Molecular Ecology Notes (2005) **5**, 177–180 doi: 10.1111/j.1471-8286.2004.00843.x

PROGRAM NOTE

PATHMATRIX: a geographical information system tool to compute effective distances among samples

NICOLAS RAY
Computational and Molecular Population Genetics Laboratory, Zoological Institute, University of Bern, Baltzerstrasse 6, 3012 Bern, Switzerland and Genetics and Biometry Laboratory, Anthropology and Ecology Department, University of Geneva, Rue Gustave Revilliod 12, 1227 Carouge, Switzerland

Abstract

PATHMATRIX is a tool used to compute matrices of effective geographical distances among samples using a least-cost path algorithm. This program is dedicated to the study of the role of the environment on the spatial genetic structure of populations. Punctual locations (e.g. individuals) or zones encompassing sample data points (e.g. demes) are used in conjunction with a species-specific friction map representing the cost of movement through the landscape. Matrices of effective distances can then be exported to population genetic software to test, for example, for isolation by distance. PATHMATRIX is an extension to the geographical information system (GIS) software ARCVIEW 3.x.

Keywords: distance matrix, ecological distance, heterogeneous landscape, isolation by distance, least-cost path, spatial genetic structure

Received 03 September 2004; revision received 17 September 2004; accepted 06 October 2004

The increasing availability of fine-scaled digital representation of landscape and the growing body of genetic markers have recently enhanced the number of applications in 'landscape genetics'. This fast-growing field primarily aims at investigating the interactions between landscape features and microevolutionary processes (Manel *et al*. 2003). Such interactions are classically detected by investigating the spatial genetic structure of a focal species in heterogeneous environments. Much of the information contained in spatial patterns may be captured in pairwise measures of genetic correlation as a function of physical distance, either considering individuals or demes (Epperson 2003). The model of isolation by distance (Wright 1943) is hence very useful in this context. Typical isolation by distance studies seek to determine whether there is a statistically significant relationship between genetic distances (or similarity) and physical distances among samples, and to assess the strength of this relationship. In that context, geographical distances measured among samples are typically straight-line Euclidean distances (e.g. Garnier *et al*. 2004; Sacks *et al*. 2004). However, recent population genetics studies have tried to incorporate information about the structure of the landscape to obtain more realistic distances regarding the movement of individuals in heterogeneous environments (e.g. Arnaud 2003; Coulon *et al*. 2004; Michels *et al*. 2001). These improved distances are called 'effective distances' (Verbeylen *et al*. 2003) and may be used to reveal the effect of landscape features on microevolutionary processes in the context of isolation by distance.

PATHMATRIX is a tool to compute effective distances among sample locations. The output of the program is a set of matrices of effective distances that may be compared to matrices of genetic distances. PATHMATRIX is an extension to the geographical information system (GIS) software ARCVIEW 3.x (*Environmental Science Research Institute, Redlands, USA*). ARCVIEW was chosen as a platform because it is widely used in academic research. The development of PATHMATRIX within an existing GIS framework (as opposed to a stand-alone tool) has numerous advantages. Main advantage is the ability to benefit from the existing GIS internal data structure and to display outputs compatible with other GIS layers. The computation of effective distances in PATHMATRIX is based on the *cost distance* algorithm implemented in the ARCVIEW module *Spatial Analyst*. This algorithm computes a deterministic least-cost path between a source point and a target point by using a friction (or resistance) layer. The friction layer is a raster map

Correspondence: Nicolas Ray. Fax: +41 31 631 48 48. E-mail: nicolas.ray@zoo.unibe.ch

© 2005 Blackwell Publishing Ltd

where each cell (landscape unit) expresses the relative difficulty (or cost) of moving through that cell for a given species. A least-cost path minimizes the sum of frictions of all cells along the path, and this sum is the least-cost distance (for detailed description and discussion of the algorithm, see Adriaensen *et al*. 2003). Especially for habitat specialists, least-cost distances may give a more realistic measure of spatial isolation (or its inverse, connectivity) than standard Euclidean distances (e.g. Chardon *et al*. 2003; Coulon *et al*. 2004).

Although the existing *cost distance* algorithm is used, PATHMATRIX greatly enhances the tools available in ARCVIEW. First, PATHMATRIX applies the *cost distance* algorithm in a pairwise fashion among a set of sample locations (a number of points or polygons), while ARCVIEW (and extensions such as *Cost Distance Grid Tools*, ESRI 1998) may compute least-cost paths only among a few single pairs of points. Second, in addition to the least-cost distance (in cost units), PATHMATRIX can also output the length of the least-cost path (in geographical distance units, e.g. meters), a type of effective distance that is not available in ARCVIEW. The length of the least-cost path has so far received little attention in the literature, possibly because of the lack of appropriate tools to compute it. However, the ecological significance of this type of distance, simply representing the length of the most likely path an individual may follow, is in some cases more straightforward than a sum of distances weighted by arbitrary costs (Thomas Broquet, pers. comm.). Third, by providing an integrated user interface and multiple tailored formats for input data and output matrices, PATHMATRIX facilitates the use of the least-cost approach by users that are not familiar with ARCVIEW.

To obtain a friction map, the environmental heterogeneity must be translated into cost units. Cost of movements are usually difficult to derive from available data on the ecology and behaviour of species, and expert knowledge must often be used (e.g. Ray *et al*. 2002). Moreover, the number of friction classes and their relative weight may have a substantial impact on the results (Verbeylen *et al*. 2003). It is therefore important to consider several friction scenarios. In that purpose, PATHMATRIX may be run using several friction grids at once. These grids may represent alternative ecological hypotheses or dispersal pathways to be tested, or they can be part of a sensitivity analysis around best estimates of friction values. Note that even in the absence of obvious environmental heterogeneity (i.e. a uniform land cost map), PATHMATRIX may still be very useful to compute realistic distances that circumvent large barriers to dispersal. For example, distances among samples at a continental scale may be obtained by setting oceans as complete barriers to movements, so that computed paths will follow shorelines instead of crossing sea surface.

PATHMATRIX needs two main inputs: a file describing the location of samples and a grid map of friction values. Location of samples are imported either directly from an existing point *shapefile* (the open format for vector data in ARCVIEW), or through a simple <x,y> coordinates text file (ASCII). The format of this text file is similar to the sample file of the program SPLATCHE (SPatiaL And Temporal Coalescences in Heterogeneous Environments, Currat *et al*. 2004). Coordinates may be in decimal degrees or in meters, and they must be in the same projection as the input grids (see below). Although it is more common to consider a centroid point (average coordinates among data points) when computing distances among samples, a single coordinate may not realistically represent a spatial sampling unit at a landscape or regional scale. Amphibians, for example, are typically sampled during breeding season in large water bodies. In such cases, a more appropriate way of computing distances between breeding habitats is to consider the zones defined by the edge of each water bodies, and not the coordinates of their centroids. PATHMATRIX allows to compute closest edge-to-edge Euclidean and effective distances based on a polygon shapefile depicting zones in which individuals were sampled (see Fig. 1).

The second main input, the friction map, must be an existing grid file (ESRI grid format) representing friction values. Alternatively, PATHMATRIX may be run on several friction grids at a time. Input grids may be in any type of projection, but to avoid bias it is important to use a projection that minimizes distortions of areas and distances when computing effective distances based on least-cost paths (for discussion of projection bias, see Steinwand *et al*. 1995).

Apart from standard Euclidean distances, two types of effective distances can be generated with PATHMATRIX: (a) the accumulative cost distance of the least-cost path (in cost units), (b) the length of the least-cost path (in geographical distance units, e.g. meters). Distances are computed between all pairs of sample locations (points or polygons), and choice is given to save output matrices in six different formats: dBase format (to be opened in common spreadsheet applications), simple tab-delimited text format, simple single column text format, IBD single column text format (Bohonak 2002), SPAGEDI matrix format (Hardy & Vekemans 2002), and FSTAT single column text format (Goudet 1995). With the three later formats, an effective distance matrix may be imported into the corresponding software and used with a genetic distance matrix to test for isolation by distance. Choice is also given to output the logarithm of distance. In term of visualization of the least-cost paths, an option allows to display the whole set of paths as a polyline shapefile (see Fig. 1), which may also be saved. This can be very useful to gain a better understanding of the variations in direction and length of the paths with alternative friction scenarios.

A preliminary version of PATHMATRIX has already been used to investigate the phylogeography of the long-toed salamander in the Cordilleran glacial valleys of North

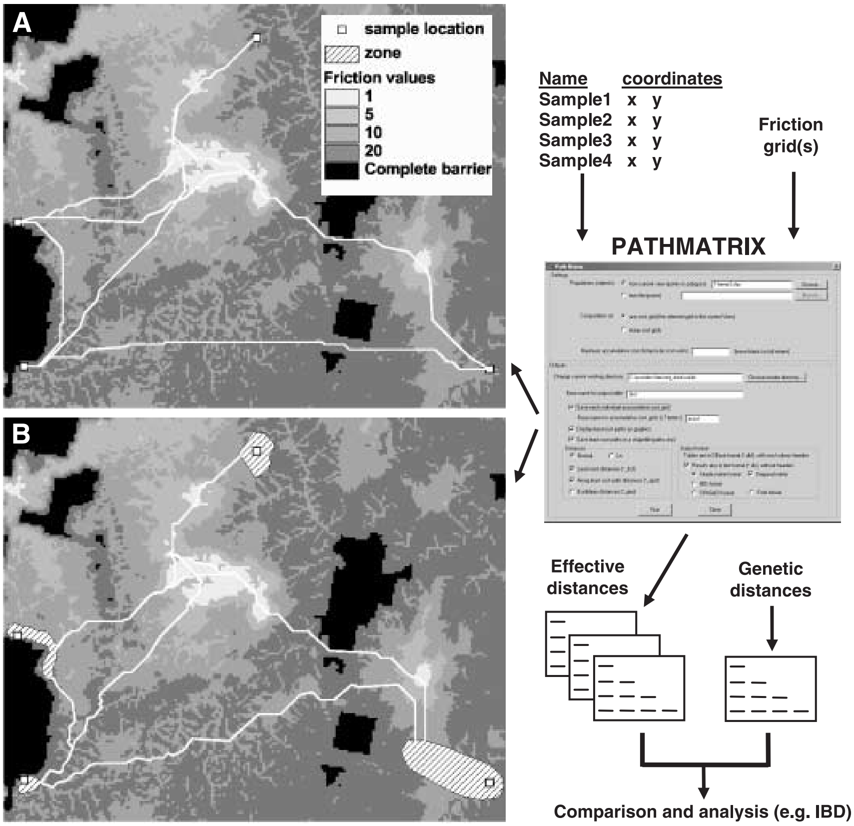

Fig. 1 Schematic view of the main inputs and outputs of PATHMATRIX, showing least-cost paths (white lines) computed among (A) four punctual locations and (B) four zones. Depending on the underlying friction values (depicted as different intensities of grey), the paths may be very different than simple Euclidean (straight line) distances. Output matrices of effective distances can then be compared to a matrix of genetic distances using, for example, IBD (Bohonak 2002).

America (Thompson *et al*. in press). In that context, the use of least-cost effective distances based on topographical constraints helped to decipher some of the observed regional mitochondrial signatures. PATHMATRIX is the only currently available tool that computes multiple effective distance matrices based on least-cost paths, and it should be of great utility for population geneticists wishing to obtain more realistic intersample distances.

PATHMATRIX 1.0 is written in *Avenue* language, and is available as an ARCVIEW 3.x extension for Windows. The extension, user guide and example files can be downloaded from http://cmpg.unibe.ch/software/pathmatrix.

Acknowledgements

I am grateful to Thomas Broquet, Jane Elith, David Duncan, and Vincent Castric for constructive comments on an earlier version of the manuscript. This work was partly supported by a Swiss NSF postdoctoral grant (n° PBGEA-101314) while I was working at the Environmental Science Group of University of Melbourne, Australia.

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.

Arnaud J-F (2003) Metapopulation genetic structure and migration pathways in the land snail *Helix aspersa*: influence of landscape heterogeneity. *Landscape Ecology*, **18**, 333–346.

Bohonak AJ (2002) IBD (isolation by distance): a program for analyses of isolation by distance. *Journal of Heredity*, **93**, 153–154.

Chardon JP, Adriaensen F, Matthysen E (2003) Incorporating landscape elements into a connectivity measure: a case study for the speckled wood butterfly (*Pararge aegeria* L.). *Landscape Ecology*, **18**, 561–573.

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

Currat M, Ray N, Excoffier L (2004) SPLATCHE: a program to simulate genetic diversity taking into account environmental heterogeneity. *Molecular Ecology Notes*, **4**, 139–142.

Epperson BK (2003) *Geographical Genetics*. Princeton University Press, Princeton, NJ.

ESRI (1998) Cost Distance Grid Tools: extension to ARCVIEW 3.x. Available at: http://arcscripts.esri.com/details.asp?dbid=10928.

Garnier S, Alibert P, Audiot P, Prieur B, Rasplus J-Y (2004) Isolation by distance and sharp discontinuities in gene frequencies: implications for the phylogeography of an alpine insect species, *Carabus solieri*. *Molecular Ecology*, **13**, 1883–1897.

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

Hardy OJ, Vekemans X (2002) SPAGEDI: a versatile computer program to analyse spatial genetic structure at the individual or population levels. *Molecular Ecology Notes*, **2**, 618–620.

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

Michels E, Cottenie K, Neys L *et al.* (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.

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.

Sacks BN, Brown SK, Ernest HB (2004) Population structure of California coyotes corresponds to habitat-specific breaks and illuminates species history. *Molecular Ecology*, **13**, 1265–1275.

Steinwand DR, Hutchinson JA, Snyder JP (1995) Map projections for global and continental data and an analysis of pixel distortion caused by reprojection. *Photogrammetric Engineering and Remote Sensing*, **61**, 1487–1497.

Thompson MD, Ray N, Russell AP (in press) Phylogeographical analysis of the long-toed salamander (*Ambystoma macrodactylum*): cross-validation of mtDNA loci and heuristic spatial statistics.

Verbeylen G, De Bruyn L, Adriaensen F, Matthysen E (2003) Does matrix resistance influence red squirrel (*Sciurus vulgaris* L 1758) distribution in an urban landscape? *Landscape Ecology*, **18**, 791–805.

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