Lab 22: Biodiverse

Introduction

Biodiverse is a program for spatial analyses of biological diversity that incorporates taxonomic, phylogenetic, and genetic distance relationships. Additionally, the program incorporates environmental and temporal variations into analyses. If you are interested in the following, you should explore Biodiverse:

1. Linked visualization of data distributions in geographic, taxonomic, phylogenetic, and matrix spaces.
2. Analyses for endemism, phylogenetic diversity, and beta diversity
3. Interactive visualizations of turnover patterns such as beta diversity

Today we will be working through a tutorial largely based off of the information found here: https://code.google.com/p/biodiverse/wiki/SampleSession

The goals of the lab today are:

I. Importing Data into Biodiverse
II. Visualizing your Data
III. Running a Cluster Analysis
IV. Running a Spatial Analysis

Exercise 1: Importing Data into Biodiverse

You were given a brief introduction to Biodiverse last week and today we will be working through some sample data so you can see how the program operates. Open the Biodiverse GUI and you will see the first menu has several options at the top "File", "Basedata", "Matrices", "Trees", "Analyses", and "Help".

Basedata

Select Basedata and Import. We'll be working with some example files, which should come pre-installed with your download. There are two generally accepted file formatting for entering into Biodiverse. We will only cover the more common one today, which is a list of one record per observation, for example a list with one line for each species/coordinate combination. Let's take a look at our formatted file before we import the data into Biodiverse so we know what we're working with. Find the file named "example_site_data.csv" and open this in a text editor. Make sure that the format and columns in this file make sense to you.

Find the folder labeled "biodiverse_0.17" → data → example_site_data.csv → Next

Another window will pop up with various options.

Keep these options as they are (options are explained with the online tutorial).

Another window will pop up with various options.

From the drop-down menu, select "Label" for both your genus and species columns.

Select "Group" for your x and y columns. When a Group column has been set you will be given the option to select its cell size (in the same units as the group data is stored) and origin. These default to 100,000.00 and 0.00 respectively so make sure they are appropriate to your data. [Note: I am actually not sure how to decide what is appropriate for your data.].

Select OK.

Matrices

Select Matrices and Import. Open the file named "Example_matrix.txt" and select OK. The next window asks how the columns in the matrix file should be read. You must select at least one label column, and one (only one!) "Matrix start" column. Any other columns in the file should be set to "Ignore". For this example data set we want to "Ignore" Column: 35, "Label" Column: Genus and Column: sp1, and "Matrix Start" Column: -

Select OK.

Trees

A tree must be in nexus format. Select Trees and Import. Open the file named "Example_tree.nex" and select OK.

Remap tree labels → No

Alright all your data should now be loaded into Biodiverse. In the menu bar at the top of the screen you should see "Selected Basedata", "Matrix", and "Tree" now filled in with the example files.

Exercise 2: Visualizing Your Data

Now that you've imported your data into a BaseData object, the first step is to view the data in order to visualize the relationships between the groups and labels. To accomplish this select "Basedata" → "View Labels"

You will now see four different panels. The top left contains two lists of the labels in the selected BaseData object. Each list can be sorted by any of the columns (go back and maybe talk about this more). The bottom left contains a grid of the groups in the current BaseData object. The top right contains a grid of the currently selected matrix elements. The bottom right contains the currently selected tree. The size of the panes can be adjusted by clicking and dragging the dividing bars between them.

The great thing about this interface is that every panel is interactive and what you click on in one panel will be reflected in the others. Click around the interface to see what I mean and explore for yourself.

Question #1: **What happens when you select elements of the upper-left panel? The upper-right? What about the phylogeny in the lower right?**

Exercise 3: Running a Cluster Analysis

Before you run a cluster or spatial analysis (see below), you have the option to exclude labels or groups that you think are redundant, irrelevant, or wrong. We're going to skip this step in today's lab, but the options can be accessed using "Basedata" → "Run Exclusions". See the online documentation for more information.

Cluster analyses are used to identify clusters in the data. Biodiverse supports agglomerative clustering of groups based on their labels or some function of their labels as the values of a linked matrix and tree.

Analyses → Cluster

A new panel will open with two main sections. The upper section determines the parameters used in the clustering to generate a tree. The lower section determines what subsequent calculations will be run for each node in this tree using the groups it contains to define the spatial sample.

Within the "Paremeters" section, you will see a box marked "Spatial Conditions". Spatial conditions are used to define **neighborhoods** of spatial analyses and also for **definition queries** used to constrain the set of groups used in analyses. A neighborhood defines the set of groups to be considered in an analysis. Definition queries assess only the processing group to determine if calculations should be run for it or not. The processing group is the group being considered in the analysis at some iteration, and to which the results of the iteration are assigned. The neighborhood and definition queries are at the core of the functionality of Biodiverse. Your analysis can be as simple or as complex as you'd like and there as there is great system flexibility. Great! So what's the bad news here? To enable this amazing flexibility, Biodiverse uses Perl code. To see several functions, variables, and examples of each of these, visit this page: https://code.google.com/p/biodiverse/wiki/SpatialConditions

You may not know Perl, but you've worked with R enough so that this won't be completely foreign to you. There are many options here and I won't go into them today, but keep in mind that if you have specific questions about your data, these are the options you would use to address them.

The "Calculations" tab is where you select the different calculations to be performed on the data. Click on the plus sign to the left of each set to view the specific options available.

Question #2: **Select 'Taxonomic Dissimilarity and Comparison' in the Calculations section. What analyses are available here?**

Select all the Taxonomic Dissimilarity and Comparison calculations. To see additional details about the indices of what each of these options calculates, press the plus button to the left of each individual option. Press "Go" and when the program asks to display results, select "Yes".

The left pane will display the group grid and the right pane will display the tree representing the clustering. Again, the interactions between the two are linked. You will see a thick blue line to the right of the tree, which you can use to drag across the tree and color the nodes and groups at

certain levels. The colors will turn red when the maximum number of contrasting colors is reached.

Question #3: **What is the maximum number of contrasting colors?**

Towards the bottom of the window, select the drop-down menu that currently says "Cluster" and instead set this to "SPATIAL_RESULTS". You will now be able to select an additional drop-down menu to the right of this. Select "Shannon_H". The Shannon Diversity Index (H) accounts for species abundance and evenness of the species present.

Question #4: **Take a screen shot showing the current data in four clusters for the Shannon Diversity Index.**

Exercise 4: Running a Spatial Analysis

Spatial analyses are used to identify spatial patterns in the data. To run a spatial analysis on the currently selected dataset:

Analyses → Spatial

The window that pops up will look very similar to what you saw during the Clustering Analysis. In the Calculations menu, check off the box for the Phylogenetic Indices. Press "Go" and when the program asks to display results, select "Yes".

[Note: If your spatial analyses take a very long time there is an option to create a Spatial Index in Biodiverse, which will apparently increase the speed. Again, I will not be covering these options today, but wanted to mention them in case you need them in the future.]

The results are displayed as a map in a single pane. You can change the color from Hue to Sat if you would like a single color scheme version. In the pull-down menu next to Spatial Results, find the Phylogenetic Distance measure (PD). Examine the results for PD and for some of the other measures (you may have to look up what they mean back in the table).

Question #5: **Examine the map and take a screen shot. Given your current data, if you were to convince a policy-maker about certain areas to protect for conservation efforts, what would you say?**