11.1 Phylogenetic community ecology in R

In this exercise we will use R to estimate some simple community phylogenetics measures.

Learning goals

- Use *R* to input a community dataset and phylogeny.
- Display community data on a phylogeny.
- Estimate PD, MPD, MNTD, NRI and NTI.

Inputting community data into R

To begin, install and load the R package picante:

```
install.packages ("picante", dependencies = TRUE)
library(picante)
```

It is also worth setting your working directory at this point:

```
setwd("C:/Desktop/MyWorkingDirectory")
```

Next we need to load in our community data. picante uses the same data structure as the community ecology package vegan (a very useful package for all kinds of community ecology calculations). The data is structured as follows, where numbers are the abundance of each of 7 species in 3 communities:

```
           Alouatta1 Ateles1 Ateles2 Cebus1 Cebus2 Saimiri1 Saimiri2
community1         0       0       0      1      2        2        1
community2         1       1       0      2      0        1        0
community3         0       0       1      3      2        1        0
```

However, getting your data into this format is not easy. Luckily picante provides a function which converts data automatically inot the correct format. For this function to work, each row of your data needs to contain the community ID, abundance, and species ID separated by tabs with no column headings, for example:

```
community1  1  Cebus1
community1  2  Cebus2
community1  2  Saimiri1
community1  1  Saimiri2
community2  2  Cebus1
community2  1  Saimiri1
community2  1  Aloutta1
community2  1  Ateles1
community3  3  Cebus1
community3  2  Cebus2
community3  1  Saimiri1
community3  1  Ateles2
```

To continue with this example, enter these data into an Excel spreadsheet and then save as a tab delimited text file. The data can then be read into R as follows:

```
comm.ds<-readsample("community.data.from.above.txt")
comm.ds
           Alouatta1 Ateles1 Ateles2 Cebus1 Cebus2 Saimiri1 Saimiri2
community1         0       0       0      1      2        2        1
community2         1       1       0      2      0        1        0
community3         0       0       1      3      2        1        0
```

Remember that you will need to set your working directory to the correct location in order to read in any files. As usual, make sure that the species names in your community dataset match those in your phylogeny.

Note that picante automatically enters zeroes for species which are not found within a given community (e.g. Cebus2 in community 2) so it is not necessary to enter this data yourself.

Next we need to load our phylogeny. For this example we will create a simple phylogeny as follows:

```
cat("primates(((Saimiri1:2, Saimiri2:2):1, (Cebus1:2,Cebus2:2):1):4,
((Alouatta1:2, Alouatta2:2):1,(Ateles1:2, Ateles2:2):1):4);", file =
"ex.tre", sep = "\n")
tree <- read.tree("ex.tre")
str(tree)
plot(tree)
```

There are a few things we need to do to the tree before we can use it. Firstly we need to prune out any species which are not in our community dataset. picante has a function for doing this, however you can also use drop.tip (see other sections).

```
tree<-prune.sample(comm.ds, tree)
```

picante assumes that the taxa in your phylogeny and the taxa in your dataset are in the SAME ORDER. This is easy to do:

```
comm.ds<-comm.ds[,tree$tip.label]
comm.ds
           Saimiri1 Saimiri2 Cebus1 Cebus2 Alouatta1 Ateles1 Ateles2
community1        2        1      1      2         0       0       0
community2        1        0      2      0         1       1       0
community3        1        0      3      2         0       0       1
```

We can now look at the distribution of species within our three communities across the phylogeny.

```
par(mfrow = c(2, 2))
plot(tree)
for (i in rownames(comm.ds)) {



        plot(tree, show.tip.label = FALSE, main = i)
        tiplabels(tip = which(comm.ds[i, ] > 0), pch = 19, cex = 2, col
        ="red")




        legend("topleft" , i, bty = "n")

        }
```

Community phylogenetics measures

1. Phylogenetic diversity (Faith 1992)

Phylogenetic diversity (PD) calculates the total branch length spanned by the species within a community. To calculate PD for our communities:

```
monkey.pd <- pd(comm.ds, tree, include.root = TRUE)
monkey.pd
           PD SR
community1 14  4
community2 20  4
community3 19  4
```

PD = phylogenetic diversity; SR = species richness.

2. Mean pairwise distance (MPD) and mean nearest taxon distance (MNTD) (Webb et al. 2002)

Mean pairwise distance (MPD) is the mean phylogenetic distance (i.e. branch length) among all pairs of species within a community. Mean nearest taxon distance (MNTD), or mean nearest neighbor distance (MNND), is the mean distance between each species within a community and its closest relative.

MPD is thought to reflect phylogenetic structuring across the whole phylogeny, whereas MNTD reflects phylogenetic structure closer to the tips. Both require the phylogeny to be represented as a phylogenetic distance matrix:

```
phy.dist<-cophenetic(tree)
```

MPD and MNTD can then be calculated as follows:

```
monkey.mpd<-mpd(comm.ds, phy.dist)
monkey.mpd
    [1] 5.333333 11.333333 9.666667
monkey.mntd<-mntd(comm.ds, phy.dist)
 monkey.mntd
    [1] 4 6 7
```

3. Net-relatedness index (NRI) and nearest taxon index (NTI)

MPD and MNND are useful summary statistics, but in order to compare values among different communities we need to standardize the values. The net-relatedness index (NRI) and nearest taxon index (NTI) do this. The code works by estimating MPD or MNND for N randomly assembled communities. The observed value of MPD or MNND is then used with the values obtained from this null distribution to calculate NRI or NTI:

$$NRI = -1*(MPDobs\text{-}mean\ MPDnull/\ sdMPDnull)$$
$$NTI = -1*(MNTDobs\text{-}mean\ MNTDnull/\ sdMNTDnull)$$

In picante, rather than calculating NRI and NTI, standardized effect sizes (SES) are reported. These values are equivalent to -1 times NRI or NTI. To calculate SESmpd or SESmntd:

```
monkey.ses.mpd<-ses.mpd(comm.ds, phy.dist, null.model = "taxa.labels", abundance.weighted = FALSE, runs = 100)

monkey.ses.mpd
           ntaxa    mpd.obs mpd.rand.mean mpd.rand.sd mpd.obs.rank  mpd.obs.z
community1     4   5.333333      10.39333   0.7345567          1.0 -6.8885086
community2     4  11.333333      10.31000   0.8892108         90.5  1.1508332
community3     4   9.666667      10.29333   1.1363596         26.0 -0.5514686
           mpd.obs.p runs
community1 0.00990099  100
community2 0.89603960  100
community3 0.25742574  100
```

mpd.obs = MPD for the community; mpd.obs.z = standardized MPD (equivalent to -NRI); mpd.obs.p = p value of observed mpd vs. null communities.

```
monkey.ses.mntd<-ses.mntd(comm.ds, phy.dist, null.model = "taxa.labels", abundance.weighted =
FALSE, runs = 100)

monkey.ses.mntd
```

```
              ntaxa mntd.obs mntd.rand.mean mntd.rand.sd mntd.obs.rank   mntd.obs.z
community1      4       4          5.92        1.070070            6  -1.79427459
community2      4       6          6.07        1.037236           43  -0.06748705
community3      4       7          6.01        1.029808           79   0.96134401
              mntd.obs.p runs
community1 0.05940594   100
community2 0.42574257   100
community3 0.78217822   100
```

mpd.obs = MPD for the community; mpd.obs.z = standardized MPD (equivalent to -NRI); mpd.obs.p = p value of observed mpd vs. null communities.

Positive values of mpd.obs.z (or mntd.obs.z) and high p values (> 0.95) indicate phylogenetic evenness, i.e., species within the community are more distantly related than expected by chance. Negative values of mpd.obs.z (or mntd.obs.z) and low p values (< 0.05) indicate phylogenetic clustering, i.e., species within the community are more closely related than expected by chance.

Information on the abundance of species within each community can also be incorporated as follows:

```
monkey.abund.ses.mpd<-ses.mpd(comm.ds, phy.dist, null.model="taxa.labels", abundance.weighted=
TRUE, runs = 100)

monkey.abund.ses.mpd
```

```
              ntaxa  mpd.obs mpd.rand.mean mpd.rand.sd mpd.obs.rank mpd.obs.z
community1      4 3.888889      7.425556    0.9923080          3.5 -3.564082
community2      4 8.160000      7.422400    0.8450798         91.0  0.872817
community3      4 5.632653      7.120816    1.1924824         10.5 -1.247954
              mpd.obs.p runs
community1 0.03465347   100
community2 0.90099010   100
community3 0.10396040   100
```

Note that this changes the interpretation of the results because the method uses the mean phylogenetic distances among individuals rather than the mean phylogenetic distances among species.

References

Faith, D. P. 1992 Conservation evaluation and phylogenetic diversity. *Biological Conservation*. **61**, 1-10.
Kembel, S.W., Cowan, P.D., Helmus, M.R., Cornwell, W.K., Morlon, H., Ackerly, D.D., Blomberg, S.P. and Webb, C.O. 2010. Picante: R tools for integrating phylogenies and ecology. *Bioinformatics*.**26**:1463-1464.
Webb, C. O., Ackerly, D. D., McPeek, M. A. &Donoghue, M. J. 2002 Phylogenies and community ecology. *Annual Reviews of Ecology and Systematics*. **33**, 475-505.

Contributed by Natalie Cooper