

Phylogenetic Diversity - Communities

Student Name; Z620: Quantitative Biodiversity, Indiana University

23 February, 2017

OVERVIEW

Complementing taxonomic measures of α - and β -diversity with evolutionary information yields insight into a broad range of biodiversity issues including conservation, biogeography, and community assembly. In this assignment, you will be introduced to some commonly used methods in phylogenetic community ecology.

After completing this assignment you will know how to:

1. incorporate an evolutionary perspective into your understanding of community ecology
2. quantify and interpret phylogenetic α - and β -diversity
3. evaluate the contribution of phylogeny to spatial patterns of biodiversity

Directions:

1. Change “Student Name” on line 3 (above) with your name.
2. Complete as much of the assignment as possible during class; what you do not complete in class will need to be done outside of class.
3. Use the handout as a guide; it contains a more complete description of data sets along with the proper scripting needed to carry out the exercise.
4. Be sure to **answer the questions** in this exercise document; they also correspond to the handout. Space for your answer is provided in this document and indicated by the “>” character. If you need a second paragraph be sure to start the first line with “>”.
5. Before you leave the classroom, **push** this file to your GitHub repo.
6. When you are done, **Knit** the text and code into a PDF file.
7. After Knitting, please submit the completed assignment by creating a **pull request** via GitHub. Your pull request should include this file *PhyloCom_assignment.Rmd* and the PDF output of **Knitr** (*PhyloCom_assignment.pdf*).

1) SETUP

Typically, the first thing you will do in either an R script or an RMarkdown file is setup your environment. This includes things such as setting the working directory and loading any packages that you will need.

In the R code chunk below, provide the code to:

1. clear your R environment,
2. print your current working directory,
3. set your working directory to your **/Week7-PhyloCom** folder,
4. load all of the required R packages (be sure to install if needed), and
5. load the required R source file.

2) DESCRIPTION OF DATA

We will revisit the data that was used in the Spatial Diversity module. As a reminder, in 2013 we sampled ~ 50 forested ponds located in Brown County State Park, Yellowwood State Park, and Hoosier National Forest in southern Indiana. See the handout for a further description of this week’s dataset.

3) LOAD THE DATA

In the R code chunk below, do the following:

1. load the environmental data for the Brown County ponds (*20130801_PondDataMod.csv*),
2. load the site-by-species matrix using the `read.otu()` function,
3. subset the data to include only DNA-based identifications of bacteria,
4. rename the sites by removing extra characters,
5. remove unnecessary OTUs in the site-by-species, and
6. load the taxonomic data using the `read.tax()` function.

Next, in the R code chunk below, do the following:

1. load the FASTA alignment for the OTUs,
2. rename the OTUs by removing everything before the tab (`\t`) and after the bar (`|`),
3. import the *Methanosarcina* outgroup FASTA file,
4. convert both FASTA files into the DNAbin format and combine using `rbind()`,
5. visualize the sequence alignment,
6. using the alignment (with outgroup), pick a DNA substitution model and create a distance matrix,
7. using the distance matrix above, make a neighbor joining tree,
8. remove any tips (OTUs) that are not in the community data set,
9. plot the rooted tree.

4) PHYLOGENETIC ALPHA DIVERSITY

A. Faith's Phylogenetic Diversity (PD)

In the R code chunk below, do the following:

1. calculate Faith's D using the `pd()` function.

In the R code chunk below, do the following:

1. plot species richness (S) versus phylogenetic diversity (PD),
2. add the trend line, and
3. calculate the scaling exponent.

Question 1: Answer the following questions about the PD-S pattern.

- a. Describe the relationship between taxonomic richness and phylodiversity?
- b. Based on what you know about how PD is calculated, why should they be correlated?
- c. When would you expect these two estimates of diversity to deviate from one another?

Answer 1a:

Answer 1b:

Answer 1c:

i. Randomizations and Null Models

In the R code chunk below, do the following:

1. estimate the standardized effect size of PD using the `richness` randomization method.

Question 2: Using `help()` and the table above, run the `ses.pd()` function using two other null models and answer the following questions:

- a. What are the null and alternative hypotheses you are testing via randomization when calculating `ses.pd`?
- b. How did your choice of null model influence your observed `ses.pd` values? Briefly, explain why this happens.

Answer 2a:

Answer 2b:

B. Phylogenetic Dispersion Within a Sample

Another way to assess phylogenetic α -diversity is to look at dispersion within a sample.

i. Phylogenetic Resemblance Matrix

In the R code chunk below, do the following:

1. calculate the phylogenetic resemblance matrix for taxa in the Indiana ponds data set.

ii. Net Relatedness Index (NRI)

In the R code chunk below, do the following:

1. Calculate the NRI for each site in the Indiana ponds data set.

iii. Nearest Taxon Index (NTI)

In the R code chunk below, do the following: 1. Calculate the NTI for each site in the Indiana ponds data set.

Question 3:

- a. In your own words describe what you are doing when you calculate the NRI.
- b. In your own words describe what you are doing when you calculate the NTI.
- c. Interpret the NRI and NTI values you observed for this dataset.
- d. In the NRI and NTI examples above, the arguments “abundance.weighted = FALSE” means that the indices were calculated using presence-absence data. Modify and rerun the code so that NRI and NTI are calculated using abundance data. How does this affect the interpretation of NRI and NTI?

Answer 3a:

Answer 3b:

Answer 3c:

Answer 3d:

5) PHYLOGENETIC BETA DIVERSITY

A. Phylogenetically Based Community Resemblance Matrix

In the R code chunk below, do the following:

1. calculate the phylogenetically based community resemblance matrix using Mean Pair Distance, and
2. calculate the phylogenetically based community resemblance matrix using UniFrac distance.

In the R code chunk below, do the following:

1. plot mean pair distance versus UniFrac distance and compare.

Question 4:

- a. In your own words describe Mean Pair Distance, UniFrac distance, and the difference between them.
- b. Using the plot above, describe the relationship between Mean Pair Distance and UniFrac distance.
Note: we are calculating unweighted phylogenetic distances (similar to incidence based measures). That means that we are not taking into account the abundance of each taxa in each site.
- c. Why might MPD show less variation than UniFrac?

Answer 4a: Answer 4b: Answer 4c:

B. Visualizing Phylogenetic Beta-Diversity

Now that we have our phylogenetically based community resemblance matrix, we can visualize phylogenetic diversity among samples using the same techniques that we used in the β -diversity module.

In the R code chunk below, do the following:

1. perform a PCoA based on the UniFrac distances, and
2. calculate the explained variation for the first three PCoA axes.

Now that we have calculated our PCoA, we can plot the results.

In the R code chunk below, do the following:

1. plot the PCoA results,
2. include the appropriate axes,
3. add and label the points, and
4. customize the plot.

In the following R code chunk: 1. perform another PCoA on taxonomic data using an appropriate measure of dissimilarity, and 2. calculate the explained variation on the first three PCoA axes.

Question 5: Using a combination of visualization tools and percent variation explained, how does the phylogenetically based ordination compare or contrast with the taxonomic ordination?

Answer 5:

C. Hypothesis Testing

i. Categorical Approach

In the R code chunk below, do the following:

1. test the hypothesis that watershed has an effect on the phylogenetic diversity of bacterial communities.

ii. Continuous Approach

In the R code chunk below, do the following: 1. from the environmental data matrix, subset the variables related to physical and chemical properties of the ponds, and

2. calculate environmental distance between ponds based on the Euclidean distance between sites in the environmental data matrix (after transforming and centering using `scale()`).

In the R code chunk below, do the following:

1. conduct a Mantel test to evaluate whether or not UniFrac distance is correlated with environmental variation.

Last, let's conduct a distance-based Redundancy Analysis (dbRDA).

In the R code chunk below, do the following:

1. conduct a dbRDA to test the hypothesis that environmental variation effects the phylogenetic diversity of bacterial communities,
2. use a permutation test to determine significance, and 3. plot the dbRDA results

Question 6: Based on the multivariate procedures conducted above, describe the phylogenetic patterns of β -diversity for bacterial communities in the Indiana ponds.

Answer 6:

6) SPATIAL PHYLOGENETIC COMMUNITY ECOLOGY

A. Phylogenetic Distance-Decay (PDD)

First, we will calculate distances for geographic data, taxonomic data, and phylogenetic data among all unique pair-wise combinations of ponds.

In the R code chunk below, do the following:

1. calculate the geographic distances among ponds,

2. calculate the taxonomic similarity among ponds,
3. calculate the phylogenetic similarity among ponds, and
4. create a dataframe that includes all of the above information.

Now, let's plot the DD relationships:

In the R code chunk below, do the following:

1. plot the taxonomic distance decay relationship,
2. plot the phylogenetic distance decay relationship, and
3. add trend lines to each.

In the R code chunk below, do the following:

1. test if the trend lines in the above distance decay relationships are different from one another.

Question 7: Interpret the slopes from the taxonomic and phylogenetic DD relationships. If there are differences, hypothesize why this might be.

Answer 7:

B. Phylogenetic diversity-area relationship (PDAR)

i. Constructing the PDAR

In the R code chunk below, do the following:

1. write a function to generate the PDAR.

ii. Evaluating the PDAR

In the R code chunk below, do the following:

1. calculate the area for each pond,
2. use the PDAR() function you just created to calculate the PDAR for each pond,
3. calculate the Pearson's and Spearman's correlation coefficients,
4. plot the PDAR and include the correlation coefficients in the legend, and
5. customize the PDAR plot.

Question 8: Compare your observations of the microbial PDAR and SAR in the Indiana ponds? How might you explain the differences between the taxonomic (SAR) and phylogenetic (PDAR)?

Answer 8:

SYNTHESIS

Ignoring technical or methodological constraints, discuss how phylogenetic information could be useful in your own research. Specifically, what kinds of phylogenetic data would you need? How could you use it to answer important questions in your field. In your response, feel free to consider not only phylogenetic approaches related to phylogenetic community ecology, but also those we discussed last week, or any other concepts that we have not covered in this course.