9.Phylogenetic Diversity - Communities

Atalanta Ritter; Z620: Quantitative Biodiversity, Indiana University

27 February, 2023

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 worksheet, 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. In the Markdown version of this document in your cloned repo, change "Student Name" on line 3 (above) with your name.
2. Complete as much of the worksheet as possible during class.
3. Use the handout as a guide; it contains a more complete description of data sets along with examples of proper scripting needed to carry out the exercises.
4. Answer questions in the worksheet. Space for your answers is provided in this document and is indicated by the ">" character. If you need a second paragraph be sure to start the first line with ">". You should notice that the answer is highlighted in green by RStudio (color may vary if you changed the editor theme).
5. Before you leave the classroom today, it is *imperative* that you **push** this file to your GitHub repo, at whatever stage you are. This will enable you to pull your work onto your own computer.
6. When you have completed the worksheet, **Knit** the text and code into a single PDF file by pressing the `Knit` button in the RStudio scripting panel. This will save the PDF output in your '9.PhyloCom' folder.
7. After Knitting, please submit the worksheet by making a **push** to your GitHub repo and then create a **pull request** via GitHub. Your pull request should include this file *9.PhyloCom_Worksheet.Rmd* and the PDF output of `Knitr` (*9.PhyloCom_Worksheet.pdf*).

The completed exercise is due on **Wednesday, March 1st, 2023 before 12:00 PM (noon)**.

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 `/9.PhyloCom` folder,
4. load all of the required R packages (be sure to install if needed), and
5. load the required R source file.

```
rm(list = ls())
getwd()
```

```
## [1] "/Users/Atalanta/GitHub/QB2023_Ritter/2.Worksheets/9.PhyloCom"
```

```
setwd("~/GitHub/QB2023_Ritter/2.Worksheets/9.PhyloCom")
package.list <- c('picante', 'ape', 'seqinr', 'vegan', 'fossil',
'reshape', 'devtools', 'BiocManager', 'ineq',
'labdsv', 'matrixStats', 'pROC')
for (package in package.list) {
  if (!require(package, character.only = TRUE, quietly = TRUE)) {
    install.packages(package, repos='http://cran.us.r-project.org')
    library(package, character.only = TRUE)
  }
}
```

```
## This is vegan 2.6-4
```

```
##
## Attaching package: 'seqinr'
```

```
## The following object is masked from 'package:nlme':
##
##     gls
```

```
## The following object is masked from 'package:permute':
##
##     getType
```

```
## The following objects are masked from 'package:ape':
##
##     as.alignment, consensus
```

```
##
## Attaching package: 'shapefiles'
```

```
## The following objects are masked from 'package:foreign':
##
##     read.dbf, write.dbf
```

```
##
## Attaching package: 'devtools'
```

```
## The following object is masked from 'package:permute':
##
##     check
```

```
## Bioconductor version '3.14' is out-of-date; the current release version '3.16'
##   is available with R version '4.2'; see https://bioconductor.org/install
```

```
##
## Attaching package: 'BiocManager'
```

```
## The following object is masked from 'package:devtools':
##
##     install
```

```
## This is mgcv 1.8-41. For overview type 'help("mgcv-package")'.
```

```
## Registered S3 method overwritten by 'labdsv':
##   method       from
```

```
##   summary.dist ade4

## This is labdsv 2.0-1
## convert existing ordinations with as.dsvord()

##
## Attaching package: 'labdsv'

## The following object is masked from 'package:stats':
##
##     density

##
## Attaching package: 'matrixStats'

## The following object is masked from 'package:seqinr':
##
##     count

## Type 'citation("pROC")' for a citation.

##
## Attaching package: 'pROC'

## The following objects are masked from 'package:stats':
##
##     cov, smooth, var
```

```
source("./bin/MothurTools.R")
```

2) DESCRIPTION OF DATA

need to discuss data set from spatial ecology!

We sampled >50 forested ponds in Brown County State Park, Yellowood State Park, and Hoosier National Forest in southern Indiana. In addition to measuring a suite of geographic and environmental variables, we characterized the diversity of bacteria in the ponds using molecular-based approaches. Specifically, we amplified the 16S rRNA gene (i.e., the DNA sequence) and 16S rRNA transcripts (i.e., the RNA transcript of the gene) of bacteria. We used a program called `mothur` to quality-trim our data set and assign sequences to operational taxonomic units (OTUs), which resulted in a site-by-OTU matrix.
In this module we will focus on taxa that were present (i.e., DNA), but there will be a few steps where we need to parse out the transcript (i.e., RNA) samples. 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 from the source-code file.

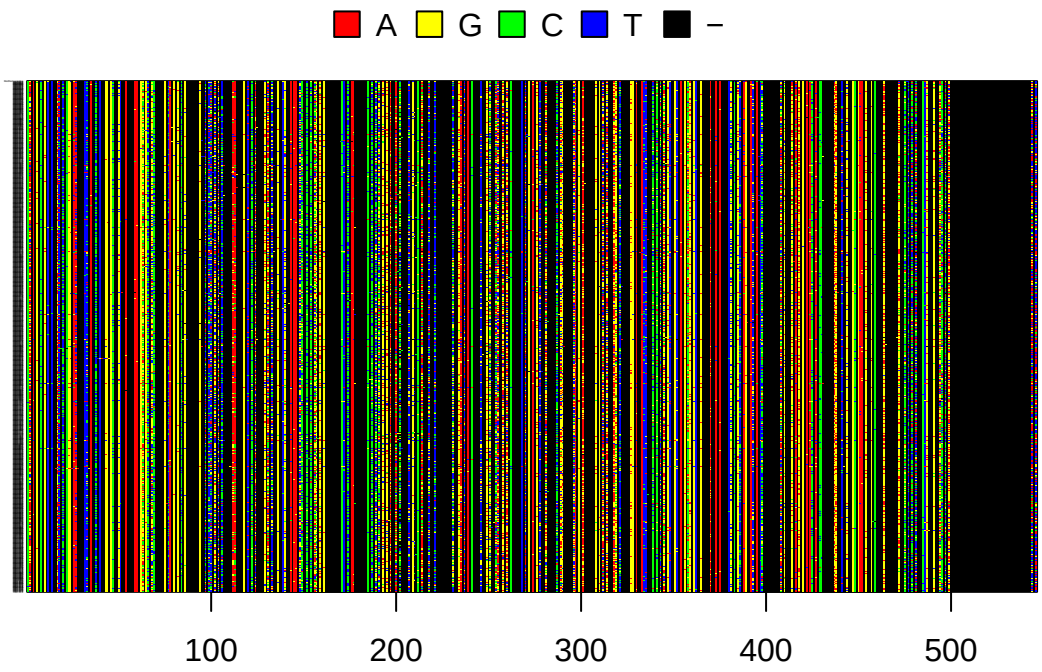
```
env <- read.table("data/20130801_PondDataMod.csv", sep = ",", header = TRUE)
env <- na.omit(env)
# Load Site-by-Species Matrix
comm <- read.otu(shared = "./data/INPonds.final.rdp.shared", cutoff = "1")
# Select DNA data using `grep()`
```

```
comm <- comm[grep("*-DNA", rownames(comm)), ]
# Perform replacement of all matches with `gsub()`
rownames(comm) <- gsub("\\-DNA", "", rownames(comm))
rownames(comm) <- gsub("\\_", "", rownames(comm))
# Remove sites not in the environmental data set
comm <- comm[rownames(comm) %in% env$Sample_ID, ]
# Remove zero-abundance OTUs from data set
comm <- comm[ , colSums(comm) > 0]
# load the taxonomic data using the `read.tax()` function
tax <- read.tax(taxonomy = "./data/INPonds.final.rdp.1.cons.taxonomy")
```

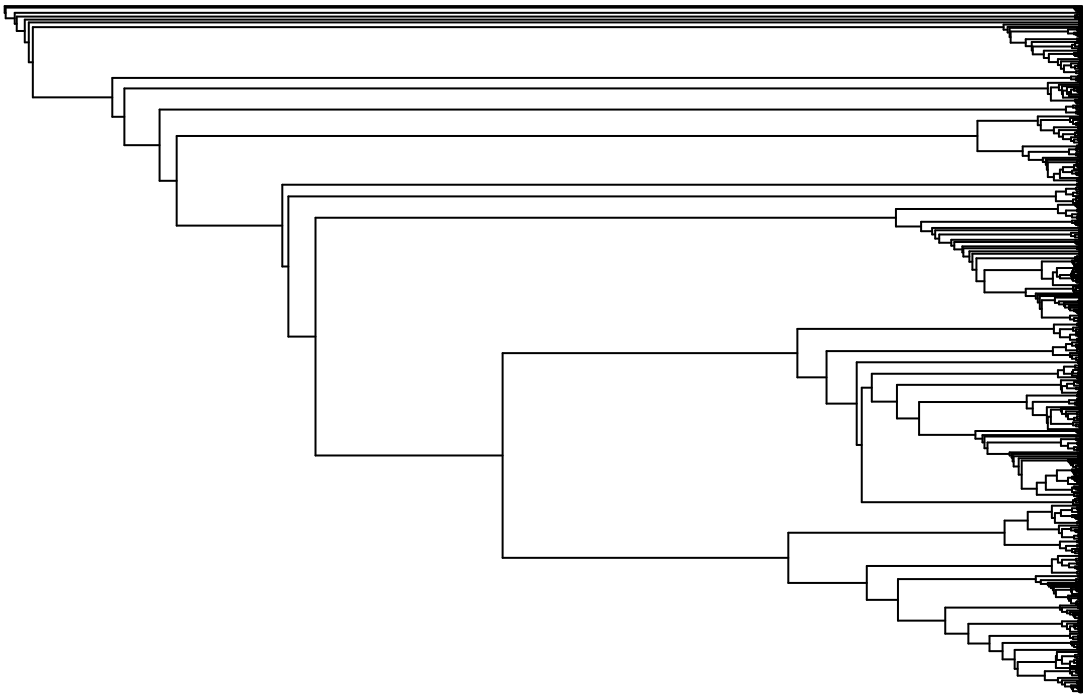
```
## Warning in type.convert.default(as.character(x)): 'as.is' should be specified
## by the caller; using TRUE

## Warning in type.convert.default(as.character(x)): 'as.is' should be specified
## by the caller; using TRUE

## Warning in type.convert.default(as.character(x)): 'as.is' should be specified
## by the caller; using TRUE

## Warning in type.convert.default(as.character(x)): 'as.is' should be specified
## by the caller; using TRUE

## Warning in type.convert.default(as.character(x)): 'as.is' should be specified
## by the caller; using TRUE

## Warning in type.convert.default(as.character(x)): 'as.is' should be specified
## by the caller; using TRUE
```

Next, in the R code chunk below, do the following:
1. load the FASTA alignment for the bacterial operational taxonomic units (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 phylogenetic 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.

```
# Import the alignment file (`seqinr`)
ponds.cons <- read.alignment(file = "./data/INPonds.final.rdp.1.rep.fasta",
format = "fasta")
# Rename OTUs in the FASTA File
ponds.cons$nam <- gsub(".*\t", "", ponds.cons$nam)
ponds.cons$nam <- gsub("\\|.*", "", ponds.cons$nam)
# Import outgroup sequence
outgroup <- read.alignment(file = "./data/methanosarcina.fasta", format = "fasta")
# Convert alignment file to DNAbin
DNAbin <- rbind(as.DNAbin(outgroup),as.DNAbin(ponds.cons))
# Visualize alignment
image.DNAbin(DNAbin, show.labels = T, cex.lab = 0.05, las = 1)
```




```
# Make distance matrix (`ape`)
seq.dist.jc <- dist.dna(DNAbin, model = "JC", pairwise.deletion = FALSE)
# Make a neigbor-joining tree file (`ape`)
phy.all <- bionj(seq.dist.jc)
# Drop tips of zero-occurrence OTUs (`ape`)
phy <- drop.tip(phy.all, phy.all$tip.label[!phy.all$tip.label %in%
                c(colnames(comm), "Methanosarcina")])
# Identify outgroup sequence
outgroup <- match("Methanosarcina", phy$tip.label)
# Root the tree {ape}
phy <- root(phy, outgroup, resolve.root = TRUE)
# Plot the rooted tree {ape}
par(mar = c(1, 1, 2, 1) + 0.1)
plot.phylo(phy, main = "Neighbor Joining Tree", "phylogram",
           show.tip.label = FALSE, use.edge.length = FALSE,
           direction = "right", cex = 0.6, label.offset = 1)
```



4) PHYLOGENETIC ALPHA DIVERSITY

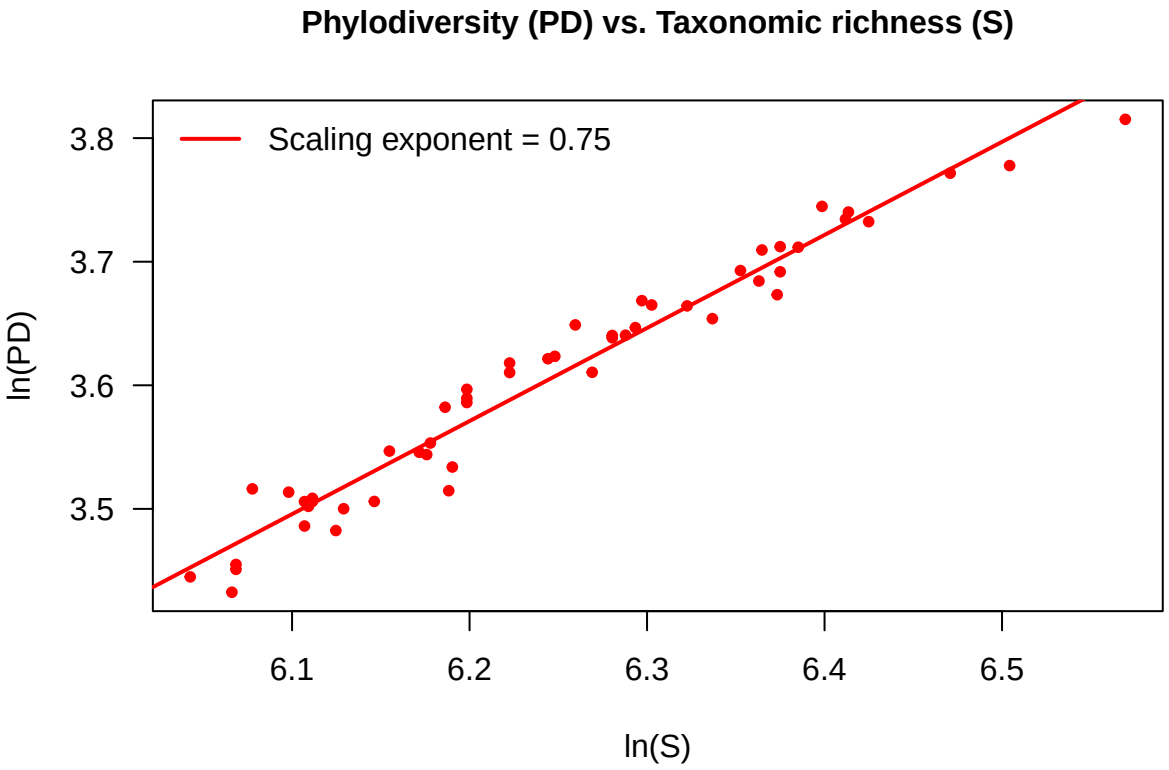
A. Faith's Phylogenetic Diversity (PD)

In the R code chunk below, do the following:
1. calculate Faith's D using the `pd()` function.

```
# Calculate PD and S {picante}
pd <- pd(comm, phy, include.root = FALSE)
```

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.

```
# Biplot of S and PD
par(mar = c(5, 5, 4, 1) + 0.1)
plot(log(pd$S), log(pd$PD),
    pch = 20, col = "red", las = 1,
    xlab = "ln(S)", ylab = "ln(PD)", cex.main = 1,
    main="Phylodiversity (PD) vs. Taxonomic richness (S)")
# Test of power-law relationship
fit <- lm('log(pd$PD) ~ log(pd$S)')
abline(fit, col = "red", lw = 2)
exponent <- round(coefficients(fit)[2], 2)
legend("topleft", legend=paste("Scaling exponent = ", exponent, sep = ""),
        bty = "n", lw = 2, col = "red")
```



Question 1: Answer the following questions about the PD-S pattern.
a. Based on how PD is calculated, why should this metric be related to taxonomic richness? b. Describe the relationship between taxonomic richness and phylodiversity. c. When would you expect these two estimates of diversity to deviate from one another? d. Interpret the significance of the scaling PD-S scaling exponent.

> ***Answer 1a***: PD has to calculate the number of species in an assemblage, aka richness, in order to determine which ones are present/absent.
> ***Answer 1b***: There is a strong positive relationship; as taxonomic richness increases, phylodiversity increases as well. ***Answer 1c***: I would expect deviation if, for instance, there was a new environment that was colonized by a single ancestral species that later branched out into a huge adaptive radiation. In that case, there would be high taxonomic richness but low phylodiversity.
> ***Answer 1d***: Because the exponent is less than 1, this means that phylodiversity increases slower than taxonomic richness. This makes intuitive sense because every time you find new species, it is not guaranteed that all of them will be phylogenetically distinct from each other.

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.

```
# Estimate standardized effect size of PD via randomization (`picante`)
ses.pd <- ses.pd(comm[1:2,], phy, null.model = "richness", runs = 25,
                 include.root = FALSE)
# null model "frequency"
ses.pd.freq <- ses.pd(comm[1:2,], phy, null.model = "frequency", runs = 25,
                 include.root = FALSE)
# null model "sample pool"
ses.pd.sampool <- ses.pd(comm[1:2,], phy, null.model = "sample.pool",
                         runs = 25, include.root = FALSE)
```

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? Explain why this choice affected or did not affect the output.

Answer 2a: I'm testing the null hypotheses that 1) species richness is the same across all communities (richness), 2) species occurrence frequency is the same across all communities (frequency), and 3) the distribution of species in communities is such that every species is equally likely to occur in the community (sample.pool). The alternative hypotheses are that 1) species richness is not the same across communities, 2) species occurrence frequency is not the same across communities, and 3) the distribution of species in communities is uneven, such that there is an unequal chance of some species being part of the community, respectively. ***Answer 2b***: The kind of null model did not seem to affect the results very much. The sample pool and richness models gave pretty much the exact same output, which made sense because species richness defines the sample pool. The frequency model gave a slightly different output, probably because it is accounting for species occurrence, so the difference between rare vs. common species may have an effect.

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.

```
# Create a Phylogenetic Distance Matrix (`picante`)
phydist <- cophenetic.phylo(phy)
```

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.

```
# Estimate standardized effect size of NRI via randomization (`picante`)
ses.mpd <- ses.mpd(comm, phydist, null.model = "taxa.labels",
abundance.weighted = TRUE, runs = 25)
# Calculate NRI
NRI <- as.matrix(-1 * ((ses.mpd[,2] - ses.mpd[,3]) / ses.mpd[,4]))
rownames(NRI) <- row.names(ses.mpd)
colnames(NRI) <- "NRI"
```

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.

```
# Estimate Standardized Effect Size of NRI via Randomization {picante}
ses.mntd <- ses.mntd(comm, phydist, null.model = "taxa.labels",
                     abundance.weighted = TRUE, runs = 25)
# Calculate NTI
NTI <- as.matrix(-1 * ((ses.mntd[,2] - ses.mntd[,3]) / ses.mntd[,4]))
rownames(NTI) <- row.names(ses.mntd)
colnames(NTI) <- "NTI"
```

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: NRI calculates the average phylogenetic distance between taxa in your ssample, then uses the difference between what your observed values and the average to determine whether your sample is over- or underdispersed based on if it falls above or below your null expectation. ***Answer 3b***: NTI uses the avaerage phylogenetic distance between ALL taxa in your sample and their most closely related "neighbor" (sister group?), and uses that to determine if your sample fall above or below the null expectation. ***Answer 3c***: All the NRI values are negative, so communities tend to be overdispersed. NTI values are less consistent, with some samples being positive and some negative. ***Answer 3d***: This changed the NTI values a lot; now all are positive, suggestive of phylogenetic clustering. It also made the NRI values less consistent.

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.

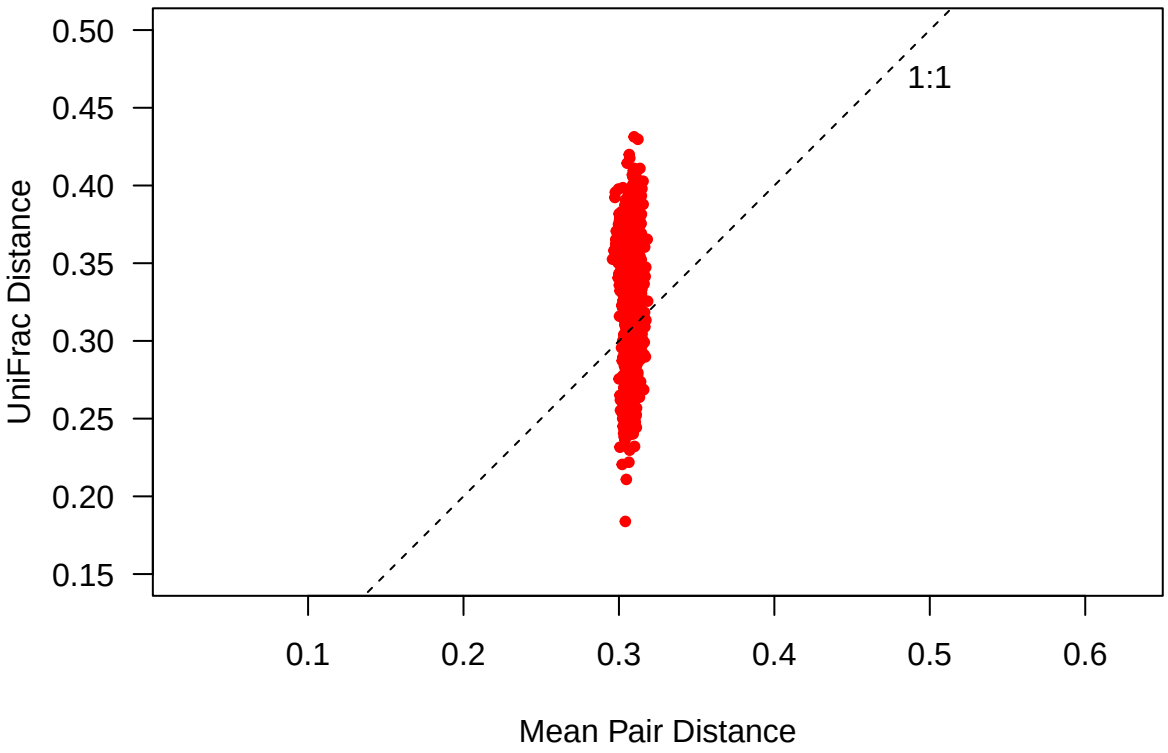
```
# Mean Pairwise Distance
dist.mp <- comdist(comm, phydist)
```

```
## [1] "Dropping taxa from the distance matrix because they are not present in the community data:"
## [1] "Methanosarcina"
```

```
# UniFrac Distance (Note: this takes a few minutes; be patient)
dist.uf <- unifrac(comm, phy)
```

In the R code chunk below, do the following:
1. plot Mean Pair Distance versus UniFrac distance and compare.

```
par(mar = c(5, 5, 2, 1) + 0.1)
plot(dist.mp, dist.uf,
pch = 20, col = "red", las = 1, asp = 1, xlim = c(0.15, 0.5), ylim = c(0.15, 0.5),
xlab = "Mean Pair Distance", ylab = "UniFrac Distance")
abline(b = 1, a = 0, lty = 2)
text(0.5, 0.47, "1:1")
```



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 taxon in each site.
c. Why might MPD show less variation than UniFrac?

Answer 4a: Mean Pair Distance is the average phylogenetic dissimilarity between pairs of taxa when comparing two samples (e.g. communities). UniFrac distance is the proportion of unique branch lengths found when comparing two trees of two different samples. ***Answer 4b***: The linear relationship between MPD and UniFrac is vertical. MPD values converge around 0.3 while UniFrac values span from 0.2-0.43. There is significantly more variation in UniFrac than in MPD. ***Answer 4c***: MPD is a general average and thus should converge around a similar value. UniFrac is a proportion, which can vary depending on what trees you are comparing.

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 from earlier in the course.

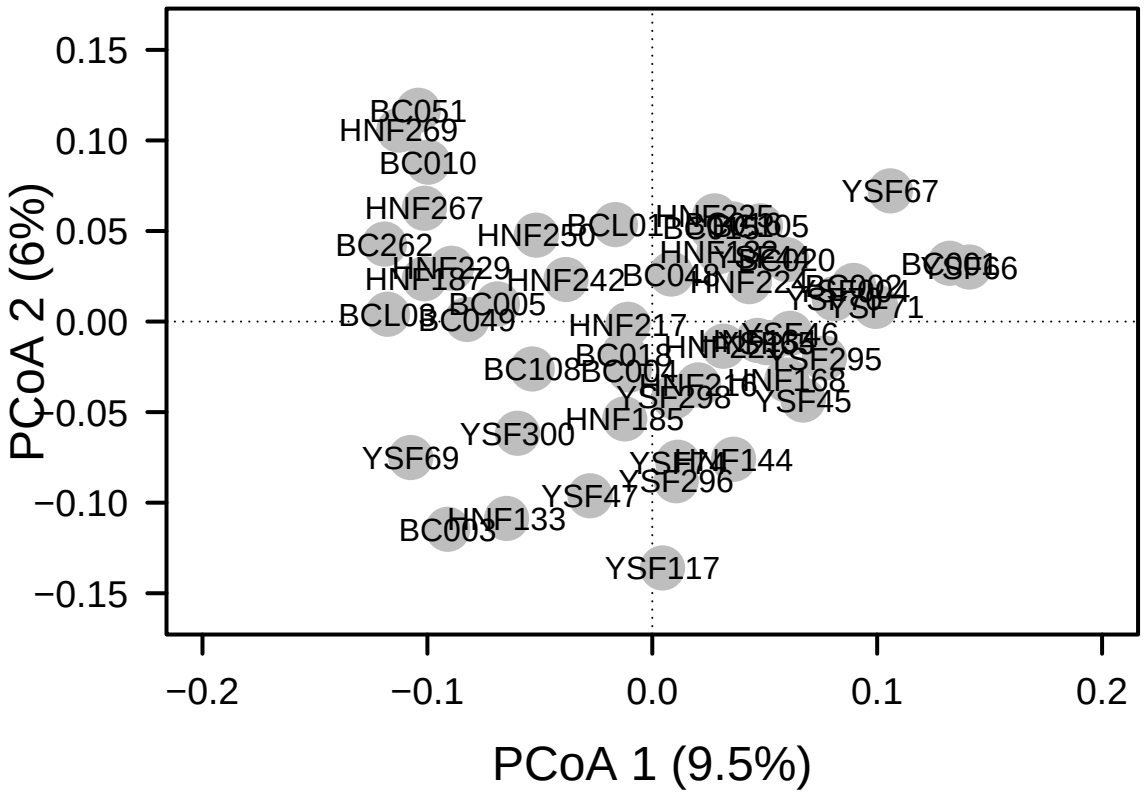
In the R code chunk below, do the following:
1. perform a PCoA based on the UniFrac distances, and
2. calculate the explained variation for the first three PCoA axes.

```
pond.pcoa <- cmdscale(dist.uf, eig = T, k = 3)
explainvar1 <- round(pond.pcoa$eig[1] / sum(pond.pcoa$eig), 3) * 100
explainvar2 <- round(pond.pcoa$eig[2] / sum(pond.pcoa$eig), 3) * 100
explainvar3 <- round(pond.pcoa$eig[3] / sum(pond.pcoa$eig), 3) * 100
sum.eig <- sum(explainvar1, explainvar2, explainvar3)
```

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 using either the R base package or the `ggplot` package,
2. include the appropriate axes,
3. add and label the points, and
4. customize the plot.

```
# Define Plot Parameters
par(mar = c(5, 5, 1, 2) + 0.1)
# Initiate Plot
plot(pond.pcoa$points[ ,1], pond.pcoa$points[ ,2],
     xlim = c(-0.2, 0.2), ylim = c(-.16, 0.16),
     xlab = paste("PCoA 1 (", explainvar1, "%)", sep = ""),
     ylab = paste("PCoA 2 (", explainvar2, "%)", sep = ""),
     pch = 16, cex = 2.0, type = "n", cex.lab = 1.5, cex.axis = 1.2, axes = FALSE)
# Add Axes
axis(side = 1, labels = T, lwd.ticks = 2, cex.axis = 1.2, las = 1)
axis(side = 2, labels = T, lwd.ticks = 2, cex.axis = 1.2, las = 1)
abline(h = 0, v = 0, lty = 3)
box(lwd = 2)
# Add Points & Labels
points(pond.pcoa$points[ ,1], pond.pcoa$points[ ,2],
pch = 19, cex = 3, bg = "gray", col = "gray")
text(pond.pcoa$points[ ,1], pond.pcoa$points[ ,2],
labels = row.names(pond.pcoa$points))
```



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.

```
# bray curtis similarity
tax.dist <- vegdist(decostand(comm, method = "log"), method = "bray")
```

```
tax.pcoa <- cmdscale(tax.dist, eig = T, k = 3)
var1 <- round(tax.pcoa$eig[1] / sum(pond.pcoa$eig), 3) * 100
var2 <- round(tax.pcoa$eig[2] / sum(pond.pcoa$eig), 3) * 100
var3 <- round(tax.pcoa$eig[3] / sum(pond.pcoa$eig), 3) * 100
tax.sum.eig <- sum(var1, var2, var3)
```

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? What does this tell you about the importance of phylogenetic information in this system?

> ***Answer 5***: The taxonomic ordination explains slightly more variation in the data (28.9%) than the phylogenetically based ordination (20.9%). Overall, neither of these explain a majority of the variation, so phylogenetic information may not be super important in this system.

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.

```
# Define Environmental Category
watershed <- env$Location
# Run PERMANOVA with `adonis()` Function {vegan}
phylo.adonis <- adonis2(dist.uf ~ watershed, permutations = 999)
# We can compare to PERMANOVA results based on taxonomy
tax.adonis <- adonis2(vegdist( # create a distance matrix on
            decostand(comm, method = "log"), # log-transformed relative abundances
            method = "bray") ~ watershed, # using Bray-Curtis dissimilarity metric
              permutations = 999)
# P = 0.005
```

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()`).

```
# Define environmental variables
envs <- env[, 5:19]
# Remove redundant variables
envs <- envs[, -which(names(envs) %in% c("TDS", "Salinity", "Cal_Volume"))]
# Create distance matrix for environmental variables
env.dist <- vegdist(scale(envs), method = "euclid")
```

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.

```
# Conduct Mantel Test (`vegan`)
mantel(dist.uf, env.dist)
```

```
##
## Mantel statistic based on Pearson's product-moment correlation
##
## Call:
## mantel(xdis = dist.uf, ydis = env.dist)
```

```
##
## Mantel statistic r: 0.1604
##       Significance: 0.069
##
## Upper quantiles of permutations (null model):
##   90%   95% 97.5%   99%
## 0.136 0.172 0.203 0.235
## Permutation: free
## Number of permutations: 999
```

Last, 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

```
# Conduct dbRDA (`vegan`)
ponds.dbrda <- vegan::dbrda(dist.uf ~ ., data = as.data.frame(scale(envs)))
# Permutation tests: axes and environmental variables
anova(ponds.dbrda, by = "axis")
```

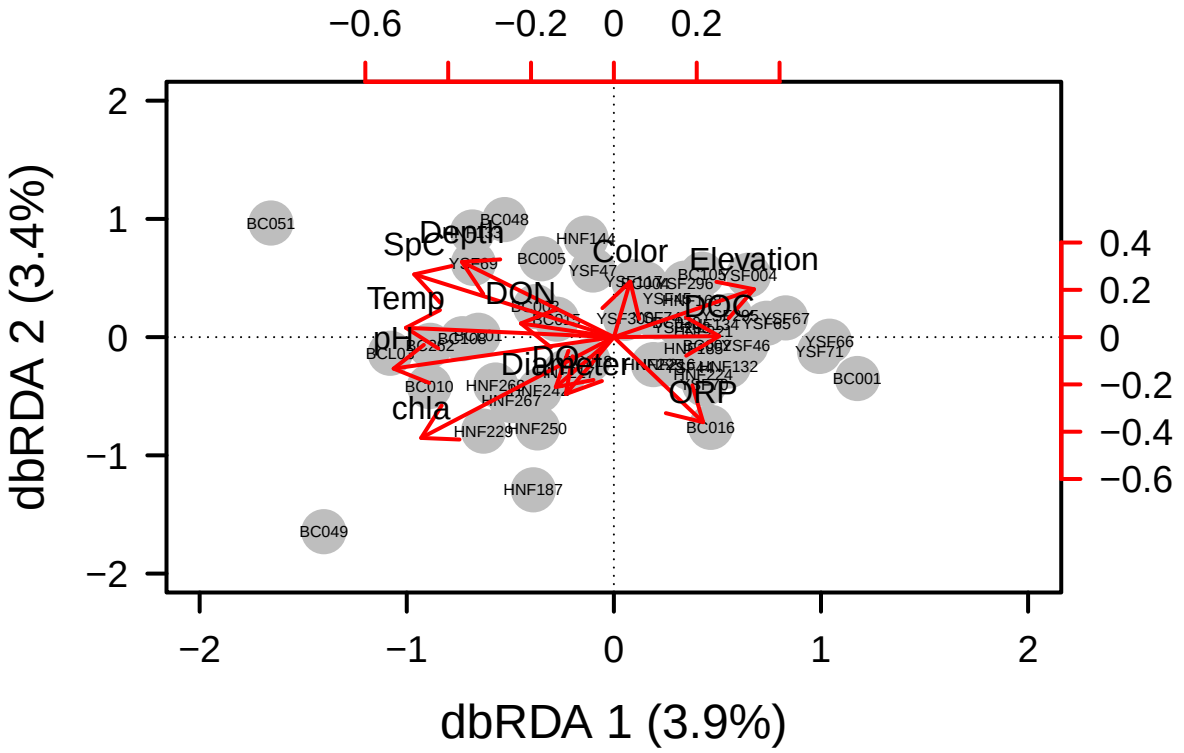
```
## Permutation test for dbrda under reduced model
## Forward tests for axes
## Permutation: free
## Number of permutations: 999
##
## Model: vegan::dbrda(formula = dist.uf ~ Elevation + Diameter + Depth + ORP + Temp + SpC + DO + pH + C
##          Df SumOfSqs      F Pr(>F)
## dbRDA1    1  0.10566 2.0152  0.428
## dbRDA2    1  0.09258 1.7658  0.629
## dbRDA3    1  0.07555 1.4409  0.975
## dbRDA4    1  0.06677 1.2735  0.995
## dbRDA5    1  0.05666 1.0807  1.000
## dbRDA6    1  0.05293 1.0095  1.000
## dbRDA7    1  0.04750 0.9059  1.000
## dbRDA8    1  0.03941 0.7517  1.000
## dbRDA9    1  0.03775 0.7201  1.000
## dbRDA10   1  0.03280 0.6256  1.000
## dbRDA11   1  0.02876 0.5485  1.000
## dbRDA12   1  0.02501 0.4770  1.000
## Residual 39  2.04482
```

```
ponds.fit <- envfit(ponds.dbrda, envs, perm = 999)
ponds.fit
```

```
##
## ***VECTORS
##
##             dbRDA1   dbRDA2     r2 Pr(>r)
## Elevation  0.77670  0.62986 0.0959  0.089 .
## Diameter  -0.27972 -0.96008 0.0541  0.244
## Depth     -0.63137  0.77548 0.1756  0.013 *
## ORP        0.41879 -0.90808 0.1437  0.019 *
## Temp      -0.98250  0.18628 0.1523  0.022 *
## SpC       -0.77101  0.63682 0.2087  0.003 **
```

```
## DO        -0.39318 -0.91946 0.0464  0.295
## pH        -0.96210 -0.27270 0.1756  0.020 *
## Color      0.06353  0.99798 0.0464  0.297
## chla      -0.60392 -0.79704 0.2626  0.005 **
## DOC        0.99847 -0.05526 0.0382  0.369
## DON       -0.91633  0.40042 0.0339  0.440
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## Permutation: free
## Number of permutations: 999
```

```r
# Calculate explained variation
dbrda.explainvar1 <- round(ponds.dbrda$CCA$eig[1] /
sum(c(ponds.dbrda$CCA$eig, ponds.dbrda$CA$eig)), 3) * 100
dbrda.explainvar2 <- round(ponds.dbrda$CCA$eig[2] /
sum(c(ponds.dbrda$CCA$eig, ponds.dbrda$CA$eig)), 3) * 100
# Make dbRDA plot
# Define plot parameters
par(mar = c(5, 5, 4, 4) + 0.1)
# Initiate plot
plot(scores(ponds.dbrda, display = "wa"), xlim = c(-2, 2), ylim = c(-2, 2),
    xlab = paste("dbRDA 1 (", dbrda.explainvar1, "%)", sep = ""),
    ylab = paste("dbRDA 2 (", dbrda.explainvar2, "%)", sep = ""),
    pch = 16, cex = 2.0, type = "n", cex.lab = 1.5, cex.axis = 1.2, axes = FALSE)
# Add axes
axis(side = 1, labels = T, lwd.ticks = 2, cex.axis = 1.2, las = 1)
axis(side = 2, labels = T, lwd.ticks = 2, cex.axis = 1.2, las = 1)
abline(h = 0, v = 0, lty = 3)
box(lwd = 2)
# Add points & labels
points(scores(ponds.dbrda, display = "wa"),
pch = 19, cex = 3, bg = "gray", col = "gray")
text(scores(ponds.dbrda, display = "wa"),
labels = row.names(scores(ponds.dbrda, display = "wa")), cex = 0.5)
# Add environmental vectors
vectors <- scores(ponds.dbrda, display = "bp")
#row.names(vectors) <- c("Temp", "DO", "chla", "DON")
arrows(0, 0, vectors[,1] * 2, vectors[, 2] * 2,
lwd = 2, lty = 1, length = 0.2, col = "red")
text(vectors[,1] * 2, vectors[, 2] * 2, pos = 3,
labels = row.names(vectors))
axis(side = 3, lwd.ticks = 2, cex.axis = 1.2, las = 1, col = "red", lwd = 2.2,
at = pretty(range(vectors[, 1])) * 2, labels = pretty(range(vectors[, 1])))
axis(side = 4, lwd.ticks = 2, cex.axis = 1.2, las = 1, col = "red", lwd = 2.2,
at = pretty(range(vectors[, 2])) * 2, labels = pretty(range(vectors[, 2])))
```



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***: According to the PERMANOVA test, watershed does have a significant effect on phylogenetic distribution of species ($p = 0.005$). However, the Mantel test found only a marginally significant correlation between phylogenetic distance and environmental distance ($p = 0.056$). Looking closer with the dbRDA, there are significant effects of depth, ORP, temperature, SpC, pH, and chlorophyll a on phylogenetic diversity. In summary, there is support for the hypothesis that watersheds act as barriers of gene flow between bacterial communities. There is also some evidence that local environmental factors may shape the phylogenetic diversity of communities, such that bacterial communities evolve in response to local conditions, but this is not a strong trend.

6) SPATIAL PHYLOGENETIC COMMUNITY ECOLOGY

A. Phylogenetic Distance-Decay (PDD)

A distance decay (DD) relationship reflects the spatial autocorrelation of community similarity. That is, communities located near one another should be more similar to one another in taxonomic composition than distant communities. (This is analagous to the isolation by distance (IBD) pattern that is commonly found when examining genetic similarity of a populations as a function of space.) Historically, the two most common explanations for the taxonomic DD are that it reflects spatially autocorrelated environmental variables and the influence of dispersal limitation. However, if phylogenetic diversity is also spatially autocorrelated, then evolutionary history may also explain some of the taxonomic DD pattern. Here, we will construct the phylogenetic distance-decay (PDD) relationship

First, 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.

```
# Geographic distances (kilometers) among ponds
long.lat <- as.matrix(cbind(env$long, env$lat))
coord.dist <- earth.dist(long.lat, dist = TRUE)
# Taxonomic similarity among ponds (Bray-Curits distance)
bray.curtis.dist <- 1 - vegdist(comm)
# Phylogenetic similarity among ponds (UniFrac)
unifrac.dist <- 1 - dist.uf
# Transform all distances into pairwise long format with the melt function from {reshape}:
unifrac.dist.mlt <- melt(as.matrix(unifrac.dist))[melt(upper.tri(as.matrix(unifrac.dist)))$value,]
```

```
## Warning in type.convert.default(X[[i]], ...): 'as.is' should be specified by
## the caller; using TRUE

## Warning in type.convert.default(X[[i]], ...): 'as.is' should be specified by
## the caller; using TRUE
```

```
bray.curtis.dist.mlt <- melt(as.matrix(bray.curtis.dist))[melt(upper.tri(as.matrix(bray.curtis.dist)))$
```

```
## Warning in type.convert.default(X[[i]], ...): 'as.is' should be specified by
## the caller; using TRUE

## Warning in type.convert.default(X[[i]], ...): 'as.is' should be specified by
## the caller; using TRUE
```

```
coord.dist.mlt <- melt(as.matrix(coord.dist))[melt(upper.tri(as.matrix(coord.dist)))$value,]
```

```
## Warning in type.convert.default(X[[i]], ...): 'as.is' should be specified by
## the caller; using TRUE

## Warning in type.convert.default(X[[i]], ...): 'as.is' should be specified by
## the caller; using TRUE
```

```
env.dist.mlt <- melt(as.matrix(env.dist))[melt(upper.tri(as.matrix(env.dist)))$value,]
```

```
## Warning in type.convert.default(X[[i]], ...): 'as.is' should be specified by
## the caller; using TRUE

## Warning in type.convert.default(X[[i]], ...): 'as.is' should be specified by
## the caller; using TRUE
```

```
# Create a data frame from the lists of distances
df <- data.frame(coord.dist.mlt, bray.curtis.dist.mlt[, 3], unifrac.dist.mlt[, 3],
env.dist.mlt[, 3])
names(df)[3:6] <- c("geo.dist","bray.curtis", "unifrac", "env.dist")
```

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.

```
# Set initial plot parameters
par(mfrow=c(2, 1), mar = c(1, 5, 2, 1) + 0.1, oma = c(2, 0, 0, 0))
# Make plot for taxonomic DD
plot(df$geo.dist, df$bray.curtis, xlab = "", xaxt = "n", las = 1, ylim = c(0.1, 0.9),
```

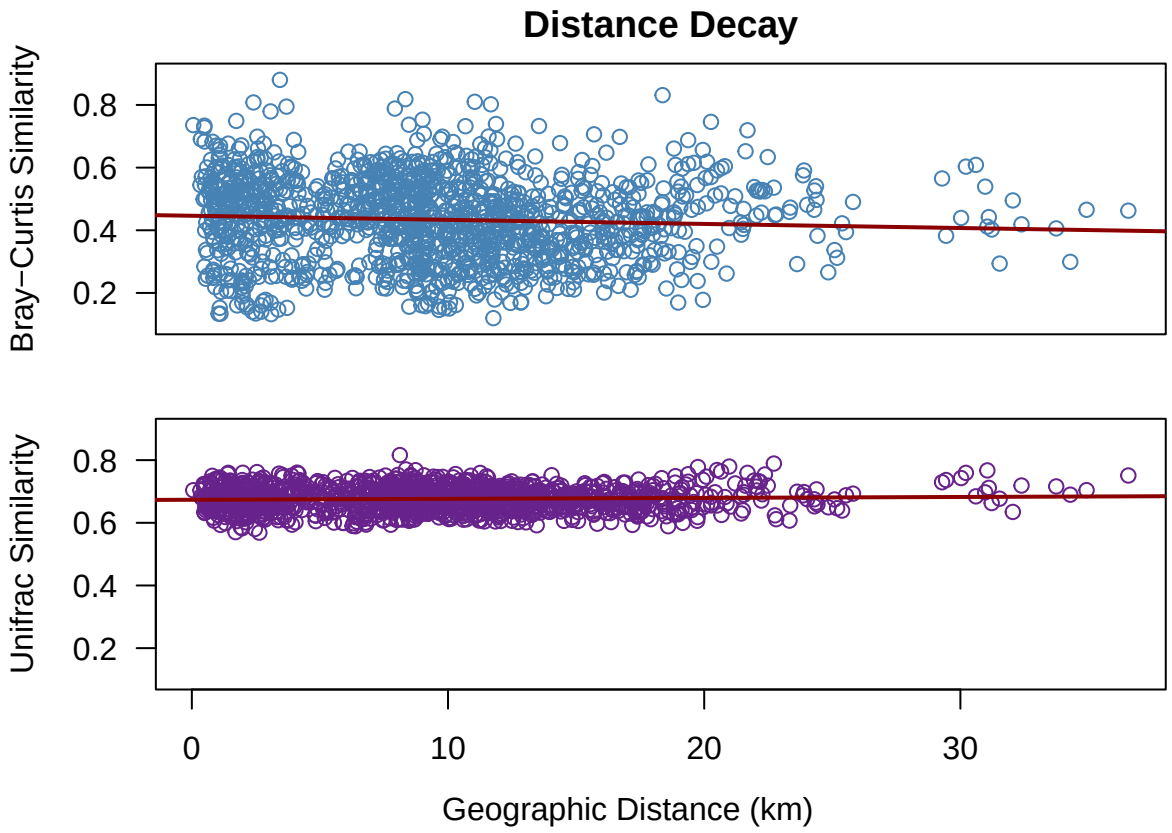
```
      ylab="Bray-Curtis Similarity",
      main = "Distance Decay", col = "SteelBlue")
# Regression for taxonomic DD
DD.reg.bc <- lm(df$bray.curtis ~ df$geo.dist)
summary(DD.reg.bc)
```

```
##
## Call:
## lm(formula = df$bray.curtis ~ df$geo.dist)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.31151 -0.08843  0.00315  0.09121  0.43817
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.4463453  0.0066883  66.735   <2e-16 ***
## df$geo.dist -0.0013051  0.0005864  -2.226   0.0262 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1303 on 1324 degrees of freedom
## Multiple R-squared:  0.003728,   Adjusted R-squared:  0.002975
## F-statistic: 4.954 on 1 and 1324 DF,  p-value: 0.0262
```

```
abline(DD.reg.bc , col = "red4", lwd = 2)
# New plot parameters
par(mar = c(2, 5, 1, 1) + 0.1)
# Make plot for phylogenetic DD
plot(df$geo.dist, df$unifrac, xlab = "", las = 1, ylim = c(0.1, 0.9),
ylab = "Unifrac Similarity", col = "darkorchid4")
# Regression for phylogenetic DD
DD.reg.uni <- lm(df$unifrac ~ df$geo.dist)
summary(DD.reg.uni)
```

```
##
## Call:
## lm(formula = df$unifrac ~ df$geo.dist)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.105629 -0.027107 -0.000077  0.026761  0.140215
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.6735186  0.0019206 350.677   <2e-16 ***
## df$geo.dist 0.0002976  0.0001684   1.767   0.0774 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.03741 on 1324 degrees of freedom
## Multiple R-squared:  0.002354,   Adjusted R-squared:  0.0016
## F-statistic: 3.124 on 1 and 1324 DF,  p-value: 0.07738
```

```
abline(DD.reg.uni, col = "red4", lwd = 2)
# Add x-axis label to plot
mtext("Geographic Distance (km)", side = 1, adj = 0.55,
line = 0.5, outer = TRUE)
```



In the R code chunk below, test if the trend lines in the above distance decay relationships are different from one another.

```
source("./bin/diffslope.R")
diffslope(df$geo.dist, df$unifrac, df$geo.dist, df$bray.curtis)
```

```
## $call
## diffslope(x1 = df$geo.dist, y1 = df$unifrac, x2 = df$geo.dist,
##     y2 = df$bray.curtis)
##
## $slope.diff
## [1] 0.001602746
##
## $signif
## [1] 0.001
##
## $permutations
## [1] 1000
##
## $perms
##    [1] -7.608235e-05  1.072677e-03 -5.136278e-04 -8.830568e-04 -6.825808e-04
##    [6] -1.663072e-04  3.599489e-04  9.129036e-05 -9.064678e-05 -3.280636e-04
##   [11]  1.381090e-03 -2.233007e-04  5.157540e-05  4.248881e-05  3.474947e-04
```

```
##   [16]  6.959921e-04 -5.744818e-04  1.209796e-04 -5.273843e-04 -2.970008e-04
##   [21] -1.624457e-03  1.067163e-03  2.442568e-04 -6.036163e-04 -4.535910e-04
##   [26]  6.048200e-04 -5.795168e-04  5.732941e-04  3.143217e-04 -3.747208e-04
##   [31] -2.781849e-04 -7.542916e-04 -7.367625e-04  5.706078e-04  4.770933e-04
##   [36]  2.766856e-05 -6.503995e-04 -7.560676e-04  4.013029e-04  9.853255e-05
##   [41]  9.158397e-05  3.278668e-04 -6.338537e-05 -1.109530e-03  6.645630e-05
##   [46] -9.016301e-04  3.582807e-04  7.852873e-05  5.617182e-04  4.901090e-04
##   [51] -2.569713e-04 -7.122369e-04 -6.290721e-04  1.008976e-03  4.124599e-04
##   [56]  1.012102e-04  1.123072e-03 -2.546284e-04 -7.478236e-05 -4.589472e-04
##   [61] -6.511698e-05 -8.067425e-04  3.038464e-04 -1.075730e-03  3.139511e-04
##   [66] -3.055798e-04 -6.145987e-04 -6.628799e-04  9.498796e-04 -6.672889e-06
##   [71] -1.084051e-03 -7.595780e-04 -5.577076e-04  5.912220e-04 -9.321025e-04
##   [76]  5.386138e-04 -1.851406e-04  3.588401e-04 -2.200681e-04  2.912928e-04
##   [81]  7.089266e-04  5.234914e-04  8.974720e-04 -9.110947e-04  1.068274e-03
##   [86] -5.435858e-05  8.258820e-05 -7.385197e-04 -2.212906e-04  1.255332e-03
##   [91]  7.507220e-04  1.490188e-04  3.639407e-04 -2.413981e-04 -8.538245e-05
##   [96] -9.487287e-04  3.358244e-04  8.756743e-04 -1.247044e-04 -1.323017e-04
##  [101] -5.672772e-04 -8.174918e-04 -6.672862e-04 -1.686311e-04  1.562817e-04
##  [106] -2.147810e-04  7.006101e-04 -6.383644e-04  6.167337e-04 -1.100305e-04
##  [111] -2.583492e-04  1.322753e-04 -2.782421e-04 -2.383213e-04  6.720765e-04
##  [116]  3.485831e-04 -3.282322e-05 -7.176265e-04  2.844034e-04 -5.344450e-04
##  [121] -7.493095e-04 -4.873608e-04  4.011784e-04  8.053619e-05  1.202972e-03
##  [126]  7.195434e-04  9.324508e-04 -1.197731e-03 -4.271475e-04 -4.023497e-04
##  [131] -2.375682e-04  3.664339e-04  6.886123e-04  5.238769e-04 -2.434095e-04
##  [136]  3.002281e-04 -8.247325e-05 -7.099928e-05  4.211812e-04 -1.081393e-03
##  [141]  7.031077e-04  3.662806e-05  4.450713e-04 -6.441411e-05  8.503406e-04
##  [146]  1.578883e-04  4.954502e-04 -3.816246e-04 -1.450166e-03 -7.006915e-04
##  [151]  1.411218e-04  6.069217e-04  8.578095e-04 -5.507625e-04 -2.158925e-04
##  [156] -7.059741e-05 -8.327640e-04  4.785445e-04 -7.326886e-04  5.537056e-04
##  [161]  4.091702e-05  2.257156e-04  5.230899e-04  5.479911e-04  3.891535e-04
##  [166] -1.842593e-04 -2.334140e-04 -4.709775e-04  5.645310e-04 -4.995053e-05
##  [171] -2.761154e-04 -7.823741e-04  7.695510e-04  5.213303e-04  2.579176e-04
##  [176] -4.277571e-04  3.560542e-04  1.659569e-04 -4.438039e-04 -3.520754e-04
##  [181]  8.989791e-04  7.244280e-04 -4.016496e-04  6.637356e-04 -6.838079e-04
##  [186] -9.518036e-04  6.777649e-05 -3.311380e-04  2.692621e-04 -1.089369e-03
##  [191]  7.895226e-04  2.705599e-04 -2.639143e-04 -9.576757e-04 -1.086638e-03
##  [196] -7.057736e-04 -2.014712e-04 -7.015245e-04 -2.340570e-04 -3.968939e-04
##  [201] -8.410441e-04 -4.235355e-05  4.879728e-04 -7.195965e-04  4.124721e-04
##  [206] -2.656224e-04 -5.939506e-04 -6.917264e-04 -9.879646e-05 -6.011565e-04
##  [211] -4.739025e-05  4.578354e-04  7.616728e-04 -3.540263e-04 -7.589846e-05
##  [216]  8.622745e-04  3.101585e-04 -1.931295e-03  7.641552e-04  6.614158e-05
##  [221] -9.780503e-04 -2.322430e-04 -7.803386e-04  7.150229e-04  4.250414e-04
##  [226] -1.169569e-04 -3.170041e-04 -5.729901e-04 -2.486736e-04  3.990178e-04
##  [231]  4.971887e-04 -7.664411e-04 -2.207223e-04 -5.624923e-04  4.526181e-04
##  [236]  7.899564e-04  9.786473e-05  4.595380e-04 -2.023063e-04 -8.471780e-05
##  [241] -6.763013e-04  7.346436e-04 -8.237613e-05  1.339797e-03 -1.614308e-04
##  [246]  1.773671e-04 -2.347114e-04  7.942233e-04  1.143460e-04  1.183575e-03
##  [251] -4.733995e-04  3.593105e-05  3.033481e-04  5.440132e-04  6.745067e-04
##  [256]  5.117926e-04 -5.767749e-04 -1.681851e-04  1.835989e-04 -4.321264e-05
##  [261]  1.050993e-03 -2.891498e-04  5.222306e-04 -5.890751e-05  7.232059e-04
##  [266]  1.483609e-04 -5.654903e-04 -6.309397e-04  1.613892e-04  1.720293e-04
##  [271] -7.575225e-05  9.238033e-04 -1.100663e-04 -7.916193e-04 -8.469765e-04
##  [276]  7.177699e-04  1.424325e-04  1.536699e-03  6.336892e-04  1.437490e-04
##  [281]  8.739213e-04  1.044518e-03 -2.054471e-04 -3.043936e-04  1.081466e-04
```

```
##  [286] -5.368952e-04 -4.946258e-04  1.849454e-04  2.881402e-04  2.591506e-04
##  [291] -1.293352e-04 -5.820114e-04 -4.756697e-04 -8.878111e-04 -3.882433e-04
##  [296]  1.498820e-03 -3.703418e-04  7.733094e-04 -2.603258e-04 -2.037855e-04
##  [301] -1.464295e-03 -6.600886e-04 -1.977723e-05 -2.549859e-04 -6.485506e-04
##  [306] -2.186249e-04 -1.873784e-03  3.641970e-04 -9.667172e-04  1.727559e-04
##  [311] -7.925533e-04 -2.749717e-04  9.266444e-05  2.989438e-04  6.852845e-05
##  [316] -4.315659e-04  2.532260e-05  3.040440e-04  2.018122e-04  6.920251e-04
##  [321]  2.389365e-04  8.069111e-04  3.514116e-04 -8.488547e-04 -5.610079e-04
##  [326]  4.806248e-04  4.760031e-04  8.749073e-04  1.120182e-03  6.580930e-04
##  [331]  4.814249e-04 -1.060997e-04 -7.936850e-04 -1.330060e-04 -1.809600e-04
##  [336]  4.314977e-04 -2.971004e-04 -3.011680e-04  9.747958e-04  7.424617e-05
##  [341]  5.788270e-04 -1.073211e-04  4.130563e-04 -1.573177e-04  4.478440e-05
##  [346]  1.575564e-04 -5.716420e-04  2.147361e-04  3.014525e-04  2.830015e-04
##  [351] -1.097746e-03 -1.149651e-04  5.503104e-05 -1.410668e-03 -3.887274e-05
##  [356] -9.093090e-04 -1.976461e-04 -5.425170e-04  1.514972e-04 -1.023826e-04
##  [361]  4.964320e-04  9.976164e-06 -2.613820e-04  3.756593e-04 -1.137213e-04
##  [366] -2.887999e-04 -9.148209e-04 -4.577069e-04  7.843767e-04 -5.255008e-04
##  [371]  3.722266e-04  3.796190e-04  9.559783e-04  1.454389e-04 -9.490329e-05
##  [376]  3.710843e-04 -5.243194e-04  1.019142e-03  3.022274e-04  4.705778e-05
##  [381]  3.458270e-04 -6.760599e-05 -3.358772e-04 -3.345379e-04  5.131182e-04
##  [386] -5.423461e-04  5.588680e-05  4.961624e-04  9.501569e-06  4.070025e-04
##  [391] -1.078258e-03  4.216147e-04  1.208706e-03 -3.354244e-04  5.583433e-04
##  [396]  4.619715e-04  1.660420e-05  1.241921e-03  1.057175e-04 -1.620725e-04
##  [401]  1.898523e-04  2.377318e-04  2.624532e-04  3.677116e-04  3.842907e-04
##  [406] -1.129672e-04 -2.061504e-04  3.553894e-04  9.343463e-04  1.797427e-03
##  [411]  9.527636e-05  3.279726e-04  5.914168e-04  6.724134e-05 -7.638233e-04
##  [416]  1.907202e-04 -5.374353e-04  2.810281e-04  1.858185e-04  4.979499e-04
##  [421]  3.778666e-04  2.613722e-05  3.461655e-04  4.849842e-05  1.852173e-04
##  [426]  1.249952e-03 -9.003368e-04  4.975479e-04 -3.198813e-04 -3.111773e-04
##  [431] -1.678451e-03  4.446577e-04  5.982324e-04 -2.546464e-04 -3.119131e-05
##  [436] -1.108924e-03 -9.724389e-04 -3.443471e-05  8.651031e-04 -1.212511e-03
##  [441] -1.014920e-03 -1.679327e-04 -4.159979e-04 -9.108883e-05 -2.592910e-04
##  [446] -8.189515e-04 -9.021546e-04  1.669047e-04 -1.542385e-04 -6.122032e-05
##  [451] -1.000638e-03  5.496957e-05  6.130722e-04 -7.831934e-06 -1.854725e-04
##  [456]  1.442503e-04 -2.761379e-04  3.201993e-04 -5.764159e-04 -2.977294e-04
##  [461]  8.668386e-04 -4.832142e-04  5.886558e-04 -7.631026e-04 -8.096913e-05
##  [466] -3.410586e-04 -6.784572e-04  3.683144e-04  6.435784e-04  4.998668e-04
##  [471]  1.406665e-03  3.178783e-04  2.363622e-05 -8.320673e-04  1.666306e-04
##  [476]  4.134115e-04  9.691321e-04 -2.602623e-04  9.477247e-05 -5.886633e-04
##  [481]  5.216462e-04  5.165573e-04  2.894675e-04 -5.052156e-05 -9.547566e-06
##  [486]  4.821863e-04 -4.704502e-04 -1.110843e-04 -2.385903e-04 -4.532038e-04
##  [491]  5.046911e-04 -1.297711e-04  3.023477e-04 -1.813833e-04  9.065881e-05
##  [496]  2.807872e-05 -2.562902e-04  6.413217e-04 -1.053230e-03 -7.545319e-04
##  [501] -2.213274e-04 -4.389793e-04 -1.289298e-03  7.157403e-05 -4.614322e-04
##  [506] -1.003791e-03  3.835497e-04  6.274987e-05 -8.441119e-05  1.040474e-03
##  [511] -7.953651e-04 -9.159956e-04  5.071534e-04 -5.430221e-04  3.927825e-04
##  [516]  4.526788e-04  1.205589e-04 -1.819414e-03  8.677577e-04 -1.225103e-03
##  [521]  7.205108e-06 -2.145563e-05  1.051292e-05  1.342101e-05 -3.903722e-04
##  [526] -2.661819e-04 -7.503445e-04 -1.351845e-03 -8.112856e-04 -5.233001e-05
##  [531]  1.910157e-04 -3.965251e-04  1.302706e-05  3.177824e-04 -5.212459e-04
##  [536]  1.007300e-04 -2.251919e-04  4.297709e-04  6.229604e-04 -1.439791e-03
##  [541] -1.820010e-04 -3.234388e-05 -7.608776e-04  4.813858e-04 -4.726241e-04
##  [546]  2.755941e-04 -4.323159e-05  2.860845e-04 -2.032467e-04  1.429745e-04
##  [551]  7.106054e-04  2.112161e-04 -1.085298e-03 -9.258680e-04  1.085236e-03
```

```
##  [556] -2.583623e-04  1.081605e-03 -9.124526e-04  1.339341e-03 -6.564218e-04
##  [561] -2.921602e-04 -1.369107e-03 -4.842337e-04 -3.013868e-04  1.783823e-04
##  [566] -1.099836e-04  3.383429e-04  9.646173e-04 -1.681976e-04  9.206889e-05
##  [571]  1.258117e-03  8.048505e-04  1.373087e-03 -5.779526e-04  3.143623e-04
##  [576]  3.773158e-04 -8.349454e-04 -7.353461e-04  1.446382e-03  2.277883e-04
##  [581] -3.599254e-04  9.458136e-04  3.754791e-04  6.734339e-04 -2.634628e-05
##  [586]  1.507771e-03 -8.099075e-04 -5.760492e-04 -4.980387e-04 -1.500078e-05
##  [591]  4.526973e-04 -1.843147e-04 -3.835535e-04 -9.396368e-04  2.039048e-04
##  [596] -3.019958e-04 -1.315321e-03 -8.809971e-05  3.527816e-04  6.714451e-04
##  [601]  1.215877e-05 -6.797240e-05 -1.958872e-04 -4.251177e-04 -3.244488e-04
##  [606]  2.578352e-04  5.438888e-04  1.865324e-05  3.121239e-04 -6.109814e-04
##  [611] -1.578303e-03 -8.215512e-04 -8.690509e-04 -9.345968e-04  1.589110e-04
##  [616]  1.982396e-04  3.962063e-04 -6.133375e-04  9.743979e-04 -4.341497e-04
##  [621]  7.387800e-04 -7.379689e-04 -3.642192e-05 -5.160052e-04 -8.448848e-04
##  [626]  1.533588e-03 -6.547405e-04  1.774988e-04 -9.161960e-04 -5.495901e-04
##  [631] -5.273536e-04 -3.004728e-04  6.565463e-04 -1.598167e-04 -4.895147e-04
##  [636] -2.626841e-04 -3.141866e-04 -7.149272e-04  4.003318e-04  4.136797e-04
##  [641] -4.952027e-04  1.403233e-05 -1.974380e-04 -1.326189e-04 -1.889423e-05
##  [646]  5.905212e-04  8.965457e-04 -5.788753e-04  6.462870e-05  4.553994e-04
##  [651] -1.290596e-05 -3.356790e-04 -4.651354e-04 -1.861169e-04  5.091090e-04
##  [656] -1.992280e-04  4.429791e-05 -5.347461e-04  6.062446e-04  1.115970e-03
##  [661]  4.024644e-04  5.634551e-04  1.723809e-04 -2.294976e-04 -4.528417e-04
##  [666] -1.804156e-04 -3.567459e-04  6.628206e-04  1.181817e-03 -6.904468e-04
##  [671]  2.267934e-04 -1.024709e-03  8.374206e-05  2.937667e-04  4.415270e-04
##  [676] -4.157574e-04  3.915923e-04 -1.416212e-04  6.556934e-04 -4.029468e-05
##  [681] -2.918334e-04  1.101357e-03  1.941743e-04  9.330876e-04  3.406503e-04
##  [686] -6.558695e-04 -1.777491e-04  4.569452e-04  3.343395e-04  1.542559e-04
##  [691] -4.365972e-04 -2.897584e-04 -7.858646e-04 -2.935393e-04 -7.752708e-04
##  [696] -1.631793e-04  1.002332e-03 -1.212675e-03  2.877969e-04 -3.092179e-04
##  [701]  9.144750e-05  4.036475e-04 -2.191919e-04  2.621361e-04 -1.841444e-04
##  [706]  5.552060e-04 -5.102921e-04 -8.543374e-04  7.108411e-04  6.615618e-04
##  [711]  1.717536e-04 -1.974545e-05  7.125724e-05 -8.940963e-04 -1.420342e-03
##  [716]  2.927498e-04  6.253352e-04 -4.987206e-05  3.653510e-04  1.763290e-04
##  [721] -5.283380e-04 -3.716447e-04  4.420412e-04  1.229386e-05 -2.608588e-04
##  [726]  1.600587e-04  4.213940e-04  9.358195e-04 -5.512390e-04 -4.412430e-04
##  [731] -9.753665e-04 -1.125409e-03  3.614926e-04  1.396174e-03 -2.015170e-04
##  [736] -3.748003e-04  8.057605e-04  4.398864e-04 -9.207950e-04  8.234306e-04
##  [741]  4.744832e-04  5.031348e-04  4.734368e-04  1.720505e-04 -4.399036e-04
##  [746]  9.338194e-05  6.254217e-04  2.751208e-04  7.198138e-04  4.663991e-04
##  [751] -9.239956e-05 -1.472036e-03  5.735041e-05 -6.537363e-04  4.742131e-04
##  [756] -4.973847e-04  3.147945e-05  3.012234e-04 -1.133861e-03  1.553531e-05
##  [761]  2.061002e-04  9.667013e-04 -1.330494e-03 -2.826358e-04 -1.369061e-03
##  [766]  1.055312e-04 -2.500021e-04  7.316262e-04 -5.947647e-04 -5.496563e-04
##  [771] -1.129867e-03 -6.430773e-04  1.295807e-03 -8.430541e-04 -4.622044e-05
##  [776]  5.182771e-04 -5.555775e-04 -3.271064e-04 -6.651305e-04 -6.327811e-05
##  [781] -1.128751e-03  1.960311e-04 -1.028098e-04 -5.215565e-04 -4.266301e-04
##  [786]  1.131821e-03 -3.261488e-04 -1.073459e-04 -2.211909e-04  2.053661e-04
##  [791] -2.390400e-04 -6.023195e-04 -1.178076e-04  1.939970e-04 -7.472713e-05
##  [796] -2.692005e-04 -1.257707e-03 -8.828672e-04  7.023798e-04  2.592474e-04
##  [801]  2.021328e-04  7.436855e-04 -4.540490e-04 -8.655999e-04 -1.173345e-04
##  [806] -1.231284e-05 -1.588531e-04  7.653516e-05 -3.628072e-04 -1.177639e-03
##  [811] -5.002580e-05 -4.688432e-04 -4.195647e-04  6.385533e-04 -5.305912e-04
##  [816] -6.613548e-04 -1.407915e-04 -5.079500e-05  9.654422e-05  9.869218e-04
##  [821] -5.917473e-04  5.614507e-04 -2.585676e-04 -1.765672e-04  4.044150e-04
```

```
##  [826] -3.257252e-04 -2.895692e-05  1.272997e-04 -1.051883e-03 -3.086981e-04
##  [831] -1.493174e-04  3.781597e-04  2.217198e-04 -3.161116e-04 -3.169357e-04
##  [836] -8.434880e-05 -3.765052e-04  6.982837e-05 -3.519440e-04 -5.747961e-04
##  [841]  5.188148e-04  1.233910e-03 -9.142512e-04 -5.606789e-04 -7.698006e-04
##  [846] -1.085907e-04 -6.884854e-05 -5.576292e-04  1.083207e-03  8.388294e-04
##  [851]  4.667642e-04 -1.808179e-04  3.022389e-04 -3.079421e-04 -5.394937e-04
##  [856]  1.986515e-04 -7.120940e-04  4.840019e-04 -1.232439e-03 -6.053906e-05
##  [861]  7.590475e-05  1.390232e-04 -4.175130e-04  5.178173e-04 -1.162743e-03
##  [866] -4.026044e-04  6.916178e-05 -3.713436e-04 -3.201419e-04 -1.830748e-03
##  [871]  2.742175e-04  2.956886e-04 -4.827289e-04 -6.221151e-04  8.967980e-05
##  [876]  4.431445e-04  1.331946e-04  1.281108e-04 -8.552048e-04 -2.615668e-04
##  [881] -1.387971e-04 -2.011152e-04  1.877272e-04  1.237895e-04  5.608281e-04
##  [886]  6.260727e-04 -4.636317e-04 -4.407391e-04  5.548725e-04  9.646060e-05
##  [891]  1.874852e-04 -8.496724e-04  5.219388e-04  5.820220e-04 -4.420047e-04
##  [896]  3.901055e-04 -8.370342e-04  2.611944e-04  7.416864e-04 -3.832042e-04
##  [901]  2.138217e-04 -9.600219e-04 -1.069404e-03 -5.154679e-04 -2.261568e-04
##  [906] -1.236220e-03 -9.647104e-04 -3.970510e-04 -1.639900e-04 -2.504010e-04
##  [911] -4.253802e-04  3.076297e-04 -6.132235e-04 -7.485873e-04 -8.162225e-04
##  [916] -1.042514e-03 -4.893866e-04 -9.996299e-04 -8.298586e-04  7.622703e-05
##  [921]  2.066616e-04  3.791241e-04 -2.392782e-04  8.195463e-04 -1.359847e-04
##  [926]  4.677378e-04 -1.053565e-03 -1.181328e-04  6.068531e-04 -5.592724e-04
##  [931] -5.572161e-04 -6.551823e-04 -1.405124e-04 -3.844939e-04  2.424331e-04
##  [936] -8.152624e-04 -1.773056e-04  2.313354e-04 -7.044347e-06 -2.612711e-04
##  [941]  3.894773e-04  6.740448e-04  5.182530e-04  2.997776e-04 -6.715435e-04
##  [946]  1.132287e-03 -7.356721e-04  3.264220e-04  7.266472e-04  3.358123e-04
##  [951] -2.869321e-04  1.123982e-03 -1.694659e-04 -1.210033e-04 -1.093553e-05
##  [956] -2.600216e-04 -4.530267e-04 -2.631801e-04 -5.722535e-04 -5.165255e-05
##  [961] -3.596779e-04  3.024537e-04 -2.049236e-04  1.904164e-04  7.789409e-04
##  [966] -3.050290e-05 -1.497764e-04  5.538448e-04 -7.828907e-05 -1.156939e-03
##  [971] -1.914628e-05 -5.019318e-04 -4.438232e-05  5.999864e-04  5.859199e-04
##  [976] -8.639411e-04  7.689626e-05  5.456508e-04  4.658895e-04 -2.074612e-04
##  [981] -7.346804e-04 -3.723128e-04  9.754164e-04  4.133390e-04 -1.095691e-03
##  [986]  3.566219e-04 -1.433561e-04 -5.517597e-04  4.708347e-04  7.829296e-04
##  [991] -1.937057e-04 -1.994875e-04 -1.853431e-04 -1.282481e-04  9.862834e-04
##  [996]  5.199983e-04 -3.402842e-04  9.208453e-05  5.318132e-04 -4.442731e-04
##
## attr(,"class")
## [1] "dsl"
```

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

> ***Answer 7***: There is a significant difference in slope (0.001602746, p = 0.004). The slope for Bray Curtis similarity is negative; as geographic distance increases, communities become more dissimilar. However, the slope of the unifrac similarity is very close to 0, so geographic distance does not correlate with phylogenetic distance. My hypothesis as to why this might be the case is that the bacteria communities in these lakes come from very similar evolutionary backgrounds so even if they have diverged in community composition, phylogenetically there would not be that much of a difference.

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 in the PhyloTraits module, or any other concepts that we have not covered in this course.

> I am broadly interested in how animals cope with environmental stressors and how these underlying mechanisms evolve over time. I think a cool way to apply phylogenetics to this would be to use a comparative genomic/transcriptomic approach to compare heat shock protein (HSP) gene alignment in various species that are both heat-adapted and non-heat adapted. I would need nucleotide sequences of the same HSP gene in multiple species to form a gene phylogeny, and then see how that matches up with the species information. I think this could address the question of whether there are differences in a highly conserved gene involved in cellular stress response across species, and if there are differences, what their relationship is with other heat-adaptive phenotypes.