Phylogenetic Diversity - Traits

Evgeniya Polezhaeva; Z620: Quantitative Biodiversity, Indiana University

22 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] "D:/Jane/GitHub/QB2017_Polezhaeva/Week6-PhyloTraits"
```

```
setwd('D:/Jane/GitHub/QB2017_Polezhaeva/Week6-PhyloTraits')
package.list <- c('ape', 'seqinr', 'phylobase', 'adephylo', 'geiger', 'picante', 'stats', 'RColorBrewer
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.

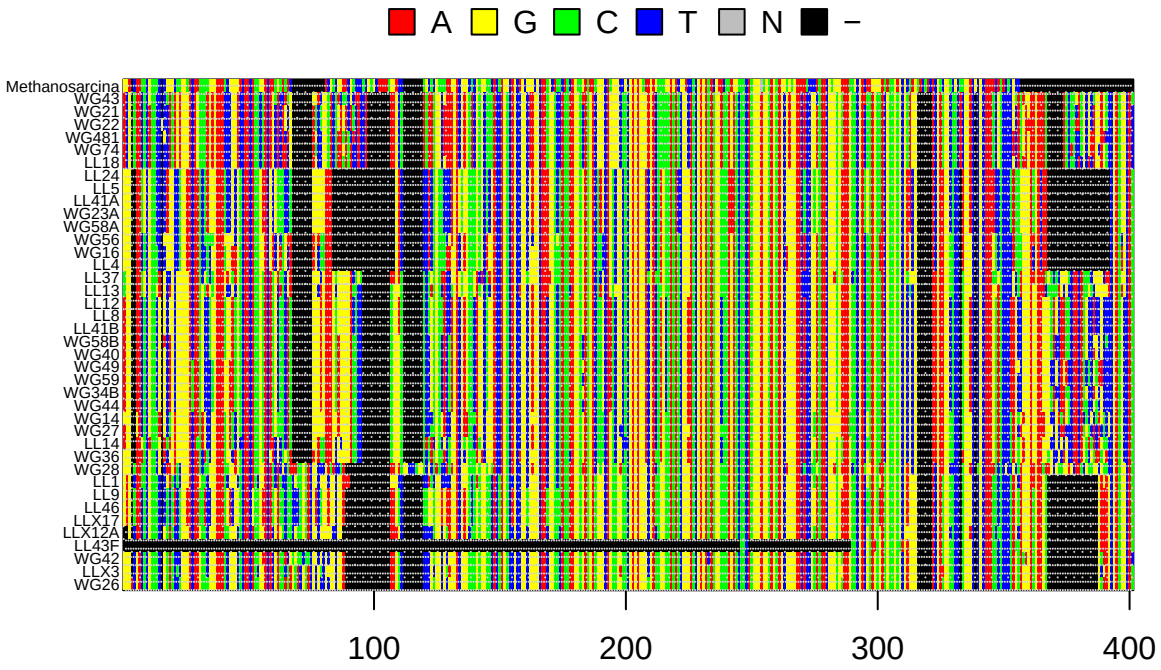
3) SEQUENCE ALIGNMENT

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***: *.fasta file contains only sequences and their names, while* .afa file contains aligned sequences.

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)
grid(ncol(window), nrow(window), col = "lightgrey")
```



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: Methanosarcina read is about 1400 bp long, the rest sequences are from 350 to 900 bp long. ***Answer 2b***: We need to analyse non-conserved regions , e.g. ~0-150 or ~300-400.

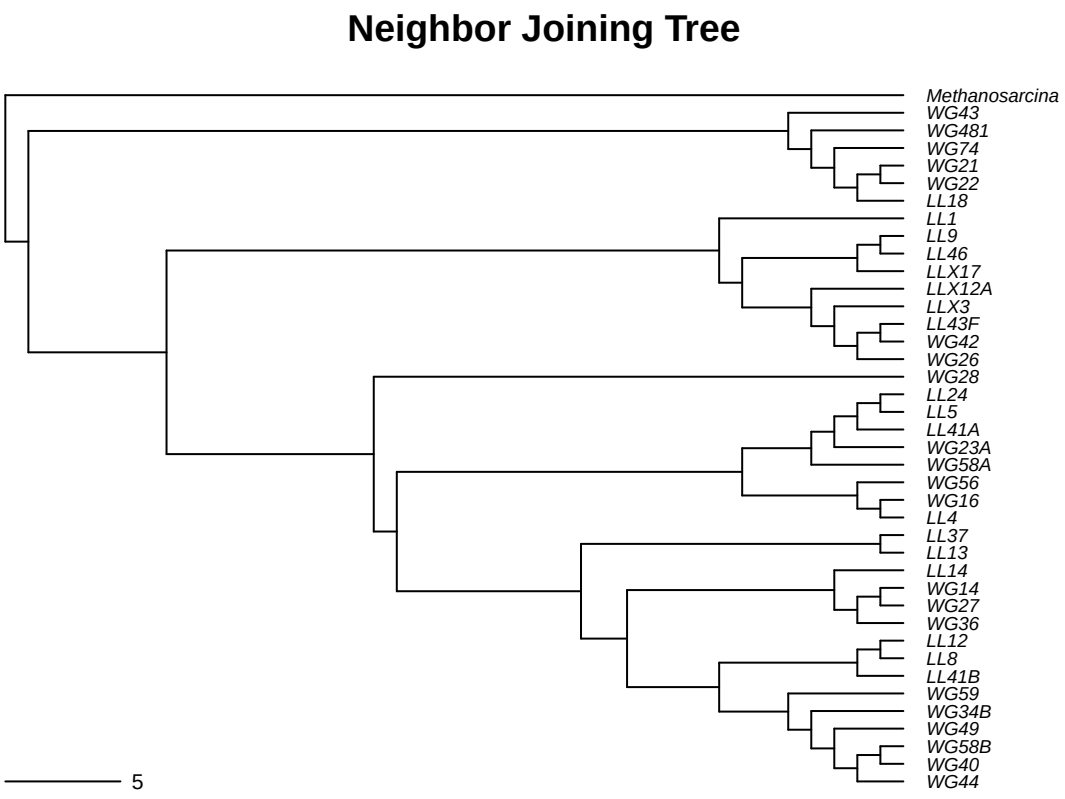
4) MAKING A PHYLOGENETIC TREE

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 = "ri
add.scale.bar(cex = 0.7)
```



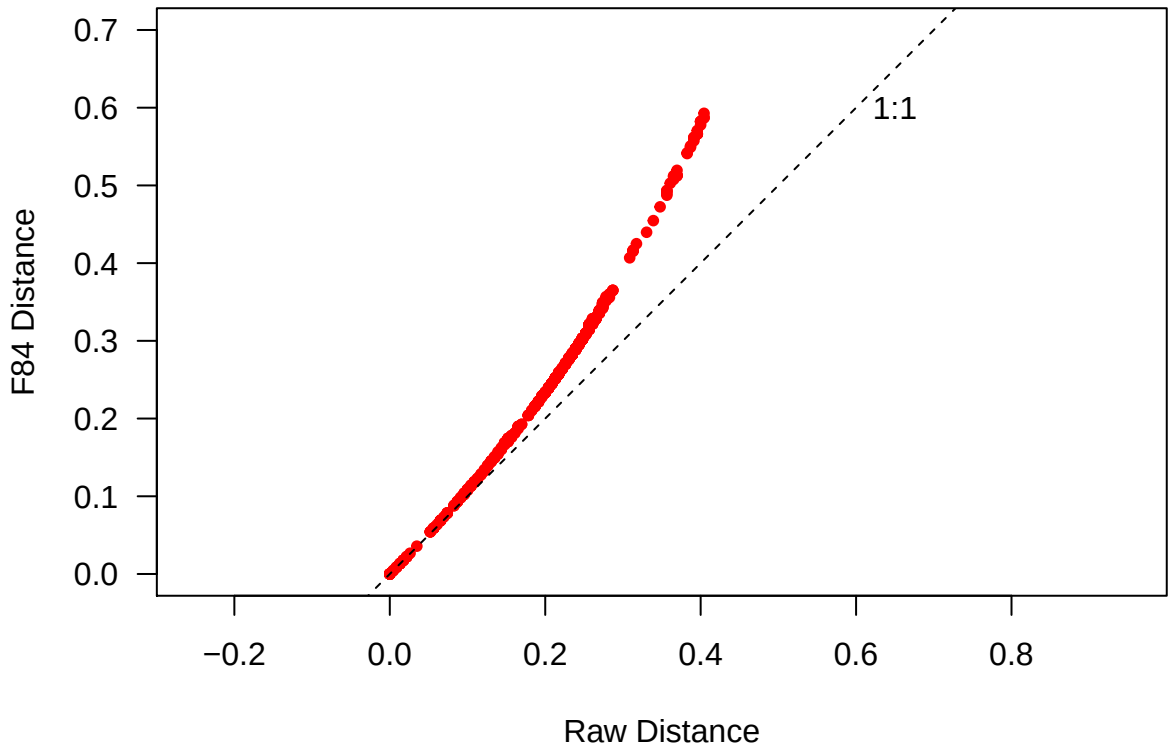
Question 3: What are the advantages and disadvantages of making a neighbor joining tree?

> ***Answer 3***: The main advantage of making a neighbor joining tree is that it is fast, - returns the correct tree for a correct distnce matrix (what is not always the case),
> - it does not assume all lineages evolve at the same rate but - it only uses a distance matrix, which doesn'y take into account specific nucleotide state - it only gives one tree, meaning it is not a statistical method - it can be sensitive to a the underlying substitution model - it may assigns negative lengths to some of the branches. - It does not lways accurately reconstruct the real phylogeny.

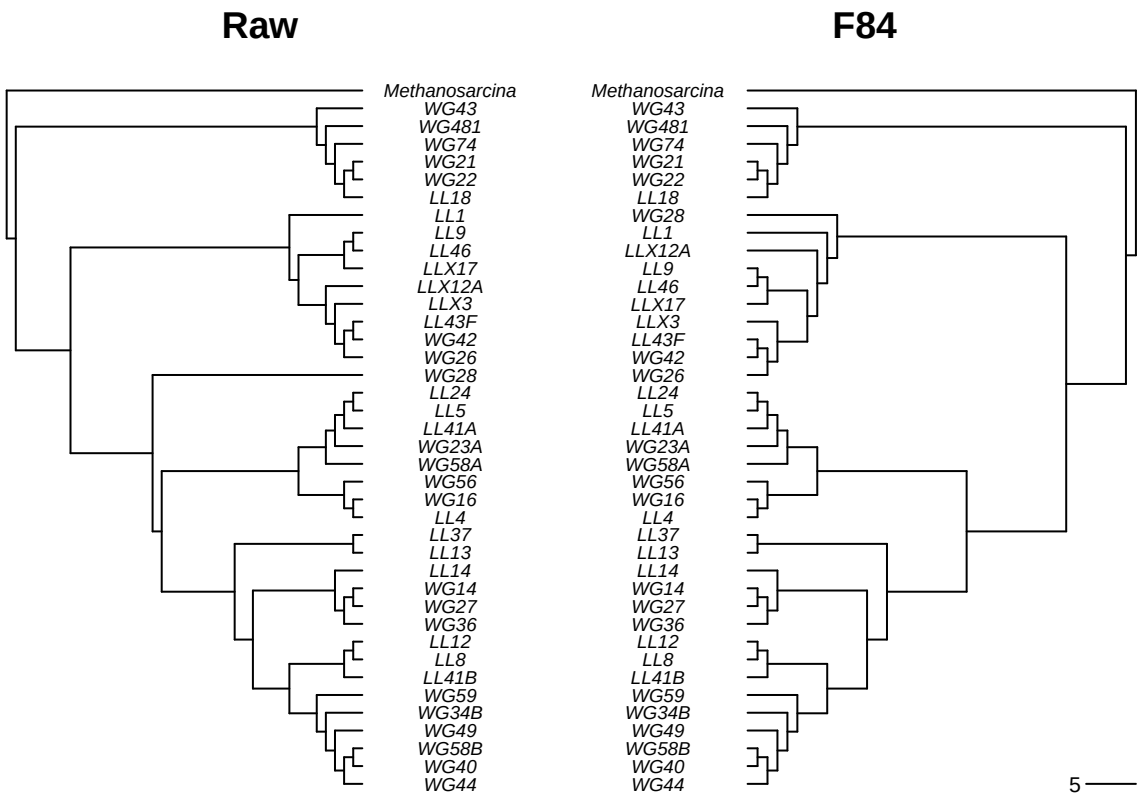
B) SUBSTITUTION MODELS OF DNA EVOLUTION

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
abline(b =1, a = 0, lty = 2)
text(0.65, 0.6, '1:1')
```

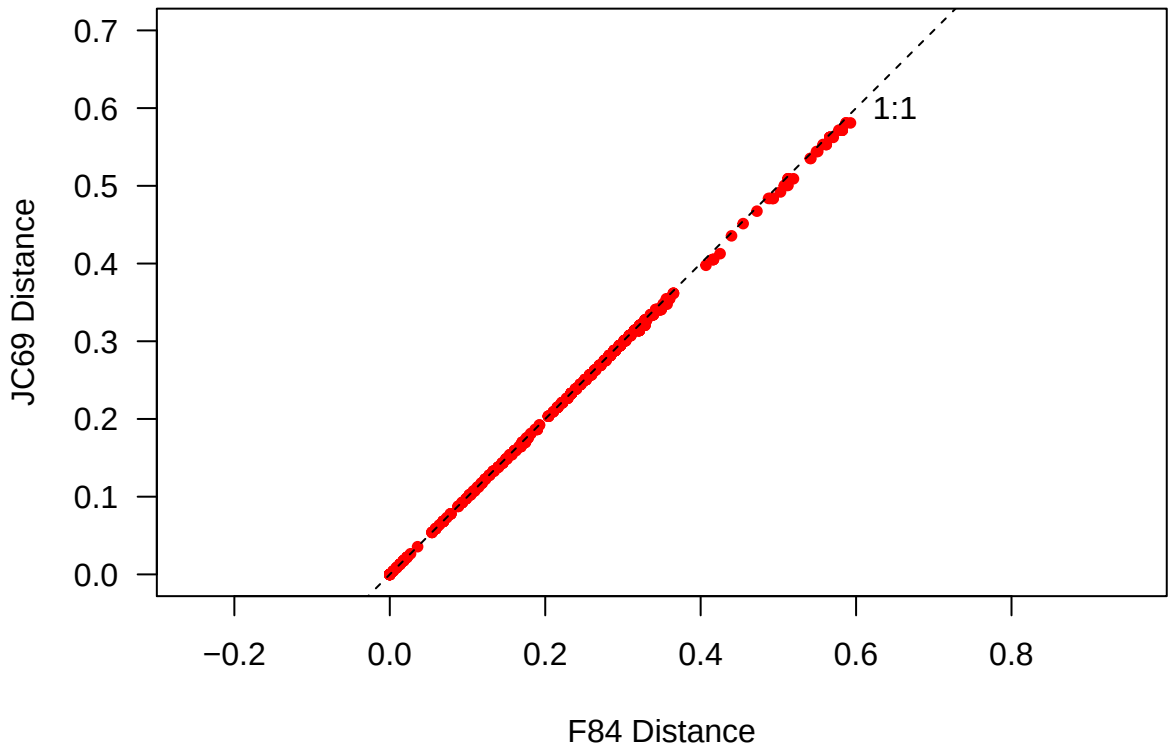



```
raw.tree <- bionj(seq.dist.raw)
F84.tree <- bionj(seq.dist.F84)

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, 1))
plot.phylo(raw.rooted, type = "phylogram", direction = "right", show.tip.label = T, use.edge.length = F

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

add.scale.bar(cex = 0.7)
```

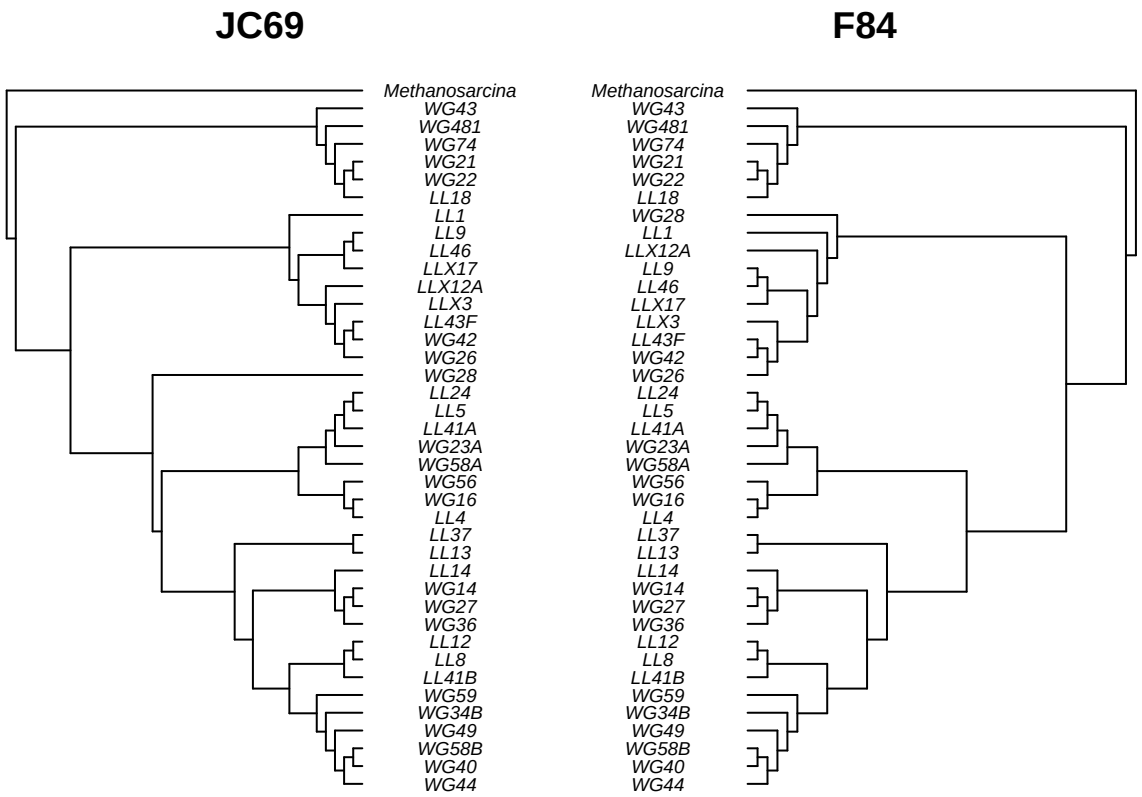


In the R code chunk below, do the following:
1. pick another substitution model,
2. create and distance matrix and tree for this model,
3. make a saturation plot that compares that model to the *Felsenstein (F84)* model,
4. make a cophylogenetic plot that compares the topologies of both models, and
5.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

```
seq.dist.JC69 <- dist.dna(p.DNAbin, model = "JC69", pairwise.deletion = F)

par(mar = c(5, 5, 2, 1) + 0.1)
plot(seq.dist.F84, seq.dist.JC69, pch = 20,
     col = 'red', las = 1, asp = 1, xlim = c(0, 0.7), ylim = c(0, 0.7), xlab="F84 Distance", ylab = "JC
abline(b =1, a = 0, lty = 2)
text(0.65, 0.6, '1:1')
```




```
JC69.tree <- bionj(seq.dist.JC69)
F84.tree <- bionj(seq.dist.F84)

JC69.outgroup <- match("Methanosarcina", JC69.tree$tip.label)
F84.outgroup <- match("Methanosarcina", F84.tree$tip.label)

JC69.rooted <- root(raw.tree, JC69.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, 1))
plot.phylo(JC69.rooted, type = "phylogram", direction = "right", show.tip.label = T, use.edge.length =

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



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?

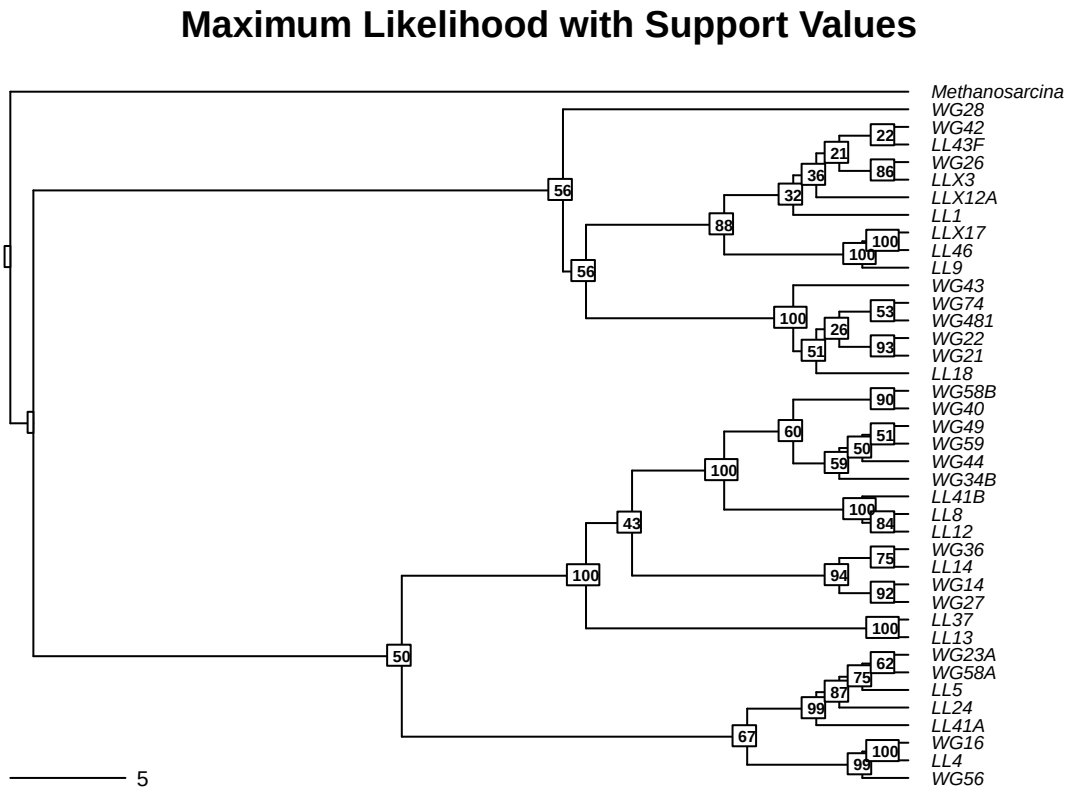
Answer 4a: Jukes-Cantor model is one of the simplest models. It assumes that all nucleotides occur at equal frequences and that these nucleotides can mutate from one to another with equal probability. F84 assumes different rates of base transitions and transversions while allowing for differences in base frequences. ***Answer 4b***: The saturation plot for F84 and JC69 models shows that they are very close to each other while appliing for our dataset. The phylogenetic trees look very similar, for several groups of branches even identical. ***Answer 4c***: Non-crutial differnces of phylogenetic trees structure allows to assume that substitution rates are very close but not equal.

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

```
ml.bootstrap <- read.tree("data/ml_tree/RAxML_bipartitions.T1")
par(mar = c(1, 1, 2, 1) + 0.1)
plot.phylo(ml.bootstrap, type = "phylogra", 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: ML tree is inconsistent with NJ tree. ML is the estimator method that is the least affected by sampling error, takes into account nucleotide states ***Answer 5b***: Trees are inconsistent with one another nad bootstraping is a way to anylise if we can trust the tree we built.Bootstraping of phylogenetic tree is the way to measure probability of errors in phylogenetic trees. ***Answer 5c***: The bootstrap values a relative metrics of how well the model we are using to generate a phylogenetic tree describes each node of it. If the value is higher than 95% it is considered as operationally correct. ***Answer 5d***: The earliest nodes have low support (~50%) and the whole upper (right) part of the tree ***Answer 5e***: We must try another models.

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.

```
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.

```
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

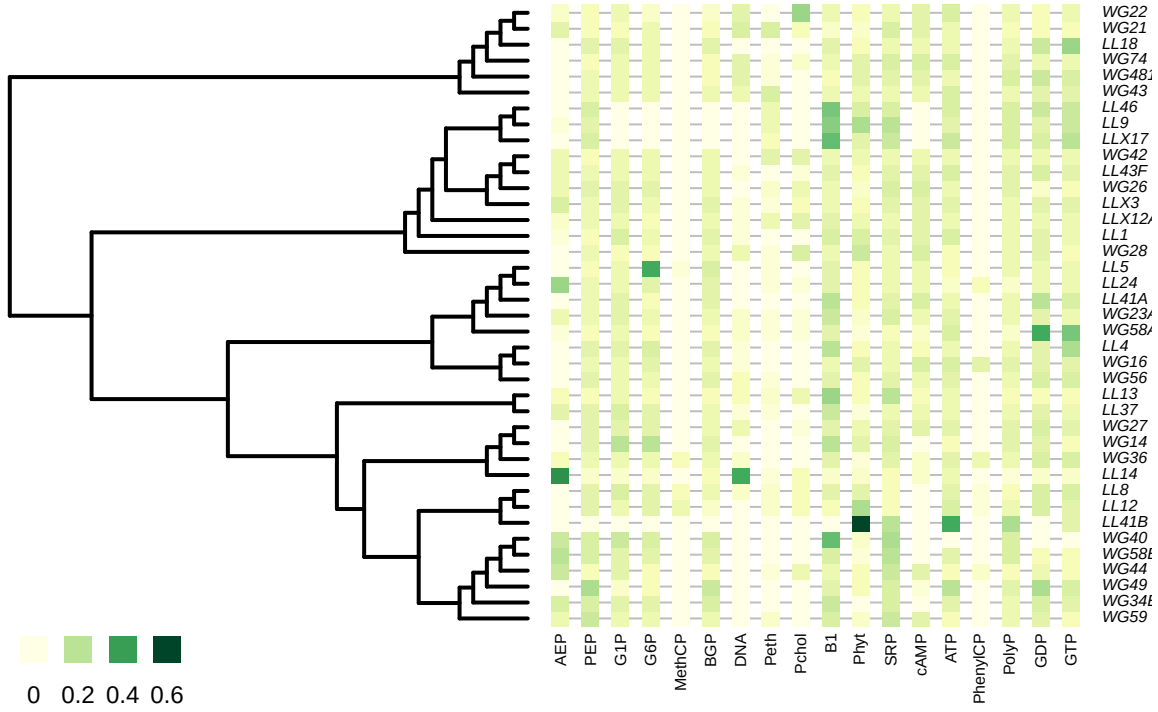
In the R code chunk below, do the following:
1. pick your favorite substitution model and make a Neighbor Joining tree,
2. define your outgroup and root the tree, and
3. remove the outgroup branch.

```
nj.tree <- bionj(seq.dist.F84)
outgroup <- match("Methanosarcina", nj.tree$tip.label)
nj.rooted <- root(nj.tree, outgroup, resolve.root=T) %>% drop.tip("Methanosarcina")
```

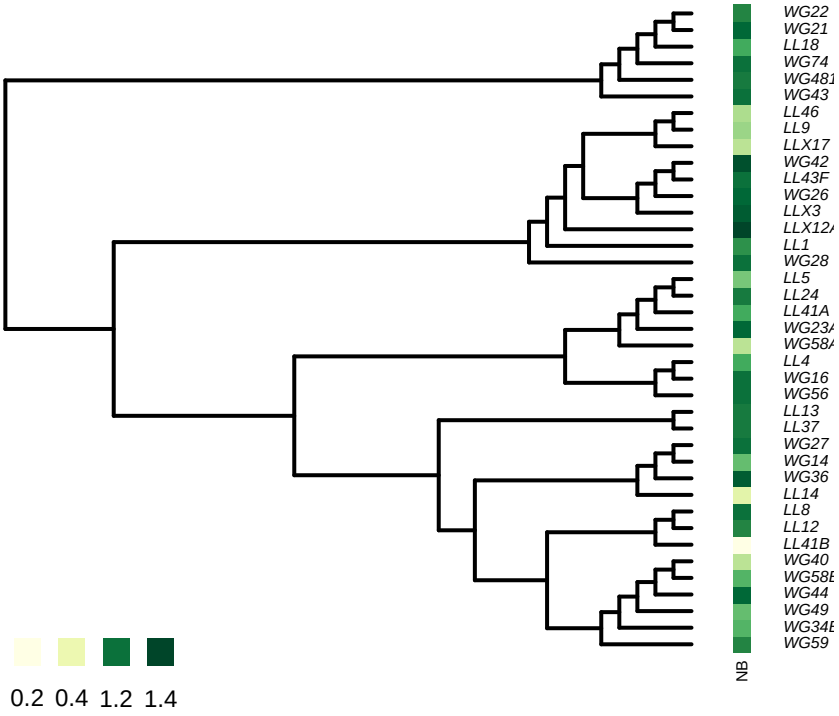
In the R code chunk below, do the following:
1. define a color palette (use something other than "YlOrRd"),
2. map the phosphorus traits onto your phylogeny,
3. map the nb trait on to your phylogeny, and
4. customize the plots as desired (use `help(table.phylo4d)` to learn about the options).

```
par(mar = c(1, 1, 1, 1) + 0.1)
mypalette <- colorRampPalette(brewer.pal(9, "YlGn"))
x <- phylo4d(nj.rooted, p.growth.std)
table.phylo4d(x, treetype = "phylo", symbol = "colors", show.node = T,
              cex.label = 0.5, scale = F, use.edge.length = F,
              edge.color = "black", edge.width = 2, box =F,
```

```
            col = mypalette(25), pch = 15, cex.symbol = 1.25, ratio.tree = 0.5,
            cex.legend  = 1.5, 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 = "black", edge.width = 2, box =F,
              col = mypalette(25), pch = 15, cex.symbol = 1.25, var.label=("NB"),
              ratio.tree = 0.9, cex.legend  = 1.5, 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: There is two groups of organisms. The first one (specialists) is able to use a limited number of resourser but uses them efficiently. The second group is able to moderatly consume wide range of resouses. ***Answer 6b***: Specialist should have low nich breadth and one or a few maximums of growth rate. Generalist should have a high value of nb but no maximums and minimums of growth rate.

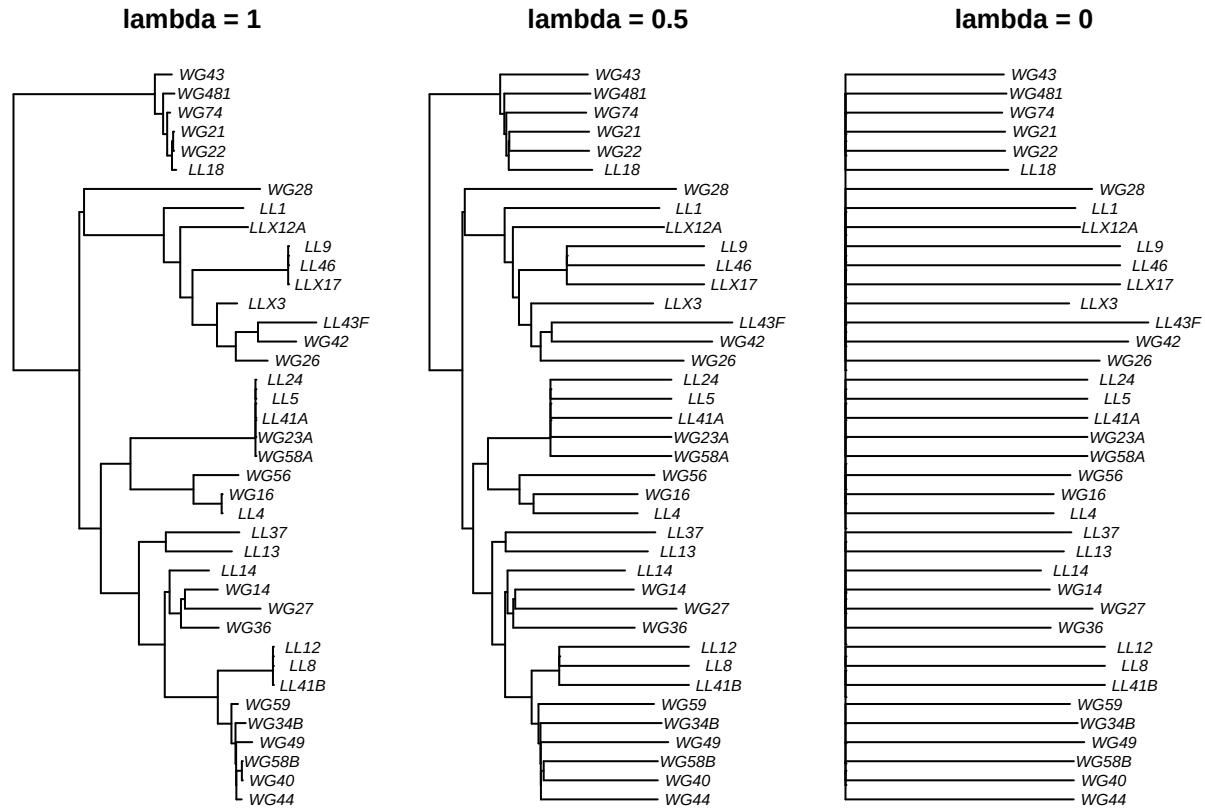
6) HYPOTHESIS TESTING

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.

```
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)
```

```
plot(nj.rooted, main = "lambda = 1", cex = 0.7, adj = 0.5)
plot(nj.lambda.5, main = "lambda = 0.5", cex = 0.7, adj = 0.5)
plot(nj.lambda.0, main = "lambda = 0", 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.

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

```
## GEIGER-fitted comparative model of continuous data
##  fitted 'lambda' model parameters:
##  lambda = 0.020848
##  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 = 59
##  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 = 0.82
##
##  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 = 0.021 (Untransformed tree) and 0 (transformed tree) ***Answer 7b***: Difference between AICs is less than 1, so we can consider these models equivalent. ***Answer 7c***: this result suggests that there is no 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              49269.44          0.553
##   PIC.variance.Z
## 1     0.03564687
```

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: Only 4 resourses assosiated with low p-value that can allows us to consider is as significant phylogenetic signal. ***Answer 8b***: For all resorses K < 1. It means that traits (growth rate and niche breadth) are overdispersed with closely related species less similar than expected by chance.

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 = 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 :  -0.2075986
## Probability of E(D) resulting from no (random) phylogenetic structure :  0.008
## Probability of E(D) resulting from Brownian phylogenetic structure    :  0.659
```

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

```
##
## Calculation of D statistic for the phylogenetic structure of a binary variable
##
##   Data :  p.growth.pa
##   Binary variable :  BGP
##   Counts of states:  0 = 4
##                      1 = 35
##   Phylogeny :  nj.rooted
##   Number of permutations :  1000
##
## Estimated D :  -0.3675144
## Probability of E(D) resulting from no (random) phylogenetic structure :  0.002
## Probability of E(D) resulting from Brownian phylogenetic structure    :  0.617
```

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

```
##
## Calculation of D statistic for the phylogenetic structure of a binary variable
##
##   Data :  p.growth.pa
##   Binary variable :  DNA
##   Counts of states:  0 = 20
##                      1 = 19
##   Phylogeny :  nj.rooted
##   Number of permutations :  1000
##
## Estimated D :  0.6019577
## Probability of E(D) resulting from no (random) phylogenetic structure :  0.036
## Probability of E(D) resulting from Brownian phylogenetic structure    :  0.004
```

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.

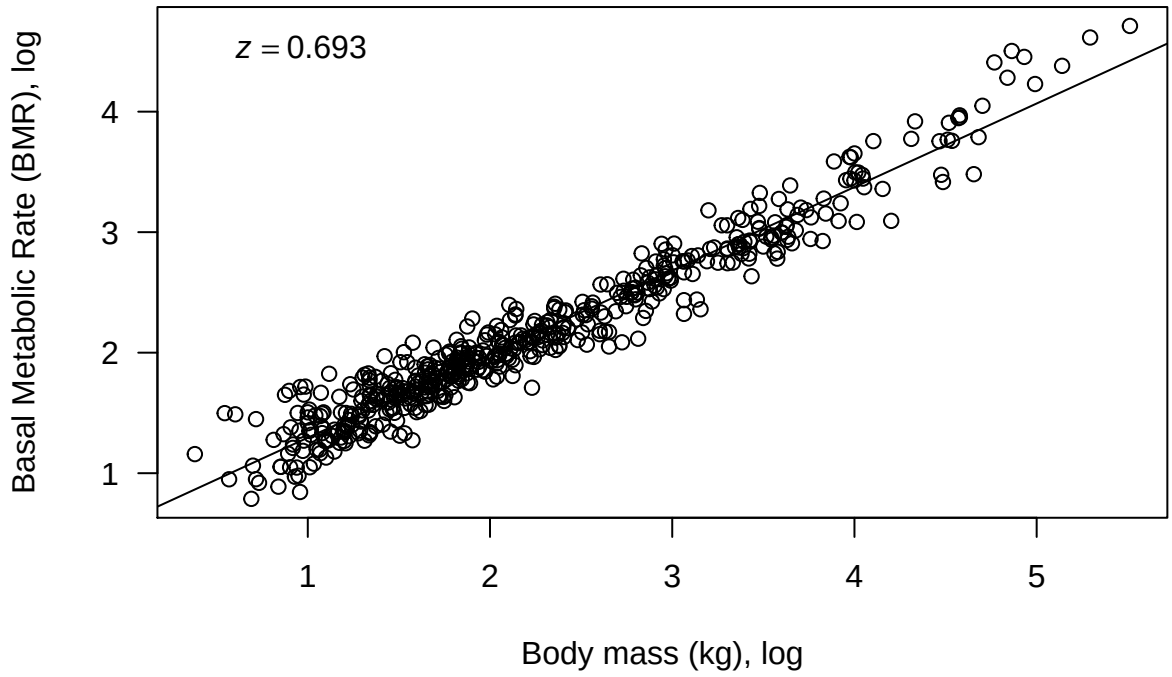
Answer 9a: MethCP and BGP growth trait are clustered or randomly clumped, DNA - overdispersed. ***Answer 9b***: Results fron the Blomberg's K analysis are inconsistent with Dispersion of traits. ***Answer 9c***:

7) PHYLOGENETIC REGRESSION

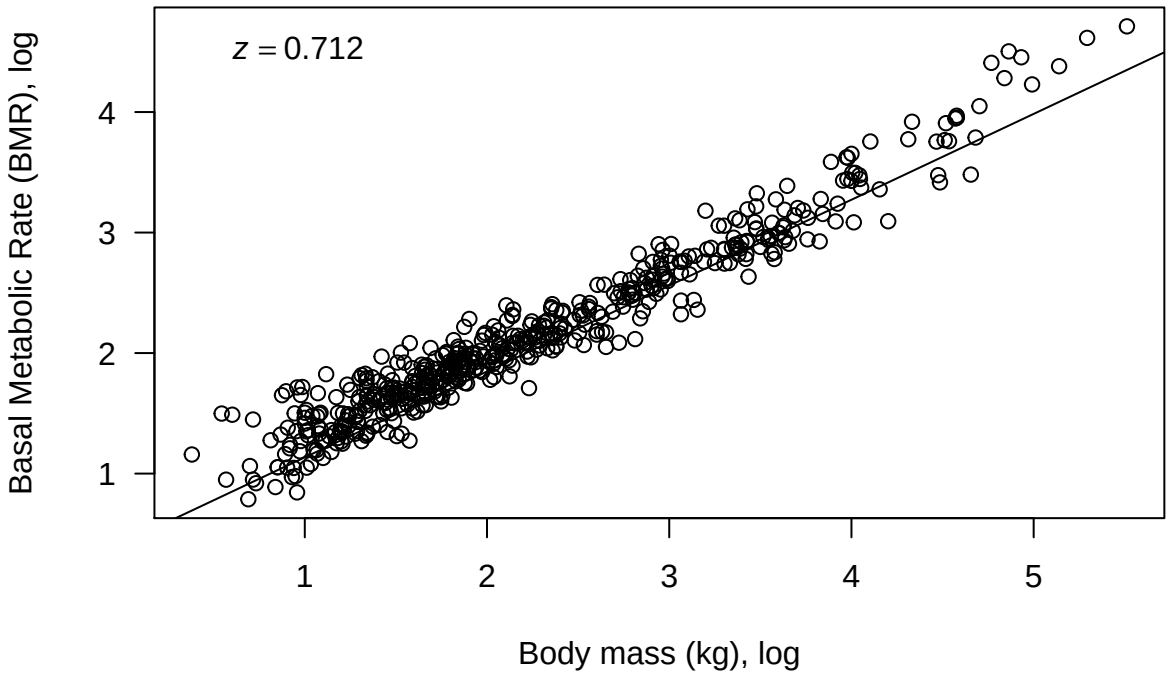
In the R code chunk below, do the following:
1. Load and clean the mammal phylogeny and trait dataset, 2. Fit a linear model to the trait dataset,

examining the relationship between mass and BMR, 2. Fit a phylogenetic regression to the trait dataset, taking into account the mammal supertree

```
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[, c("Species", "BMR_.mlO2.hour.", "Body_mass_for_BMR_.gr.")]
mammal.species <- array(mammal.data$Species)
pruned.mammal.tree <- drop.tip(mammal.Tree, mammal.Tree$tip.label[-na.omit(match(mammal.species, mammal
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.),
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.),
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)
```

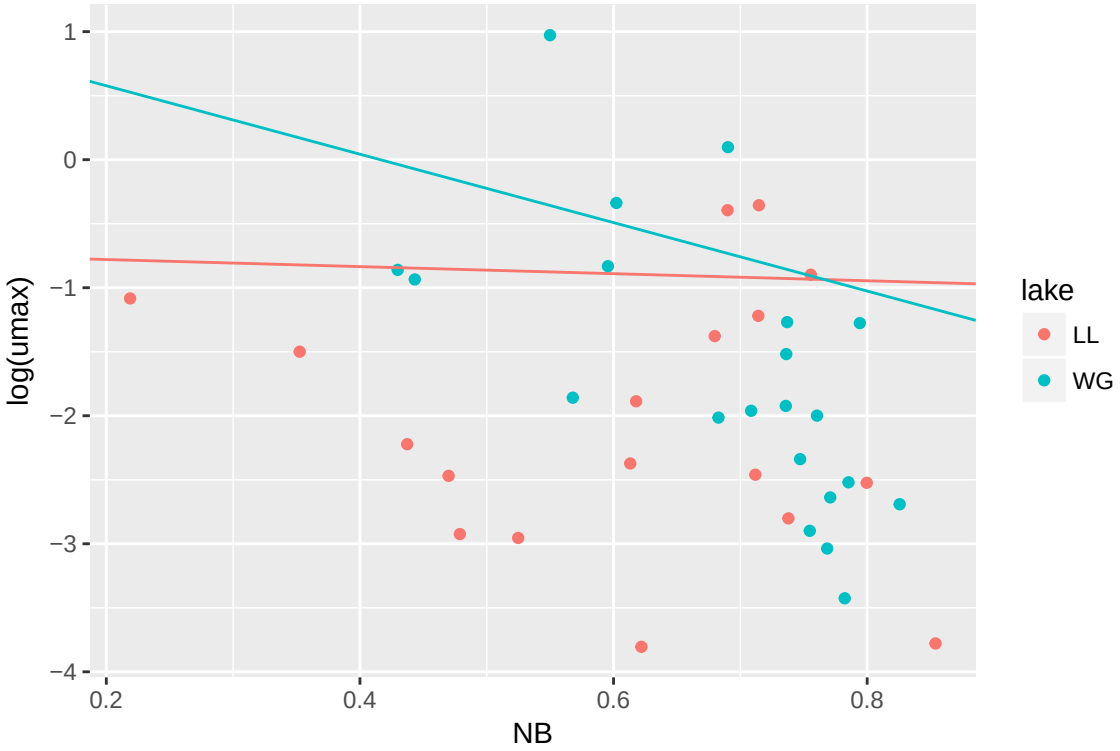


a. Why do we need to correct for shared evolutionary history?
b. How does a phylogenetic regression differ from a standard linear regression?
c. Interpret the slope and fit of each model. Did accounting for shared evolutionary history improve or worsten the fit?
d. Try to come up with a scenario where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

Answer 10a: To perform a simple regression our samples must be independent, but they are not, because of shared evolutionary history. ***Answer 10b***: In simple linear regression the residual errors, given an aindependent variable are assumes to be independent and identically distributed random variables that follow a normal distribution wuth a mean of 0. In a phylogenetig regression the variance of residual errors are described by a covariance matrix that takes into account the branch length of the underlying phylogeny. ***Answer 10c***: Slope in this case is the power in the power law : BMR = c$M\hat{}z$ **Answer 10d***:

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***: For both lakes we can see the trend that samples with the highest niche breadth have the lowest max growth rate.