12.Phylogenetic Diversity - Communities

Joshua Jones; Z620: Quantitative Biodiversity, Indiana University

06 May, 2021

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 '12.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 *12.PhyloCom_Worksheet.Rmd* and the PDF output of `Knitr` (*12.PhyloCom_Worksheet.pdf*).

The completed exercise is due on **Monday, May 10th, 2021 before 09:00 AM**.

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 `/12.PhyloCom` folder,
4. load all of the required R packages (be sure to install if needed), and
5. load the required R source file.

```
# Clearning Environment
rm(list = ls())
dev.off ()
```

```
## null device
##           1
```

```
# Setting Working Directory
getwd()
```

```
## [1] "C:/Users/joshu/quantbio/QB2021_Jones/2.Worksheets/12.PhyloCom"
```

```
setwd("C:/Users/joshu/quantbio/QB2021_Jones/2.Worksheets/12.PhyloCom/")

# Installing and Loading Packages
package.list <- c('ape', 'seqinr', 'picante', 'vegan', 'fossil', 'reshape', 'simba')
for (package in package.list){
  if (!require(package, character.only = TRUE, quietly = TRUE)) {
    install.packages(package)
    library(package,character.only = TRUE)
  }
}
```

```
## Warning: package 'ape' was built under R version 4.0.5
```

```
## Warning: package 'seqinr' was built under R version 4.0.5
```

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

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

```
## Warning: package 'picante' was built under R version 4.0.5
```

```
## Warning: package 'vegan' was built under R version 4.0.4
```

```
## Warning: package 'permute' was built under R version 4.0.4
```

```
##
## Attaching package: 'permute'
```

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

```
## This is vegan 2.5-7

##
## Attaching package: 'nlme'

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

## Warning: package 'fossil' was built under R version 4.0.5

## Warning: package 'sp' was built under R version 4.0.5

## Warning: package 'maps' was built under R version 4.0.5

##
## Attaching package: 'shapefiles'

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

## Warning: package 'reshape' was built under R version 4.0.5

## Warning: package 'simba' was built under R version 4.0.5

## This is simba 0.3-5

##
## Attaching package: 'simba'

## The following object is masked from 'package:picante':
##
##     mpd

## The following object is masked from 'package:stats':
##
##     mad
```

```
# Loading the required R source file
source("./bin/MothurTools.R")
```

2) DESCRIPTION OF DATA

need to discuss data set from spatial ecology!

In 2013 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

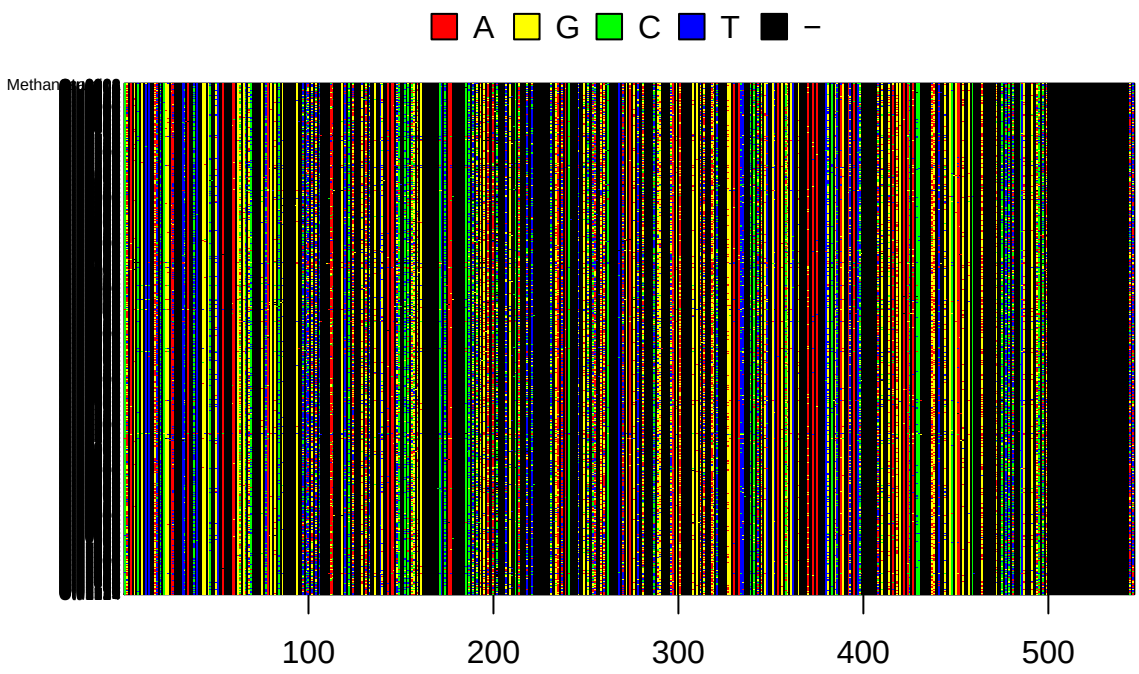
In the R code chunk below, do the following:
1. load the environmental data for the Brown County ponds (*20130801_PondDataMod.csv*),
2. load the site-by-species matrix using the `read.otu()` function,
3. subset the data to include only DNA-based identifications of bacteria,
4. rename the sites by removing 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.

```
# Loading Environment and site-by-species data
env <- na.omit(read.table("data/20130801_PondDataMod.csv", sep = ",", header = TRUE))
comm <- read.otu(shared = "./data/INPonds.final.rdp.shared", cutoff = 1)

# Subsetting data to include only DNA-based identification of bacteria
comm <- comm[grep("*-DNA", rownames(comm)), ]

# Renaming Sites
rownames(comm) <- gsub("\\-DNA", "", rownames(comm))
rownames(comm) <- gsub("\\_", "", rownames(comm))

# Remove sites not in env data set
comm <- comm[rownames(comm) %in% env$Sample_ID,]

# Removing Unnecessary OTUs
comm <- comm[, colSums(comm) > 0]

# Loading taxonomic data
tax <- read.tax(taxonomy = "./data/INPonds.final.rdp.1.cons.taxonomy")
```

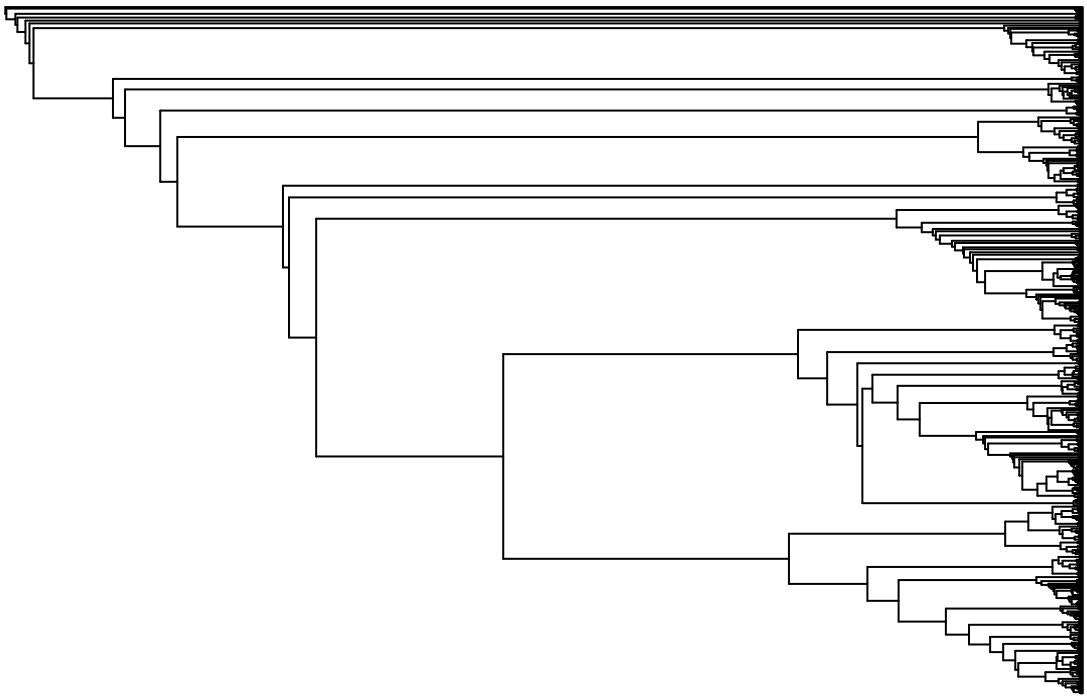
Next, in the R code chunk below, do the following:
1. load the FASTA alignment for the bacterial operational taxonomic units (OTUs),
2. rename the OTUs by removing everything before the tab (\t) and after the bar (|),
3. import the *Methanosarcina* outgroup FASTA file,
4. convert both FASTA files into 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.

```
# Importing alignment file
ponds.cons <- read.alignment(file = "./data/INPonds.final.rdp.1.rep.fasta", format = "fasta")

# Renaming the OTUs
ponds.cons$nam <- gsub("\\|.*$", "", gsub("^.*?\t", "", ponds.cons$nam))

# Importing outgroup "Methanosarcina"
outgroup <- read.alignment(file = "./data/methanosarcina.fasta", format = "fasta")

# Converting FASTA files into the DNAbin format and combining
DNAbin <- rbind(as.DNAbin(outgroup), as.DNAbin(ponds.cons))
```

```
# Visualize the sequence alignment
image.DNAbin(DNAbin, show.labels = T, cex.lab = .5, las = 1)
```




```
# Making a Distance Matrix
seq.dist.jc <- dist.dna(DNAbin, model = "JC", pairwise.deletion = FALSE)

# Making a neighbor joining tree
phy.all <- bionj(seq.dist.jc)

# Remove any tips that are not in community data set
phy <- drop.tip(phy.all, phy.all$tip.label[!phy.all$tip.label %in% c(colnames(comm), "Methanosarcina")]

# Plotting Rooted tree
outgroup <- match("Methanosarcina", phy$tip.label)
phy <- root(phy, outgroup, resolve.root = TRUE)

par(mar = c(1,1,2,1) + .1)
plot.phylo(phy, main = "Neighbor Joining Tree", "phylogram", show.tip.label = FALSE, use.edge.length = I
```



4) PHYLOGENETIC ALPHA DIVERSITY

A. Faith's Phylogenetic Diversity (PD)

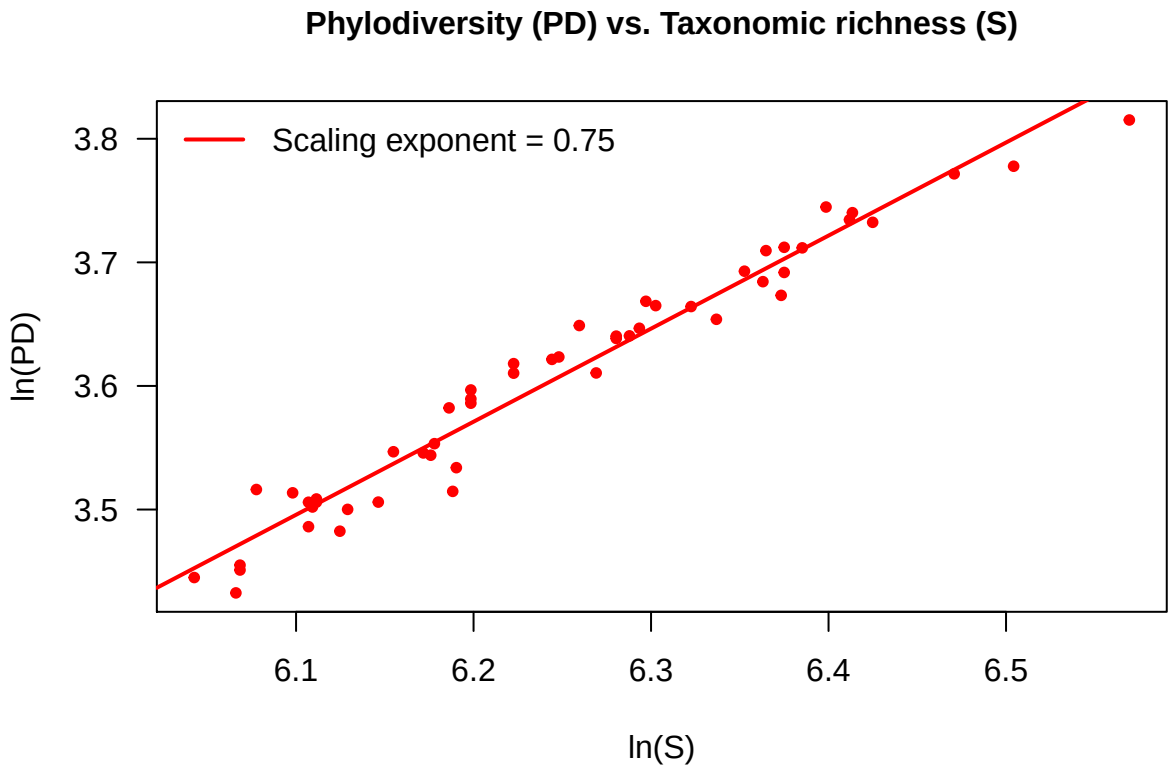
In the R code chunk below, do the following:
1. calculate Faith's D using the `pd()` function.

```
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.

```
par(mar = c(5,5,4,1) + .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)")

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, ep = ""),
       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 taxonmic 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***: Because it sums the branch length for all species, so the value of PD is directly dependent on total richness of the community. ***Answer 1b***: They are positvely correlated meaning that as richeness increases so does phylodiversity ***Answer 1c***: I would expect them to deviate if we had a large clade of closely related organisms, so we would have a large richness but a low PD ***Answer 1d***: The scaling relationship represents how our total PD increases as we add more taxa from our samples into the analysis. Since it is positive we can say that our taxa are pretty divergent since adding in new taxa we also add more phylodiversity.

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.

```
ses.pd <- ses.pd(comm[1:2,], phy, null.model = "richness", runs = 25, include.root = FALSE)
ses.pd1 <- ses.pd(comm[1:2,], phy, null.model = "taxa.labels", runs = 25, include.root = FALSE)
ses.pd2 <- ses.pd(comm[1:2,], phy, null.model = "phylogeny.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***: When calculating ses.pd we are asking if the community we observe in our sample is more or less divergent than expected by change. With the null hypotheses being that the sample is clustered, or that if the topology is randomly rearranged in some way than the resulting topologies will be show a more divergent phylogeny than ours. ***Answer 2b***: The observed values are unchanging because the model determines how it randomizes the taxa to create the expected value not how it measures the observed.

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.

```
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.

```
ses.mpd <- ses.mpd(comm, phydist, null.model = "taxa.labels",
                   abundance.weighted = TRUE, runs = 25)

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.

```
ses.mntd <- ses.mntd(comm, phydist, null.model = "taxa.labels",
                     abundance.weighted = TRUE, runs = 25)

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: In calculating NRI we are determining if the samples are clustered or overdispersed in relation to the phylogenetic distance of the taxa contained within the sample compaired to all the samples collected. This is accomplished by subtracting your randomized MPD from your observed MPD and dividing that by your expected(randomized) standard deviation. Therefore, if the value is negative (observed > expected) we can conclude that the sample is overdispersed, ie. we observe a higher MPD than random so there must be a force limiting closely related taxa within samples, and if the value is positive (observed < expected) the sample is clustered, there is some force maintaining closely related taxa within the sample. ***Answer 3b***: In calculating NTI we are determining if samples are clustered or overdispersed by determining how close they are phylogenetically to the nearest related taxa in their sample compared to randomized distances. It the math is overall math is the same as NRI but instead of MPD it utilizes MNND (how taxonomically distant are you from your nearest relative instead of average distance) ***Answer 3c***: The majority of values across both datasets are negative meaning that the taxa are mostly overclustered. ***Answer 3d***: That changes all the values to positive, meaning that within the sites there are most likely a few closely related taxa representing a major fraction of the overall community, so when abundance is factored in the overall distance across so many taxa is diminished by these few that are closely related.

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.

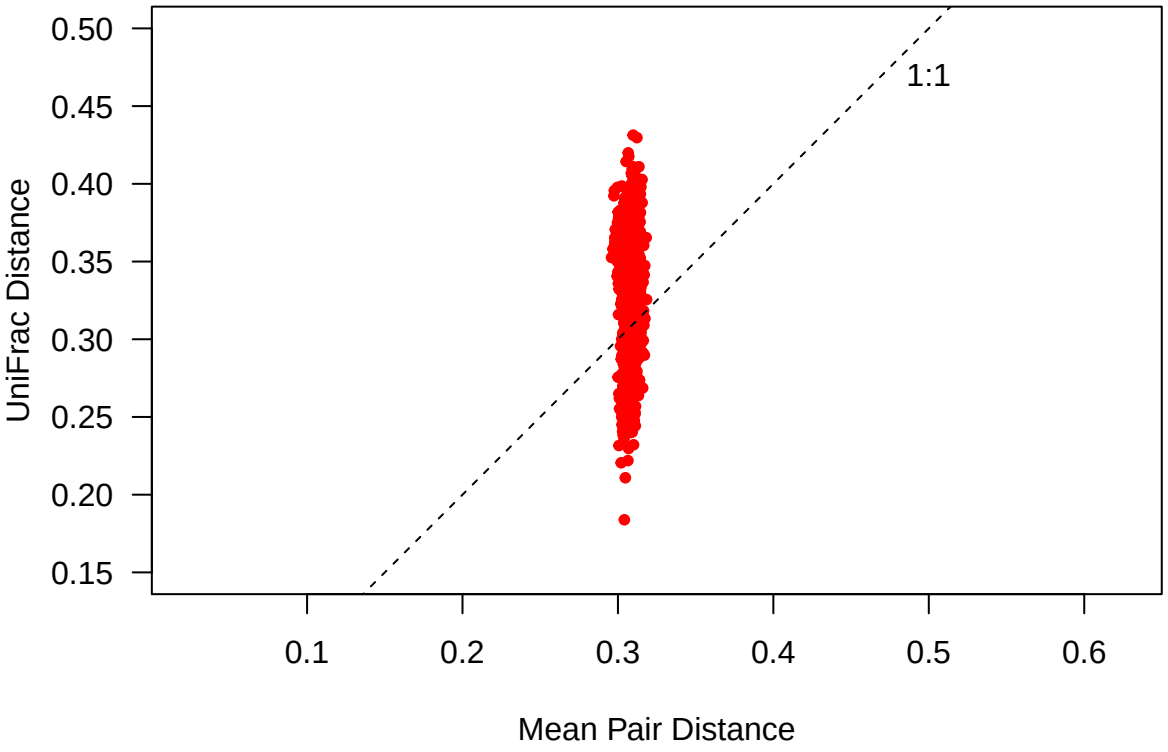
```
dist.mp <- comdist(comm, phydist)
```

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

```
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) + .1)
plot(dist.mp, dist.uf,
     pch = 20, col = "red", las = 1, asp = 1, xlim = c(.15, .5), ylim = c(.15, .5),
     xlab = "Mean Pair Distance", ylab = "UniFrac Distance")
abline(b = 1, a = 0, lty = 2)
text(.5, .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 calculating the distance between taxa based on their phylogenetic distance compared to the mean distance across all taxa while UniFrac calculates pairwise distances samples based on how much shared evolution the taxa within them have. So the major difference is utilizing mean phylogenetic distance verses using total shared evolutionary history. ***Answer 4b***: It looks as if UniFrac has a higher range of values across the samples, whiel Mean Pair Distance is pretty consistent at ~.3. Additionally, Unifrac seems to have a higher average value than Mean Pair Distance. ***Answer 4c***: Because Mean Pair Distance is determining the distance from a standardized average that is used in all measures while UniFrac generates values independent of each other based on the evolutionary history unique to each comparison.

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.

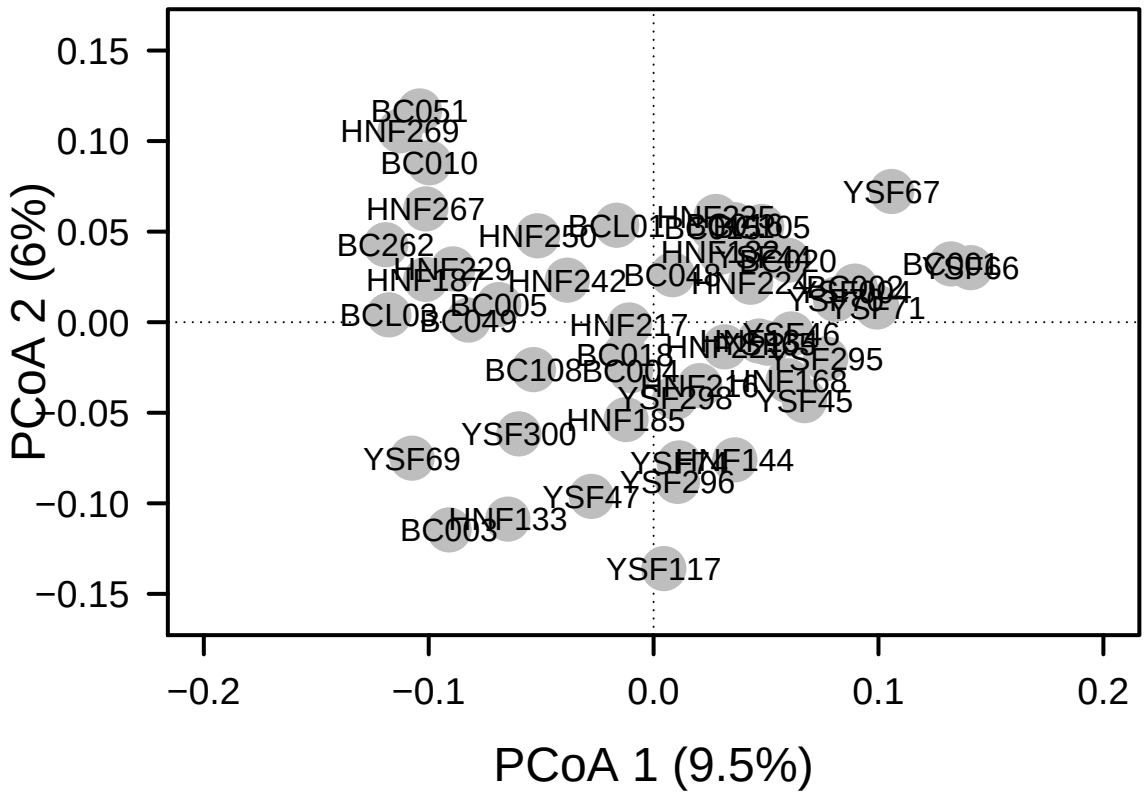
In the R code chunk below, do the following:
1. perform a PCoA based on the UniFrac distances, and
2. calculate the explained variation for the first three PCoA axes.

```
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.

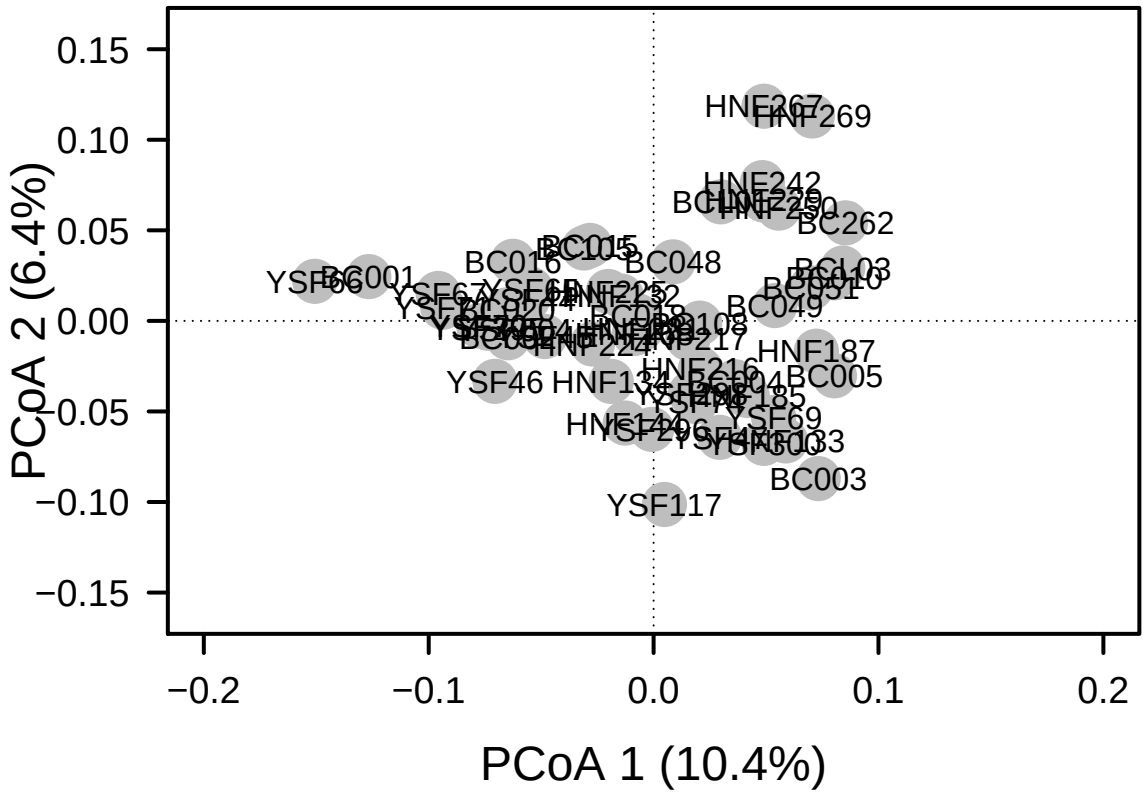
```
par(mar = c(5,5,1,2) + .1)

plot(pond.pcoa$points[,1], pond.pcoa$points[,2],
     pch = 16, cex = 2.0, type = "n", cex.lab = 1.5, cex.axis = 1.2, axes = FALSE,
     xlab = paste("PCoA 1 (", explainvar1, "%)", sep = ""),
     ylab = paste("PCoA 2 (", explainvar2, "%)", sep = ""),
     xlim = c(-.2, .2), ylim = c(-.16, .16))

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)

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.

```
#PCoA using Sorensen
dist.S <- vegdist(comm, method = "bray", binary = TRUE)

pond.pcoa <- cmdscale(dist.S, 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)

par(mar = c(5,5,1,2) + .1)

plot(pond.pcoa$points[,1], pond.pcoa$points[,2],
     pch = 16, cex = 2.0, type = "n", cex.lab = 1.5, cex.axis = 1.2, axes = FALSE,
     xlab = paste("PCoA 1 (", explainvar1, "%)", sep = ""),
     ylab = paste("PCoA 2 (", explainvar2, "%)", sep = ""),
     xlim = c(-.2, .2), ylim = c(-.16, .16))

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)
```

```
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))
```



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 difference between the percent variation explained is minimum, with Sorensen providing slightly more explanation acorss both axes but UniFrac has a bit more visual dispersion across the samples. From this I would assume that phylogenetic information isn't critical for understanding this system since the addition of it doesn't largly alter the results. (It does alter the results but not in a way that clearly, visually explain variation)

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.

```
watershed <- env$Location

adonis(dist.uf ~ watershed, permutations = 999)
```

```
##
## Call:
## adonis(formula = dist.uf ~ watershed, permutations = 999)
##
## Permutation: free
## Number of permutations: 999
##
## Terms added sequentially (first to last)
##
##           Df SumsOfSqs  MeanSqs F.Model     R2 Pr(>F)
## watershed  2   0.13316 0.066579  1.2679 0.0492  0.027 *
## Residuals 49   2.57305 0.052511         0.9508
## Total     51   2.70621                  1.0000
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
adonis(
  vegdist(
    decostand(comm, method = "log"),
    method = "bray") ~ watershed,
  permutations = 999
)
```

```
##
## Call:
## adonis(formula = vegdist(decostand(comm, method = "log"), method = "bray") ~      watershed, permuta
##
## Permutation: free
## Number of permutations: 999
##
## Terms added sequentially (first to last)
##
##           Df SumsOfSqs  MeanSqs F.Model      R2 Pr(>F)
## watershed  2   0.16601 0.083003  1.5689 0.06018  0.007 **
## Residuals 49   2.59229 0.052904         0.93982
## Total     51   2.75829                  1.00000
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

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()`).

```
envs <- env[,5:19]
envs <- envs[, -which(names(envs) %in% c("TDS", "Salinity", "Cal_Volume"))]
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.

```
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: 0.058
##
## Upper quantiles of permutations (null model):
##   90%   95% 97.5%   99%
## 0.129 0.171 0.202 0.229
## 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

```
ponds.dbrda <- dbrda(dist.uf ~ ., as.data.frame(scale(envs)))

anova(ponds.dbrda, by = "axis")
```

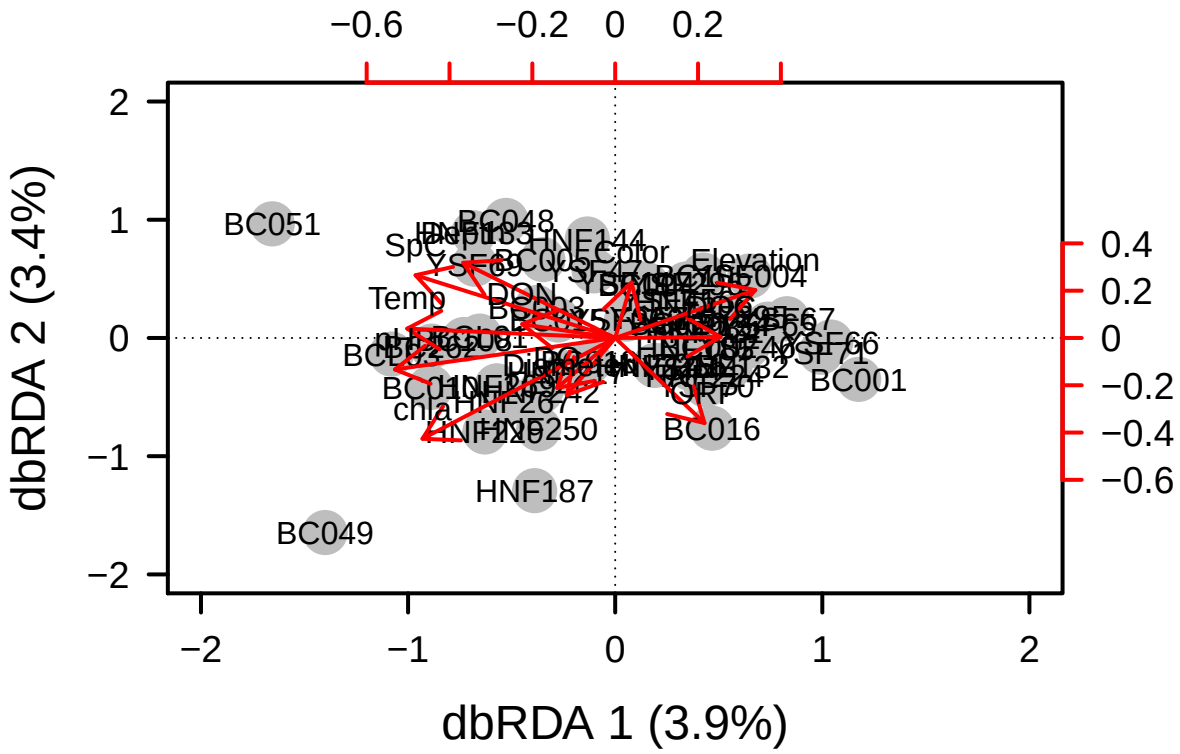
```
## Permutation test for dbrda under reduced model
## Forward tests for axes
## Permutation: free
## Number of permutations: 999
##
## Model: dbrda(formula = dist.uf ~ Elevation + Diameter + Depth + ORP + Temp + SpC + DO + pH + Color +
##          Df SumOfSqs      F Pr(>F)
## dbRDA1    1  0.10566 2.0152  0.461
## dbRDA2    1  0.09258 1.7658  0.624
## dbRDA3    1  0.07555 1.4409  0.968
## dbRDA4    1  0.06677 1.2735  0.996
## 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, permutations = 999)
ponds.fit
```

```
##
## ***VECTORS
##
##             dbRDA1   dbRDA2     r2 Pr(>r)
## Elevation  0.77670  0.62986 0.0959  0.092 .
## Diameter  -0.27972 -0.96008 0.0541  0.247
## Depth     -0.63137  0.77548 0.1756  0.011 *
## ORP        0.41879 -0.90808 0.1437  0.023 *
## Temp      -0.98250  0.18628 0.1523  0.016 *
## SpC       -0.77101  0.63682 0.2087  0.009 **
## DO        -0.39318 -0.91946 0.0464  0.306
## pH        -0.96210 -0.27270 0.1756  0.009 **
## Color      0.06353  0.99798 0.0464  0.325
## chla      -0.60392 -0.79704 0.2626  0.005 **
## DOC        0.99847 -0.05526 0.0382  0.363
## 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
```

```
#Calculating explained variation on axes
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

#Plotting constrained ordination results
par(mar = c(5,5,4,4) + .1)

plot(scores(ponds.dbrda, display = "wa"), xlim = c(-2, 2.), ylim = c(-2.0, 2.0),
     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
     )

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)

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")))

#Plotting vectors for influence of environmental factors
vectors <- scores(ponds.dbrda, display = "bp")
arrows(0, 0, vectors[,1] * 2, vectors[,2] * 2,
       lwd = 2, lty = 1, length = .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***: It seems as if phylogenetic relationship across ponds can be predicted by watershed and several other environmental variables (Depth, oxidation reduction potential, Temp, Specific conductivity of water, pH, and chlorophyll concentration)

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.

> ***Synthesis***: For my research, specifically host-microbe interactions, phylogenetic comparisons are very common and simple to create since amplicon sequincing is the most common method used

for identification of different taxa and can also be used to measure phylogenetic distance between taxa. UniFrac is frequently used to measure differentiation across samples, but considering the massive diversity that can be hidden within taxa with similar amplicon sequences conclusions from these aren't as strait forward in macroorganisms. Phylogenetics of microbes can be used interestingly is phylosymbiosis studies, or for answering questions of how microbes are associating with hosts across evolutionary distance. This could be done by measuring how different communities are are generally across host phylogeny or how evolutionary divergent specific microbial taxa are compared to the evolutionary divergence of their hosts.