

COMPUTER PROGRAM NOTE

Genetic landscapes GIS Toolbox: tools to map patterns of genetic divergence and diversity

AMY G. VANDERGAST,* WILLIAM M. PERRY,† ROBERTO V. LUGO‡ and STACIE A. HATHAWAY*

*U.S. Geological Survey, Western Ecological Research Center, San Diego Field Station, 4165 Spruance Road Suite 200, San Diego, CA 92101, USA, †U.S. Geological Survey, Dixon Field Station, 6924 Tremont Road, Dixon, CA 95620, USA, ‡U.S. Geological Survey, Western Region Geospatial Information Office, 3020 State University Drive East, Modoc Hall, Suite 4003, Sacramento CA 95819, USA

Abstract

The Landscape Genetics GIS Toolbox contains tools that run in the Geographic Information System software, ArcGIS®, to map genetic landscapes and to summarize multiple genetic landscapes as average and variance surfaces. These tools can be used to visualize the distribution of genetic diversity across geographic space and to study associations between patterns of genetic diversity and geographic features or other geo-referenced environmental data sets. Together, these tools create genetic landscape surfaces directly from tables containing genetic distance or diversity data and sample location coordinates, greatly reducing the complexity of building and analyzing these raster surfaces in a Geographic Information System.

Keywords: ArcGIS, genetic distance, mapping

Received 21 June 2010; revision accepted 6 July 2010

The rapidly growing field of landscape genetics aims to investigate the influences of landscape features on micro-evolutionary processes such as gene flow and genetic differentiation (Manel *et al.* 2003). A promising new analysis approach is the creation of 'genetic landscapes' to visualize the distribution of genetic diversity across geographic space (Miller *et al.* 2006). A potentially powerful application of this approach is visualization and analysis of genetic landscape surfaces in combination with other geo-referenced environmental data layers in a Geographic Information System (GIS). Using this approach, one can evaluate hypothesized geographic barriers to movement and gene flow and other landscape features associated with patterns of genetic diversity. Additionally, it is possible to overlay genetic landscapes from multiple sources (e.g. different types of genetic data: mtDNA/nuDNA sequences, microsatellites or multiple species or taxa) to examine patterns of concordance and to locate geographic regions that are important reservoirs of genetic diversity across taxa. A multiple species approach may have great conservation value in identify-

ing geographic areas of evolutionary or adaptive potential. Examining the spatial distribution of genetic variability within species can help identify vicariant and/or adaptive processes. Maintenance of adequate genetic variability ultimately facilitates these evolutionary processes as the biotic and abiotic environments change. Across multiple taxa, comparative phylogenetic and population genetic approaches can reveal the geological and ecological processes that have shaped contemporary genetic and species biodiversity patterns and provide insight into the future (Avice 1992; Humphries *et al.* 1995; Moritz & Faith 1998; Riddle & Hafner 2006). To be incorporated into conservation planning efforts, the spatial components of such evolutionary processes must be mapped and presented in a standardized framework (Prendergast *et al.* 1999; Cowling & Pressey 2001). Overlaying and summarizing genetic landscapes from multiple species in a GIS platform can provide just such a framework (Vandergast *et al.* 2008).

We have developed a series of GIS tools to map genetic landscapes and to summarize multiple genetic landscapes as average and variance surfaces in the Geographic Information Systems software package, ArcGIS (Environmental Systems Research Institute, Redlands, CA, USA). Together, the tools contained within this

Correspondence: Amy G. Vandergast, Fax: 619 225 6436; E-mail: avandergast@usgs.gov

2 COMPUTER PROGRAM NOTE

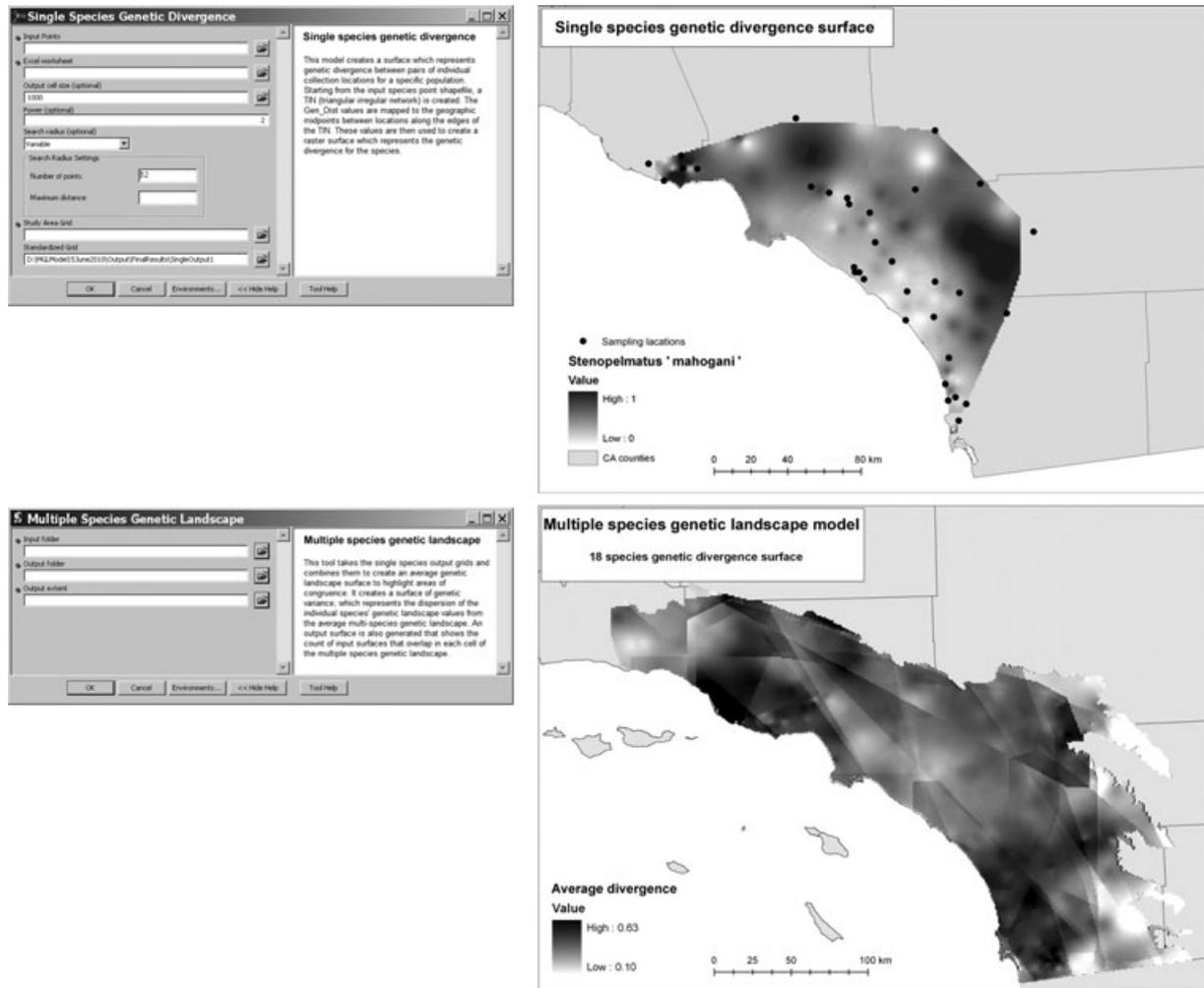


Fig. 1 Input GUIs for the Single Species Divergence Tool and the Multiple Species Genetic Landscapes Tool and corresponding output files created in ArcGIS.

toolbox automate a series of calculations and data manipulations to create genetic landscape surfaces directly from tables containing genetic distance or diversity data and sample location coordinates. This allows users with little GIS experience to create and analyze these complex raster surfaces with efficiency and accuracy. The development of these tools within an existing GIS framework is more advantageous than stand-alone tools because they utilize existing GIS data structure and standards and produce outputs that are compatible with other GIS layers. ArcGIS software was chosen as a platform because it is widely used in research and applied landscape conservation.

There are four tools included in the Genetic Landscapes GIS Toolbox. The 'Single Species Genetic Divergence Tool' creates a genetic divergence raster surface from pairwise population genetic distances for a single species or marker-type/locus (Fig. 1). The 'Single Species

Genetic Diversity Tool' creates a raster surface based on intrapopulation genetic diversity for a single species or marker. The 'Multiple Species Genetic Landscape Tool' summarizes multiple genetic divergence or diversity surfaces with geographic overlap as three raster surfaces; an average, variance and a count of input surfaces per cell. The 'Create Shapefile from Table Tool' is a utility that creates a shapefile from a table of sample location GPS coordinates. Each tool is described in more detail in the following section.

Single species genetic divergence tool

The single species divergence tool creates a genetic landscape based on pairwise population (or individual-based) genetic divergence as measured among multiple collection point locations. The method is similar to that implemented in the programs Alleles in Space (Miller

2005) and Barriers (Manni *et al.* 2004). A network connecting all collection points to their nearest neighbours with nonoverlapping edges is drawn. The midpoints between each connected edge are mapped, and genetic divergence values are attached to these points. A spatial interpolation algorithm, inverse distance weighted (IDW) interpolation, is used to generate a surface from the mapped genetic distance values. IDW interpolation is a commonly used technique for interpolation of scatter points. IDW methods are based on the assumption that the interpolating surface should be influenced most by nearby points and less by more distant points. The interpolating surface is a weighted average of the scatter points, and the weight assigned to each scatter point decreases as the distance from the interpolation point to the scatter point increases. To avoid extrapolating beyond the spatial extent of collection points, the genetic landscape is clipped to the extent of the original network (sampling extent) and to the boundaries of the region of analysis (as determined by the user). The tool will create a raster surface with the raw genetic divergence values and another raster surface with values scaled between 0 and 1 to allow equal weighting of multiple surfaces for inclusion in the Multiple Species Genetic Landscape Tool.

Input files needed to run this tool include a shapefile of collection points (which can be generated from a table using the fourth tool in this package), a table containing pairwise genetic divergence values and a grid file representing the extent of the analysis area. (A grid file is the raster GIS file format utilized in ARCGIS software). While any pairwise genetic divergence estimates can be used, residual values from a regression of geographic distance on genetic divergence may be particularly informative. It is commonly found that genetic differences accumulate as a function of increasing geographic distance (Slatkin 1993). Utilizing residual values from a regression of geographic distance on genetic divergence allows one to remove the effects of simple distance on genetic divergence to reveal regions of unusually high or low divergence, regardless of distance (Manni *et al.* 2004).

Single species genetic diversity tool

The single species diversity tool creates a genetic landscape similar to the genetic divergence tool, but it is based on within-site population genetic diversity for multiple populations or collection points within a species. From a starting table containing collection point coordinates and a genetic diversity measure for each population, IDW interpolation is used to generate a surface from the mapped genetic diversity values. To avoid extrapolating beyond the original collection locations, the genetic landscape is clipped to the extent of the input

collection points (sampling extent) and to the boundaries of the region of analysis (as determined by the user). As with the single species divergence tool, two output surfaces will be created, one with the raw values and another with the values scaled between 0 and 1. Two input files are necessary to run the single species genetic diversity tool, a table containing collection location coordinates and a corresponding genetic diversity estimate and a grid file representing the analysis extent.

Multiple species genetic landscape tool

This tool summarizes multiple single species divergence or diversity surfaces (as calculated with the previous tools) that have geographic overlap. There are three output surfaces. From the input genetic surfaces, a raster calculator is employed to calculate a simple arithmetic mean in each grid cell, creating a mean surface. To better explore data, a second surface, sample variance, which represents the dispersion of the individual species' genetic landscape values from the average multiple species genetic landscape, is calculated. Finally, a third output surface is generated that shows the count of input surfaces that overlap in each cell of the multiple species genetic landscape. This count surface allows the user to visually assess how many individual surfaces are contributing to the multiple species average and variance genetic landscapes for any given cell, in cases where individual surfaces do not show complete overlap.

To run this tool, all single species divergence or diversity landscape surfaces to be summarized are placed in a single folder. This folder of landscape surfaces and a grid representing analysis extent are necessary to run the tool.

Create shapefile from table tool

The final tool is included to assist in creating the shapefile that is necessary for running the Single Species Genetic Divergence Tool. A shapefile is the open specification for vector data in ARCGIS software. The tool creates a shapefile from an input table containing a unique numerical code representing sampling points and corresponding geographic coordinates. The table can be an ASCII file, dbf or a MICROSOFT EXCEL[®] worksheet. The input coordinate system is defined by the user, and the output shapefile coordinate system is specified.

The analysis steps implemented in the Landscape Genetics GIS Toolbox were originally developed for an ecoregional assessment of genetic divergence and diversity patterns of multiple species, aimed at determining geographic areas of high evolutionary potential (Vander-gast *et al.* 2008). Automation of these analyses should greatly enhance the ability to conduct similar studies, both in terms of increased efficiency (the processes run in

4 COMPUTER PROGRAM NOTE

a matter of seconds on a PC with 2.66 GHz dual core processor) and accuracy (automation of analysis steps reduces chance for user error). The simple graphical interface and formatting requirements for input files reduce the complexity of the analyses and should minimize the amount of GIS experience required to create these genetic landscapes. The Landscape Genetics GIS Toolbox was developed using ModelBuilder™ (an ArcGIS application to create, edit and manage models) and Python scripting in ArcGIS 9.3. ArcMAP™ Spatial analyst and 3D analyst extensions are needed to run the tools. The toolbox, user guide and example files are freely available and can be downloaded as a zip file from the USGS Western Ecological Research Center website at: http://www.werc.usgs.gov/products/MGL_toolbox.

Acknowledgements

We are grateful to Steve Predmore for advice on developing these analyses in ArcGIS. We also thank Mark P. Miller and Kelly R. Barr for testing the tools and for providing insightful comments to improve implementation.

References

- Awise JC (1992) Molecular population structure and the biogeographic history of a regional fauna: a case history with lessons for conservation biology. *Oikos*, **63**, 62–76.
- Cowling RM, Pressey RL (2001) Rapid plant diversification: planning for an evolutionary future. *Proceedings of the National Academy of Sciences of the United States of America*, **98**, 5452–5457.
- Humphries CJ, Williams PH, Vane-Wright RI (1995) Measuring biodiversity value for conservation. *Annual Review of Ecology and Systematics*, **26**, 93–111.
- Manel S, Schwartz MK, Luikart G, Taberlet P (2003) Landscape genetics: combining landscape ecology and population genetics. *Trends in Ecology and Evolution*, **18**, 189–197.
- Manni F, Guérard E, Heyer E (2004) Geographic patterns of (genetic, morphologic, linguistic) variation: how barriers can be detected by using Monmonier's algorithm. *Human Biology*, **76**, 173–190.
- 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.
- Miller MP, Bellinger MR, Forsman ED, Haig SM (2006) Effects of historical climate change, habitat connectivity, and vicariance on genetic structure and diversity across the range of the red tree vole (*Phenacomys longicaudus*) in the Pacific Northwestern United States. *Molecular Ecology*, **15**, 145–159.
- Moritz C, Faith DP (1998) Comparative phylogeography and the identification of genetically divergent areas for conservation. *Molecular Ecology*, **7**, 419–429.
- Prendergast JR, Quinn RM, Lawton JH (1999) The gaps between theory and practice in selecting natural reserves. *Conservation Biology*, **13**, 484–492.
- Riddle BR, Hafner DJ (2006) A step-wise approach to integrating phylogeographic and phylogenetic biogeographic perspectives on the history of a core North American warm deserts biota. *Journal of Arid Environments*, **66**, 435–461.
- Slatkin M (1993) Isolation by distance in equilibrium and non-equilibrium populations. *Evolution*, **47**, 264–279.
- Vandergast AG, Hathaway SA, Fisher RN, Boys J, Bohonak AJ (2008) Are hotspots of evolutionary potential adequately protected in southern California? *Biological Conservation*, **141**, 1648–1664.