Genetic Landscapes GIS Toolbox

Using Model Builder and Python to Create Custom Tools

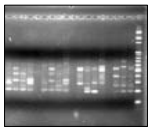
Jerusalem Cricket

Wrentit

William M. Perry [1], Roberto V. Lugo [3], Stacie A. Hathaway [2], Amy G. Vandergast [2]

U.S. Geological Survey, Western Ecological Research Center, Dixon Field Station [1], San Diego Field Station [2]

U.S. Geological Survey, Western Region Geospatial Information Office [3]



An agarose gel image of raw genetic data

Working in the lab

Introduction

A new analysis approach in the field of landscape genetics and phylogeography is the creation of "genetic landscapes" to visualize the distribution of genetic diversity across geographic space. A 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).

The **Genetic Landscapes GIS Toolbox** contains three tools to map genetic landscapes and to summarize multiple genetic landscapes as average and variance surfaces in ArcGIS ® (Environmental Science Research Institute, Redlands, CA, USA). Together, these tools 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 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 (Vandergast *et al.*, 2008). A fourth tool, **Create Shapefile from Table Tool**, is included as a utility.

Methods

- The initial sequence of steps were written out listing the input/output data and the geoprocessing operations.
- Model Builder was used to simultaneously draw the flow chart and actually create the model.
- Work was done in small pieces at a time, testing the processes and verifying the output data.
- New steps were added to previous work to build up the model.
- We used Python for the Multiple Species Genetic Landscape Tool because it was robust for handling map algebra and processing a list of grids.

Process Steps

Early test model

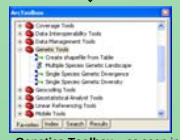
Intermediate data

Black points = collection locations.

Green lines = TIN edges connecting each collection point to its nearest neighbors.

Red points = midpoints between collection locations.

The raster surface is the Inverse Distance Weighted (IDW) interpolation of genetic values.



Genetics Toolbox – as seen in ArcGIS

Which tool to use?

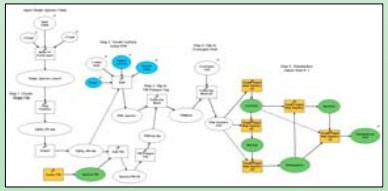
Single Species Genetic Diversity Tool

This tool creates a genetic landscape based on intra-population genetic diversity for a single species or marker. It estimates where the highest levels of genetic diversity are across a species range. The steps listed below are repeated for each species file.

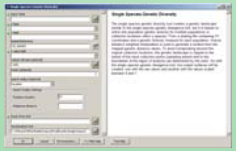
- Create a triangular irregular network (TIN) connecting all collection locations to their nearest neighbors.
- Run the inverse distance weighted (IDW) spatial interpolator to generate a surface from the mapped genetic distance values.
- Clip the raster surface to the sampling extent to avoid extrapolating beyond the original collection locations.
- Clip the raster surface to the boundaries of the region of analysis (as determined by the user).
- Scale the values between 0 and 1 to allow equal weighting of multiple surfaces for inclusion in the Multiple Species Genetic Landscape Tool.

Inputs – start with tables with collection coordinates and genetic values

Inputs – point locations



Model – geoprocessing steps to create the genetic diversity surface



Tool Dialog – presented to the user

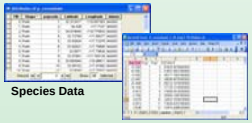
Model Output – genetic diversity surface

Single Species Genetic Divergence Tool

This tool creates a genetic landscape based on pairwise population genetic divergence as measured among multiple collection locations or populations. The steps listed below are repeated for each species file.

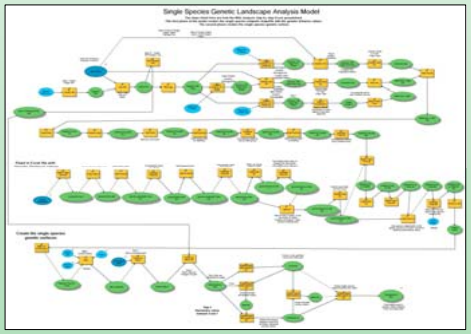
- Create a TIN connecting all collection locations to their nearest neighbors.
- Map the genetic divergence values to the midpoint edges of the TIN.
- Generate a surface using IDW spatial interpolator from the mapped genetic distance values.
- Clip the raster surface to the sampling extent to avoid extrapolating beyond the original collection locations.
- Clip the raster surface to the boundaries of the region of analysis (as determined by the user).
- Scale the values between 0 and 1 to allow equal weighting of multiple surfaces for inclusion in the Multiple Species Genetic Landscape Tool.

Inputs – start with tables with collection coordinates and genetic values



Species Data

Genetic Values



Model – geoprocessing steps to create the genetic divergence surface

Tool Dialog – presented to the user

Model Output – genetic divergence surface

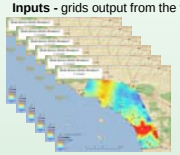
Then

Multiple Species Genetic Landscape Tool

This tool summarizes multiple single species genetic divergence or diversity surfaces with geographic overlap and displays these as a multiple species genetic landscapes.

- Reads in the genetic surfaces from either the diversity or divergence tools.
- Using the raster calculator, the arithmetic mean is calculated.
- Sample variance, which represents the dispersion of the individual species' genetic landscape values from the average multiple species genetic landscape, is also calculated.
- A third output surface is generated that shows the count of input surfaces that overlap in each cell.

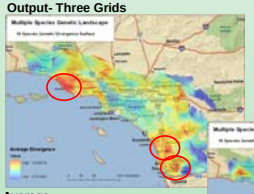
Inputs - grids output from the first two tools



Python Script

Tool Dialog

Output- Three Grids



Average

(Circled areas identify potential evolutionary 'hotspots')

Variance

Number of Overlapping Species Grids

Conclusions/Future Directions

- A GIS approach is useful in evaluating genetic landscapes in association with other georeferenced data, allowing evaluation of geographic barriers to gene flow, and other landscape features associated with patterns of genetic diversity.
- Building custom tools speeds up your work (time went from days to minutes!) and reduces the potential for user error when conducting repetitive, complex workflows.
- We felt Model builder was best for linear geoprocessing steps and is an easy to use visual development environment which allows you to drag and drop tools and data.
- Python is a full scripting language and is better for batch processing, looping and branching, and has add on modules available for increased functionality.
- We plan to add a randomization algorithm to assess confidence in multiple species genetic landscapes.
- The tool box is available for download; **http://www.werc.usgs.gov/products/MGL_toolbox**

Acknowledgments

We would like to thank Steve Predmore, USGS California Water Science Center, for his work on the initial phases of this project.

Literature Cited

Vandergast, A. G., S. A. Hathaway, R. N. Fisher, J. Boys, and A. J. Bohonak. 2008. Are hotspots of evolutionary potential adequately protected in southern California? Biological Conservation **141**:1648-1664.