Measuring phylogenetic diversity within communities

Jesús N. Pinto-Ledezma and Jeannine Cavender-Bares

The main goal of this practice is to present basic understanding about measuring phylogenetic diversity with in communities or best known as the analysis of community phylogenetics. The community phylogenetics integrates ecological and evolutionary concepts and explores the mecanisms (e.g., biotic interactions or environemntal filters) governing the assembly of ecological communities.

There are different sources of information and web pages with a lot of information about this field. The most common and useful are the web pages of the books: "Modern Phylogenetic Comparative Methods and Their Application in Evolutionary Biology" (Garamszegi, 2014) (http://www.mpcm-evolution.org/) and "Phylogenies in Ecology" (Cadotte and Davies, 2016) (https://www.utsc.utoronto.ca/~mcadotte/page-3/).

Install and load packages

Check if you are in the correct working directory.

```
getwd()
```

Now install the load the necessary packages.

```
package.names <- c('picante', 'pez', 'car', 'vegan', 'MASS', 'ecodist', 'FD', 'adephylo', 'phytools')

if ( ! (package.names[1] %in% installed.packages())) {install.packages(package.names[1], dependencies =
```

Or you can use a conditional!

```
missing_pkgs <- package.names[which(!package.names %in% installed.packages())]

install.packages(missing_pkgs)
```

Or you can use a conditional inside the for loop!

```
package.names <- c('ape', 'picante', 'pez', 'car', 'vegan', 'MASS', 'ecodist', 'FD', 'adephylo')

for (pkg in package.names) {
  if (!require(pkg, character.only = TRUE, quietly = TRUE)) {
    install.packages(pkg)
    library(pkg, character.only = TRUE)
  }
}
```

Prepare data

Load phylogentic and community data

```
trMB <- ape::read.tree("Data/ALLMB.CCESR.tre")
#trMB$tip.label <- gsub(" ", "_", phy$tip.label)
#trMB$tip.label <- gsub("sp.", "sp", trMB$tip.label)
```

Let's inspect the tree.

```
length(trMB$tip.label)
head(trMB$tip.label)
```

Now load the community data.

```
BBSraw <- read.csv("Data/BBSpecies.biomass.2014.csv")[2:5]
bio.dat <- BBSraw
head(bio.dat)
```

```
##   Year Plot Biomass.g.m2                Species
## 1 2014    2         1.42     Aristida basiramea
## 2 2014    2         0.92            Cyperus sp.
## 3 2014    2         0.02          Digitaria sp.
## 4 2014    2         0.02 Euphorbia glyptosperma
## 5 2014    2       182.12     Lespedeza capitata
## 6 2014    2        22.03   Miscellaneous litter
```

```
bio.dat$Species <- gsub(" ", "_", bio.dat$Species)
bio.dat$Plot.Year <- paste(bio.dat$Plot, bio.dat$Year, sep = ";", collapse = NULL)   #join plot and yea
bio.dat <- bio.dat[, -c(1, 2)]
bio.dat <- bio.dat[, c(3, 1, 2)]
```

Most of the time we work with community data from different sources or collected by our peers, friends, students, this cause that some species names in the community data are not updated or are wrong. We can solve this issue manually by inspecting the spreadsheet or rely the task to R, in any case you can solve the issue species by species...

```
bio.dat$Species <- gsub("Petalostemum_purpureum", "Dalea_purpurea", bio.dat$Species)
bio.dat$Species <- gsub("Petalostemum_candidum", "Dalea_candida", bio.dat$Species)
bio.dat$Species <- gsub("Petalostemum_villosum", "Dalea_pulchra", bio.dat$Species)
bio.dat$Species <- gsub("Taraxicum_officinalis", "Taraxacum_croceum", bio.dat$Species)
bio.dat$Species <- gsub("Leptoloma_cognatum", "Digitaria_ciliaris", bio.dat$Species)
bio.dat$Species <- gsub("Artemisia_.caudata._campestris", "Artemisia_caudata", bio.dat$Species)
bio.dat$Species <- gsub("Achillea_millefolium.lanulosa.", "Achillea_millefolium", bio.dat$Species)
bio.dat$Species <- gsub("Euphorbia_.supina._maculata", "Euphorbia_supina", bio.dat$Species)
bio.dat$Species <- gsub("Tragopogon_dubius_.major.", "Tragopogon_dubius", bio.dat$Species)
bio.dat$Species <- gsub("Ambrosia_artemisiifolia_elatior", "Ambrosia_artemisiifolia", bio.dat$Species)
bio.dat$Species <- gsub("Andropogon_gerardi", "Andropogon_gerardii", bio.dat$Species)
bio.dat$Species <- gsub("Erigeron_canadensis", "Erigeron_canadense", bio.dat$Species)
```

Or, update the species names using a **for loop**:

```
oldsp <- c("Petalostemum_purpureum", "Petalostemum_candidum", "Petalostemum_villosum",
           "Taraxicum_officinalis", "Leptoloma_cognatum", "Artemisia_.caudata._campestris",
           "Achillea_millefolium.lanulosa.", "Euphorbia_.supina._maculata",
           "Tragopogon_dubius_.major.",
           "Ambrosia_artemisiifolia_elatior","Andropogon_gerardii", "Erigeron_canadensis")

newsp <- c("Dalea_purpurea", "Dalea_candida", "Dalea_pulchra",
           "Taraxacum_croceum", "Digitaria_ciliaris", "Artemisia_caudata",
           "Achillea_millefolium", "Euphorbia_supina", "Tragopogon_dubius",
           "Ambrosia_artemisiifolia", "Andropogon_gerardii", "Erigeron_canadense")

for(i in 1:length(oldsp)){
  cat("FROM", oldsp[i], "TO", newsp[i], "\n")
```

```
  bio.dat$Species <- gsub(oldsp[i], newsp[i], bio.dat$Species)
}
```

```
## FROM Petalostemum_purpureum TO Dalea_purpurea
## FROM Petalostemum_candidum TO Dalea_candida
## FROM Petalostemum_villosum TO Dalea_pulchra
## FROM Taraxicum_officinalis TO Taraxacum_croceum
## FROM Leptoloma_cognatum TO Digitaria_ciliaris
## FROM Artemisia_.caudata._campestris TO Artemisia_caudata
## FROM Achillea_millefolium.lanulosa. TO Achillea_millefolium
## FROM Euphorbia_.supina._maculata TO Euphorbia_supina
## FROM Tragopogon_dubius_.major. TO Tragopogon_dubius
## FROM Ambrosia_artemisiifolia_elatior TO Ambrosia_artemisiifolia
## FROM Andropogon_gerardii TO Andropogon_gerardii
## FROM Erigeron_canadensis TO Erigeron_canadense
```

Ok, until now we loaded our phylogeny and updated the species names in the community data, however, the phylogeny is the complete phylogeny of Smith and Brown (2018) and includes a total of 356.305 species (length(trMB$tip.label)). Now, we will prepare a phylogeny to include only the species that are present in our community.

```
#Get list of species in big bio
spnames <- unique(bio.dat$Species)
trMBcom <- drop.tip(trMB, setdiff(trMB$tip.label, spnames))
setdiff(spnames, trMBcom$tip.label)
```

And now we will prepare the community data matrix that we will use for further analyses.

```
bio.dat <- data.frame(na.omit(bio.dat))
head(bio.dat)

BBScom <- data.frame(sample2matrix(bio.dat))
BBScom[1:10, 1:10]
```

Finally, we have all data necessary for calculating different phylogenetic diversity metrics, Yay!

Phylogenetic diversity metrics

Before to continue lets check again if our data (phylogeny and community) match! To do this we will use the awesome function **match.phylo.com()** from the package **picante**.

Now we can clean our **R environment** and only maintain the information needed for analyses.

```
ls() # obtain the names of objects stored in the environment
rem <- ls()
rem
rm(list = rem[2:10])
```

```
matched <- picante::match.phylo.comm(phy = trMBcom, comm = BBScom)
```

Ok, now lets inspect the data that were stored in the object matched.

```
matched$comm[1:10, 1:10]
```

```
plot(matched$phy, show.tip.label = FALSE)
```

Explore diversity metrics

Awesome, we are now ready to explore some the of the **jungle** of metrics for the evaluation of phenotypic and phylogenetic structure of communities (Pausas and Verdú 2010).

Phylogenetic diversity

```
sum(matched$phy$edge.length) #  sum of the total branch lengths in the community
```

```
BBSpd <- pd(matched$comm, matched$phy, include.root = FALSE) # Faith's PD
head(BBSpd)
```

```
cor.test(BBSpd$SR, BBSpd$PD)
```

```
plot(BBSpd$SR, BBSpd$PD, xlab = "Species richness", ylab = "PD (millions of years)", pch = 16)
```

Mean pairwise distance (MPD) and mean nearest-pairwise distance (MNTD)

Other common metrics are MPD and MNTD. As in PD, lets calculate MPD and MNTD mannualy.

```
# MPD
dist.trMB <- cophenetic(matched$phy)
dist.trMB <- dist.trMB[lower.tri(dist.trMB, diag = FALSE)]

mean(dist.trMB)
```

```
# MNTD
dist.trMB2 <- cophenetic(matched$phy)
diag(dist.trMB2) <- NA
apply(dist.trMB2, 2, min, na.rm = TRUE)

mean(apply(dist.trMB2, 2, min, na.rm = TRUE))
```

And now using the package **picante**

```
BBSmpd <- mpd(matched$comm, cophenetic(matched$phy)) # MPD
head(BBSmpd)
```

```
BBSmntd <- mntd(matched$comm, cophenetic(matched$phy)) # MPD
head(BBSmntd)
```

Community diversity metrics

The analyses of community phylogenetic started making inferences about the mechanisms structuring the local communities through the evaluation of phylogenetic arrangements in local communities (see Cavender-Bares et al. 2009 for an initial criticism). However, new methods are now available, such that more complex balance between ecological and historical processes at local and regional scales can be incorporated into the analyses (Pigot and Etienne 2015, Pinto-Ledezma et al. 2019).

Now, lets calulate some of the most common metrics.

PD - phylogenetic diversity is the sum of the total phylogenetic branch length for one or multiple samples.

```
# We can also calculate the standardized effect size of PD in each community
BBScdm <- ses.pd(matched$comm, matched$phy, runs = 99)
BBScdm <- BBScdm[, c(1, 2, 6, 7)]

head(BBScdm)
```

Rao's quadratic entropy (Rao 1982) is a measure of diversity in ecological communities that can optionally take species differences (e.g. phylogenetic dissimilarity) into account.

```
require(phytools)
# Simpson's
BBSraoD <- raoD(matched$comm, force.ultrametric(matched$phy))

BBScdm$RaoD <- BBSraoD$Dkk
```

MPD - Mean pairwise distance separating taxa in a community

```
# SESMPD
BBSsesmpd <- ses.mpd(matched$comm, cophenetic(matched$phy), runs = 99)

BBScdm$MPD <- BBSsesmpd[, c(2)]
BBScdm$sesMPD <- BBSsesmpd[, c(6)]
BBScdm$MPDpval <- BBSsesmpd[, c(7)]
```

MNTD - Mean nearest taxon distance for taxa in a community

```
# SESMNTD
BBSsesmntd <- ses.mntd(matched$comm, cophenetic(matched$phy), runs = 99)

BBScdm$MNTD <- BBSsesmntd[, c(2)]
BBScdm$sesMNTD <- BBSsesmntd[, c(6)]
BBScdm$MNTDpval <- BBSsesmntd[, c(7)]
```

Phylogenetic species variability (PSV) quantifies how phylogenetic relatedness decreases the variance of a hypothetical unselected/neutral trait shared by all species in a community.

```
# PSV or phylogenetic species variability
BBSpsv <- psv(matched$comm, matched$phy, compute.var = TRUE)

BBScdm$PSV <- BBSpsv[, 1]
```

Phylogenetic species richness (PSR) is the number of species in a sample multiplied by PSV.

```
# PSR or phylogenetic species richness
BBSpsr <- psr(matched$comm, matched$phy, compute.var = TRUE)

BBScdm$PSR <- BBSpsr[, 1]
```

Phylogenetic species evenness (PSE) is the metric PSV modified to incorporate relative species abundances.

```
# PSR or phylogenetic species evenness
BBSpse <- pse(matched$comm, matched$phy)

BBScdm$PSE <- BBSpse[, 1]
```

qD(p) is a metric that measure the variation in species' divergences within communities. This metric is a modification of the Hill index, weighting a species' proportional abundance by its relative share of phylogenetic information.

```r
# Scheiner 2012 qD(p)
source("R-Functions/qDp.R")

BBSqDp <- qDp(matched$phy, matched$comm, q = 2)

BBScdm$qDP <-   BBSqDp
```

```r
head(BBScdm, 10)
```

Now, rearrange the metric calculated in the previous steps.

```r
BBScdm2 <- BBScdm[, c(1, 2, 5, 6, 9, 12, 13, 14, 15)]
names(BBScdm2) <- c("SR", "PD", "RaoD", "MPD", "MNTD", "PSV", "PSR", "PSE", "qDP")
head(BBScdm2)
```

Compare the metrics

```r
scatterplotMatrix(BBScdm)
```

```r
cor.table(na.omit(BBScdm))
```

```r
plot(BBScdm2$MPD, BBScdm2$PSV, xlab = "MPD", ylab = "PSV", pch = 17)
```

```r
BBSmds <- metaMDS(na.omit(BBScdm2), trace = FALSE)

ordiplot(BBSmds, type = "t", display = "species")
```

What do you think? Which metric is better?

References

Cadotte, M. W. and Davies, T. J. (2016). Phylogenies in Ecology: A Guide to Concepts and Methods. Princeton: Princeton University Press.

Cavender-Bares, J., Kozak, K. H., Fine, P. V. A. and Kembel, S. W. (2009). The merging of community ecology and phylogenetic biology. Ecology letters 12, 693–715.

Garamszegi, L. Z. (2014). Modern Phylogenetic Comparative Methods and Their Application in Evolutionary Biology. (ed. Garamszegi, L. Z.) Berlin: Springer-Verlag.

Helmus, M. R. 2007. Phylogenetic measures of biodiversity. American Naturalist 169:E68-E83.

Pausas, J. G. and Verdú, M. (2010). The Jungle of Methods for Evaluating Phenotypic and Phylogenetic Structure of Communities. BioScience 60, 614–625.

Pinto-Ledezma, J. N., Jahn, A. E., Cueto, V. R., Diniz-Filho, J. A. F., & Villalobos, F. (2019). Drivers of Phylogenetic Assemblage Structure of the Furnariides, a Widespread Clade of Lowland Neotropical Birds. The American Naturalist, E000–E000.

Rao, C. R. 1982. Diversity and dissimilarity coefficients: a unified approach. Theoretical Population Biology 21:2443.

Smith, S. A., & Brown, J. W. (2018). Constructing a broadly inclusive seed plant phylogeny. American Journal of Botany, 105(3), 302–314.

Scheiner, S. M. (2012). A metric of biodiversity that integrates abundance, phylogeny, and function. Oikos, 121(8), 1191–1202.