

## Lab 14:

### Spatial phylogenetics with BIODIVERSE

By *Ischel González-Ramírez*

## 1 Before you begin

Today's lab will require the software BIODIVERSE. Please install it from <http://shawnlaffan.github.io/biodiverse/>.

## 2 Introduction

### 2.1 Phylogenetic diversity in space

As we have been discussing in lecture, phylogenetic diversity (PD) is an important component of biodiversity. It has been increasingly used as a conservation-relevant metric since it is thought to correlate with morphological and functional diversity. By prioritizing the conservation of phylogenetic diverse areas, we are protecting non-redundant species with complementary biological traits. From the evolutionary perspective, the study of phylogenetic diversity in space allows to identify diversity patterns and to make inferences about the processes that gave rise to them. For example, a clustered pattern of relative PD [more closely related species co-occurring than what you would expect by random assembly] is congruent with a recent diversification event. Alternatively, an over-dispersed pattern of PD [distant-related species co-occurring than what you would expect by random assembly]

### 2.2 Mexican cool beans

In this lab we are going to use an empirical case to walk through the implementation of a spatial phylogenetic analysis in BIODIVERSE and exemplify the kind of questions we can explore with this tool. We will focus on a subset of a recent study on the phylogenetic diversity of vascular plants of Mexico by Sosa et al. (2018). Mexico is a very diverse country, hosting more than 26,000 species of vascular plants with a very incomplete diversity inventory. Fabaceae is the family of beans and it is the second most diverse family of plants in this country with ~1900 spp. Fabaceae species occur in different types of vegetation and display an amazing morphological diversity. Let's explore the diversity patterns of this cool plants.

## 3 The data

Decompress the folder with the data for this lab. Open the *fabaceae\_mex.csv*. Open the tree file *fabaceae.tre* in a text editor and then in FigTree. Explore the data and answer the following question.

### Question 1

- What information contains the file *fabaceae\_mex.csv*? What does each record represent and how many records are there?

- What format type is the tree written in? When you open the tree in FigTree do you notice something in the topology? Hint: is it likely that this tree is the output of a phylogenetic inference analysis?

## 4 BIODIVERSE

Biodiverse is an open source tool for the spatial analysis of diversity using indices based on taxonomic, phylogenetic, trait and matrix-based (e.g. genetic distance) relationships, as well as related environmental and temporal variations (Laffan et al., 2010). This software has a lot of functions and I encourage you to explore it if this is something that interests you <http://shawnlaffan.github.io/biodiverse/>.

Open BIODIVERSE software. On the top of the window, you will observe three main object-type menus : *Basedata*, *Matrix*, *Tree*. In this lab we won't use the *Matrix* type, but it is very useful when you need to clean databases or establish relation between different data sets (For example, if you want to establish a relationship between names in the tip of your branches and names in your geographic data set).

### 4.1 Loading the data

#### 4.1.1 Basedata

Click on the "plus" sign in the *Basedata* menu and load the `fabaceae_mex.csv` file.

- Click next.
- It will ask you what type of data is each column. We want to use the **species name as a label**. The latitude and longitude will allow us not only to map the observations but also to assign them to grid cells based on their position. Therefore, for Latitude and Longitude columns we will select the type **group** and specify a cell size of **0.5 degrees**, so we obtain a long-lat grid of 0.5 by 0.5 degrees. Don't forget to specify "is\_lat", "is\_long" as corresponds. Yo can ignore the rest of the columns.
- In the next window make sure you order Longitud as X axis and Latitud as Y axis.
- We don't want to Remap the labels.

The file is uploaded and it appears now in your screen. On the right panel clic on **Show** and explore the new window.

#### Question 2

- Explain the meaning of the metrics "Variety", "Samples" and "Redundancy"
- What does having a Redundancy = 1 mean for a taxon?

### 4.1.2 Tree data

Back on the top menu, click on the "plus sign" of the tree objects. We will import our tree file, so select **No** for the tabular tree format question. Select the file containing the tree and open it. We don't want to Remap the node names. Now, on the top tree menu, your tree have appeared as an option.

Having geographic and phylogenetic data, we can start asking questions about what areas of Mexico have a greater phylogenetic diversity.

#### Question 3

Click on several of the grid cells and observe what happens.

- How do you interpret the red color on the map?
- What does the red color means in the phylogeny?
- Reminding that we want to assess phylogenetic diversity patterns of Fabaceae, can you identify a geographic pattern by playing with this platform?

### 4.2 The analysis

As you might have noticed, it is hard to infer general patterns of phylogenetic diversity distribution from the map. BIODIVERSE has more than 300 indexes implemented for community phylogenetics and spatial phylogenetic analyses. You can explore more details about the indexes here: <https://github.com/shawnlaffan/biodiverse/wiki/Indices>. We will focus on the indexes we discussed in lecture.

On the very top menu, click on *Analyses* and select *Spatial* and select the following indexes:

- Under **Lists and counts**: Richness
- Under **Phylogenetic Indices**: Phylogenetic Diversity.
- Under **Phylogenetic Indices (relative)**: Relative Phylogenetic Diversity
- Under **Phylogenetic Endemism Indices**: Corrected wighted phylogenetic endemism.

And click on **Go!**

### 4.3 The results

Explore the results by changing the index that is being displayed. Can you identify spatial patterns on the diversity of this group? Do you notice anything strange that you would like to double check?

#### Question 4

Briefly summarize your results. Contrast the results of the different metrics. If you were to define a conservation area that aim to preserve Fabaceae biodiversity, what area would you pick based on these results?

## 5 Exercise

Re-run the analysis, but this time pick a finer grid cell. You can also choose different metrics in the analysis. Pick the ones that you consider are more valuable for conservation decisions.

1. Does the patterns you observed before hold in this new analysis?
2. Imagine you are a decision taker and choose of a new protected area that aim to preserve Fabaceae biodiversity based on your new analysis. Write a short paragraph providing **biological arguments** based on your metrics to justify your choice.

**Please email me the following:**

1. Your answers of questions 1-4.
2. Your exercise answers with screenshots of your results.

## References

Laffan, S.W., Lubarsky, E. Rosauer, D.F. (2010) Biodiverse, a tool for the spatial analysis of biological and related diversity. *Ecography*. Vol 33, 643-647

Sosa, V., De-Nova, J. A. and Vásquez-Cruz, M. (2018) 'Evolutionary history of the flora of Mexico: Dry forests cradles and museums of endemism', *Journal of Systematics and Evolution*, 56(5), pp. 523–536. doi: 10.1111/jse.12416.