**PICANTE Package (selected functions) in R 2.1.9**

**Maria Mercedes Gavilanez (mgavil2@lsu.edu)**

**30 October 2009**

Package Picante (**P**hylocom **i**ntegration **c**ommunity **a**nalyses, **n**ull-models, **t**raits and **e**volution) in R is a valuable tool for integrating phylogenies and ecology, a tendency that has gained much interest in the last decade or so (some valuable references: Webb *et al.* 2002; Helmus *et al.* 2007; Cavender-Bares *et al.* 2009). The first version of this package was launched in June 2008, and is currently available as version 0.7-1. The developers of this package are Peter Cowan (graduate student U.C.Berkely), Matthew Helmus (graduate student U.Washington Madison), Steven Kembel (post doc U. Oregon), and many other researches have contributed different functions in the package. Development of Picante is supported by [NSERC](http://nserc.ca/), [NESCent](http://www.nescent.org/index.php), and the [Google Summer of Code](http://code.google.com/soc/2008/).

This package integrates all functions provided by Phylocom, a software for the analysis of phylogenetic community structure and character evolution. However, Picante also incorporates new functions that have been developed for the analysis of community structure for different taxa (e.g. functions for generating estimates of phylogenetic beta diversity and phylogenetic community-environment regressions), and has a more flexibility in the manipulation of data and display of phenotypic and phylogenetic data.

There are different ways to obtain the package. However I would definitely suggest going to the CRAN webpage for the package where you can obtain the pdf file of the reference manual which is sometimes helpful when you are using many functions at the same time and need to look up information about each.

Here are the ways in which you can obtain Picante

1. Looking it up in [www.rseek.org](http://www.rseek.org) or within CRAN webpage in the packages link. There you can obtain either the .zip file (Windows) or .tgz file (MAC OS). Which you can later source into R
2. Typing directly into your R console install.packages("picante",repos="http://R-Forge.R-project.org")

This will install the package. Remember you need to load it either through load package (Windows) or Package Manager (Mac), or just type

library(picante) directly in the R console after you install the packages.

Notice that Picante requires you to have other packages installed on which some of the functions depend: **ape**, **vegan** and **nlme**. So make sure you have all them installed and loaded into R before you install Picante. To do this on a PC you have to follow the same procedure as the one described above for each of these packages. Mac has the option of installing them along with the main package by clicking on the box of “install dependencies”.

Some notes about the packages Picante depends on:

Ape is a package for the analysis of phylogenetics and evolution, which offers functions for reading and manipulating phylogenies, a key component in many of the functions of picante. Vegan is a package for multivariate analysis of ecological communities. It contains many ordination methods, dissimilarity indices, species richness and abundance models required within some functions of picante (e.g. Species co-occurrence distances).

For this tutorial I will focus on some of the functions I find more useful for my own research (exploring aspects of phylogenetic diversity across communities, phylogenetic beta diversity and examining phylogenetic community structure under different null models). Much if the information given here was compiled using the help files fro the functions I am describing. However, there are many other interesting functions (such as integrating trait evolution in analysis of community structure, determining phylogenetic signal in community composition, etc.) that I have yet to explore and you can find out all about them by going through the reference manual (you can download it from here <http://cran.r-project.org/web/packages/picante/index.html>), as well as on Steve Kembel’s document walkthrough picante (<http://www.bio.utk.edu/fesin/MSA2009/Phylocom/picante-walkthrough.pdf>).

Another source I have found very helpful for exploring this package and some others related to phylogenetics and community ecology in R are the archives of R-sig-ecology <https://stat.ethz.ch/pipermail/r-sig-ecology/> and R-sig-phylo where you can search among different questions that R users have posted on these areas.

NOTE:

R packages will run only with the version with which they were created or an older one, so if you are using a package that was created in version 2.9.1 but if your R version is 2.8.1 it will not recognize it.

#First we need to install the packages:

MAC: Go to *packages and data*

Select *package installer*

Search for *picante*, select it from the list

Click on *install dependencies*

Click on *Install selected*

Now picante and its dependent packages should be included in your library.

Next step is loading it into your R session

library(picante)

*all packages will be opened simultaneously*

PC: Go to *packages*

Select *install packages*. Selec a CRAN mirror

Search for *picante*, *nlme, vegan* and *ape* select them from the list (hold Crtl for multiple selection)

Click on *OK*

Now picante and its dependent packages should be included in your library.

Next step is loading it into your R session

library(picante)

*all packages will be opened simultaneously*

For the analysis I will explore in this tutorial I will use data on Neotropical primate community structure of a set of 23 communities (presence absence data only) and a species level dated phylogeny for Neotropical primates obtained from Cooper *et al.* 2008

Some important considerations in terms of data structure:

* There cannot be blank spaces or special characters on either the sample community names or species names, so replace them all with underscores before importing the files into R. (i.e. parque\_nacional\_yasuni instead of Parque Nacional Yasuní; Alouatta\_seniculus instead of Alouatta seniculus)
* Species names in the phylogeny have to match exactly with species names in the community data matrix (capitalization is also important).
* In the community data matrix species should be placed in the columns and the sites/communities as rows

*Now we will read the data files needed for the analyses.*

**First the community sample matrix:**

comun =read.table("/Users/mariamercedesgavilanez/Desktop/R/comun.txt", header=T, row.names=1)

dim(comun) *here I am checking if all rows and colums were read correctly. The data frame table should have 1 column and 1 row less than the original file (which are representing species and sites names respectively but are not included in the computations)*

[1] 23 35 *so, this is right*

class(comun) *now check what class of object was created (we need a data frame for community composition data)*

[1] "data.frame"

**Phylogeny**

There are different formats that are accepted as input into R, .nex , .tre or simply newick format copied on a txt file, or copied and pasted directly to the console. The package that allows you to incoporate phylogenies into r is ape.

In this case I will use the function to read a tree in newick format – read.tree():

An example of newick format : ((B:5.0,(A:2.0,C:3.0,E:2.0):5.0,D:8.0);)

tree\_nc =read.tree("/Users/mariamercedesgavilanez/Desktop/newick\_tree\_COOPER.txt")

or

tree\_nc=read.tree()

1: *## copy here the entire newick file*

2:

Now lets just call the tree and see the properties it has (is it rooted or not, how many tips, nodes, includes branch lengths, etc) and plot it.

class(tree\_nc) *checking the class of file we created (we need a phylo for all our analysis)*

[1] "phylo"

tree\_nc *Just by calling the phylo object we created we can check the properties of our phylogeny*

Phylogenetic tree with 86 tips and 83 internal nodes.

Tip labels:

Canis\_lupus, Rattus\_norvegicus, Alouatta\_caraya, Alouatta\_pigra, Alouatta\_fusca, Alouatta\_belzebul, ...

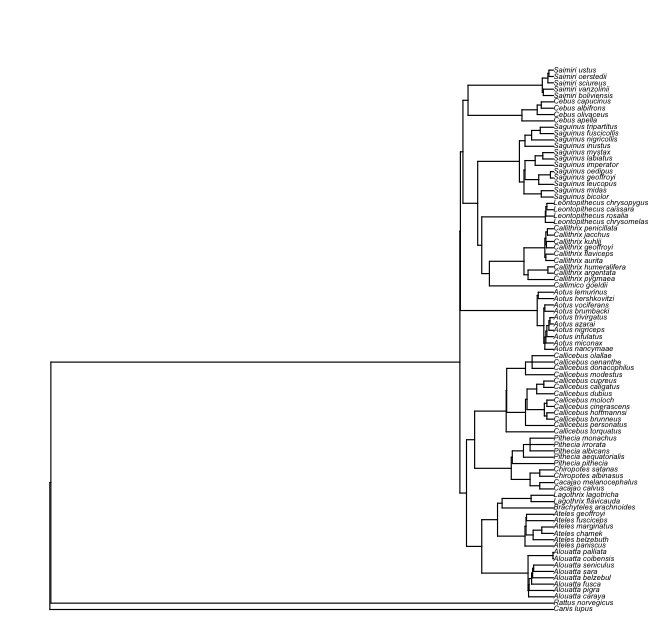
Node labels:

'1', '2', '3', '4', '5', '6',...

Rooted; includes branch lengths.

Finally we can plot our phylogeny to check how it looks, we do this using the function plot()

plot(tree\_nc, cex=0.4) *# cex argument determines the size of tip and node labels*



Now that we have our tree and community matrix the last ”preparation” step is to create a matrix of phylogenetic distances, which is used to calculate some phylogenetic diversity measures.

We can do this by calling the function cophenetic(x), where x will be or phylogenetic tree (object class phylo). This will generate a symmetric matrix with the phylogenetic distances among all species included in the phylogeny.

dm\_tree.nc =cophenetic (tree\_nc)

dim(dm\_tree.nc) *here I am checking the number of rows/columns matches the number of species in the phylogeny*

[1] 86 86

Note:

If you phylogeny has a lot of species it might be a good idea to prune your phylogeny to include only species that are present in at least one of your sample communities by using the function prune.sample before generating the distance matrix.

prunedT\_nc =prune.sample(comun, tree\_nc) *this generates a pruned phylo object with the same number of species as columns we have in our community data matrix*

prunedT\_nc

Phylogenetic tree with 35 tips and 33 internal nodes.

Tip labels:

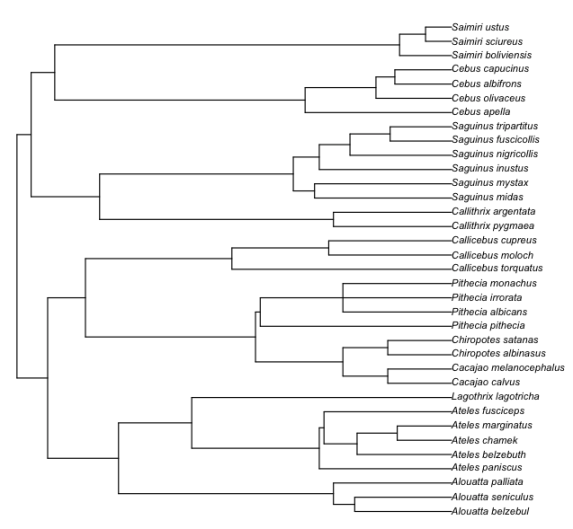
Alouatta\_belzebul, Alouatta\_seniculus, Alouatta\_palliata, Ateles\_paniscus, Ateles\_belzebuth, Ateles\_chamek, ...

Node labels:

'3', '4', '5', '6', '10', '13',...

Rooted; includes branch lengths.

plot(prunedT\_nc, cex=0.4) *we can compare the new tree with the original phylogeny to check the differences*



Finally we need to create a matrix of phylogenetic distances for the pruned phylogeny

dm\_comun =cophenetic(prunedT\_nc)

dim(dm\_comun)

[1] 35 35

This is an important step when carrying out many of the randomization procedures provided by this package, as this pruned tree/distance matrix will represent your species source pool.

Some general data visualization:

We can visually represent in the phylogeny how taxa in a community are arranged on the tree to get an idea of the pattern of distribution of coexisting species across the phylogeny.

par(mfrow=c(2,3)) *# how many columns and rows the graphic window should be divided in.*

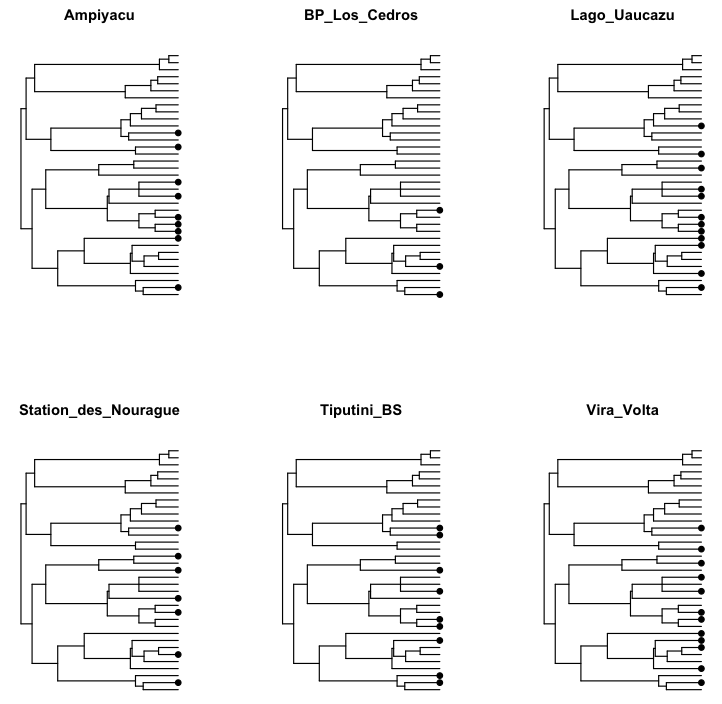
for (i in row.names(comun\_6)) {

plot(prunedT\_nc, show.tip.label = FALSE, main = i)

tiplabels(tip = which(comun\_6[i, ] >0), pch = 19, cex =1)

}

*this is just plotting the phylogeny we want, and having it assign a symbol for all species that have a value >1 in for a community*



Now to the functions:

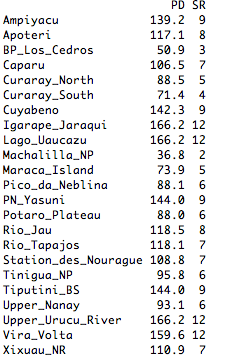
**Phylogenetic diversity and phylogenetic community structure metrics:**

**PD:** Faith’s Phylogenetic Diversity (Faith 1992). Index developed for conservation biologists which incorporates phylogenetic relationships and hence, evolutionary history within taxa. It is determined by summing all branch lengths of species within a community across the phylogeny. Lengths of branches shared by two taxa in the same clade are counted only once

The function pd() requires a sample community and a phylo object, and generates a data frame of each site’s PD value and species richness

pd\_comun= pd(comun, prunedT\_nc)

pdcomun



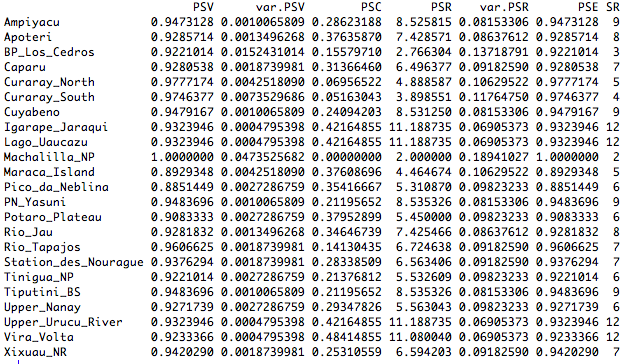
Matt Helmus’s metrics of phylogenetic diversity and community structure (Helmus *et al*. 2007) can also be obtained fairly easily. We can call the function psd- *phylogenetic species diversity metrics* which creates a data frame with six metrics and their respective variances (as compute.var=TRUE is the default).

?psd *this is calling for the help file for the function here you can find a description of each of the metrics calculated by the function.*

The arguments required by the function are the same as for the pd() function, so a community matrix and a phylo object. The data frame generated by this function includes the calculated values of the different phylogenetic diversity metrics (psv, psc, psr, pse) for each community, variance estimates for psv and psr, and the species richness for each community

psd.comun = psd(comun, prunedT\_nc, compute.var=TRUE)

psd.comun



If you are interested in only a particular diversity metric (e.g. PSR phylogenetic species richness) you can call the individual function, i.e.

psr(comun, prunedT\_nc)

**Phylogenetic structure**

Webb’s standard measures of community phylogenetic over and under dispersion: net relatedness index (*NRI*) and nearest taxon index (*NTI*) (Webb *2000*) reflect the level of clustering or evenness of the taxa present in a sample (community) with respect to the pool phylogeny based on the mean branch length distance among sample taxa.

This metrics are based on calculation of the mean pairwise distance (MPD for NRI) and the mean nearest taxon distance (MNTD for NTI) among species in a community and compares them to the values of MPD/MNTD for randomly generated samples (null communities) or phylogenies by calculating the standardized effect size of MPD and MNTD, using the function ses.mpd and ses.mntd respectively.

The functions are ses.mpd and ses.mntd respectively, and both require:

* Sample community matrix
* Phylogenetic distance matrix, created with the function cophenetic()
* Select a null model
* Determine if you want to weight phylogenetic distances by abundance.
* Determine the number of randomizations and iterations to do for each randomization.

A brief description of the currently implemented null models:

* *taxa.labels* : Shuffle distance matrix labels (across all taxa included in distance matrix)
* *sample.pool* : Randomize community data matrix by drawing species from pool of species occurring in at least one community (sample pool) with equal probability
* *phylogeny.pool* : Randomize community data matrix by drawing species from pool of species occurring in the distance matrix (phylogeny pool) with equal probability
* *independentswap*: Randomize community data matrix with the independent swap algorithm (Gotelli 2000) maintaining species occurrence frequency and sample species richness.
* *trialswap*: Randomize community data matrix with the trial-swap algorithm (Miklos & Podani 2004) maintaining species occurrence frequency and sample species richness

The object created by this function is a data frame containing the following values for each community:

ntaxa Number of taxa in community

mpd.obs Observed mpd in community

mpd.rand.mean Mean mpd in null communities

mpd.rand.sd Standard deviation of mpd in null communities

mpd.obs.rank Rank of observed mpd vs. null communities

mpd.obs.z Standardized effect size of mpd vs. null communities (= (mpd.obs

- mpd.rand.mean) / mpd.rand.sd, equivalent to -NRI)

mpd.obs.p P-value (quantile) of observed mpd vs. null communities (=

mpd.obs.rank / runs + 1)

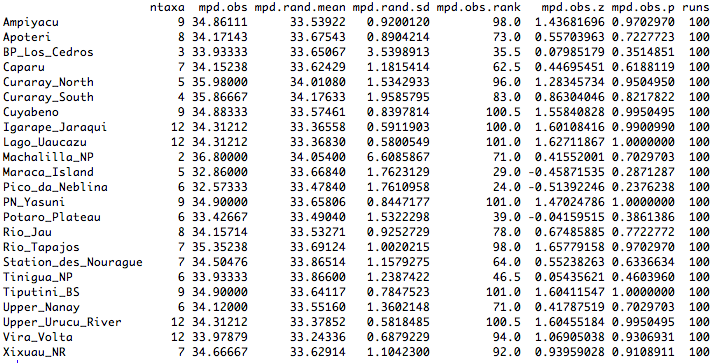
runs Number of randomizations

Positive values of mpd.obs.z and high p values (mpd.obs.p>0.95) indicate phylogenetic evenness (i.e. greater phylogenetic distance among co-occuring species than expected). While low, negative values represent phylogenetic clustering (small phylogenetic distances among co-occuring species than expected).

Now the function,

ses.mpd.comun=ses.mpd(comun, dm\_comun, null.model= "taxa.labels", runs=100, iterations=100)

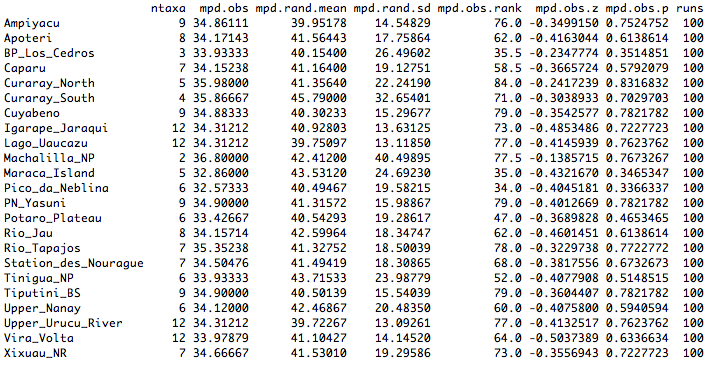
ses.mpd



Here I will run the same function, but changing the phylogenetic distance matrix to the one that includes all neotropical primate species, and the null model that generates new sample communities randomizing all species in the distance matrix to show how results can substantially vary depending on the species pool source.

ses.mpd.comun.complete tree=ses.mpd(comun, dm\_tree.nc, null.model= "taxa.labels", runs=100, iterations=100)

ses.mpd



You can then use this information to determine overall patterns of phylogenetic community structure.

To show how using a different specie spool source I will make a series of histograms of mpd.random.mean and mpd obseverd.p value distributions for a trial using the pruned tree, and the complete tree\_nc, but the same null model.

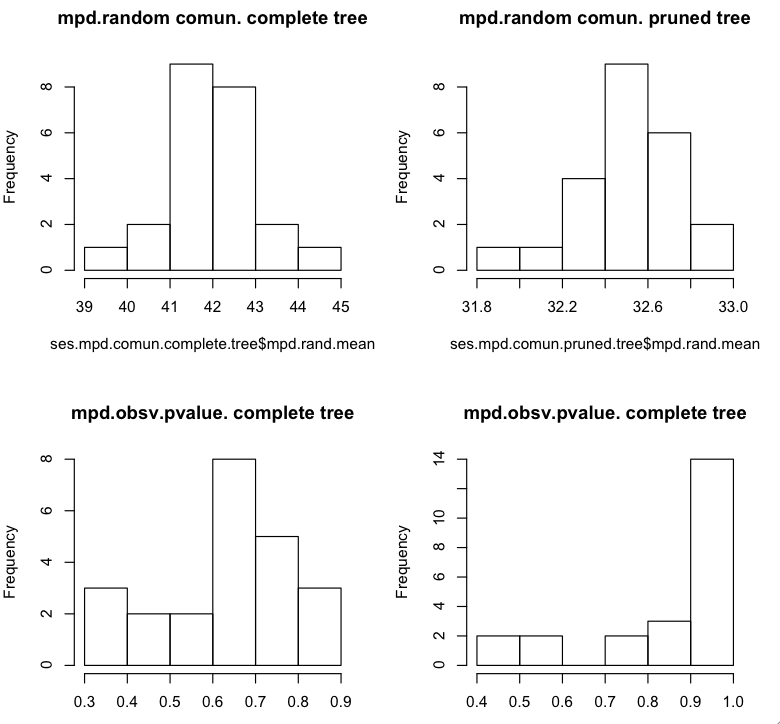
par(mfrow=c(2,2))

hist(ses.mpd.comun.complete.tree$mpd.rand.mean, main="mpd.random comun. complete tree")

hist(ses.mpd.comun.pruned.tree$mpd.rand.mean, main="mpd.random comun. pruned tree")

hist(ses.mpd.comun.complete.tree$mpd.obs.p, main="mpd.obsv.pvalue. complete tree")

hist(ses.mpd.comun.pruned.tree$mpd.obs.p, main="mpd.obsv.pvalue. pruned tree")



Notes:

I used only the function of mpd for this exercise, but the function for mntd is requires the same set of arguments.

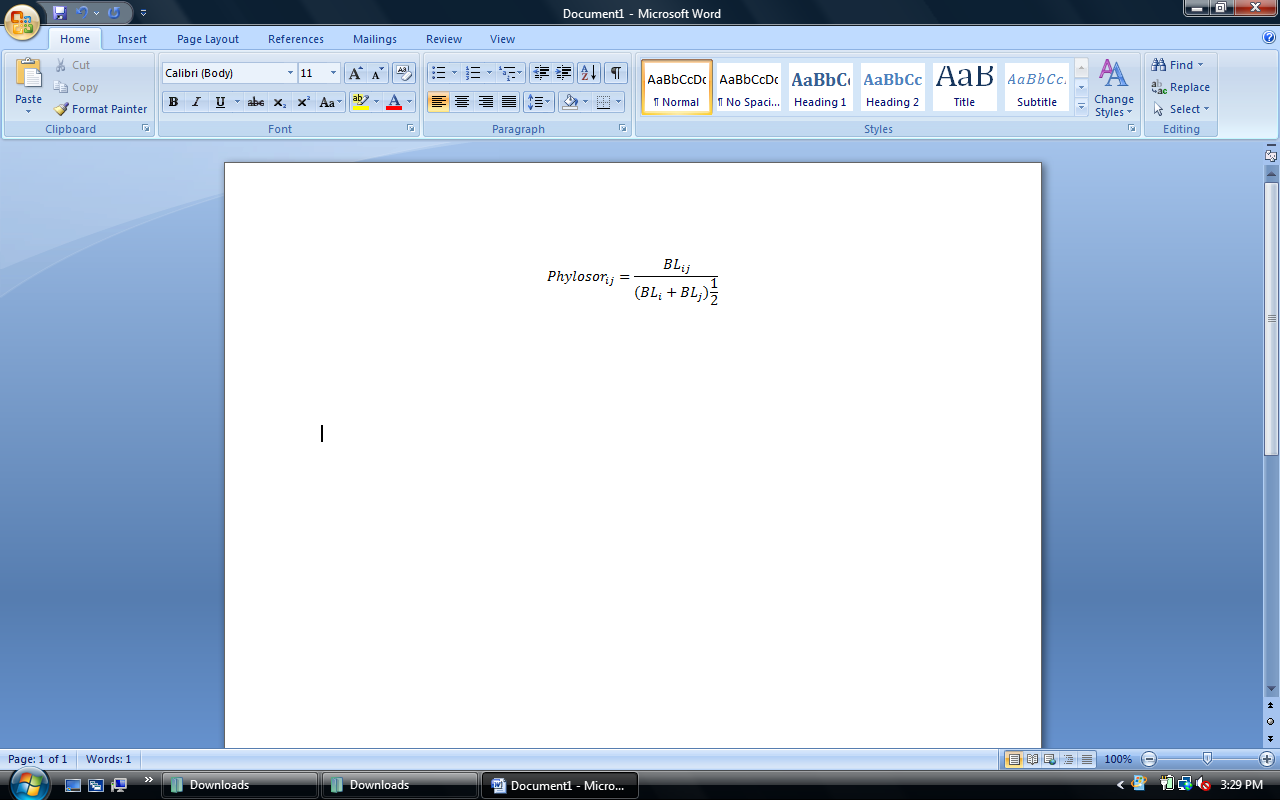
You can also analyze PD values under the different null models using the function ses.pd()

**Measuring phylogenetic betadiversity**

Comparing change in phylogenetic composition across sites (beta diversity or turnover) provides valuable insights about the ecological, historical and evolutionary processes that structure communities (Hardy & Senterre 2007). Analyzing patterns of phylogenetic beta diversity across communities can help us elucidate if any given lineage is driving turnover patterns between regions and during which time periods communities appear to be structured, as well as providing insights regarding the relative importance of processes such as in situ diversification vs. differential extinction in driving patterns of extant diversity and species compositions (Graham & Fine 2008).

Picante offers some functions to incorporate phylogenetic relationships into the calculations of beta diversity.

**Phylosor** (Bryant *et al.* 2008) – Sorensen’s index of phylogenetic beta diversity quantifies how phylogenetic similarity between pairwise communities varies, and is calculated using the following formula:



where, *BLij* is the branch length common to both communities *i* and *j*, and *BLi* and *BLj* are the total branch lengths of community *i* and *j*, respectively. The lower values (close to 0) represent two communities that have species which only share a very small root), while values close to 1 represent communities which are composed by the same taxa.

The function requires a community data matrix and a phylogeny as arguments, and generates a distance object of the PhyloSor index of similarity between communities.

phylosor.comun=phylosor(comun, prunedT\_nc)

phylosor.comun

Using PhyloSor, one can test whether two communities are phylogenetically more or less similar than what is expected given their taxa similarity using the same randomization procedures (null models) as those for estimating MPD and MNTD using the function phylosor.rnd()

**Comdist and comdistnt**

These functions also calculate measures of phylogenetic dissimilarity across communities (beta diversity), but these are based on the mean phylogenetic distance separating two taxa drawn randomly from different communities (MPD) and the average closest phylogenetic distance to the most similar taxon in the other community for taxa in two communities (MNTD).

Both require a community data matrix, a interspecific distance matrix generated from the phylogenetic tree, a decision about whether the measure should be weighted by abundance (false is the default) and whether conspecifics should be excluded (this option is useful when you have unresolved polytomies for certain groups in the phylogeny, the default is false). The results is a distance object of MPD/MNTD values separating each pair of communities

comdist.comun=comdist(comun, dm\_comun, abundance.weighted=FALSE, exclude.conspecifics=FALSE)

comdist.comun

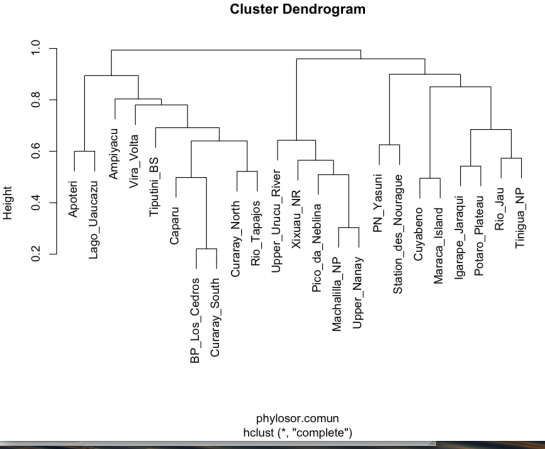
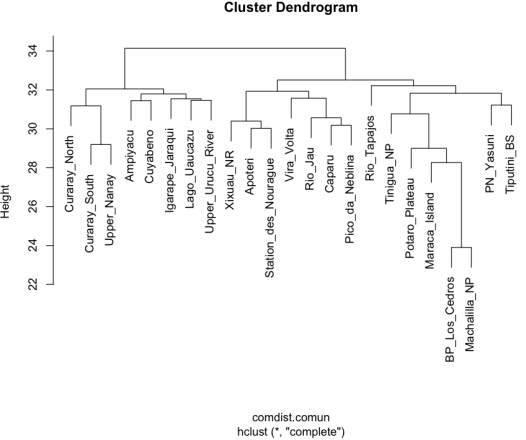
We can then use any of the distance matrices generated by these functions in hierarchical cluster analysis of phylogenetic dissimilarities among communities, and compare them with clusters formed with taxonomic dissimilarities (e.g. Sorensen’s index - function betadiver or vegdist from package vegan))

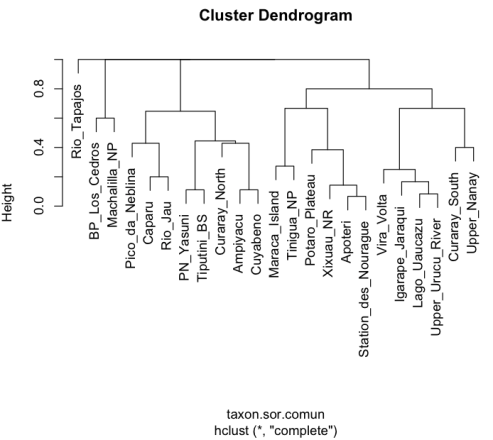
library(cluster)

plot(hclust(phylosor.comun))

plot(hclust(comdist.comun))

plot(hclust(taxon.sor.comun))





**References:**

Bryant, J.B., Lamanna, C., Morlon, H., Kerkhoff, A.J., Enquist, B.J. & Green, J.L. 2008. Microbes on mountainsides: Contrasting elevational patterns of bacterial and plant diversity. *Proceedings of the National Academy of Science* 105:11506–11511.

Cavender-Bares J., Kozak K., Fine P. & Kembel S. 2009. The merging of community ecology and phylogenetic biology. *Ecology Letters* 12:693–715.

Cooper, N., Rodriguez, J. & Purvis, A. 2008. A common tendency for phylogenetic overdispersion in mammalian. *Proceedings of the Royal Society B*. 275:2031–2037.

Faith, D.P. 1992. Conservation evaluation and phylogenetic diversity. *Biological Conservation* 61:1–10.

Gotelli, N.J. 2000. Null model analysis of species co-occurrence patterns. *Ecology* 81:2606–2621.

Graham, C.H. & Fine, P.V.A. 2008. Phylogenetic beta diversity: linking ecological and evolutionary processes across space in time. *Ecology Letters* 11(12):1265–1277.

Hardy, O.J. & Senterre, B. 2007. Characterizing the phylogenetic structure of communities by an additive partitioning of phylogenetic diversity. *Journal of Ecology* 95: 493–506.

Helmus, M.R., Bland, T.J., Williams, C.K. & Ives, A.R. 2007. Phylogenetic measures of biodiversity. *American Naturalist* 169:E68–E83.

Miklos I. & Podani J. 2004. Randomization of presence-absence matrices: Comments and new algorithms. *Ecology* 85:86–92.

Webb, C.O. 2000. Exploring the phylogenetic structure of ecological communities: an example for rainforest trees. *American Naturalist* 156:145–155.

Webb, C.O., Ackerly, D.D., McPeek, M.A. & Donogue, M.J. 2002. Phylogenies and community ecology. *Annual Review of Ecology and Systematics* 33:475–50.

Fucntions:

Before running the codes make sure you set the working directory to where you have all your files at

setwd("Users/mariamercedesgavilanez/Desktop/R/")

install.packages("vegan",repos="http://R-Forge.R-project.org")

install.packages("ape",repos="http://R-Forge.R-project.org")

install.packages("picante",repos="http://R-Forge.R-project.org")

library(picante)

READ SAMPLE DATA FRAME

comun =read.table("/Users/mariamercedesgavilanez/Desktop/R/comun.txt", header=T, row.names=1)

comun

dim(comun)

class(comun)

READ PHYLOGENY

tree\_nc =read.tree("/Users/mariamercedesgavilanez/Desktop/newick\_tree\_COOPER.txt")

tree\_nc=read.tree()

class(tree\_nc)

tree\_nc

plot(tree\_nc, cex=0.4)

PHYLOGENETIC DISTANCE MATRIX

dm\_tree.nc=cophenetic(tree\_nc)

dm\_tree.nc

dim(dm\_tree.nc)

PRUNING THE TREE ACCORDING TO SPECIES PRESNET IN SAMPLE COMMUNITIES

prunedT\_nc =prune.sample(comun, tree\_nc)

prunedT\_nc

plot(prunedT\_nc, cex=0.4)

PHYLOGENETIC DISTANCE MATRIX

dm\_comun =cophenetic(prunedT\_nc)

dim(dm\_comun)

PLOTTING SAMPLE SPECIES IN A PHYLOGENY

comun\_6=read.table("/Users/mariamercedesgavilanez/Desktop/comun6.txt", header=TRUE, row.names=1)

par(mfrow=c(2,3))

for(i in row.names(comun\_6)) {

plot(prunedT\_nc, show.tip.label = FALSE, main = i)

tiplabels(tip = which(comun\_6[i, ] >0), pch = 19, cex =1)

}

PHYLOGENETIC DIVERSITY METRICS

pd\_comun= pd(comun, prunedT\_nc)

pdcomun

psd.comun = psd(comun, prunedT\_nc, compute.var=TRUE)

psd.comun

psr(comun, prunedT\_nc)

PHYLOGENETIC COMMUNITY STRUCTURE

?ses.mpd

ses.mpd.comun.complete.tree=ses.mpd(comun,dm\_tree.nc,null.model="taxa.labels",runs=100,iterations=100)

ses.mpd.comun.pruned.tree=ses.mpd(comun,dm\_comun,null.model="taxa.labels",runs=100,iterations=100)

PLOTING HISTOGRAMS OF DISTRIBUTIONS OF VALUES USING DIFFERENT SPECIES POOL’S (PHYLOGENIES)

par(mfrow=c(2,2))

hist(ses.mpd.comun.complete.tree$mpd.rand.mean, main="mpd.random comun. complete tree")

hist(ses.mpd.comun.pruned.tree$mpd.rand.mean, main="mpd.random comun. pruned tree")

hist(ses.mpd.comun.complete.tree$mpd.obs.p, main="mpd.obsv.pvalue. complete tree")

hist(ses.mpd.comun.pruned.tree$mpd.obs.p, main="mpd.obsv.pvalue. pruned tree")

PHYLOGENETIC BETADIVERSITY

phylosor.comun=phylosor(comun, prunedT\_nc)

phylosor.comun

comdist.comun=comdist(comun, dm\_comun, abundance.weighted=FALSE, exclude.conspecifics=FALSE)

comdist.comun

taxon.sor.comun=vegdist(comun, method="bray")

install.packages("cluster",repos="http://R-Forge.R-project.org")

library(cluster)

plot(hclust(phylosor.comun))

plot(hclust(comdist.comun))

plot(hclust(taxon.sor.comun))