

Lab 14

Jennifer Lin

Question 1

1. What information contains the file fabaceae_mex.csv? What does each record represent and how many records are there?

There are 5 columns and 59067 rows in total (including the header).

The 5 columns are:

1. Family: The family name of the organism
2. Genus: The genus name of the organism
3. species: The species name of the organism
4. Latitude: The location of that organism
5. Longitude: The location of that organism

There are $59067 - 1$ (header line) = 59066 records.

2. 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?

The format of the tree file is Nexus. I noticed that there are a lot of branches look like consensus trees. In addition to the fact that the format is Nexus format, I think this phylogeny is an output from MrBayes.

Question 2

1. Explain the meaning of the metrics "Variety", "Samples" and "Redundancy"

The **Variety** column shows the number of groups the label occurs in.

The **Samples** column lists the number of times this label occurs across all groups.

The **Redundancy** column shows the sample redundancy for each label.

2. What does having a Redundancy = 1 mean for a taxon?

Redundancy is calculated as $(1 - \text{variety} / \text{samples})$. A value close to one represents a good overall sample of a label relative to the number of groups it occurs in (many redundant samples). On the other hand, a value of zero means that there is only one sample per group the label occurs in, and it is therefore not well sampled.

Question 3

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

1. How do you interpret the red color on the map?

When we click a grid on the map, it will highlight it in dark red. Other groups may also be highlighted depending on the number of labels they share with the selected group (darker red for more shared labels, lighter red for less, white for no labels in common).

2. What does the red color means in the phylogeny?

The species that is contained in that grid will be marked as red in the phylogenetic tree.

3. Reminding that we want to assess phylogenetic diversity patterns of Fabaceae, can you identify a geographic pattern by playing with this platform?

Most of the near grids that near to each other shared similar species contents. It also looks like similar habitats may contain similar species.

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?

Some of the indices show clear patterns that some area has higher diversity according to that index. Some of the other indices just show plain green or blue results, which may not help us to define a conservation area. According to PD_P_per_taxon, PD_per_taxon and PHYLO_RPD_1, I think the northern part or the eastern part of Mexico generally have higher biodiversity. Thus, I would define a conservation area in those places.

In PD and PD_P, most of the grids are in green or blue. However, there is one single grid is marked as red in the middle. I think there might be some data collection issues within that grid.

In PD_P_per_taxon and PD_per_taxon, most of the red grids are in the northern part or the eastern part of Mexico.

In PE_CWE, there's also only a single grid with red color while other grids are mostly green or blue. Nevertheless, I think this is more acceptable since there are at least some grids marked in orange near the red grid.

In PHYLO_RPD_1, most of the red grids are in the northern part or the eastern part of Mexico.

In PHYLO_RPD_DIFF1, PHYLO_RPD_NULL1, RICHNESS_ALL, and RICHNESS_SET1, most of the grids are in green or blue. However, there

is one single grid is marked as red in the middle. I think there might be some data collection issues within that grid.

In RICHNESS_SET2, most of the grids are just marked in blue.

Question 5

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?

I think most of the patterns still hold in this new analysis. In some indices, there are still a single red grid in the middle of the map while other grids are all green or blue. For other indices, the general patterns are maintained.

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.

When choosing a new protected area, we should make sure we use our limited budget on a place with highest biodiversity. In my analysis, the eastern part of Mexico shows high biodiversity in plenty indices. Thus, establishing a conserved area in that region would benefit the most.



