Phylogenetic Diversity - Traits

Michelle Benavidez; Z620: Quantitative Biodiversity, Indiana University

26 February, 2017

OVERVIEW

Up to this point, we have been focusing on patterns taxonomic diversity in Quantitative Biodiversity. Although taxonomic diversity is an important dimension of biodiversity, it is often necessary to consider the evolutionary history or relatedness of species. The goal of this exercise is to introduce basic concepts of phylogenetic diversity.

After completing this exercise you will be able to:

1. create phylogenetic trees to view evolutionary relationships from sequence data
2. map functional traits onto phylogenetic trees to visualize the distribution of traits with respect to evolutionary history
3. test for phylogenetic signal within trait distributions and trait-based patterns of biodiversity

Directions:

1. Change “Student Name” on line 3 (above) with your name.
2. Complete as much of the exercise as possible during class; what you do not complete in class will need to be done on your own 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. For homework, follow the directions at the bottom of this file.
7. When you are done, **Knit** the text and code into a PDF file.
8. After Knitting, please submit the completed exercise by creating a **pull request** via GitHub. Your pull request should include this file *PhyloTraits_exercise.Rmd* and the PDF output of `Knitr` (*PhyloTraits_exercise.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 “*/Week6-PhyloTraits*” folder, and
4. load all of the required R packages (be sure to install if needed).

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

```
## [1] "C:/Users/Michelle/GitHub/QB2017_Benavidez/Week6-PhyloTraits"
```

```r
setwd("C:/Users/Michelle/GitHub/QB2017_Benavidez/Week6-PhyloTraits")
package.list <- c("ape","seqinr","phylobase","adephylo","geiger","picante","stats","RColorBrewer","cape
for (package in package.list){
  if (!require(package, character.only = T, quietly = T)){
    install.packages(package)
    library(package, character.only = T)
  }
}
```

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

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

##
## Attaching package: 'phylobase'

## The following object is masked from 'package:ape':
##
##     edges

##
## Attaching package: 'adephylo'

## The following object is masked from 'package:ade4':
##
##     orthogram

##
## Attaching package: 'permute'

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

## This is vegan 2.4-2

##
## Attaching package: 'vegan'

## The following object is masked from 'package:ade4':
##
##     cca

##
## Attaching package: 'nlme'

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

##
## Attaching package: 'dplyr'

## The following object is masked from 'package:MASS':
##
##     select

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

```
##
##     collapse

## The following objects are masked from 'package:seqinr':
##
##     count, query

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union

##
## Attaching package: 'phangorn'

## The following objects are masked from 'package:vegan':
##
##     diversity, treedist
```

2) DESCRIPTION OF DATA

The maintenance of biodiversity is thought to be influenced by **trade-offs** among species in certain functional traits. One such trade-off involves the ability of a highly specialized species to perform exceptionally well on a particular resource compared to the performance of a generalist. In this exercise, we will take a phylogenetic approach to mapping phosphorus resource use onto a phylogenetic tree while testing for specialist-generalist trade-offs.

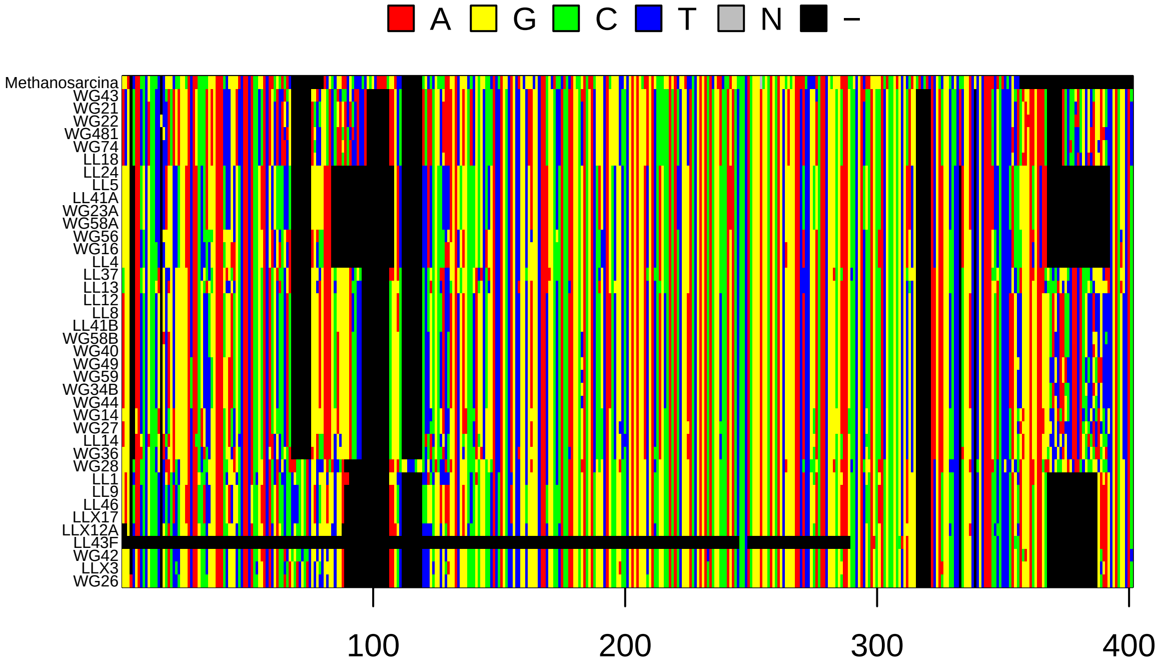
3) SEQUENCE ALIGNMENT

Question 1: Using less or your favorite text editor, compare the `p.isolates.fasta` file and the `p.isolates.afa` file. Describe the differences that you observe between the files.

> ***Answer 1***: The fasta file looked clear. Seemed to compress the sequences so that the missing fragments weren't empty data points.

In the R code chunk below, do the following: 1. read your alignment file, 2. convert the alignment to a DNAbin object, 3. select a region of the gene to visualize (try various regions), and 4. plot the alignment using a grid to visualize rows of sequences.

```
read.aln <- read.alignment(file = "./data/p.isolates.afa", format = "fasta")

p.DNAbin <- as.DNAbin(read.aln)

window <- p.DNAbin[, 100:500]

image.DNAbin(window, cex.lab = 0.50)
```



Question 2: Make some observations about the `muscle` alignment of the 16S rRNA gene sequences for our bacterial isolates and the outgroup, *Methanosarcina*, a member of the domain archaea. Move along the alignment by changing the values in the `window` object.

a. Approximately how long are our reads?
b. What regions do you think would are appropriate for phylogenetic inference and why?

> ***Answer 2a***: ~400bp ***Answer 2b***: The first ~75bp and the area before ~400bp, because this is where most of the variation seems to exist

4) MAKING A PHYLOGENETIC TREE

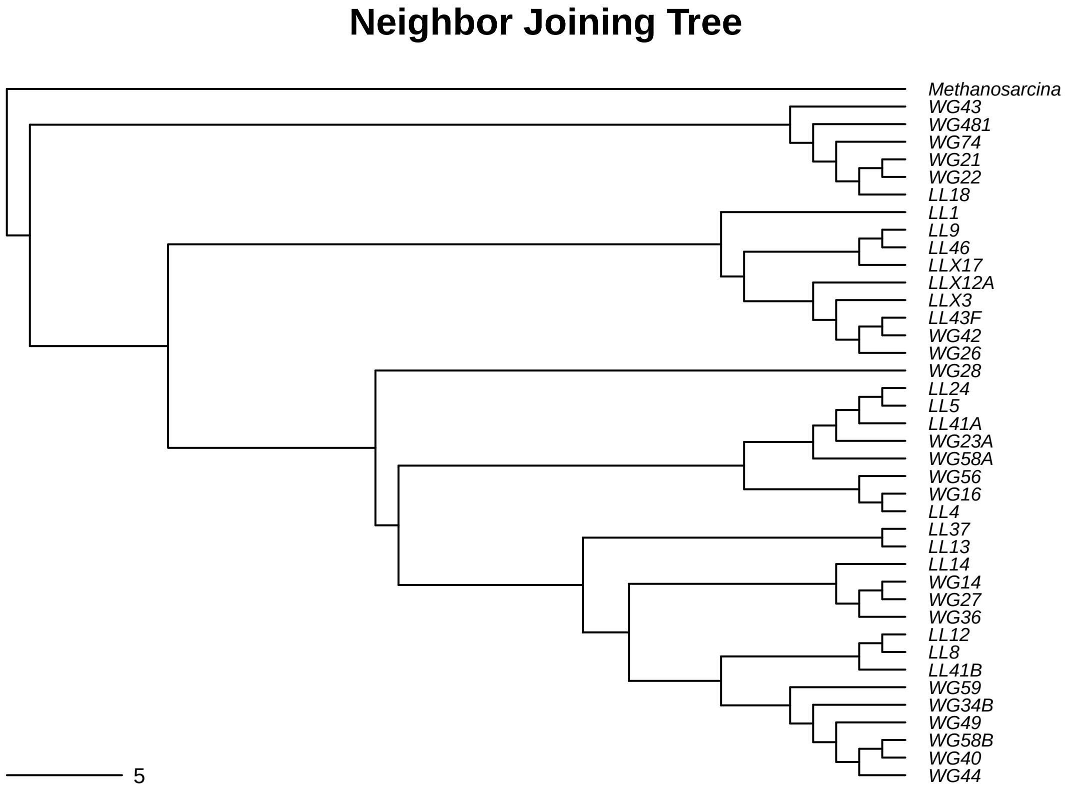
Once you have aligned your sequences, the next step is to construct a phylogenetic tree. Not only is a phylogenetic tree effective for visualizing the evolutionary relationship among taxa, but as you will see later, the information that goes into a phylogenetic tree is needed for downstream analysis.

A. Neighbor Joining Trees

In the R code chunk below, do the following:
1. calculate the distance matrix using `model = "raw"`,
2. create a Neighbor Joining tree based on these distances,
3. define "Methanosarcina" as the outgroup and root the tree, and
4. plot the rooted tree.

```
seq.dist.raw <- dist.dna(p.DNAbin, model = "raw", pairwise.deletion = F)
```

```
nj.tree <- bionj(seq.dist.raw)
outgroup <- match("Methanosarcina", nj.tree$tip.label)
nj.rooted <- root(nj.tree, outgroup, resolve.root = T)

par(mar = c(1,1,2,1) + 0.1)
plot.phylo(nj.rooted, main = "Neighbor Joining Tree", "phylogram",
           use.edge.length = F, direction = "right", cex = 0.6,
           label.offset = 1)
add.scale.bar(cex = 0.7)
```



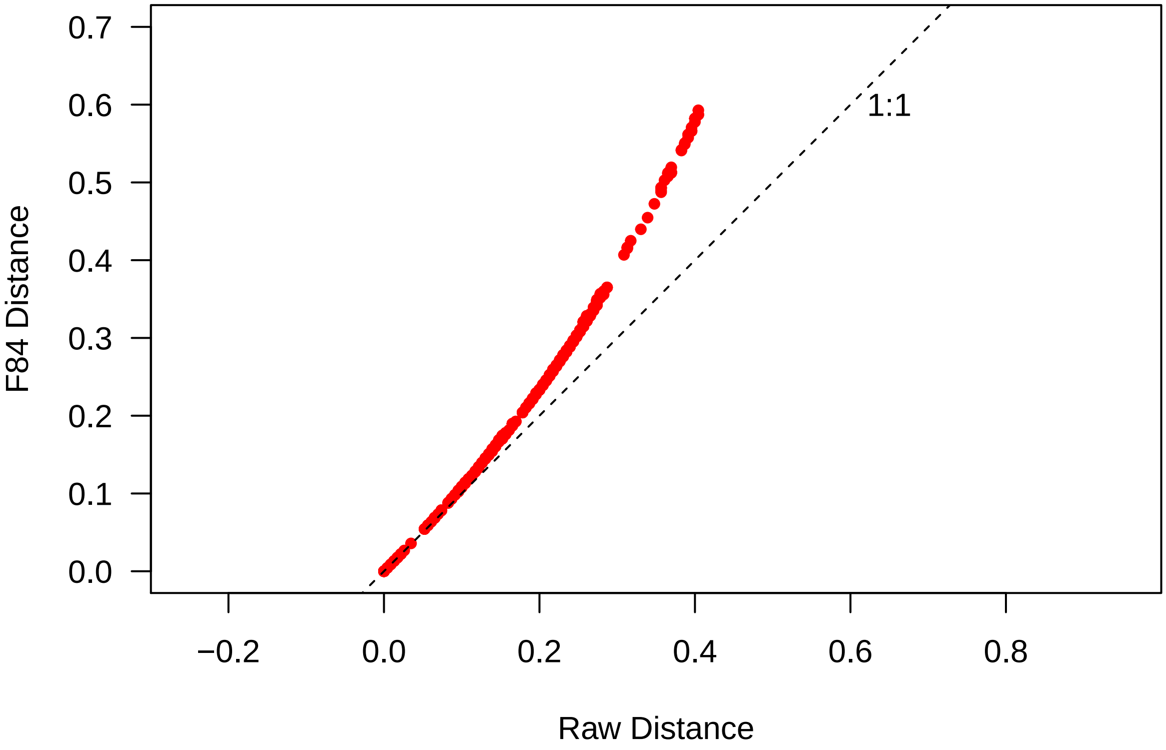
Question 3: What are the advantages and disadvantages of making a neighbor joining tree?

> ***Answer 3***: Advantages: This process analyzes a large number of sequences very quickly. Can use plain(?) reads to build the tree and it corrects for multiple substitutions. Disadvantages: Information in the reads are greatly reduced. The algorithm used for star decomposition does not exhaust all the possibilities so likley does not give the more optimal solution (but it does give a good approximation); therefore, we only get to see one tree out of several alternatives. Only serves as a starting point.

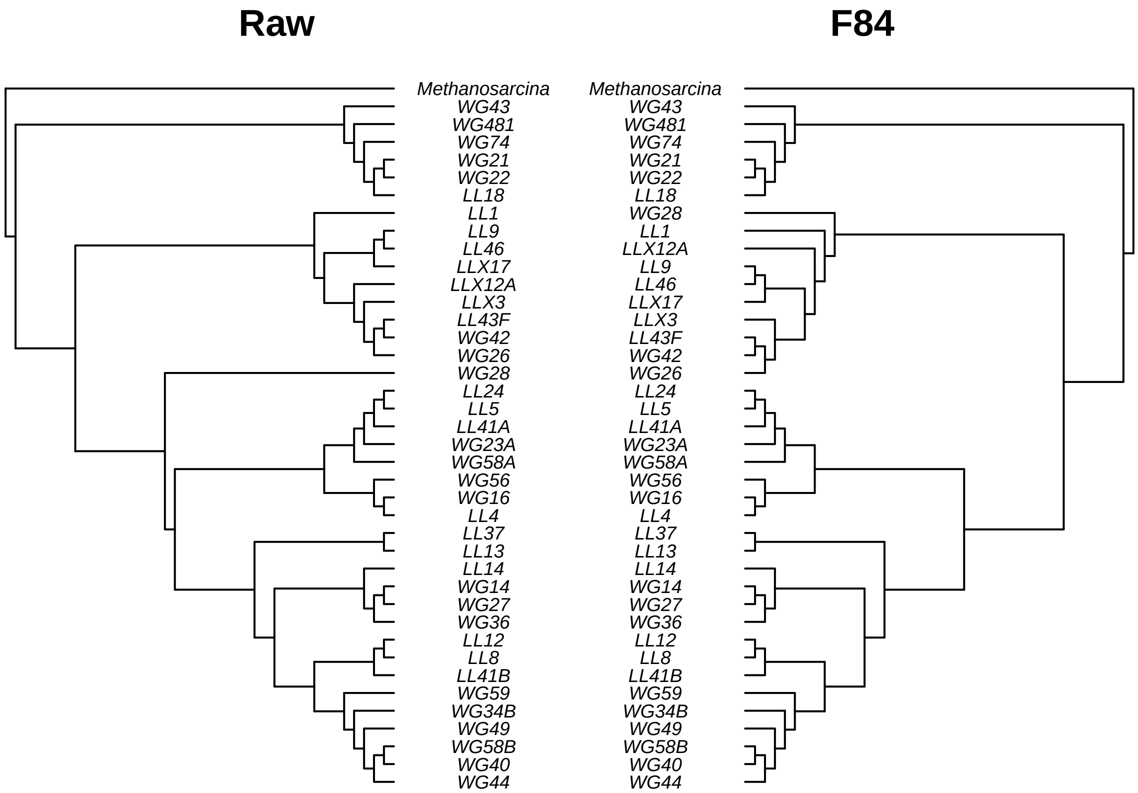
B) SUBSTITUTION MODELS OF DNA EVOLUTION

In the R code chunk below, do the following:
1. make a second distance matrix based on the Felsenstein 84 substitution model,
2. create a saturation plot to compare the *raw* and *Felsenstein (F84)* substitution models,
3. make Neighbor Joining trees for both, and
4. create a cophylogenetic plot to compare the topologies of the trees.

```
seq.dist.F84 <- dist.dna(p.DNAbin, model = "F84", pairwise.deletion = F)

par(mar = c(5,5,2,1) + 0.1)
plot(seq.dist.raw, seq.dist.F84,
     pch = 20, col = "red", las = 1, asp = 1, xlim = c(0,0.7), ylim = c(0, 0.7),
     xlab = "Raw Distance", ylab = "F84 Distance")
abline(b = 1, a = 0, lty = 2)
text(0.65, 0.6, "1:1")
```

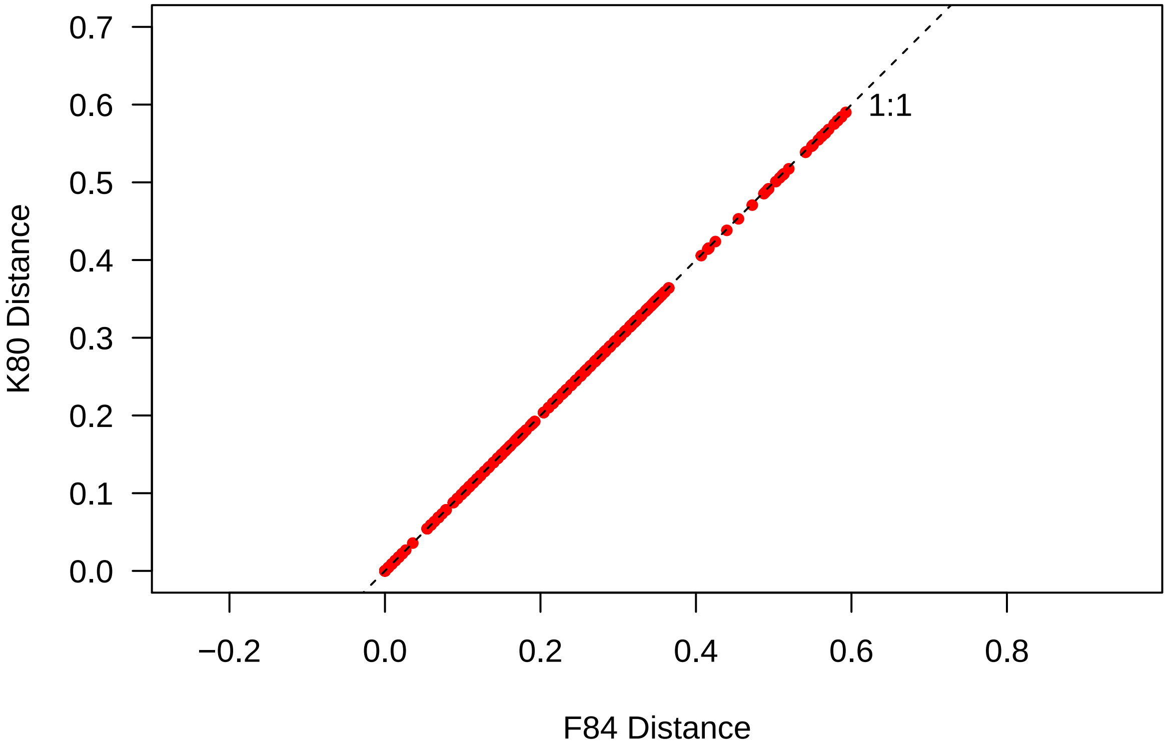


```
raw.tree <- bionj(seq.dist.raw)
F84.tree <- bionj(seq.dist.F84)

raw.outgroup <- match("Methanosarcina", raw.tree$tip.label)
F84.outgroup <- match("Methanosarcina", F84.tree$tip.label)

raw.rooted <- root(raw.tree, raw.outgroup, resolve.root = T)
F84.rooted <- root(F84.tree, F84.outgroup, resolve.root = T)

layout(matrix(c(1,2),1,2), width = c(1,1))
par(mar = c(1,1,2,0))
plot.phylo(raw.rooted, type = "phylogram", direction = "right", show.tip.label = T,
           use.edge.length = F, adj = 0.5, cex = 0.6, label.offset = 2, main = "Raw")

par(mar = c(1,0,2,1))
plot.phylo(F84.rooted, type = "phylogram", direction = "left", show.tip.label = T,
           use.edge.length = F, adj = 0.5, cex = 0.6, label.offset = 2, main = "F84")
```

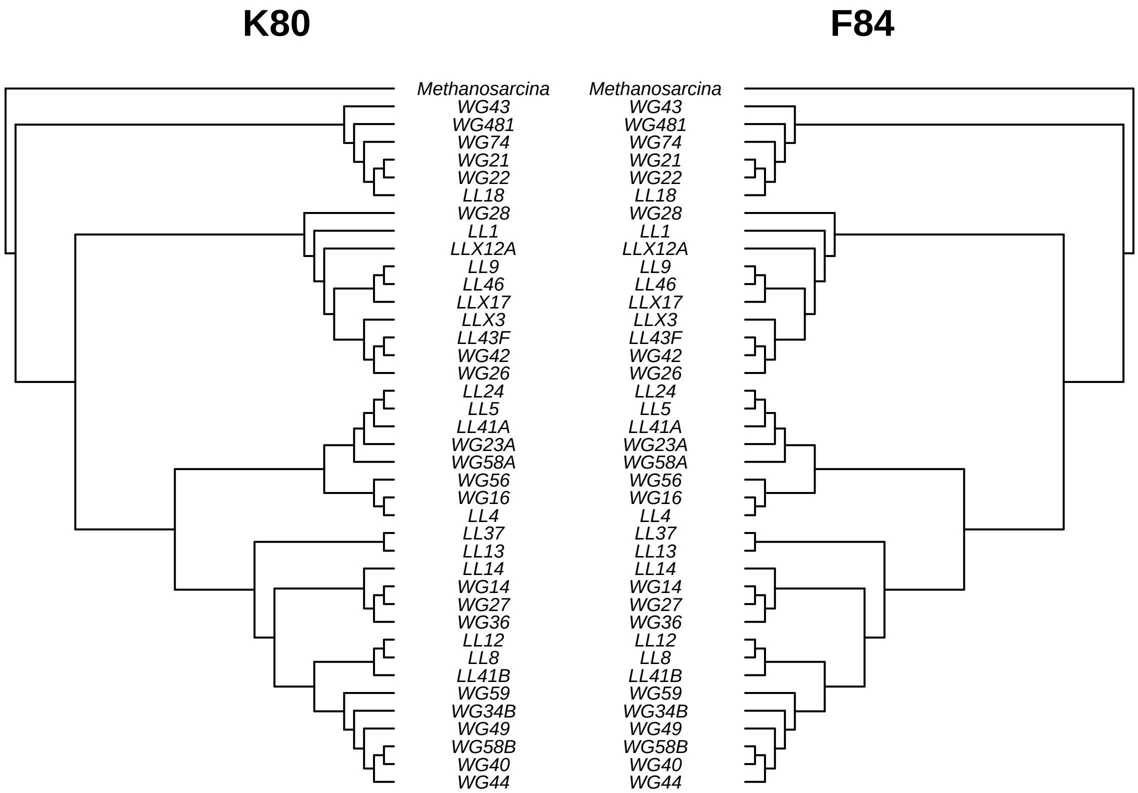


In the R code chunk below, do the following:
1. pick another substitution model,
2. create and distance matrix and tree for this model,
3. make a saturation plot that compares that model to the *Felsenstein (F84)* model,
4. make a cophylogenetic plot that compares the topologies of both models, and
5. be sure to format, add appropriate labels, and customize each plot.

```
seq.dist.K80 <- dist.dna(p.DNAbin, model = "K80", pairwise.deletion = F)

par(mar = c(5,5,2,1) + 0.1)
plot(seq.dist.F84, seq.dist.K80,
     pch = 20, col = "red", las = 1, asp = 1, xlim = c(0,0.7), ylim = c(0, 0.7),
     xlab = "F84 Distance", ylab = "K80 Distance")
abline(b = 1, a = 0, lty = 2)
text(0.65, 0.6, "1:1")
```



```r
K80.tree <- bionj(seq.dist.K80)
F84.tree <- bionj(seq.dist.F84)

K80.outgroup <- match("Methanosarcina", K80.tree$tip.label)
F84.outgroup <- match("Methanosarcina", F84.tree$tip.label)

K80.rooted <- root(K80.tree, K80.outgroup, resolve.root = T)
F84.rooted <- root(F84.tree, F84.outgroup, resolve.root = T)

layout(matrix(c(1,2),1,2), width = c(1,1))
par(mar = c(1,1,2,0))
plot.phylo(K80.rooted, type = "phylogram", direction = "right", show.tip.label = T,
           use.edge.length = F, adj = 0.5, cex = 0.6, label.offset = 2, main = "K80")

par(mar = c(1,0,2,1))
plot.phylo(F84.rooted, type = "phylogram", direction = "left", show.tip.label = T,
           use.edge.length = F, adj = 0.5, cex = 0.6, label.offset = 2, main = "F84")
```



Question 4:

a. Describe the substitution model that you chose. What assumptions does it make and how does it compare to the F84 model?
b. Using the saturation plot and cophylogenetic plots from above, describe how your choice of substitution model affects your phylogenetic reconstruction. If the plots are inconsistent with one another, explain why.
c. How does your model compare to the *F84* model and what does this tell you about the substitution rates of nucleotide transitions?

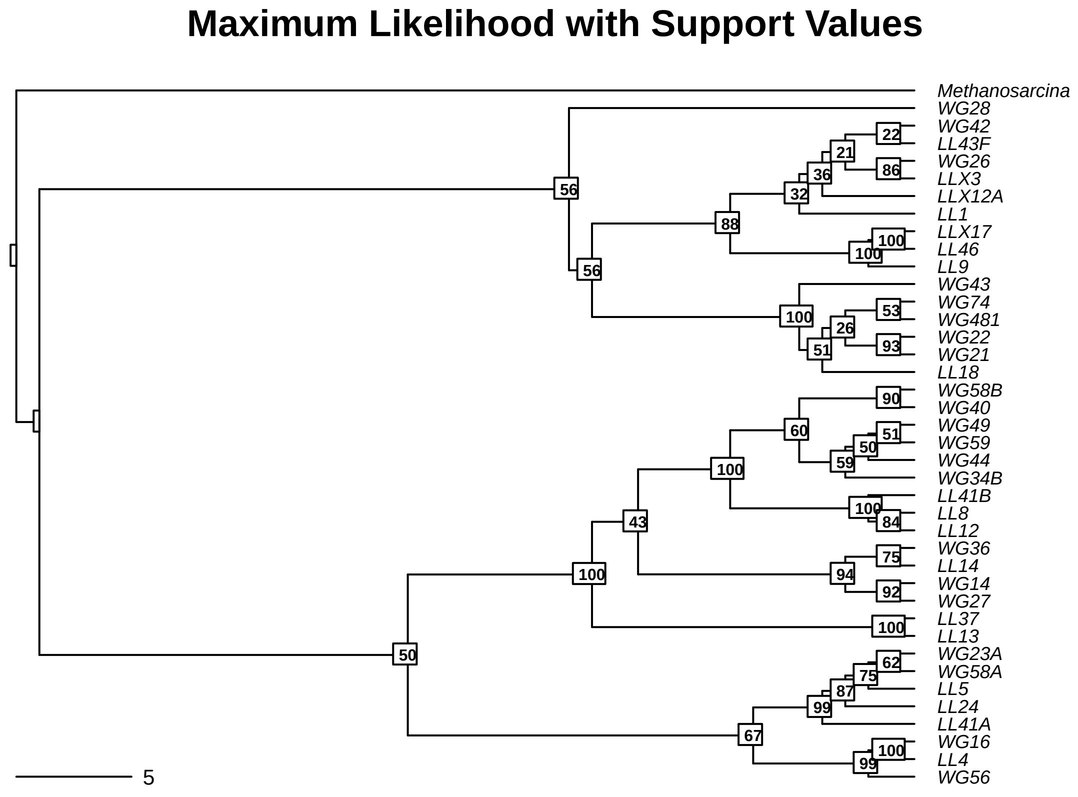
> ***Answer 4a***: I chose the K80 (Kimura Model) model. This model basically distiguishes between transition and transversions - particularly it recognizes that transversion mutations will occur with lower probability than transition mutations. The main assumption of this is equal frequencies of nucleotides - so all the bases in the sequences are equally frequent.
> ***Answer 4b***: These plots seems consistent with one another. I see no apparent differences.
> ***Answer 4c***: The only dicernable difference between K80 and F84 is that, unlike K80, F84 allows for differences in base frequencies. Since there is not a noticable difference between the two models, then perhaps there are not detectable differences in nucleotide frequences in the reads.

C) ANALYZING A MAXIMUM LIKELIHOOD TREE

In the R code chunk below, do the following:
1. Read in the maximum likelihood phylogenetic tree used in the handout. 2. Plot bootstrap support values onto the tree

```
p.DNAbin.phyDat <- read.phyDat("./data/p.isolates.afa", format = "fasta", type = "DNA")
#fit <- pml(nj.rooted, data=p.DNAbin.phyDat)
```

```r
#fitJC <- optim.pml(fit, T)

#fitGTR <- optim.pml(fitGTR, model = "GTR", optInv = T, optGamma = T,
                    #rearrangement = "NNI", control = pml.control(trace = 0))

#anova(fitJC, fitGTR)
#AIC(fitJC)
#AIC(fitGTR)

#bs = botstrap.pml(fitJC, bs = 100, optNni = T,
                  #control = pml.control(trace = 0))


ml.bootstrap <- read.tree("./data/ml_tree/RAxML_bipartitions.T1")
par(mar = c(1,1,2,1) + 0.1)
plot.phylo(ml.bootstrap, type = "phylogram", direction = "right",
           show.tip.label = T, use.edge.length = F, cex = 0.6,
           label.offset = 1, main = "Maximum Likelihood with Support Values")
add.scale.bar(cex = 0.7)
nodelabels(ml.bootstrap$node.label, font = 2, bg = "white", frame = "r",
           cex = 0.5)
```



Question 5:

a) How does the maximum likelihood tree compare the to the neighbor-joining tree in the handout? If the plots seem to be inconsistent with one another, explain what gives rise to the differences.

b) Why do we bootstrap our tree?

c) What do the bootstrap values tell you?

d) Which branches have very low support?

e) Should we trust these branches?

Answer 5a: The plots are inconsistent, primarily with placment of the tree members. Discrepancy seems to primarily appear in earlier branching, while the more recent branching are somewhat consistent. The process by which the tree if formed between the two methods gave rise to these differences.
Answer 5b: We bootstrap our tree in order to get a more robust appoximation of the relability of the tree structure.
Answer 5c: Bootstrap values give an approximation of the liklihood that, upon repeated sampling, you would get the same branching. For example, if you had a bootstrap value of 94 then upon repeated sampling you would get the same results 94% of the time. ***Answer 5d***: Branches which have low numerical numbers associated with them render the least support. Branches with a numerical value of >/= 95 can likely be interpreted as operationally correct.
Answer 5e: Some branches are more reliable than others. Overall, however, the earliest branches are only 50 - 56 % reliable. Drawing any inference from this tree should take in to consideration the bootstrap values that associated to the branching to which you generating the information. In the least, we can trust this tree more than the neighbor-joining tree.

5) INTEGRATING TRAITS AND PHYLOGENY

A. Loading Trait Database

In the R code chunk below, do the following:
1. import the raw phosphorus growth data, and
2. standardize the data for each strain by the sum of growth rates.

```r
p.growth <- read.table("./data/p.isolates.raw.growth.txt", sep = "\t", header = T,
                       row.names = 1)

p.growth.std <- p.growth / (apply(p.growth, 1, sum))
```

B. Trait Manipulations

In the R code chunk below, do the following:
1. calculate the maximum growth rate (μ_{max}) of each isolate across all phosphorus types,
2. create a function that calculates niche breadth (nb), and
3. use this function to calculate nb for each isolate.

```r
umax <- (apply(p.growth, 1, max))

levins <- function(p_xi = ""){
  p = 0
  for (i in p_xi){
    p = p + i^2
  }
  nb = 1 / (length(p_xi) * p)
  return(nb)
}
```

```
nb <- as.matrix(levins(p.growth.std))

rownames(nb) <- row.names(p.growth)
colnames(nb) <- c("NB")
```

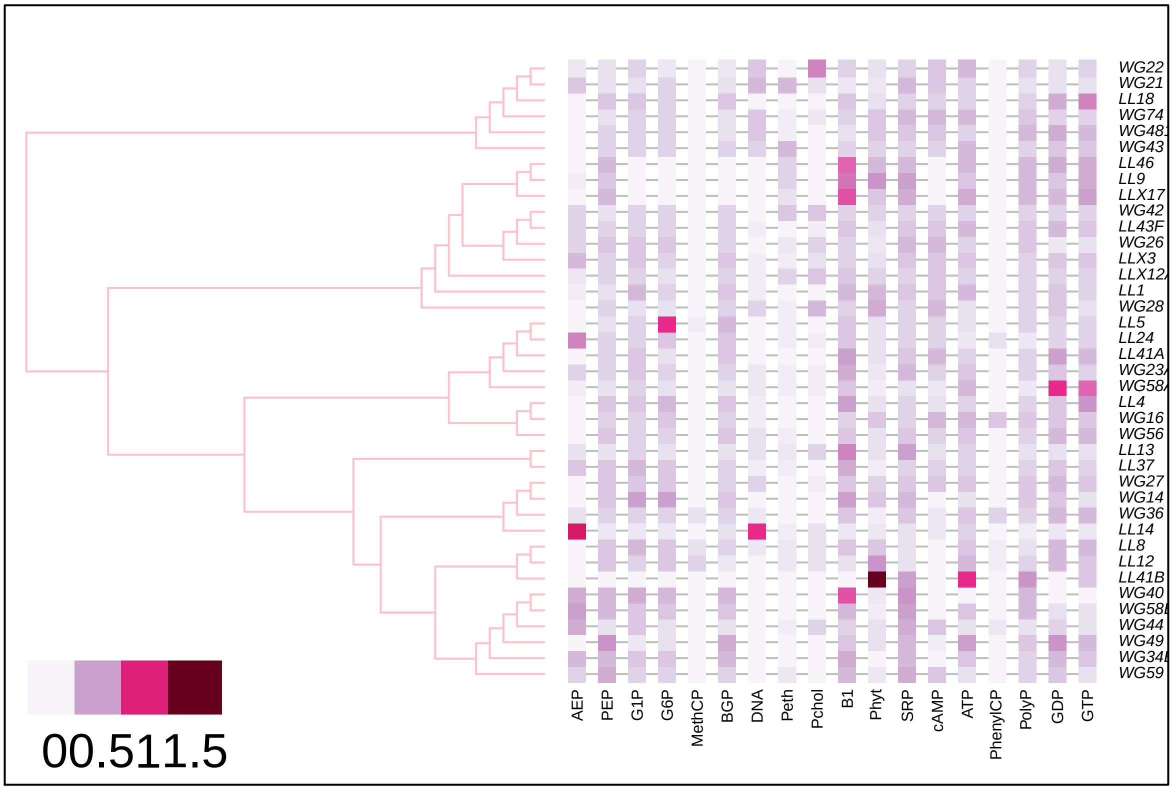
C. Visualizing Traits on Trees

In the R code chunk below, do the following:
1. pick your favorite substitution model and make a Neighbor Joining tree,
2. define your outgroup and root the tree, and
3. remove the outgroup branch.

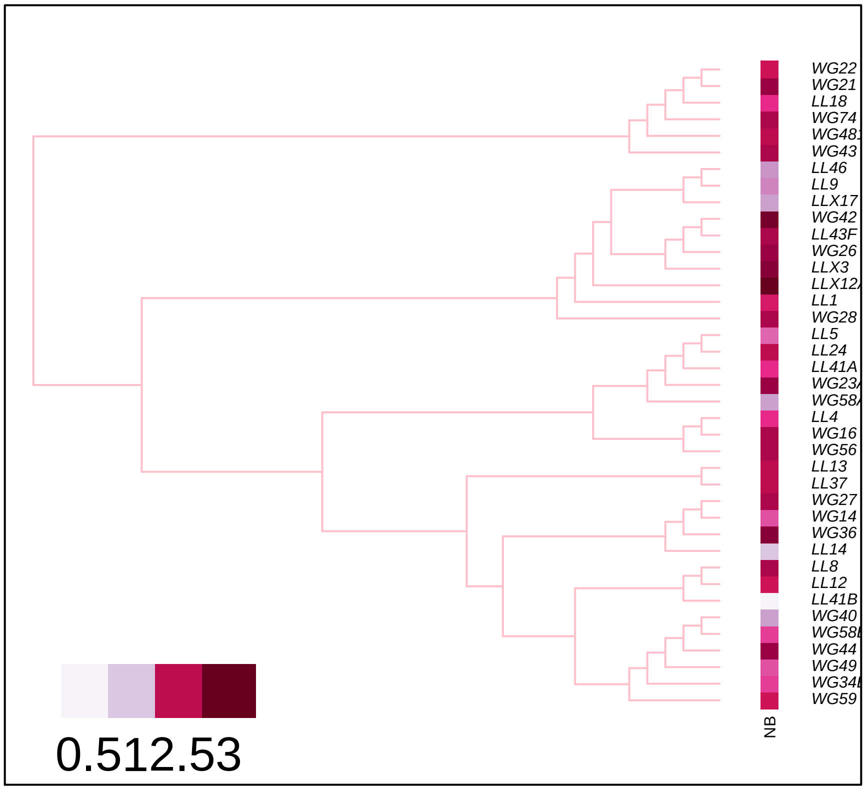
```
nj.tree <- bionj(seq.dist.F84)

outgruop <- match("Methanosarcina", nj.tree$tip.label)

nj.rooted <- root(nj.tree, outgroup, resolve.root =T)

nj.rooted <- drop.tip(nj.rooted, "Methanosarcina")
```

In the R code chunk below, do the following:
1. define a color palette (use something other than "YlOrRd"),
2. map the phosphorus traits onto your phylogeny,
3. map the *nb* trait on to your phylogeny, and
4. customize the plots as desired (use `help(table.phylo4d)` to learn about the options).

```
mypalette <- colorRampPalette(brewer.pal(9, "PuRd"))

par(mar=c(1,1,1,1) + 0.1)
x <- phylo4d(nj.rooted, p.growth.std)

table.phylo4d(x, treetype = "phylo", symbol = "colors", show.node = T,
              cex.label = 0.50, scale = F, use.edge.length = F,
              edge.color = "pink", edge.width = 1, box = T,
              col=mypalette(25), pch = 15, cex.symbol = 1.25,
              ratio.tree = 0.5, cex.legend = 3, center = F)
```



```
par(mar = c(1,5,1,5) + 0.1)
x.nb <- phylo4d(nj.rooted, nb)
table.phylo4d(x.nb, treetype = "phylo", symbol = "colors", show.node = T,
              cex.label = 0.5, scale = F, use.edge.length = F,
              edge.color = "pink", edge.width = 1, box = T,
              col=mypalette(25), pch = 15, cex.symbol = 1.25, var.label = ("NB"),
              ratio.tree = 0.9, cex.legend = 3, center = F)
```



Question 6:

a) Make a hypothesis that would support a generalist-specialist trade-off.

b) What kind of patterns would you expect to see from growth rate and niche breadth values that would support this hypothesis?

Answer 6a: If tradeoffs exist between niche breadth and growth, then those bacteria that possess a wider nidth breath will be able to process a wider variety of phosophorus types than those that have a narrow and more specialized niche breadth.
Answer 6b: I would expect that generalists (with wider niche breadth) will be smaller and grow more slowly compared to specialist species (narrow niche breadth).

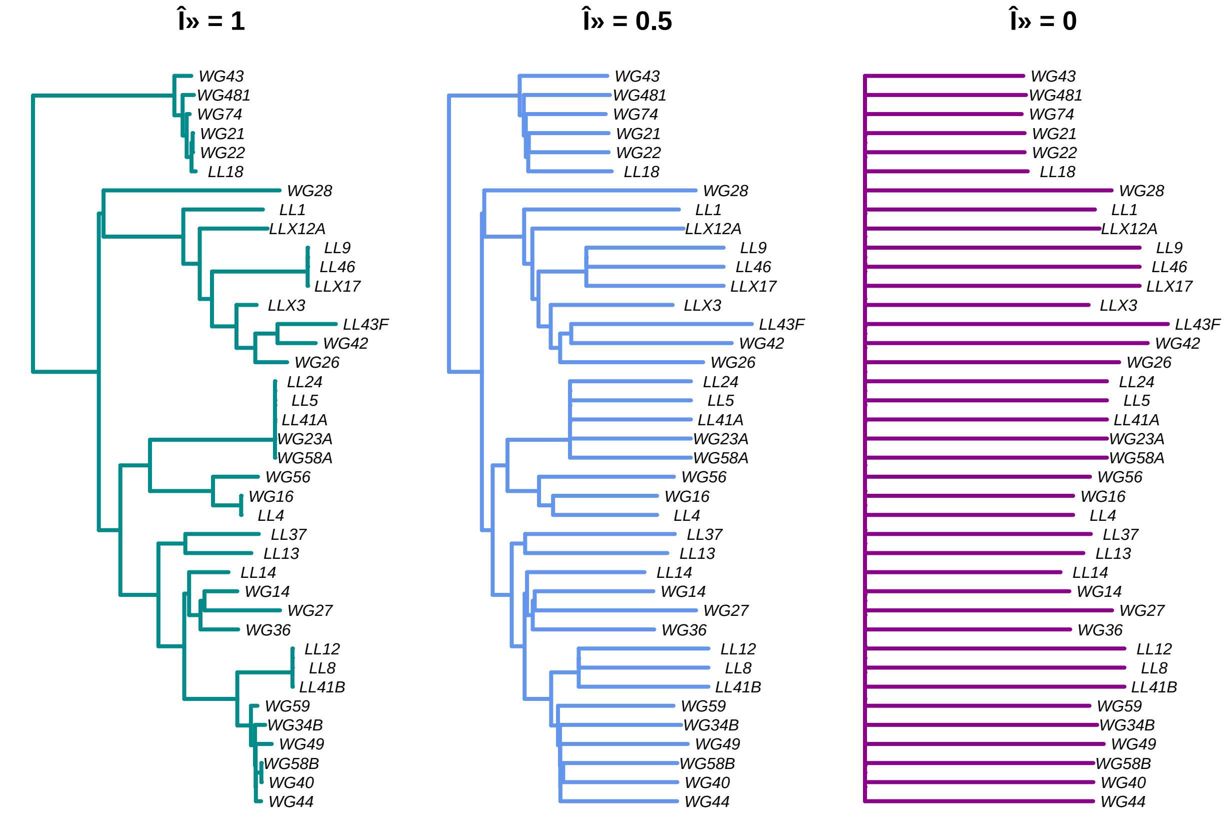
6) HYPOTHESIS TESTING

A) Phylogenetic Signal: Pagel's Lambda

In the R code chunk below, do the following:
1. create two rescaled phylogenetic trees using lambda values of 0.5 and 0,
2. plot your original tree and the two scaled trees, and
3. label and customize the trees as desired.

```r
nj.lambda.5 <- rescale(nj.rooted, "lambda", 0.5)
nj.lambda.0 <- rescale(nj.rooted, "lambda", 0)

layout(matrix(c(1,2,3), 1, 3), width = c(1,1,1))
par(mar = c(1,0.5,2, 0.5) + 0.1)
```

```r
plot(nj.rooted, main = "Î» = 1", edge.color = "darkcyan", edge.width = 2, cex = 0.7, adj = 0.5)
plot(nj.lambda.5, main = "Î» = 0.5", edge.color = "cornflowerblue", edge.width = 2, cex = 0.7, adj = 0.5
plot(nj.lambda.0, main = "Î» = 0", edge.color = "darkmagenta", edge.width = 2, cex = 0.7, adj = 0.5)
```



In the R code chunk below, do the following:
1. use the `fitContinuous()` function to compare your original tree to the transformed trees.

```r
fitContinuous(nj.rooted, nb, model = "lambda")
```

```
## GEIGER-fitted comparative model of continuous data
##  fitted 'lambda' model parameters:
##  lambda = 0.020847
##  sigsq = 0.106492
##  z0 = 0.661368
##
##  model summary:
##  log-likelihood = 21.661104
##  AIC = -37.322208
##  AICc = -36.636494
##  free parameters = 3
##
## Convergence diagnostics:
##  optimization iterations = 100
##  failed iterations = 54
##  frequency of best fit = NA
##
##  object summary:
```

```
##  'lik' -- likelihood function
##  'bnd' -- bounds for likelihood search
##  'res' -- optimization iteration summary
##  'opt' -- maximum likelihood parameter estimates
```

```
fitContinuous(nj.lambda.0, nb, model = "lambda")
```

```
## GEIGER-fitted comparative model of continuous data
##  fitted 'lambda' model parameters:
##  lambda = 0.000000
##  sigsq = 0.106395
##  z0 = 0.657777
##
##  model summary:
##  log-likelihood = 21.652293
##  AIC = -37.304587
##  AICc = -36.618872
##  free parameters = 3
##
## Convergence diagnostics:
##  optimization iterations = 100
##  failed iterations = 0
##  frequency of best fit = 0.90
##
##  object summary:
##  'lik' -- likelihood function
##  'bnd' -- bounds for likelihood search
##  'res' -- optimization iteration summary
##  'opt' -- maximum likelihood parameter estimates
```

Question 7: There are two important outputs from the `fitContinuous()` function that can help you interpret the phylogenetic signal in trait data sets. a. Compare the lambda values of the untransformed tree to the transformed (lambda = 0). b. Compare the Akaike information criterion (AIC) scores of the two models. Which model would you choose based off of AIC score (remember the criteria that the difference in AIC values has to be at least 2)? c. Does this result suggest that there's phylogenetic signal?

> ***Answer 7a***: Lambda values from the two trees are similar. While value given from the rooted tree equals approximately 0.02, the value of the transformed tree is 0. ***Answer 7b***: Based on the AIC scores, there is not a decernible difference between the two models. ***Answer 7c***: The results do not suggest that there is a phylogenetic signal.

B) Phylogenetic Signal: Blomberg's K

In the R code chunk below, do the following:
1. correct tree branch-lengths to fix any zeros,
2. calculate Blomberg's K for each phosphorus resource using the `phylosignal()` function,
3. use the Benjamini-Hochberg method to correct for false discovery rate, and
4. calculate Blomberg's K for niche breadth using the `phylosignal()` function.

```
nj.rooted$edge.length <- nj.rooted$edge.length + 10^-7

p.phylosignal <- matrix(NA, 6, 18)
colnames(p.phylosignal) <- colnames(p.growth.std)
rownames(p.phylosignal) <- c("K", "PIC.var.obs", "PIC.var.mean",
                             "PIC.var.P", "PIC.var.z", "PIC.P.BH")
```

```
for (i in 1:18){
  x <- as.matrix(p.growth.std[,i, drop = F])
  out <- phylosignal(x, nj.rooted)
  p.phylosignal[1:5, i] <- round(t(out), 3)
}

p.phylosignal[6,] <- round(p.adjust(p.phylosignal[4,], method = "BH"), 3)

signal.nb <- phylosignal(nb, nj.rooted)
signal.nb
```

```
##              K PIC.variance.obs PIC.variance.rnd.mean PIC.variance.P
## 1 3.427719e-06         49966.78              49640.23          0.557
##   PIC.variance.Z
## 1     0.01619188
```

Question 8: Using the K-values and associated p-values (i.e., “PIC.var.P”“) from the `phylosignal` output, answer the following questions:

a. Is there significant phylogenetic signal for niche breadth or standardized growth on any of the phosphorus resources?

b. If there is significant phylogenetic signal, are the results suggestive of clustering or overdispersion?

   Answer 8a: There is not a significant phylogenetic signal based on this test (p = 0.54) ***Answer 8b***: There does not appear to be a significant phylogenetic signal - IF there was, the results would suggest overdispersal, based on a K value of < 1

C. Calculate Dispersion of a Trait

In the R code chunk below, do the following:
1. turn the continuous growth data into categorical data,
2. add a column to the data with the isolate name,
3. combine the tree and trait data using the `comparative.data()` function in `caper`, and
4. use `phylo.d()` to calculate *D* on at least three phosphorus traits.

```
p.growth.pa <- as.data.frame((p.growth > 0.01) * 1)

apply(p.growth.pa, 2, sum)
```

```
##      AEP      PEP      G1P      G6P   MethCP      BGP      DNA     Peth
##       20       38       35       34        3       35       19       21
##    Pchol       B1     Phyt      SRP     cAMP      ATP PhenylCP    PolyP
##       18       38       36       39       29       38        6       39
##      GDP      GTP
##       37       38
```

```
p.growth.pa$name <- rownames(p.growth.pa)

p.traits <- comparative.data(nj.rooted, p.growth.pa, "name")
phylo.d(p.traits, binvar = PEP)
```

```
##
## Calculation of D statistic for the phylogenetic structure of a binary variable
##
```

```
##   Data :  p.growth.pa
##   Binary variable :  PEP
##   Counts of states:  0 = 1
##                      1 = 38
##   Phylogeny :  nj.rooted
##   Number of permutations :  1000
##
## Estimated D :  -0.4419386
## Probability of E(D) resulting from no (random) phylogenetic structure :  0.288
## Probability of E(D) resulting from Brownian phylogenetic structure    :  0.518
```

```
phylo.d(p.traits, binvar = MethCP)
```

```
##
## Calculation of D statistic for the phylogenetic structure of a binary variable
##
##   Data :  p.growth.pa
##   Binary variable :  MethCP
##   Counts of states:  0 = 36
##                      1 = 3
##   Phylogeny :  nj.rooted
##   Number of permutations :  1000
##
## Estimated D :  -0.1858903
## Probability of E(D) resulting from no (random) phylogenetic structure :  0.008
## Probability of E(D) resulting from Brownian phylogenetic structure    :  0.637
```

```
phylo.d(p.traits, binvar = Peth)
```

```
##
## Calculation of D statistic for the phylogenetic structure of a binary variable
##
##   Data :  p.growth.pa
##   Binary variable :  Peth
##   Counts of states:  0 = 18
##                      1 = 21
##   Phylogeny :  nj.rooted
##   Number of permutations :  1000
##
## Estimated D :  0.635566
## Probability of E(D) resulting from no (random) phylogenetic structure :  0.055
## Probability of E(D) resulting from Brownian phylogenetic structure    :  0
```

Question 9: Using the estimates for D and the probabilities of each phylogenetic model, answer the following questions:

a. Choose three phosphorus growth traits and test whether they are significantly clustered or overdispersed?

b. How do these results compare the results from the Blomberg's K analysis?

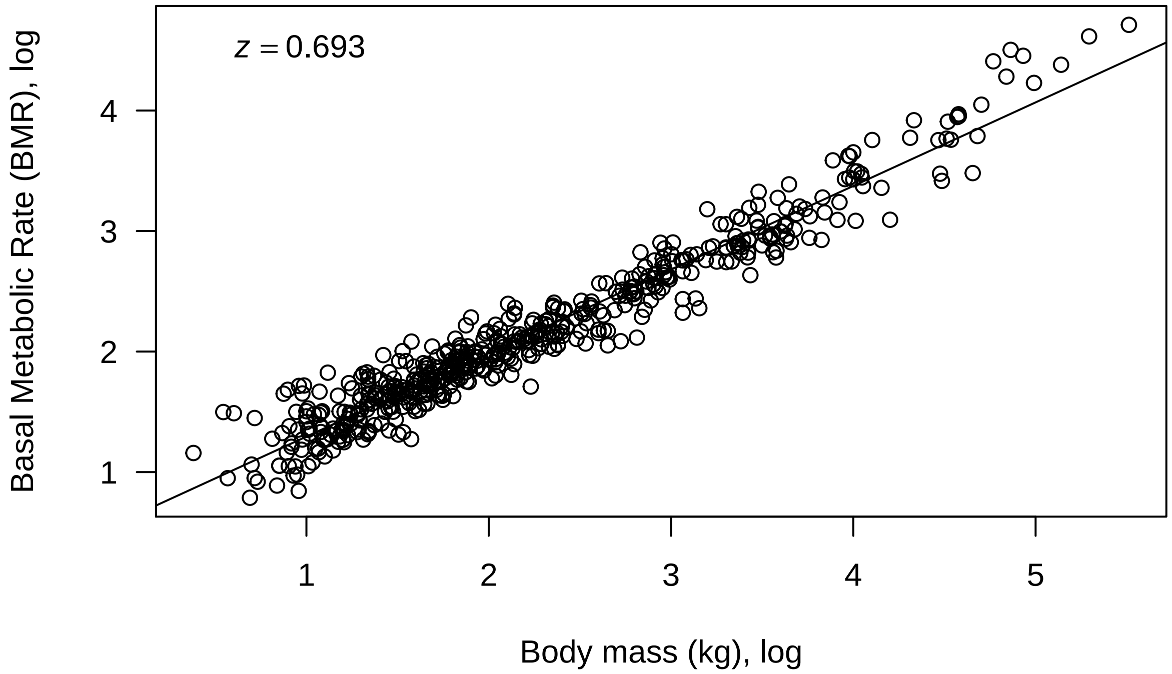
c. Discuss what factors might give rise to differences between the metrics.

   Answer 9a: PEP is clustered, MethCP is clustered, and Peth is clumped. ***Answer 9b***: There are some discrepancies between this analysis that the Blomberg K analysis for PEP and MethCP. While K values suggest that these two traits are overdispersed, D values suggest that they are clustered. For MethCP, the K value suggest that is it somewhat consistent with Brownian motion while the D value indicates that it is more randomly clumped. Peth, on the other hand, is similar between the two methods. PEP (D = -0.30; K = 0.30); MethCP (D = -0.27; K = 0.601); Peth (D = 0.63; K = 0.601) ***Answer 9c***: D and K are measuring two slightly different things to answer the same question. While D is measuring how the trait is spatially situated, K is measuring

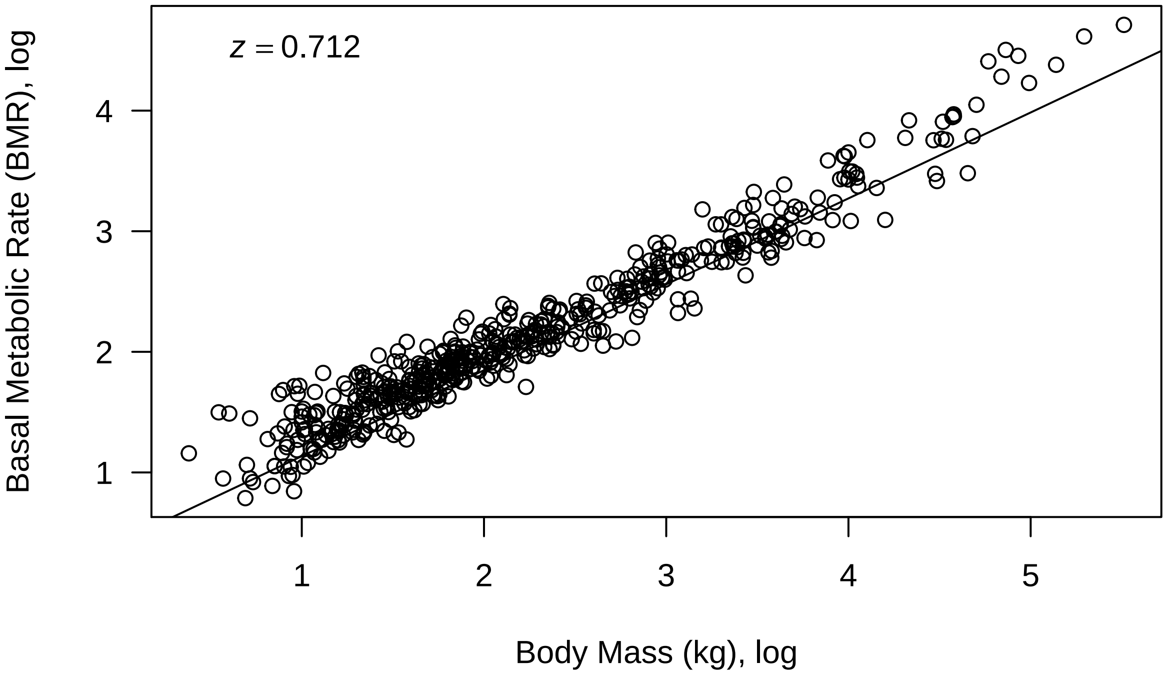
7) PHYLOGENETIC REGRESSION

In the R code chunk below, do the following:
1. Load and clean the mammal phylogeny and trait dataset, 2. Fit a linear model to the trait dataset, examining the relationship between mass and BMR, 2. Fit a phylogenetic regression to the trait dataset, taking into account the mammal supertree

```r
mammal.Tree <- read.tree("./data/mammal_best_super_tree_fritz2009.tre")
mammal.data <- read.table("./data/mammal_BMR.txt", sep = "\t",
                          header = T)

mammal.data <- mammal.data[,3:5]
mammal.species <- array(mammal.data$Species)

pruned.mammal.tree <- drop.tip(mammal.Tree, mammal.Tree$tip.label[-na.omit(match(mammal.species,
                                                                                 mammal.Tree$tip.label)]

pruned.mammal.data <- mammal.data[mammal.data$Species %in% pruned.mammal.tree$tip.label,
                                  ]

rownames(pruned.mammal.data) <- pruned.mammal.data$Species

fit <- lm(log10(BMR_.mlO2.hour.) ~ log10(Body_mass_for_BMR_.gr.),
          data = pruned.mammal.data)
plot(log10(pruned.mammal.data$Body_mass_for_BMR_.gr.), log10(pruned.mammal.data$BMR_.mlO2.hour.),
     las = 1, xlab = "Body mass (kg), log", ylab = "Basal Metabolic Rate (BMR), log")
abline(a = fit$coefficients[1], b = fit$coefficients[2])
b1 <- round(fit$coefficients[2], 3)
eqn <- bquote(italic(z) == .(b1))

text(0.5, 4.5, eqn, pos = 4)
```



```
fit.phy <- phylolm(log10(BMR_.mlO2.hour.) ~ log10(Body_mass_for_BMR_.gr.),
                   data = pruned.mammal.data, pruned.mammal.tree, model = "lambda",
                   boot = 0)

plot(log10(pruned.mammal.data$Body_mass_for_BMR_.gr.), log10(pruned.mammal.data$BMR_.mlO2.hour.),
     las = 1, xlab = "Body Mass (kg), log", ylab = "Basal Metabolic Rate (BMR), log")
abline(a = fit.phy$coefficients[1], b = fit.phy$coefficients[2])
b1.phy <- round(fit.phy$coefficients[2], 3)
eqn <- bquote(italic(z) == .(b1.phy))
text(0.5, 4.5, eqn, pos = 4)
```



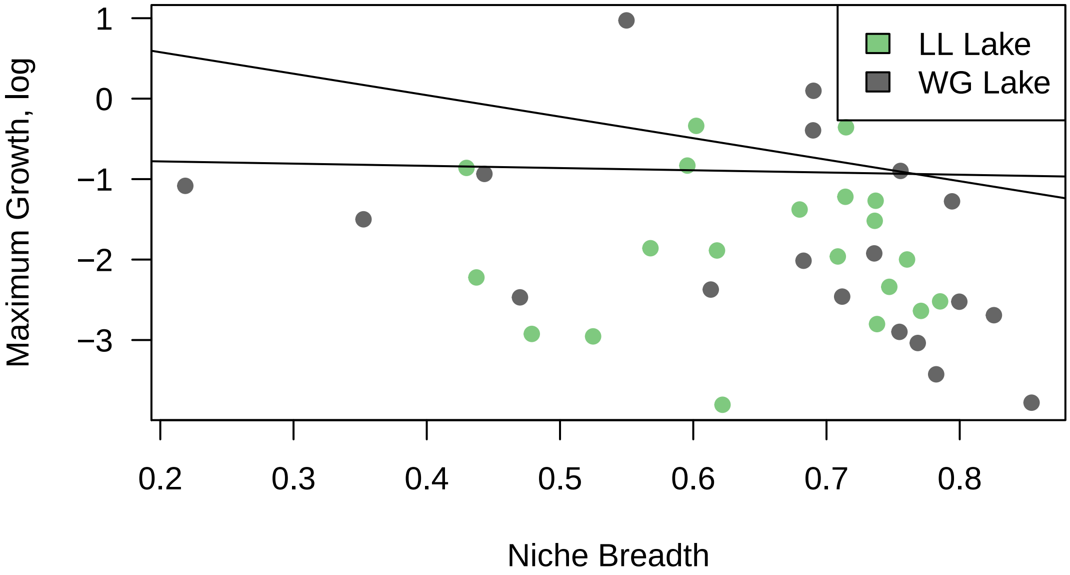
a. Why do we need to correct for shared evolutionary history?
b. How does a phylogenetic regression differ from a standard linear regression?
c. Interpret the slope and fit of each model. Did accounting for shared evolutionary history improve or worsten the fit?
d. Try to come up with a scenario where the relationship between two variables would completely disappear when the underlying phylogeny is accounted for.

Answer 10a: It is important to account for shared evolutionary history because of possible phylogenetic dependencies among traits. If the observed patterns among your study subjects is a result of a recently shared common anscestor, then you cannot treat those study subjects as independent samples. ***Answer 10b***: Phylogenetic regression differ from a simple linear regression because it accounts for potential phylogenetic dependencies among study subjects; therefore, it accounts for inherent nonindependence that may exist. ***Answer 10c***: For the simple linear regression, approximately 70% of the residual variation was accounted for by the model. For the phylogenetic model, approximately 71% of the residual variation was accounted for my the model. Accounting for the shared evolutionary hisotry slightly improved the fit of the regession. ***Answer 10d***: Forging strategy and gut morphology among cebid monkeys.

7) SYNTHESIS

Below is the output of a multiple regression model depicting the relationship between the maximum growth rate (μ_{max}) of each bacterial isolate and the niche breadth of that isolate on the 18 different sources of phosphorus. One feature of the study which we did not take into account in the handout is that the isolates came from two different lakes. One of the lakes is an very oligotrophic (i.e., low phosphorus) ecosystem named Little Long (LL) Lake. The other lake is an extremely eutrophic (i.e., high phosphorus) ecosystem named Wintergreen (WG) Lake. We included a “dummy variable” (D) in the multiple regression model (0 =

WG, 1 = LL) to account for the environment from which the bacteria were obtained. For the last part of the assignment, plot nich breadth vs. μ_{max} and the slope of the regression for each lake. Be sure to color the data from each lake differently.



Question 11: Based on your knowledge of the traits and their phylogenetic distributions, what conclusions would you draw about our data and the evidence for a generalist-specialist tradeoff?

> ***Answer 11***: There is decent evidence for a generalist-specialist tradeoff based on significant interactions between niche breadth and maximum growth rates. Additionally, growth rates were slower in the phosophorus-rich lake - so if genrealist species are taking advatage of the phosophorus here, it would support the slow-growth rate hypothesis.