, from a conservation perspective, the high inbreeding and low heterozygosity we observed with sGD in the southern portions of the study area were substantially lower than the regional sampling estimate for the region (Table 1, Fig. 5). Thus, regional sampling underestimated the threat genetic factors may pose at a local level among mountain goats in the south Cascades. This has important conservation implications given the estimated 70% decline in mountain goats throughout the range over the past 50 years (Rice & Gay 2010), the clear link between heterozygosity and fitness in this species (Mainguy *et al.* 2009) and the greater impact of inbreeding for species inhabiting extreme environments (Hedrick & Kalinowski 2000).

The study of mountain goat population structure and genetic diversity we provide here and in Shirk *et al.* (2010) provides an analytical framework that yields many inferences from a single codominant marker data set. From a collection of genotypes, we performed causal modelling to identify the mechanism of genetic isolation and a spatial model of gene flow that captured the landscape's effect on gene flow. From these same genetic data, we used spatial genetic autocorrelation approaches to infer the size of the genetic neighbourhood (as in Campbell & Dooley 1992; Waser & Elliott 1991; White & Svendsen 1992) and then evaluated the spatial complexity in diversity indices across the study area with sGD in a way that groups individuals by their genetic unit of organization. This approach has relatively small data requirements, is reproducible, is objective and potentially yields a strong spatial understanding of landscape effects on gene flow and diversity. The greater spatial resolution and fidelity of sGD may be valuable in monitoring population genetic diversity and guiding appropriate conservation plans.

There are several important considerations regarding the use and limitations of sGD that are important to note. sGD is designed to depict spatial variation in patterns of genetic diversity. It does not provide inferences regarding population genetic structure (although it requires measures of structure, the genetic neighbourhood distance and the effective landscape distance between individuals, as input parameters). The software Alleles in Space (Miller 2005) is similar in that it seeks to characterize continuous genetic patterns, but its focus is on spatial genetic structure rather than formal indices of diversity. Alleles in Space does not, for example, estimate indices of heterozygosity, allelic diversity and inbreeding. In addition, sGD is best suited for data sampled continuously in space rather than in discrete clusters. Sufficient sampling density within the genetic neighbourhood distance is also required to achieve a minimum sample size to reliably estimate diversity indices. Finally, sGD estimates of diversity are dependent upon accurate inferences regarding the effective landscape distance between individuals and the associated genetic neighbourhood distance. There are a variety of alternative approaches to inferring landscape distances and genetic neighbourhood distances available, yet the consistency and accuracy across these methods have not been formally evaluated. As such, error in these input parameters may influence the outcome. We therefore urge sGD users to perform sensitivity analysis for these parameters.

Acknowledgements

We thank Cliff Rice, Dave Wallin, Ken Warheit and three anonymous reviewers for their helpful comments.

References

Adriaensen F, Chardon JP, De Blust G *et al.* (2003) The application of 'least-cost' modelling as a functional landscape model. *Landscape and Urban Planning*, **64**, 233–247.

Amos W, Harwood J (1998) Factors affecting levels of genetic diversity in natural populations. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **353**, 177.

Baker AJ, Greenslade AD, Darling LM, Finlay JC (2008) High genetic diversity in the blue-listed British Columbia population of the purple martin maintained by multiple sources of immigrants. *Conservation Genetics*, **9**, 495–505.

Campbell DR, Dooley JL (1992) The spatial scale of genetic differentiation in a hummingbird-pollinated plant: comparison with models of isolation by distance. *American Naturalist*, **139**, 735–748.

Chambers SM (1995) Spatial structure, genetic variation, and the neighborhood adjustment to effective population size. *Conservation Biology*, **9**, 1312–1315.

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.

Eckert CG, Samis KE, Lougheed SC (2008) Genetic variation across species' geographical ranges: the central–marginal hypothesis and beyond. *Molecular Ecology*, **17**, 1170–1188.

Ernest HB, Boyce WM, Bleich VC *et al.* (2003) Genetic structure of mountain lion (Puma concolor) populations in California. *Conservation Genetics*, **4**, 353–366.

Fischer J, Lindenmayer DB (2007) Landscape modification and habitat fragmentation: a synthesis. *Global Ecology and Biogeography*, **16**, 265–280.

Gaggiotti OE (2003) Genetic threats to population persistence. *Annales Zoologici Fennici*, **40**, 155–168.

Gardner RH, Urban DL (2007) Neutral models for testing landscape hypotheses. *Landscape Ecology*, **22**, 15–29.

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

Goudet J (1995) FSTAT (Version 1.2): a computer program to calculate F-statistics. *Journal of Heredity*, **86**, 485–486.

Harte J, Ostling A, Green JL, Kinzig A (2004) Biodiversity conservation: climate change and extinction risk. *Nature*, **430**, 145–148.

Hartl DL, Clark AG, Clark AG (1997) *Principles of Population Genetics*. Sinauer associates, Sunderland, MA.

Hedrick PW, Kalinowski ST (2000) Inbreeding depression in conservation biology. *Annual Review of Ecology and Systematics*, **31**, 139–162.

Keyghobadi N (2007) The genetic implications of habitat fragmentation for animals. *Canadian Journal of Zoology-Revue Canadienne De Zoologie*, **85**, 1049–1064.
Lacy RC (1997) Importance of genetic variation to the viability of mammalian populations. *Journal of Mammalogy*, **78**, 320–335.
Laliberte AS, Ripple WJ (2004) Range contractions of North American carnivores and ungulates. *BioScience*, **54**, 123–138.
Landguth EL, Cushman SA (2010) CDPOP: a spatially explicit cost distance population genetics program. *Molecular Ecology Resources*, **10**, 156–161.
Mainguy J, CÔTÉ SD, Coltman DW (2009) Multilocus heterozygosity, parental relatedness and individual fitness components in a wild mountain goat, Oreamnos americanus population. *Molecular Ecology*, **18**, 2297–2306.
McRae BH (2006) Isolation by resistance. *Evolution*, **60**, 1551–1561.
Miller MP (2005) Alleles In Space (AIS): computer software for the joint analysis of interindividual spatial and genetic information. *Journal of Heredity*, **96**, 722–724.
Mousadik A, Petit RJ (1996) High level of genetic differentiation for allelic richness among populations of the Argan tree [*Argania spinosa* (L.) Skeels] endemic to Morocco. *Theoretical and Applied Genetics*, **92**, 832–839.
Nei M (1973) Analysis of gene diversity in subdivided populations. *Proceedings of the National Academy of Sciences of the United States of America*, **70**, 3321–3323.
Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics*, **155**, 945–959.
Rice CG, Gay D (2010) Effects of mountain goat harvest on historic and contemporary populations. *Northwestern Naturalist*, **91**, 40–57.
Rodríguez JP (2002) Range contraction in declining North American bird populations. *Ecological Applications*, **12**, 238–248.
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, Luikart G, Waples RS (2007) Genetic monitoring as a promising tool for conservation and management. *Trends in Ecology & Evolution*, **22**, 25–33.
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**, 3603–1619.
Spear SF, Balkenhol N, Fortin MJ, McRae BH, Scribner KIM (2010) Use of resistance surfaces for landscape genetic studies: considerations for parameterization and analysis. *Molecular Ecology*, **19**, 3576–3591.
Wahlund S (1928) Composition of populations and correlation appearances viewed in relation to the studies of inheritance. *Hereditas*, **11**, 65–106.
Waser PM, Elliott LF (1991) Dispersal and genetic structure in kangaroo rats. *Evolution*, **45**, 935–943.
White MM, Svendsen GE (1992) Spatial-genetic structure in the eastern chipmunk, Tamias striatus. *Journal of Mammalogy*, **73**, 619–624.
Wiegand T, Revilla E, Moloney KA (2005) Effects of habitat loss and fragmentation on population dynamics. *Conservation Biology*, **19**, 108–121.
Wright S (1922) Coefficients of inbreeding and relationship. *American Naturalist*, **56**, 330–338.
Wright S (1931) Evolution in mendelian populations. *Genetics*, **16**, 97–159.
Wright S (1943) Isolation by distance. *Heredity*, **28**, 114–138.
Wright S (1946) Isolation by distance under diverse systems of mating. *Genetics*, **31**, 39–59.

Supporting Information

Additional supporting information may be found in the online version of this article.

Table S1 Pearson's correlation between sample size (N) in the IBD, IBR, and WA (USA) landscapes and the sGD indices of genetic diersity.

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting information supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.