

# Optimizing dispersal and corridor models using landscape genetics

CLINTON W. EPPS\*, JOHN D. WEHAUSEN†, VERNON C. BLEICH‡, STEVEN G. TORRES§ and JUSTIN S. BRASHARES\*

\*Department of Environmental Science, Policy and Management, University of California Berkeley, 137 Mulford Hall, Berkeley, California 94720–3114, USA; †White Mountain Research Station, University of California, 3000 E. Line Street, Bishop, California 93514, USA; ‡California Department of Fish and Game, Sierra Nevada Bighorn Sheep Recovery Program, 407 West Line Street, Bishop, California 93514, USA; and §California Department of Fish and Game, Wildlife Investigations Laboratory, 1701 Nimbus Road, Suite D, Room # 170, Rancho Cordova, California 95670, USA

## Summary

**1.** Better tools are needed to predict population connectivity in complex landscapes. ‘Least-cost modelling’ is one commonly employed approach in which dispersal costs are assigned to distinct habitat types and the least-costly dispersal paths among habitat patches are calculated using a geographical information system (GIS). Because adequate data on dispersal are usually lacking, dispersal costs are often assigned solely from expert opinion. Spatially explicit, high-resolution genetic data may be used to infer variation in animal movements. We employ such an approach to estimate habitat-specific migration rates and to develop least-cost connectivity models for desert bighorn sheep *Ovis canadensis nelsoni*.

**2.** Bighorn sheep dispersal is thought to be affected by distance and topography. We incorporated both factors into least-cost GIS models with different parameter values and estimated effective geographical distances among 26 populations. We assessed which model was correlated most strongly with gene flow estimates among those populations, while controlling for the effect of anthropogenic barriers. We used the best-fitting model to (i) determine whether migration rates are higher over sloped terrain than flat terrain; (ii) predict probable movement corridors; (iii) predict which populations are connected by migration; and (iv) investigate how anthropogenic barriers and translocated populations have affected landscape connectivity.

**3.** Migration models were correlated most strongly with migration when areas of at least 10% slope had 1/10th the cost of areas of lower slope; thus, gene flow occurred over longer distances when ‘escape terrain’ was available. Optimal parameter values were consistent across two measures of gene flow and three methods for defining population polygons.

**4.** Anthropogenic barriers disrupted numerous corridors predicted to be high-use dispersal routes, indicating priority areas for mitigation. However, population translocations have restored high-use dispersal routes in several other areas. Known intermountain movements of bighorn sheep were largely consistent with predicted corridors.

**5. Synthesis and applications.** Population genetic data provided sufficient resolution to infer how landscape features influenced the behaviour of dispersing desert bighorn sheep. Anthropogenic barriers that block high-use dispersal corridors should be mitigated, but population translocations may help maintain connectivity. We conclude that developing least-cost models from similar empirical data could significantly improve the utility of these tools.

*Key-words:* bighorn sheep, connectivity, fragmentation, GIS, isolation, least-cost model, metapopulation, *Ovis canadensis*, roads, translocation.

*Journal of Applied Ecology* (2007) **44**, 714–724  
doi: 10.1111/j.1365-2664.2007.01325.x

## Introduction

Defining and maintaining connectivity of natural populations has become a conservation priority (Moilanen *et al.* 2005). As natural populations become increasingly fragmented by habitat destruction and the creation of dispersal barriers such as roads, extinction probabilities for some populations will increase due to demographic and genetic factors associated with reduced dispersal (Hanski 1999; Hedrick 2005). Greater recognition that isolation of protected areas will lead to faunal relaxation (the gradual loss of species, e.g. Soule, Wilcox & Holtby 1979) has resulted in world-wide efforts to link protected areas using corridors, buffer zones and mixed-use areas. Models that incorporate land use, habitat quality, human activities and other factors are often employed to aid the mapping of landscape connectivity and prioritize land acquisitions (e.g. Hunter, Fisher & Crooks 2003; Nikolakaki 2004). However, identifying the optimal locations of wildlife corridors has proved to be difficult and controversial, in part because the details of how different species disperse across landscapes are often inadequately understood.

The advent of geographical information systems (GIS) analysis as a tool for identifying corridors and defining population connectivity has led to the widespread application of techniques such as 'least-cost' modelling (Adriaensen *et al.* 2003) and 'friction' analyses (Ray, Lehmann & Joly 2002; Joly, Morand & Cohas 2003; Sutcliffe *et al.* 2003; Nikolakaki 2004). Models created through these approaches are based typically on raster maps that divide landscapes into many cells with unique values that depict different habitat or vegetation types, elevation, slope or other landscape features. Cells are given weights or 'resistance values' reflecting the presumed influence of each variable on movement of the species in question. Least-cost routines (see Adriaensen *et al.* 2003), then, are employed to: (i) calculate the relative cost of all possible routes among populations or islands of core habitat; (ii) determine the least costly route for animal movement between pairs of populations or core areas of habitat; and (iii) plot these most probable routes on maps for use in conservation planning. 'Cost' is related to probability of transit and may not be defined explicitly; energetic costs, increased risk of predation or costs associated with reduced forage availability are among the reasons why an animal might avoid or be less able to traverse a landscape feature.

Although the least-cost approach has been employed widely (e.g. Adriaensen *et al.* 2003; Beazley *et al.* 2005; Rouget *et al.* 2006), this approach has two major drawbacks. First, the underlying models of dispersal (i.e. what resistance values are assigned to different landscape categories) are based rarely on anything more than informed opinions from experts. Where empirical data are available, dispersal costs are typically inferred from presence/absence or abundance data in

different habitats, but such data may reflect habitat use rather than dispersal cost. Second, although these techniques define the most probable route according to the cost weighting system, the actual cost of a route over which dispersal can occur is unknown. Therefore, despite the increasing need and frequent application of such tools, these largely untested models are of uncertain value for conservation planning and management.

Population genetics approaches offer additional tools that can be applied to questions of dispersal and connectivity. Selectively neutral genetic markers can provide indices of gene flow derived from differences in allele frequencies between individuals or populations (Waser & Strobeck 1998). The emerging field of 'landscape genetics' uses high-resolution genetic data to determine the influence of landscape features such as fields (Vos *et al.* 2001) or highways (Keller & Largiader 2003; Epps *et al.* 2005) on gene flow and dispersal (Manel *et al.* 2003). However, developing dispersal models from genetic data entails large data sets and certain assumptions.

In particular, migration (in the sense of gene flow) operates at a different time scale than dispersal. Genetic data may reflect long-term dispersal patterns, but the time-period represented is variable and depends partly on the effective size ( $N_e$ ) of the populations. Time to equilibrium between migration and drift is proportional to  $N_e$  (Slatkin 1993). Therefore, among populations with small  $N_e$ , estimates of genetic distance or gene flow should reflect more recent dispersal patterns than estimates among populations with large  $N_e$ . Simulated data can be used to describe more clearly the time scale for a given data set (e.g. Epps *et al.* 2005), but in general the time scale represented is unknown. Furthermore, migration reflects effective dispersal, i.e. dispersal followed by reproduction. Individuals that disperse and do not reproduce will not be represented unless they are sampled directly. This could be advantageous if effective dispersal is the process of interest, but might not be as useful when considering, for instance, the role of dispersing individuals in spreading disease. Finally, sex-biased dispersal must be considered; gene flow estimates derived from nuclear DNA may largely represent movements of the least philopatric sex. Despite these possible limitations, genetic analyses may provide comprehensive pictures of dispersal that are otherwise unavailable (Koenig, VanVuren & Hooge 1996).

Efforts to develop more sophisticated models of migration from genetic data that consider species' dispersal behaviour are increasingly common. One such approach is to examine the correlation of gene flow with measures of 'effective geographical distance' (EGD) among populations, in addition to measures of geographical distance or the presence or absence of specific elements such as roads (Michels *et al.* 2001). EGD is a composite measure of dispersal distance between populations that incorporates both geographical distance and landscape features hypothesized to affect

dispersal. Recent examples of EGD include distances along riparian areas (Vignieri 2005), elevation change (Spear *et al.* 2005) and least-cost models that use a cost weighting surface based on assumed habitat value (Coulon *et al.* 2004; Spear *et al.* 2005; Vignieri 2005). EGD often explains more variation in gene flow between individuals or populations than geographical distance alone. This suggests that gene flow and dispersal patterns may not always fit a simple nearest-neighbour model, and it is important to test alternate hypotheses. However, genetic-based studies of dispersal rarely have examined more than a few alternate models of dispersal, and efforts to combine least-cost models with genetic data have been limited by a priori assumptions used to build the models. For instance, Vignieri (2005) used knowledge of preferred habitat for the Pacific jumping mouse *Zapus trinotatus* Rhoads to assign a lower dispersal cost to riparian and low-elevation habitat; however, that dispersal cost appeared arbitrary with respect to magnitude.

We propose that the effectiveness of combining least-cost and genetics-based approaches can be tested by comparing the ability of multiple least-cost models based on different landscape characteristics and a range of parameter values to explain observed variation in gene flow. Past analyses appear only to have tested hypotheses about which landscape factors affect dispersal. To translate least-cost models into effective conservation tools that identify active movement corridors and rank them according to predicted levels of gene flow, we also propose to estimate empirically how gene flow varies with EGD and determine the maximum EGD over which gene flow will occur.

In this paper we present methods to (1) test assumptions underlying least-cost connectivity models using genetic data; (2) predict landscape connectivity; and (3) test alternative management scenarios. We use estimates of gene flow among populations of desert bighorn sheep *Ovis canadensis nelsoni* Merriam to test the effectiveness of different least-cost GIS models and to optimize parameter values. We employ the following: (1) two methods for estimating gene flow among populations; (2) estimates of EGD derived from least-cost GIS models based on slope and distance with a wide range of parameter values; (3) three methods of defining population polygons used as the basis of our spatial analyses; (4) partial Mantel tests to assess correlation between gene flow estimates and EGD from alternate least-cost models; (5) regression of gene flow estimates on EGD to determine the maximum EGD over which gene flow is detectable; (6) identification and ranking of dispersal corridors using the best-fitting model of EGD; and (7) use of that model to identify probable movement corridors among populations of desert bighorn sheep while considering alternate management scenarios. Finally, we discuss the application of these techniques to conservation and management of species occupying fragmented habitats.

## DESERT BIGHORN SHEEP AND PREVIOUS DISPERSAL MODELS

Desert bighorn sheep are desert-adapted ungulates native to the south-western United States. Preferred habitat is generally steep, rocky, arid terrain. In California, desert bighorn sheep populations are typically small, often < 50 individuals (Epps *et al.* 2003) and located in small mountain ranges isolated by varying expanses of low-lying desert habitat. The metapopulation-like distribution of desert bighorn sheep results in frequent extinction and recolonization of populations (Schwartz, Bleich & Holl 1986; Bleich, Wehausen & Holl 1990), and it is recognized that appropriate management requires consideration of population connectivity (e.g. determining when translocation of bighorn sheep may be needed to re-establish recently extirpated populations; Bleich *et al.* 1996). Bleich *et al.* (1996) proposed a model of population connectivity that considered populations < 15 km apart as likely to be connected by dispersal and hypothesized that interstate highways were barriers to dispersal. That model was used to determine management units above the level of individual populations. Low-resolution genetic markers [mitochondrial DNA (mtDNA) control region restriction fragment length polymorphism (RFLP) data] were used to verify that detectable genetic differences existed between management units.

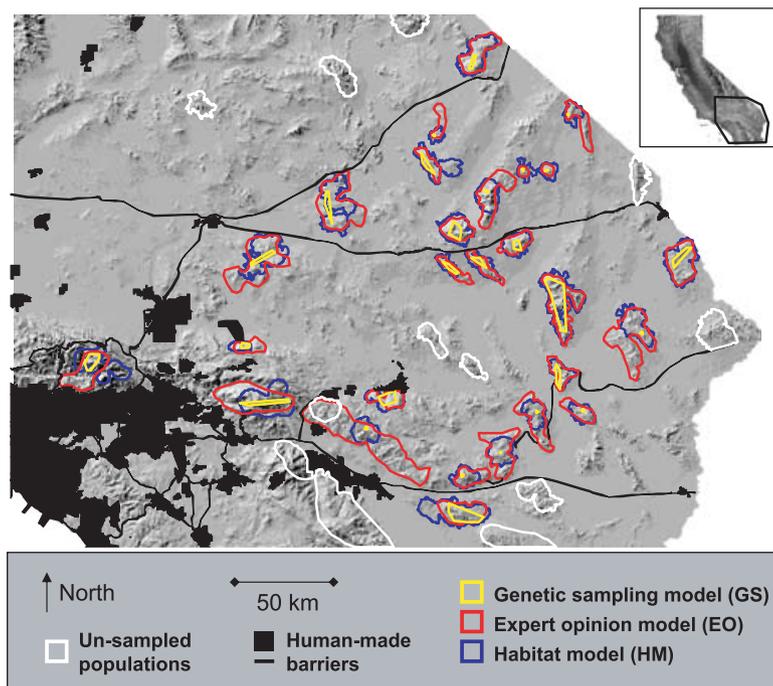
Population genetics data from 26 populations of desert bighorn sheep in the Mojave and Sonoran Desert regions of California were used to investigate the spatial scale of gene flow and the role of anthropogenic (human-made) barriers such as interstate highways, urban areas and canals (Epps *et al.* 2005). Epps *et al.* (2005) tested whether estimates of gene flow and genetic distance ( $Nm$  and  $F_{ST}$ ) were correlated with simple linear distance between populations and the presence of anthropogenic barriers. Those analyses confirmed that little or no gene flow had occurred across those barriers and that gene flow occurred primarily among populations < 15 km apart. However, habitat features expected to favour bighorn dispersal (e.g. areas with topographic relief sufficient to provide escape terrain for predator evasion) were not considered. Owing to considerable variation in the amount of escape terrain in low-lying areas among populations, we hypothesized that a least-cost model of migration based on topography could significantly improve our ability to predict the degree to which populations are linked by dispersal.

## Materials and methods

### OVERALL APPROACH: USING GENETIC DATA TO OPTIMIZE PARAMETER VALUES FOR A LEAST-COST MODEL

We used a matrix-based regression approach to test whether gene flow among populations of desert bighorn

sheep varied as a function of distance and topography or distance alone, and to identify which model of distance and topography best approximated the effect of these variables on gene flow. First, we calculated a series of matrices ( $X_1$ – $X_i$ ) of effective geographical distances (EGD) among populations. Each matrix represented estimates of EGD between all population pairs among 26 populations of desert bighorn sheep in California, USA (Fig. 1), resulting from a unique set of parameter values ( $i$  unique combinations). Next, a matrix ( $Y$ ) depicting the presence or absence of anthropogenic barriers (fenced highways, canals and urban areas) among those 26 populations was generated to control for the effect of those barriers on gene flow. Finally, a matrix ( $Z$ ) of gene flow estimates between all population pairs was developed. We used partial Mantel tests to assess the correlation of  $Z$  (gene flow) with each matrix  $X_i$  (EGD), while controlling for the effect of  $Y$  (anthropogenic barriers). In that manner parameter values for the EGD model resulting in the strongest correlation between  $X$  and  $Z$  were identified. That exercise was repeated using three different methods to define the geographical extent of each population, as well as a second method of estimating gene flow, to examine how sensitive model fitting was to those variables. The optimized model of EGD was then used in later analyses of corridor length and location. Our methods are detailed in the following sections.



**Fig. 1.** Topography (hill-shade) and distribution of desert bighorn sheep in south-eastern California, United States. Coloured polygons represent genetically sampled populations used to develop the dispersal model. GS polygons are minimum convex polygons around genetic sample locations. EO polygons were hand-drawn based on topography and expert opinion on bighorn sheep distribution. HM polygons were developed either from a GIS habitat model (described in Appendix S2) or from 95% density kernels based on radio-telemetry locations. Population polygons not used for model development (outlined in white) are based on the HM or EO models. Anthropogenic barriers indicated include fenced interstate highways, canals and urban areas.

## DEVELOPING LEAST-COST GIS MODELS TO CALCULATE EGD

We used slope as the variable for identifying the relative resistance or migration value of habitat between population polygons. We compiled 30 m Digital Elevation Model (DEM) data [US Geological Survey (USGS) 2003 series] for our study area and estimated slope for each 30 m cell using ArcGIS 9.0 (ESRI, Redlands, CA, USA). To simplify the models of bighorn migration as a function of topography and distance, we defined a ‘slope cut-off’ value for each model. Grid cells with slope greater than the cut-off value (‘slope’ cells) were considered more suitable (lower resistance) for bighorn dispersal than grid cells with slope lower than the cut-off (‘flat’ cells). We tested three slope cut-off values (5%, 10% and 15%), based on our assessment of radio telemetry data that suggested bighorn sheep are found mainly in habitat of at least 10% slope (3386 locations across the study area; unpublished data; California Department of Fish and Game). For each cut-off value tested, we generated six grids representing a wide range of different resistance values (weights) for slope cells. Thus, relative to the fixed cost of ‘1.0’ for a flat cell, slope cells were given weights of 0.7, 0.5, 0.3, 0.1, 0.05 or 0.01 for each respective cost grid, yielding 18 different least-cost models and thus 18 matrices of different estimates of EGD ( $X_i$ ). For example, the model of EGD with 15% slope cut-off and slope cell weight of 0.1 considered cells with slope < 15% as 10 times more costly to cross than cells with slope > 15%. Slope grids were resampled at 90 m resolution to reduce calculation time.

## ESTIMATING GENETIC DISTANCE AND GENE FLOW AMONG POPULATIONS

We used genetic data from 26 populations of desert bighorn sheep in California to develop the matrix of population pairwise gene flow estimates ( $Z$ ). We identified 392 different individuals from data for 14 microsatellite loci using DNA extracted from faeces, tissue or blood, using two to six replicate polymerase chain reactions (PCRs) (see Epps *et al.* 2005). We used ARLEQUIN (Schneider, Roessli & Excofier 2000) to estimate population pairwise  $F_{ST}$  values and transformed these to  $Nm$  values via the standard Wright–Fisher model  $F_{ST} = 1/(1 + 4Nm)$  as our primary index of relative gene flow. Due to the restrictive assumptions of this model,  $Nm$  is unlikely to represent the actual number of migrants per generation (Whitlock & McCauley 1999) but can indicate relative levels of gene flow, particularly when migration rates exceed mutation rates (Slatkin 1993).

As a second measure of gene flow, we estimated migration rates ( $M$ ) using MIGRATE (Beerli & Felsenstein 2001). Because computation time for the full data set of 26 populations was estimated at about 2 years, we restricted analyses to a subset of nine populations. MIGRATE estimates migration rates among populations

using maximum-likelihood Markov chain Monte Carlo (MCMC) methods, and is an effort to improve migration rate estimates beyond the usual  $F_{ST}$ -based statistics (see Appendix S1 in Supplementary material for details).

#### USING GENE FLOW ESTIMATES TO TEST ALTERNATIVE PARAMETER VALUES

We used PATHMATRIX (Ray 2005) to calculate the least-cost paths among the 26 genetically sampled populations. This extension for ArcView version 3.2 (ESRI) uses a cost grid (here, derived from a given model of EGD) to (1) calculate least-cost paths among all pairs of population polygons; (2) generate the matrix  $\mathbf{X}_i$  of EGD; and (3) map each least-cost path. Each estimate of EGD between a population pair is calculated as:

$$\text{EGD} = \sum x_j w_j \quad \text{eqn 1}$$

where  $x_j$  is the linear distance across each grid cell  $j$  and  $w_j$  is the weight for that cell (determined here by whether the slope value is above or below the slope cut-off), summed over all the cells in a given path. All possible paths are evaluated, but only the EGD of the least-costly path is reported in matrix  $\mathbf{X}_i$ . Finally, we  $\log_{10}$ -transformed values in each matrix  $\mathbf{X}_i$  to linearize the relationship of distance with  $Nm$  (Epps *et al.* 2005).

The presence of anthropogenic barriers (fenced highways, canals and urban areas) was found previously to affect gene flow strongly among these populations (Epps *et al.* 2005). We chose to correct for this effect by including barrier presence/absence as a second predictor matrix  $\mathbf{Y}$  when assessing correlation between EGD and gene flow. Otherwise, if barriers were incorporated into each least-cost grid during the model-fitting process (by assigning large cost values to any grid cell where a barrier was present), appropriate cost values would vary for each least-cost grid. Inappropriate cost values would disrupt the otherwise linear relationship between gene flow ( $Nm$ ) and  $(\log_{10})\text{EGD}$ . Moreover, those barriers have been present for only 40–60 years and have presumably affected gene flow at a different time scale than topography. Finally, barriers could be mitigated and therefore should be considered separately. We incorporated barriers formally into the underlying cost grid only when using the final best-fitting model to define active corridors (as described below). Barriers were recorded as present for any population pair with a barrier interposed; the map of barriers was compiled as described by Epps *et al.* (2005).

We used partial Mantel tests (Smouse, Long & Sokal 1986; Manly 1991) to estimate the partial correlation of matrix  $\mathbf{Z}$  ( $Nm$  or MIGRATE  $M$ ) with each matrix  $\mathbf{X}_i$ , while controlling for the presence of anthropogenic barriers by including matrix  $\mathbf{Y}$  as a second predictor matrix. Tests were conducted using XLSTAT (Addinsoft, New York, USA). Partial Mantel tests determine the correlation of a response matrix  $\mathbf{Z}$  to a predictor matrix

$\mathbf{X}$ , while removing a spurious correlation resulting from a second predictor matrix  $\mathbf{Y}$  that may be correlated with both  $\mathbf{Z}$  and  $\mathbf{X}$ . We used the value of the partial correlation coefficient  $r$  resulting for each  $\mathbf{X}_i$  to compare graphically the relative fit of each model of EGD. We also estimated  $r$  for the null model ( $\mathbf{X}_0$ ) matrix of straight-line distances ( $\log_{10}$ -transformed) between population polygons.

While partial Mantel tests are controversial due to potential underestimation of type I error (Raufaste & Rousset 2001; Rousset 2002), Castellano & Balletto (2002) argued that this concern has been overstated. Moreover, because we compared the partial correlation coefficient of distance matrices while using the same second predictor matrix  $\mathbf{Y}$  in all tests, and did not compare  $P$ -values, such underestimation is unlikely to affect our conclusions.

#### DEFINING POPULATION POLYGONS

Most metrics of gene flow use populations as the basic unit of comparison, defined theoretically as groups of freely interbreeding individuals. In practice, defining the spatial extent of populations may be difficult. To calculate accurate distances among populations, population map polygons must depict habitat used regularly by interacting individuals. To test how sensitive parameter optimization for the least-cost models was to population polygon definition, we repeated EGD calculations using three different methods to define population polygons.

Our first polygon model ['Genetic sampling' (GS); Fig. 1] used minimum convex polygons drawn around the locations in each mountain range where DNA samples were actually collected. If samples were collected at only one location such as a waterhole, we used a circle with diameter of 1 km centred on the sampling point. This approach would be useful for species where the extent of each population sampled is not defined clearly by the habitat patch and is likely to provide a conservatively small habitat area. The second polygon model ['expert opinion' (EO); Fig. 1] used the population polygons defined by Epps *et al.* (2005). These polygons were drawn on the basis of both the topographic extent of each mountain range and expert opinion regarding the distribution of bighorn sheep in each location, derived from field observations and helicopter surveys. Bleich *et al.* (1996) used a similar approach to define population polygons for management purposes. Expert opinion may often be the only available means to define populations for many species.

The final polygon model tested ['habitat model' (HO); Fig. 1] was a GIS model based on slope and distance to perennial water sources. It was designed to provide repeatable polygons depicting desert bighorn sheep distribution and to predict the probable distribution of new populations in vacant habitat. The model was developed using radio telemetry locations of desert bighorn sheep in five populations (California Department.

of Fish and Game, unpublished data) and expert opinion to inform model fit (see Appendix S2).

#### IDENTIFYING AND RANKING DISPERSAL CORRIDORS USING THE BEST-FITTING DISPERSAL MODEL

After examining graphically correlation coefficients from Mantel tests for all  $X_i$  matrices, repeated for three sets of population polygons and  $Z$  matrices based on two different estimates of gene flow, we chose the best-fitting model of EGD by selecting the values of slope cut-off and slope weight that resulted in the strongest correlation coefficients. We then used that best-fitting model to identify probable movement corridors between bighorn sheep populations, after selecting a population polygon model based both on performance and practical considerations.

To identify probable movement corridors, we used two regression-based procedures. We first estimated the maximum effective dispersal distance (the greatest effective geographical distance separating population polygons over which gene flow can be detected; hereafter,  $EGD_{MAX}$ ) for desert bighorn sheep. This was performed via regression of population pairwise estimates of  $Nm$  on estimates of EGD from the best-fitting model for population pairs without intervening barriers. Gene flow, as measured by  $Nm$  between populations, is expected to decline with increasing distance until an asymptote at a 'background' non-zero level of  $Nm$  is reached. At distances greater than this point, current gene flow is unlikely but some degree of genetic similarity exists because of descent from common ancestors or recurrent mutations (Slatkin 1993). Because we could not identify a regression model that adequately described the rapid decline of  $Nm$  to a non-zero asymptote, we used XLSTAT version 2006.2 (Addinsoft) to perform nonparametric regression (Härdle 1992) of  $Nm$  on EGD from the best-fitting dispersal model. Nonparametric regression is essentially a smoothing method for predictive purposes. We used the LOWESS method with the tri-weight kernel and bandwidth equal to the standard deviation, based on the underlying model of a second-degree polynomial. We defined our estimate of  $EGD_{MAX}$  as the point at which the predicted values from the nonparametric regression first stopped decreasing (excluding initial fluctuations at high  $Nm$ ).

We defined active dispersal corridors as those least-cost paths with total cost  $< EGD_{MAX}$ . However, because nonparametric regression does not generate a general predictive equation for gene flow as a function of EGD, we modelled this relationship with a negative exponential regression function for  $EGD < EGD_{MAX}$  (where an adequate fit could be achieved) and used the resulting equation to predict relative gene flow over active dispersal corridors.

To identify probable dispersal corridors on the current landscape, we added barriers to the cost grid of

the best-fitting migration model. Because Epps *et al.* (2005) determined that those barriers had eliminated recent gene flow, we assigned barrier cells a cost equivalent to  $EGD_{MAX}$  to make them impermeable. After adding polygons for un-sampled populations to the population map, we used PATHMATRIX to calculate and map all least-cost paths between populations with a total cost less than  $EGD_{MAX}$ . This was repeated without human-made barriers in the cost-grid to examine how mitigation of those barriers might affect landscape connectivity. To investigate the role that translocations have played in maintaining population connectivity in south-eastern California, we repeated the first analysis but removed five populations re-established by the California Department of Fish and Game through translocations. The relative strength of each corridor was assessed using the exponential decay model to estimate  $Nm$  as a function of EGD.

#### MODEL VALIDATION

Current radio-telemetry data were insufficient to validate the presence of dispersing bighorn sheep in the predicted least-cost corridor routes. Radio-telemetry locations were typically collected monthly; intermountain movements are relatively rare and time spent moving between mountain ranges may be of short duration. However, radio-collared or marked individuals have been detected after moving between mountain ranges. We compiled a list of all such movements as well as those inferred from anecdotal reports. We then evaluated whether least-cost paths from the best-fitting model linked each pair of ranges for which intermountain movements were detected.

#### Results

Effective geographical distance (EGD) based on topography was more strongly negatively correlated with gene flow (both  $Nm$ , as calculated from population pairwise  $F_{ST}$  values, and  $M$ , as estimated by MIGRATE) than straight-line distance in almost all cases, with an absolute increase of the correlation coefficient  $r$  of up to 23% (Fig. 2). EGD models based on 5% slope cut-off performed more poorly than models based on 10% or 15% slope in all cases. The 15% slope cut-off performed slightly better than the 10% cut-off over most (but not all) tests (Fig. 2). For all slope cut-off values, all population polygon models and both measures of gene flow, best-fitting models resulted when sloped terrain had 1/20th to 1/10th the cost of movement across flat terrain (Fig. 2), with the slope weight of 0.10 most often favoured. Therefore, the EGD model employing the 15% slope cut-off and slope weight of 0.10 (hereafter referred to as the 15/0.10 model) was used for further corridor modelling. Stronger correlation coefficients ( $r$ ) were observed when using EO model population polygons (Fig. 2). However, the differences in  $r$  were not large, and

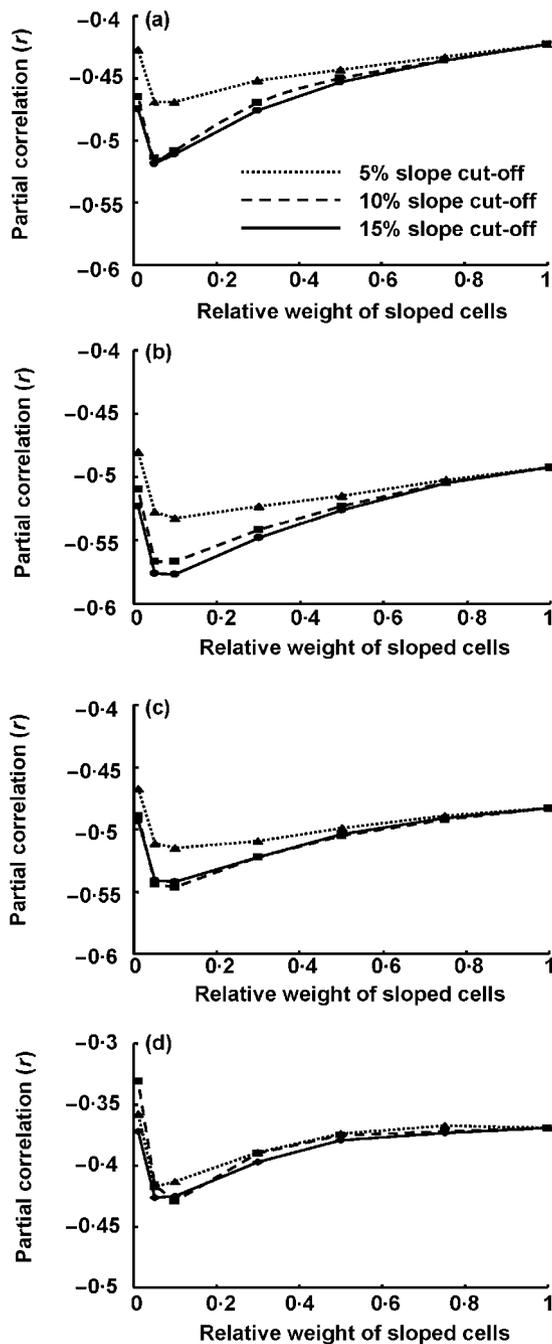


Fig. 2. Coefficients ( $r$ ) for partial correlation of gene flow ( $Nm$ ) with effective geographical distance from least-cost models, while correcting for anthropogenic barriers. Models use slope cut-off values of 5%, 10% and 15% and relative weights for slope cells of 0.01–1.0, for (a) GS polygons; (b) EO polygons; (c) HM polygons; and (d) a subset of nine populations using estimates of gene flow ( $M$ ) from MIGRATE with HM polygons. The slope weight of 1.0 represents the shortest straight-line distance between population pairs.

optimal slope cut-off values and weights were similar, indicating low sensitivity to the choice of population polygon model. We chose HM polygons to calculate  $EGD_{MAX}$  and model different corridor scenarios because this model can be used easily where bighorn sheep are currently absent or their distribution is poorly understood.

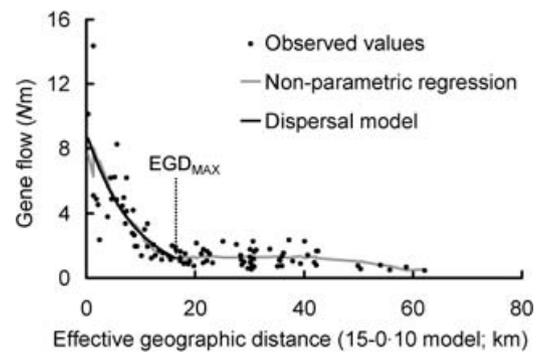


Fig. 3. Population pairwise estimates of gene flow ( $Nm$ ) (for population pairs without intervening anthropogenic barriers) plotted against effective geographical distance (EGD) from the best-fitting model. Maximum effective dispersal distance ( $EGD_{MAX}$ , indicated with dashed arrow) was defined as the smallest EGD (after initial fluctuations) at which the slope of the line of predicted values generated by the nonparametric regression (grey line) stopped decreasing. Non-linear regression (black line) was conducted on all points below  $EGD_{MAX}$  to generate a predictive model for gene flow as a function of EGD. Above  $EGD_{MAX}$ , dispersal was assumed to be negligible.

From nonparametric regression of population pairwise  $Nm$  values on estimates of EGD from the 15/0.10 model, we estimated the maximum effective dispersal distance ( $EGD_{MAX}$ ) as 16.4 km-cost-units (corresponding to 16.4 km of flat terrain or 164 km of sloped terrain; Fig. 3). From regression of  $Nm$  values on EGD (km scale) for all values  $< EGD_{MAX}$  (Fig. 3), we derived the following negative exponential model:

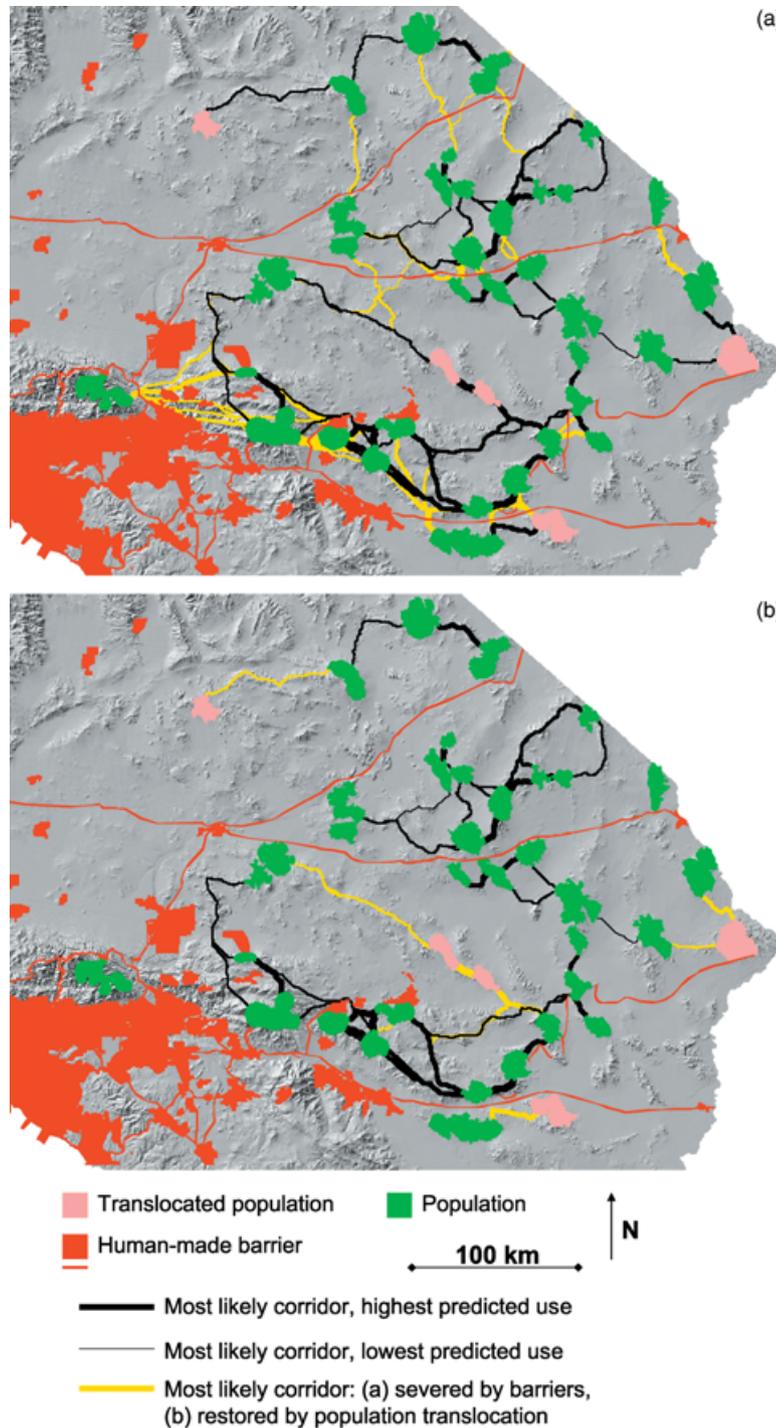
$$Nm = 9.141 * e^{-0.112 * EGD} - 0.219 \quad \text{eqn 2}$$

We used equation 2 to estimate the relative strength of gene flow across active dispersal corridors with  $EGD < EGD_{MAX}$  (Fig. 4).

The connectivity of the current landscape suggested that nearly all populations are currently linked to another population by at least one possible dispersal corridor (black lines, Fig. 4a). However, in some cases these corridors had costs nearing  $EGD_{MAX}$ , making significant gene flow unlikely (narrow-width corridor lines, Fig. 4a). Comparison with corridors mapped in the absence of human-made barriers (yellow lines, Fig. 4a) indicated that those barriers have disrupted several regions of formerly high connectivity and resulted in complete isolation for at least one population. Mapping of corridors with and without populations re-established successfully by translocation (Fig. 4b) demonstrated that those translocations have helped maintain corridors for gene flow across a large region in the centre of the study area and several other areas, thereby greatly reducing the isolation of several native populations.

We identified 31 pairs of mountain ranges in the study area between which intermountain movements of bighorn sheep have been detected or inferred (Appendix S3). Of 22 pairs between which movements were detected via radio-telemetry or observation of

marked animals, 21 pairs were linked by a predicted dispersal corridor. Of nine pairs between which movements were suggested on the basis of anecdotal reports, all were linked by predicted dispersal corridors.



**Fig. 4.** Dispersal corridors predicted by the best-fitting dispersal model (15/0-10) and the HM population model, depicted with hill-shade topography. Black lines indicate least-costly corridor routes for corridors with cost  $< EGD_{MAX}$ , yellow lines indicate least-costly corridor routes that (a) were severed by anthropogenic barriers; or (b) were re-established by translocated populations. Corridors are presented based on (a) all extant populations within the study area, with and without current anthropogenic barriers considered; and (b) extant populations with and without those successfully re-established by translocation, with current anthropogenic barriers considered.

## Discussion

Migration models that incorporated topography explained substantially more variation in gene flow than models that considered only geographical distance. While the models presented here reflect a small portion of possible models, we found that the best-fitting cost weights and slope cut-off values were consistent across different population polygon models and different measures of gene flow (Fig. 2). While time-consuming, we suggest that testing more than one type of gene flow estimate or population polygon model is important as a sensitivity analysis. Greater confidence in our results was derived from the concordance among models tested.

Inferring active dispersal corridors via the best-fitting migration model for desert bighorn sheep in California resulted in several conclusions. Most importantly, anthropogenic barriers currently fragment several regions that previously exhibited high connectivity (Fig. 4a), suggesting priority locations for the mitigation of these barriers. Additionally, mapping dispersal corridors including populations re-established by translocation (Fig. 4b) demonstrated that our models can be used to improve connectivity: if population establishment in an empty habitat patch could link existing populations by active dispersal corridors, a population translocation to that patch might receive higher priority. Potential future barriers can also be evaluated explicitly in this manner and avoided or mitigated at the time of construction. Finally, the successful restoration of several major dispersal corridors connecting otherwise isolated populations suggests that translocation could be used to restore critical nodes of population connectivity for other species.

These applications of the best-fitting migration model demonstrate the value of this tool for conservation and management. Because we parameterized this model from real data, we can have higher confidence that it models correctly the behaviour of bighorn sheep. We suggest reporting the effective geographical distance (EGD) values or predicted relative gene flow to rank corridors. Here, we scaled corridor widths by  $Nm$  to portray relative predicted corridor effectiveness (Fig. 4).

Comparison of the population polygon models suggested that, in this case at least, the definition of population extent did not affect greatly the parameterization of the migration model. Even the most restrictive polygon model (GS polygons, based on the location of the genetic samples collected) exhibited model-fitted curves of the same shape as those generated by the EO and HM polygons. This suggests that fitting least-cost dispersal models may be possible even in situations where the geographical extent of populations is difficult to define. If there is no clear basis at all for defining populations, it should also be possible to develop models in this fashion based on individual pairwise genetic comparisons (e.g. Vignieri 2005). Because this model-testing exercise was designed to examine migration, we caution against over-interpreting differences in

absolute model fit between the population polygon models.

The number of populations in the genetic data set (26) was large, and such a sample might be considered prohibitive to applying this technique for other species. However, results obtained from testing dispersal models using MIGRATE *M* estimates for the nine-population subset were entirely consistent with those from the full data set (Fig. 3d). Thus, even relatively few populations may suffice to fit such dispersal models.

The connectivity network derived from the genetic analyses confirmed that knowledge of bighorn sheep behaviour (i.e. preference for steep terrain) could be incorporated into a connectivity design, even to the extent of identifying where additional population nodes could be reintroduced to improve the overall connectivity of the bighorn sheep metapopulations. This, in turn, suggests that core and corridor analyses for other species, based on behaviour and proper weighting of landscape variables, could provide important tools for management and conservation. Many aspects of this approach bear further investigation. For instance, rather than use the cumbersome 'trial and error' testing of model parameters, it may be possible to determine the best-fitting model mathematically. However, no mathematical solution will be possible once the number of parameter estimates exceeds the number of population pairs with genetic data. Setting up a few biologically plausible alternative models for testing and exploring restricted subsets of parameter space may be the most practical strategy.

Another aspect worthy of investigation is how best to determine when one model represents a 'significant' improvement over another. Model-selection techniques such as Akaike's information criterion (AIC) may be of little value when the identity of the predictor variables does not change among models. For this reason, we selected the best models using a graphical assessment of model fit. In the end, once the appropriate range of model parameters is identified, slight variations in model fit resulting from small differences in cost weights are likely to be unimportant. In our case, fitting corridors based on slope supported the behavioural inference that bighorn sheep prefer to travel over sloped terrain offering security from predators, regardless of minor differences between 10% and 15% slope cut-offs. Small changes in model parameters may become more important when considering whether an individual corridor is likely to be used or not. For this reason, we reiterate that the relative likelihood of corridor use should be considered, rather than merely a 'corridor or not-corridor' assessment.

#### MODEL VALIDATION

Known intermountain movements by bighorn sheep correlated well with our corridor model, with the exception of one marked individual that apparently crossed an interstate highway. This observation

highlights the difference between individual dispersal events and the broad patterns of movement over time inferred by our analyses of gene flow. Occasional movements may far exceed those predicted by our migration model. Whether bighorn sheep follow routes consistent with the least costly paths among ranges is also unclear. Acquiring enough data points to verify the complete movement paths of dispersing bighorn sheep will probably require the use of GPS collars set to collect multiple locations per day. Until then, path locations predicted by our model should be considered as hypotheses for further testing.

#### LIMITATIONS OF THE APPROACH

While the field of landscape genetics is making rapid strides in developing analyses of gene flow that consider complicated landscape features, our approach still has a number of limitations. For instance, such a modelling exercise is better suited to dealing with common landscape characteristics that affect large numbers of populations, given the low statistical power of matrix correlation tests. In this analysis, the south-westernmost populations inhabit mountain ranges with thick forests and chaparral. Those habitat elements probably strongly limit movement by bighorn sheep because of increased predation risk. We did not consider those elements in model development because of the small number of populations affected; thus, connectivity in that region may have been overstated.

A second limitation to our model is that it reflects more effectively the potential for gene flow rather than colonization of empty habitat patches. Desert bighorn sheep have sex-biased dispersal: males are much more likely to travel long distances between populations, while females are probably the limiting factor in colonization events. Because the model described here is fitted using nuclear genetic markers, it represents both male- and female-mediated gene flow. A correction for the reduced movement of females possibly could be generated from radio-telemetry data or mtDNA, although the variability in estimates of gene flow from mtDNA (resulting from its behaviour as one linked locus) makes its use inherently imprecise. This limitation may be important to consider when using these models for management decisions; for example, determining when translocation may be necessary for population re-establishment.

Determining how to model landscape features such as anthropogenic barriers proved to be a complex issue. We dealt with those barriers in a separate analytical framework during model fitting and brought them back into the final model. This approach seemed appropriate because roads have been present on the landscape for only a short period of time. Moreover, road impacts can be mitigated and therefore corridor design should be assessed as a function of the mitigated landscape. A further technical limitation is that the width of interstate highway corridors and other barriers

varies; ideally, the estimated cost of the barrier should be applied to any path crossing the barrier but not on a per-pixel basis (where that cost is accumulated for each pixel encountered). Other, more integrative approaches may be of value in other systems.

Finally, an important caveat is that we used migration, a long-term process, to make inferences about current patterns of bighorn sheep dispersal. Variation in allele frequencies used to estimate migration may be affected by other factors such as population bottlenecks (Whitlock & McCauley 1999). Moreover, if past conditions are reflected more strongly than current dispersal patterns, management decisions using these models might be flawed. However, the small size of these populations and the detectable effect of barriers present for only six to seven generations (Epps *et al.* 2005) suggest that in this case we can still make useful inferences about movement of bighorn sheep on the recent landscape as well as identify factors likely to affect connectivity on the current landscape. Because dispersal is a complex process and the reasons that an individual animal does or does not disperse are unclear, and may not be reduced to simple models, fitting least-cost models using genetic data is probably most effective at identifying broad-scale patterns of gene flow resulting from landscape features that have been present for at least a few generations.

#### IMPROVING CORRIDOR MODELS AND PLANS TO MAINTAIN OR RE-ESTABLISH CONNECTIVITY

Our study suggests that developing least-cost models from genetic data can improve significantly the quality of and confidence in models of dispersal, migration and connectivity. Other types of data on movement could be used in a similar approach (e.g. Sutcliffe *et al.* 2003). Least-cost models have been employed worldwide to plan landscape-scale conservation strategies, to design reserves and to assess the effects of habitat fragmentation on many species. In some cases those models may have been applied uncritically with respect to their underlying assumptions. While developing genetic data or other data on movement may be a difficult task for many species, it may at least be possible to inform such models using data from species with similar biological characteristics.

#### Acknowledgements

We thank T. W. Epps, A. Hendricks, E. Kaufman, L. Konde and J. Thorne for assistance in developing the concepts of fitting least-cost models, G. Sudmeier and the late B. Campbell of the Society for the Conservation of Bighorn Sheep for information on water distribution, as well as F. He, P. Beier and two anonymous reviewers for helpful comments on the manuscript. GIS analyses were conducted in the Geospatial Informatics and Imaging Facility (GIIF) of the University of California,

Berkeley (UCB) and funded by the Resource Assessment Program of the California Department of Fish and Game. This is professional paper 056 of the Eastern Sierra Center for Applied Population Ecology. We also thank D. McCullough, P. Palsbøll, G. Roderick and the many individuals who assisted during the collection of the genetic data; funding for that portion of the project was provided by the National Science Foundation, the Agricultural Experiment Station of UCB, the Golden Gate Chapter of Safari Club International and the UCB chapter of Sigma Xi.

#### References

- Adriaenssens, F., Chardon, J.P., De Blust, G., Swinnen, E., Villalba, S., Gulinck, H. & Matthysen, E. (2003) The application of 'least-cost' modelling as a functional landscape model. *Landscape and Urban Planning*, **64**, 233–247.
- Beazley, K., Smandych, L., Snaith, T., MacKinnon, F., Austen-Smith, P. & Duinker, P. (2005) Biodiversity considerations in conservation system planning: map-based approach for Nova Scotia, Canada. *Ecological Applications*, **15**, 2192–2208.
- Beerli, P. & Felsenstein, J. (2001) Maximum likelihood estimation of a migration matrix and effective population sizes in *n* subpopulations by using a coalescent approach. *Proceedings of the National Academy of Sciences USA*, **98**, 4563–4568.
- Bleich, V.C., Wehausen, J.D. & Holl, S.A. (1990) Desert-dwelling mountain sheep: conservation implications of a naturally fragmented distribution. *Conservation Biology*, **4**, 383–390.
- Bleich, V.C., Wehausen, J.D., Ramey, R.R. & Rechel, J.L. (1996) Metapopulation theory and mountain sheep: implications for conservation. *Metapopulations and Wildlife Conservation* (ed. D.R. McCullough), pp. 353–373. Island Press, Covelo.
- Castellano, S. & Balletto, E. (2002) Is the partial Mantel test inadequate? *Evolution*, **56**, 1871–1873.
- Coulon, A., Cosson, J.F., Angibault, J.M., Cargnelutti, B., Galan, M., Morellet, N., Petit, E., Aulagnier, S. & Hewison, A.J.M. (2004) Landscape connectivity influences gene flow in a roe deer population inhabiting a fragmented landscape: an individual-based approach. *Molecular Ecology*, **13**, 2841–2850.
- Epps, C.W., Bleich, V.C., Wehausen, J.D. & Torres, S.G. (2003) Status of bighorn sheep in California, 2004. *Desert Bighorn Council Transactions*, **47**, 20–35.
- Epps, C.W., Palsboll, P.J., Wehausen, J.D., Roderick, G.K., Ramey, R.R.I.I. & McCullough, D.R. (2005) Highways block gene flow and cause a rapid decline in genetic diversity of desert bighorn sheep. *Ecology Letters*, **8**, 1029–1038.
- Hanski, I. (1999) *Metapopulation Ecology*. Oxford University Press, New York.
- Hardle, W. (1992) *Applied Nonparametric Regression*. Cambridge University Press, Cambridge, UK.
- Hedrick, P.W. (2005) *Genetics of Populations*. Jones and Bartlett Publishers, Boston.
- Hunter, R.D., Fisher, R.N. & Crooks, K.R. (2003) Landscape-level connectivity in coastal Southern California, USA, as assessed through carnivore habitat suitability. *Natural Areas Journal*, **23**, 302–314.
- Joly, P., Morand, C. & Cohas, A. (2003) Habitat fragmentation and amphibian conservation: building a tool for assessing landscape matrix connectivity. *Comptes Rendus Biologies*, **326**, S132–S139.
- Keller, I. & Largiader, C.R. (2003) Recent habitat fragmentation caused by major roads leads to reduction of gene flow and

- loss of genetic variability in ground beetles. *Proceedings of the Royal Society of London, Series B, Biological Sciences*, **270**, 417–423.
- Koenig, W.D., VanVuren, D. & Hooge, P.N. (1996) Detectability, philopatry, and the distribution of dispersal distances in vertebrates. *Trends in Ecology and Evolution*, **11**, 514–517.
- Manel, S., Schwartz, M.K., Luikart, G. & Taberlet, P. (2003) Landscape genetics: combining landscape ecology and population genetics. *Trends in Ecology and Evolution*, **18**, 189–197.
- Manly, B.F.J. (1991) *Randomization and Monte Carlo Methods in Biology*. Chapman & Hall, London.
- Michels, E., Cottenie, K., Neys, L., De Gelas, K., Coppin, P. & De Meester, L. (2001) Geographical and genetic distances among zooplankton populations in a set of interconnected ponds: a plea for using GIS modelling of the effective geographical distance. *Molecular Ecology*, **10**, 1929–1938.
- Moilanen, A., Franco, A.M.A., Eary, R.I., Fox, R., Wintle, B. & Thomas, C.D. (2005) Prioritizing multiple-use landscapes for conservation: methods for large multi-species planning problems. *Proceedings of the Royal Society, Series B, Biological Sciences*, **272**, 1885–1891.
- Nikolakaki, P. (2004) A GIS site-selection process for habitat creation: estimating connectivity of habitat patches. *Landscape and Urban Planning*, **68**, 77–94.
- Raufaste, N. & Rousset, F. (2001) Are partial Mantel tests adequate? *Evolution*, **55**, 1703–1705.
- Ray, N. (2005) PATHMATRIX: A geographical information system tool to compute effective distances among samples. *Molecular Ecology Notes*, **5**, 177–180.
- Ray, N., Lehmann, A. & Joly, P. (2002) Modeling spatial distribution of amphibian populations: a GIS approach based on habitat matrix permeability. *Biodiversity and Conservation*, **11**, 2143–2165.
- Rouget, M., Cowling, R.M., Lombard, A.T., Knight, A.T. & Graham, I.H.K. (2006) Designing large-scale conservation corridors for pattern and process. *Conservation Biology*, **20**, 549–561.
- Rousset, F. (2002) Partial Mantel tests: reply to Castellano and Balletto. *Evolution*, **56**, 1874–1875.
- Schneider, S., Roessli, D. & Excofier, L. (2000) *ARLEQUIN: A Software for Population Genetics Data Analysis*, version 2.000. Genetics and Biometry Laboratory, Department of Anthropology, University of Geneva, Geneva.
- Schwartz, O.A., Bleich, V.C. & Holl, S.A. (1986) Genetics and the conservation of mountain sheep *Ovis canadensis nelsoni*. *Biological Conservation*, **37**, 179–190.
- Slatkin, M. (1993) Isolation by distance in equilibrium and non-equilibrium populations. *Evolution*, **47**, 264–279.
- Smouse, P.E., Long, J. & Sokal, R. (1986) Multiple regression and correlation extensions of the Mantel test of matrix correspondence. *Systematic Zoology*, **35**, 627–632.
- Soule, M.E., Wilcox, B.A. & Holtby, C. (1979) Benign neglect: a model of faunal collapse in the game reserves of East Africa. *Biological Conservation*, **15**, 259–272.
- Spear, S.F., Peterson, C.R., Matocq, M.D. & Storfer, A. (2005) Landscape genetics of the blotched tiger salamander (*Ambystoma tigrinum melanostictum*). *Molecular Ecology*, **14**, 2553–2564.
- Sutcliffe, O.L., Bakkestuen, V., Fry, G. & Stabbetorp, O.E. (2003) Modelling the benefits of farmland restoration: methodology and application to butterfly movement. *Landscape and Urban Planning*, **63**, 15–31.
- U.S. Geological Survey (2003) 30m digital elevation model for the coterminous United States. USGS, Washington, D.C.
- Vignieri, S.N. (2005) Streams over mountains: influence of riparian connectivity on gene flow in the Pacific jumping mouse (*Zapus trinotatus*). *Molecular Ecology*, **14**, 1925–1937.
- Vos, C.C., Antonisse-De Jong, A.G., Goedhart, P.W. & Smulders, M.J.M. (2001) Genetic similarity as a measure for connectivity between fragmented populations of the moor frog (*Rana arvalis*). *Heredity*, **86**, 598–608.
- Waser, P.M. & Strobeck, C. (1998) Genetic signatures of interpopulation dispersal. *Trends in Ecology and Evolution*, **13**, 43–44.
- Whitlock, M.C. & McCauley, D.E. (1999) Indirect measures of gene flow and migration: F-ST not equal 1/(4Nm+1). *Heredity*, **82**, 117–125.

Received 24 August 2006; final copy received 13 March 2007  
Editor: Fangliang He

### Supplementary material

The following supplementary material is available for this article.

**Appendix S1.** Estimating migration rates *M* using MIGRATE software.

**Appendix S2.** Defining the Habitat Model population polygons.

**Appendix S3.** Reported intermountain movements by desert bighorn sheep in the study area.

This material is available as part of the online article from: <http://www.blackwell-synergy.com/doi/full/10.1111/j.1365-2664.2007.01325.x>

Please note: Blackwell Publishing is not responsible for the content or functionality of any supplementary materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

Copyright of *Journal of Applied Ecology* is the property of Blackwell Publishing Limited and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.