Class 10b - Tools for Phylogenetic Ecology

This is a very brief introduction to some of the R tools available for phylogenetic community ecology and trait analysis. There are lots of levels to this, with many conceptually challenging options. The goal here is to make it possible to access some of the tools available in R (Picante), Picante's parent program Phylocom, and Phylomatic. We will not delve deeply into the theory.

Needed for this Class:
Packages you need to install
picante: is the R Package that is the R implementation of Phylocom.
ape: R Package for Analysis of Phylogenetic and Evolution. Picante calls to ape.
taxize: a bunch of really useful tools for taxonomy, useful for cleaning up names
vegan: vegetation analysis package. Picante calls to vegan

Optional but Recommended Software (not R based):
Phylocom is a stand-alone program by Campbell Webb, David Ackerly, and Steven Kembel for the Analysis of Phylogenetic Community Structure and Character Evolution.
Install Phylocom 4.2 software from http://www.phylodiversity.net/phylocom/ (Mac OS X or Windows).
Phylomatic is companion software to Phylocom (included in the Phylocom 4.2 download) that takes species lists, a reference phylogenetic super tree, and creates a Newick-format tree of your species. This file is needed for analyses in Phylocom or Picante.
We will look at how you can call out from R to other programs, like phylocom

Plantminer is a website that provides updated classification information for plants
LOG ON at www.plantminer.com to get your own api key number, which is in the upper right of the page. You will use it for accessing plant names

Kembel_picante_walkthrough.pdf Steve Kembel (picante author) created a super handout for a workshop to introduce picante to fungal ecologists. The handout he is super and is posted on the web site.

ferp_taxa.txt**:** a family/genus/code file of FERP Angiosperms
ferp_taxa.new**:** a Newick file of the angiosperm woody species on the FERP
ferp_comm6ha.csv**:** a comma-delimited community matrix of FERP data, by ha (Angiosperms)
R201208290gg3eric.new: megatree based on APG3 (courtesy Phylodiversity.net), with some additional resolution in the Ericaceae

```
#Read in datasets and open libraries
require(picante); require(taxize)
samp<-read.csv
("http://people.ucsc.edu/~ggilbert/Rclass_docs/ferp_comm6ha.csv",row.names=1)
# this is a comma-delimited community matrix of FERP data, by ha
(Angiosperms)
phy<-read.tree("http://people.ucsc.edu/~ggilbert/Rclass_docs/ferp_taxa.new")
# this is a Newick file of the angiosperm woody species on the FERP
traits<-
read.csv("http://people.ucsc.edu/~ggilbert/Rclass_docs/ferp_traits.csv",row.n
ames=1)
# a file with some traits of the angiosperm woody species on the FERP

#Set api.key from Plantminer (upper right) to your api key number
api.key <- 3080788976 # an integer with your api key

#create a vector with the names of your plants
#these are the woody angiosperms from the FERP
plants <- c ("Acer macrophyllum","Arbutus menziesii","Arctostaphylos
andersonii","Arctostaphylos tomentosa","Baccharis pilularis","Ceanothus
thyrsiflorus","Corylus cornuta","Cotoneaster franchetii","Cotoneaster
pannosus","Crataegus monogyna","Eucalyptus globulus","Hedera
helix","Heteromeles arbutifolia","Ilex aquifolium","Notholithocarpus
densiflorus","Lonicera hispidula","Morella californica","Pyracantha
angustifolia","Quercus agrifolia","Quercus parvula","Rhamnus
californica","Ribes divaricatum","Salix lasiandra","Sambucus
nigra","Toxicodendron diversilobum","Umbellularia californica","Vaccinium
ovatum")
##Done reading in datasets
```

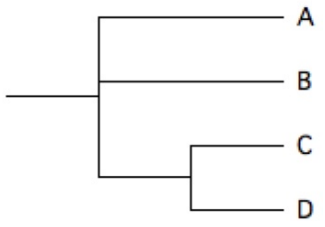
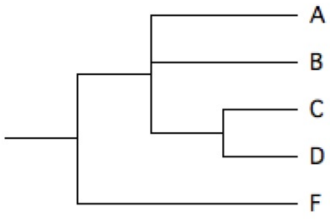
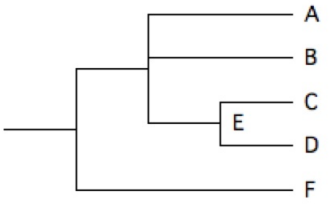
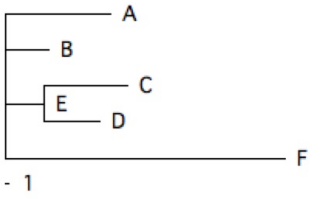
Setting up phylocom

Within the phylocom folder is a mac (for macs) and a w32 (for windows) folder that contain the engines to run phylocom and phylomatic. Choose the one that makes sense for you for setting up your paths (see later).

To use the bladj and phylomatic functions, you need to include an appropriate ages file and an appropriate supertree phylogeny in the same folder as the phylocom and phylomatic engines. One approach for plants is to use the wikstrom.ages file found in the bladj_example folder (see the README). Copy that file, put it in the mac (or w32) folder, and rename it "ages".

For a supertree, for plants you can use the R20120829.new newick file that is the latest APGIII-based supertree, or any other tree that is appropriate for your needs. This also needs to be in the folder. I've posted a slightly modified version of the (with added resolution for ericaceae) on the website (R20120829gg3eric.new).

Newick Tree Format. This is a common notation used to represent phylogenetic trees. At a minimum it shows the relationships among leaf nodes; it can also include intermediate named nodes (here E), and distances for the nodes.

Just include the leaf nodes (A,B,(C,D));	
Root it on F ((A,B,(C,D)),F);	
Label internal node E ((A,B,(C,D)E),F);	
Add distances ((A:19,B:8,(C:15,D:10)E:7),F:50);	

To take a quick look at trees, make an object using read.tree

```
atree<-read.tree(text="((A,B,(C,D)E),F);")
plot(atree, show.node.label=TRUE)
```

Getting a tree for a list of Angiosperms of interest to you.
There are a number of Supertrees available, from which you can create derive a tree specific to your species of interest. Note that these supertrees usually work at the level of family, sometimes genus. So this is a rough pass. If you need more resolution, you need to supply your own trees. You can find trees at TreeBase: http://www.treebase.org

The most recent Angiosperm tree is the APG3 megatree (R201208290.new). This tree does not have node ages associated with it, but just topology.

How to make sure you have the most recent accepted names for plants?

Here we will take a list of plants of interest (Genus species), use Plantminer to get the family names and prepare a phylocom-ready species list,
Then call out directly to phylocom and phylomatic to create a dated tree.
Our names are in the object plants, above
#use plantminer function from taxize to call out to plantminer website
#and return the correct plant family, genus, and species

```
ctn<- plantminer(plants,api.key)   #returns the names

#take the returned names and create a variable with names in phylomatic
 format
mytaxa<-
 tolower(paste(ctn$fam,ctn$genus,paste(ctn$genus,ctn$sp,sep="_"),sep="/"))

mytaxa

#THIS CREATES A FILE IN THE FORM USED BY PHYLOCOM
# [1] "sapindaceae/acer/acer_macrophyllum"
# [2] "ericaceae/arbutus/arbutus_menziesii"
# [3] "ericaceae/arctostaphylos/arctostaphylos_andersonii"
# [4] "ericaceae/arctostaphylos/arctostaphylos_tomentosa"
# [5] "asteraceae/baccharis/baccharis_pilularis"
# [6] "rhamnaceae/ceanothus/ceanothus_thyrsiflorus"
#…
#[25] "anacardiaceae/toxicodendron/toxicodendron_diversilobum"
#[26] "lauraceae/umbellularia/umbellularia_californica"
#[27] "ericaceae/vaccinium/vaccinium_ovatum"

#NOW CALL OUT FROM R TO USE PHYLOMATIC AND PHYLOCOM FUNCTIONS
#TO CREATE AND DATE A TREE
```

```
#Set file and working directory names
mypath<-"/Users/ggilbert/Desktop/"
PMpath<-"/Users/ggilbert/Documents/phylocom-4.2/mac/" #wd for phylocom
mytaxafile<- paste(PMpath,"mytaxa.txt",sep="") #name for phylocom taxa file
mymastertree<-paste(PMpath,"R20120829gg3eric.new",sep="") #supertree name
myoutfile<-paste(PMpath,taxafile,".new",sep="") #name for output newick file
mycleanoutfile<- paste(PMpath,taxafile,".clean.new",sep="") #name for cleaned
 newick file
mydatedoutfile<-paste(PMpath,taxafile,".dated.new",sep="") #dated newick

setwd(PMpath) #set the working directory to the phylocom folder

#save taxafile into the phylocom folder
write.table(phylotaxa, mytaxafile,
 row.names=FALSE,col.names=FALSE,quote=FALSE)

system(paste(PMpath,"phylomatic -f ", mymastertree," -t ",mytaxafile, " >
 ",myoutfile,sep="")) #extract a tree from the mastertree

system(paste(PMpath,"phylocom cleanphy -f ", myoutfile," -e >
 ",mycleanoutfile,sep=""))  #cleanphy to remove one-daughter nodes

system(paste(PMpath,"phylocom bladj -f ", mycleanoutfile," > ",
 mydatedoutfile,sep=""))  #date the nodes using Wikstrom dates

#newick file has a head and tail that do not work in picante.
#need to get rid of the tail )euphyllophyte:1.00000 and the leading (, and
 any space in the newick tree.
phytext<-scan(mydatedoutfile,what="character",nmax=-1,sep="]")
phytext<-sub(pattern=")euphyllophyte:1.000000",replacement="",x=phytext)
phytext<-sub(pattern="\\(",replacement="",x=phytext)
phytext<-gsub(pattern=" ",replacement="",x=phytext)
write(x=phytext,file=mydatedoutfile)
#should now have a nice clean dated file

mytree<-read.tree(mydatedoutfile) #read in the file
plot(mytree,cex=.5,show.node.label=TRUE) #take a look at it
axisPhylo() #include dated axis

#write the species list and the newick file to the desktop
setwd(mypath)
write(x=phytext,file= "mytaxa.txt.dated.new") #save dated newick file to
 desktop
write(x=mytaxa,file= "mytaxa.txt")
```

Playing with the FERP data in R Picante
Picante uses three kids of files:
(1) sample file (vegan community matrix) **(samp)**
(2) phylogeny file (a Newick tree) **(phy)**
(3) traits file (traits of all each species) **(traits)**

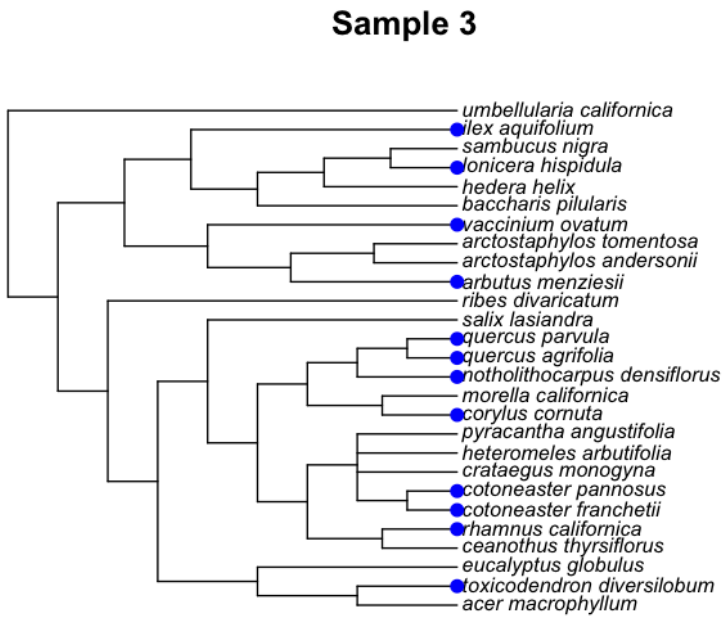
Make sure that the samp, phy, and traits files have the same taxa, in the same order!

```
phy<-prune.sample(samp,phy) #phylogeny only includes taxa in samp
samp<-samp[,phy$tip.label] #samp is in same order as phy
traits<-traits[phy$tip.label,] #traits are in same order as phy
```

Looking at the phylogenetic tree
Use plot to show trees. Use ?plot.phylo to get help
plot(phy) #simple tree in phylogram style

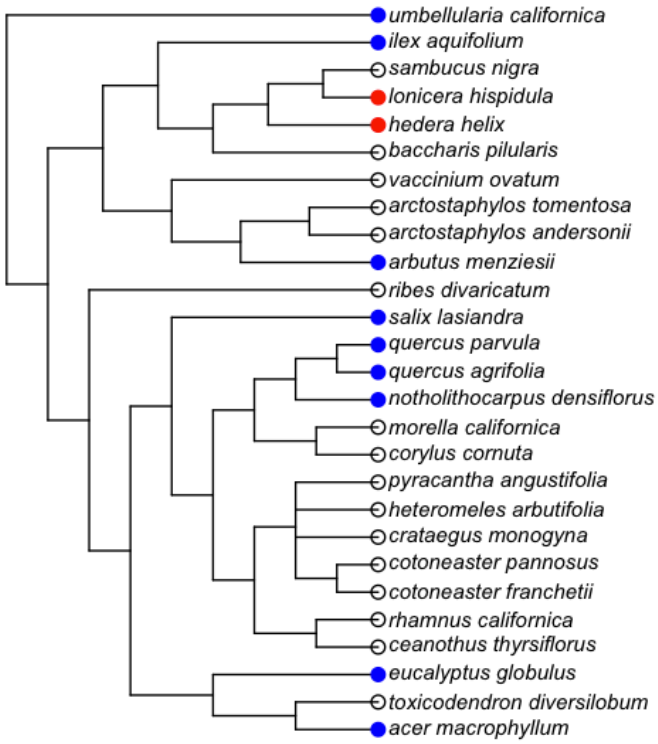
par(mfrow=c(2,2)) #a few variations
plot(x=phy, show.node.label=TRUE, cex=.75) #show internal nodes and make labels smaller
plot(phy,type="cladogram",cex=.75) #unrooted tree
plot(phy,type="fan", cex=.75) #fan style
plot(phy,type="radial", cex=.75) #radial style
par(mfrow=c(1,1))

Which species are in a particular sample?
plot(phy, show.tip.label=TRUE, main="Sample 3",label.offset=4,cex=.75) #draw tree with labels
tiplabels(tip=which(samp[3,]>0),pch=19,col="blue") #add blue dots for species in sample 3



Which species are trees, shrubs, or vines from the traits data?

```
#labels each leaf node by a trait
# in this case blue trees, open circle shrubs, red vines
plot(phy, show.tip.label=TRUE, label.offset=10,cex=.7)
tiplabels(tip=which(traits$habit=="tree"),pch=19,cex=1,col="blue")
tiplabels(tip=which(traits$habit=="shrub"),pch=1,cex=1,col="black")
tiplabels(tip=which(traits$habit=="vine"),pch=19,cex=1,col="red")
```



#or given the labels themselves informative colors
plot(phy, type="fan", show.tip.label=TRUE,
tip.color=as.numeric(traits$habit),edge.width=3,cex=.6)

Calculate the phylogenetic distance between taxa

```
phydist<-as.data.frame(cophenetic(phy))
```

Some Basic Phylogenetic Diversity measures

Calculate the phylogenetic distance between taxa

```
phydist<-cophenetic(phy)  #pairwise time of independent evolution
phydist
```

Phylogenetic diversity within samples

```
#PD is Faiths phylogenetic diversity
#SR is Species Richness
 pd(samp,phy,include.root=TRUE) #measures within each sample
```

Mean pairwise phylogenetic distance among all species in each sample. Shuffles labels across tip 999 times to get p value

phydist<-cophenetic(phy)
ses.mpd(samp,phydist,null.model="taxa.labels",abundance.weighted=FALSE,runs=999)

Mean nearest taxon distance measure

ses.mntd(samp,phydist,null.model="taxa.labels",abundance.weighted=FALSE,runs=999)

```
#positive z values and high p values (p>.95) indicate phylogenetic evenness.
#negative z values and low quantile (p<.05) indicate phylogenetic clustering.
```

There are a number of other null models available:

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
frequency: Randomize community data matrix abundances within species (maintains species occurence frequency)
richness: Randomize community data matrix abundances within samples (maintains sample species richness)
*independentswa*p: 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

Nearest taxon distance among all samples

comdistnt(samp,phydist)

	1	2	3	4	5
2	50.50505				
3	61.14943	81.90476			
4	68.95238	112.94118	96.88889		
5	54.11765	91.71717	36.32184	75.80952	
6	33.93939	52.08333	48.09524	68.62745	55.75758