

Lab 16: Community Ecology and Phylogenetics

Today's lab we will utilize the *picante* package of *R* to calculate community phylogenetic statistics, and visualize the results on the phylogeny. This package has many of the operations found in the program *phylocom* available from: www.phylodiversity.net/phylocom. The two statistics we will focus on are NRI and NTI, defined in today's lecture handout. There are many other functions in the *picante* package for phylogenetic analysis of community ecology. If you are interested in these types of analyses then I would suggest that you read over the *picante* manual.

Open *R* and load the *picante* package.

We will use some of the data that is already found in the *picante* package. It is the same data that comes with *phylocom*.

data(phylocom)

To look at the elements of this list:

plot(phylocom\$phylo)

This is the made up tree for our example.

phylocom\$sample

Is a matrix of abundance data for the taxa in this tree from 6 different locations. *phylocom\$traits* is another matrix with discrete character data.

Plotting community data

Let's look at how the taxa found in those communities are distributed on our tree. First we need to create a vector matching the columns of the sample matrix to the taxa in our tree.

taxa<-match(colnames(phylocom\$sample),phylocom\$phylo\$tip.label)

Then translate those taxa into the branches that end with them.

branch<-match(taxa,phylocom\$phylo\$edge[,2])

The next step is to create an array for which the rows of the array correspond to the communities in our sample matrix and the columns are the branches of the tree. The elements will then be 1=black for all the branches that aren't in our community and 2=red for all that are.

colors.comm<-array(1,dim=c(6,dim(phylocom\$phylo\$edge)[1]))

colors.comm[,branch][phylocom\$sample>0]<-2

“[,branch]” will limit what we're changing to only those columns that correspond to the

appropriate branches. “[phylocom\$sample>0]” will only change the elements of those columns that are greater than 0 or in other words the elements that have some taxa.

Finally we will plot the tree with each of the communities shown at the tips.

```
plot(phylocom$phylo,edge.color=colors.comm[1,],edge.width=2)
```

Look at the names of the rows of *phylocom\$sample* matrix. Look through all six communities by using the different rows of the *colors.comm* matrix. Do the names of the rows seem to make sense based on what you see in these trees?

Calculating NRI and NTI

Now we will calculate NRI and NTI for our different communities. First we need to make a phylogenetic distance matrix.

```
phy.dist<-cophenetic(phylocom$phylo)
```

Take a look at *phy.dist*. This is a matrix where the rows and columns are the taxa and the elements of the matrix are the phylogenetic distance between those pairs of taxa. To calculate the NRI:

```
ses.mpd(phylocom$sample, phy.dist,null.model="taxa.labels")
```

The rows are the communities. The first four columns should be pretty straight forward given the definition of the NRI given in today’s lecture; the fifth column is the rank of that score against randomized communities; the sixth column is the negative NRI; and the seventh column is the one tailed p-value for significantly high NRI. So that communities 1, 2 and 3 re significantly clustered and community 5 is significantly spread out.

You can also calculate NTI.

```
ses.mntd(phylocom$sample, phy.dist,null.model="taxa.labels")
```

The explanations for the column are as above, except substitute NTI for NRI.

Question 1. Compare the results to the figures we made for each community – do the NRI and NTI values make sense? Explain how.

Null Distributions

You can also use different null models to calculate these values. “*taxa.labels*” shuffles the distance matrix across the taxa in the community. Let’s try “*phylogeny.pool*”, which randomly draws species from the phylogeny for its null distribution.

```
ses.mpd(phylocom$sample, phy.dist,null.model="phylogeny.pool")
```

Did that have much affect on our estimate of NTI and our p-values?

There are several other null distributions available. Look at the documentation for these two functions to see your options.

Reading Data

Open the 'oakphylo' file in a text editor. This is the default name for the phylogeny read in by phylocom, and it is a simple newick string (with or without branch lengths). To open it in *R* drag the file onto the terminal and edit around the file name to use the *ape* function *read.tree*.

```
tree.oak <- read.tree("filename")  
plot(tree.oak)
```

Open the 'oaksample' file. This is the format to input species lists for individual communities. As you can see the first column is the community the second column is the abundance and the third column is the taxon. Drag it into *R*, and then edit around it to read:

```
sample.oak<-read.table("filename")  
sample.oak<-sample2matrix(sample.oak)  
sample.oak
```

There's your sample matrix.

Question 2. Plot each community on the tree. Calculate NRI and NTI for each community. Compare results to Fig. 4 from Cavender-Bares et al. 2004 (see lecture handout). Do they tell the same story? Explain.

Question 3. Class challenge: To help you get an intuition for NRI and NTI, your challenge is to construct communities where these are as different as possible. Using the default 32-taxon tree in '*phylocom\$phylo*', create hypothetical communities and calculate NRI and NTI. It may be easiest to create a matrix in the form of "*oaksample*" and convert that to a sample matrix using *sample2matrix*. Try to create a community with $NRI < NTI$ and the largest possible difference between them. Ditto for $NTI < NRI$. Bonus points to the lab group with the best example of each one.