An introduction to plant community phylogenetic and trait patterns, and a tutorial on how to analyze them

By Ty Taylor

This tutorial contains:

Part 1

Understanding and analyzing phylogenetic patterns in community assembly

Taxonomic Name Resolution Service

- TNRS "Taxon Scrubber", by iPlant collaborative, cleans plant taxonomic lists
  - Matches names from your list to database at Missouri Botanical Gardens
  - Fixes spelling errors
  - Recognizes synonims (one species that has been given >1 name) and provides current accepted name
  - Provides current accepted genus and family names, and correct author(s) (author of first published species description with that name)
  - Located at: http://tnrs.iplantcollaborative.org/
- What to expect: mostly, family names will change, spellings will be fixed, and where synonyms occur, the current accepted name will be provided.
- Watch out for
  - Mispelled genera: if specific epithet is not matched (or was a "sp"), might match to different genus (usually match < 0.5)
    - → be suspicious and check family.
  - Warning "Higher scoring names were found"
    - Download results with "all matches" and check what alternative names were provided
- Search settings: Take a look at these in upper right of application, but typically leave them as default.
  - For example "constrain best match by higher taxonomy" will ensure that a misspelled genus name will be matched instead of looking for best overall match to genus_species combo. That can produce the warning stated above, which should be investigated.

Using TNRS Taxon Scrubber

- Copy list of **species names** and paste into Taxon Scrubber window, click "submit list" (file for this tutorial is "list for TNRS.xlsx")
  - List should be a column with hard returns after each taxon, but format of names can vary: e.g.,

    Ceiba_pentandra
    Clitoria racemosa
    MUSA TEXTILIS
- Can click on individual "accepted name" links for details
- Click "Download results"
  - I like to download twice: once with "Best matches only" in "simple format" and a second time with "All matches" in "simple format"
- Open text files and copy and paste each one into a separate worksheet in your excel book with your original lists.

Using TNRS Taxon Scrubber

- You will primarily use your "best matches" file, which has the same number of rows as your original taxon list, so it can directly replace that list.
- Use the "all matches" output to see details where multiple suggestions were provided.

Using TNRS Taxon scrubber

- First: Scan the "warnings" column in your "best matches" sheet for "partial match – higher scoring names were found"
  - Genus name probably did not match an accepted species name, and it may suggest something a completely different family.
  - Look at your "all matches" sheet to see all of the suggestions, and use your brain to pick the best match (one way is to compare the suggested family name with the original family name from your list)
  - Example: see Zizyphus_jujuba from tutorial file. TNRS "best matches" prioritizes a genus match "Zizyphus" for which it has no opinion of taxonomic status.
  - But in "all matches" you see option "Ziziphus jujuba", which has "accepted" status. You should choose that one.

Using TNRS Taxon scrubber

- Second: Look out for match scores <0.5
  - Specific epithet did not match, and genus did not match 100%. It may suggest the wrong family entirely. So look at options in "all matches", compare suggested family names to original family, and make your best judgment.
- Third: family names are often changed
  - This is normal, but if you have the taxonomic knowledge, compare original with suggested families side by side. If e.g., "Fabaceae" (peas) became "Pteridaceae" (a fern family) you probably got a funky name match.
  - If you are suspicious, go to www.tropicos.org , search both family names, and compare where they sit in the taxonomic hierarchy. If the orders differ, or worse, the classes, TNRS probably gave you a poor name suggestion.
- Fourth: synonyms are corrected
  - If a single species has been assigned different names (synonyms), the "status" column of the output will say "synonym"; use the "accepted name" provided.
- Finally: copy taxon names from "Accepted name", "Accepted genus", and "Accepted family" columns
  - Save both original and scrubbed names to keep a record of changes

Generating a phylogenetic tree with Phylomatic

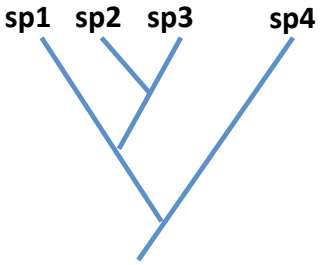
- Go to: http://www.phylodiversity.net/phylomatic/phylomatic.html
- Enter your scrubbed list (see TNRS section this tutorial) into the window. (for this tutorial, use file "B2TF_origin")
  - Phylomatic is picky, so format must be as follows, including symbols, with all lowercase letters (even though the site suggests first letter caps): family/genus/genus_species RETURN
  - With a good text editor (Notepadd ++ for Windows or Text Wrangler for Mac) you can change case of a long list easily. In Notepad ++, paste your list, select all and hit ctrl+u.
  - With everything in the right case back in Excel, concatenate the names together
    - e.g., in a new cell, type =concatenate(A1,"/",B1,"/",C1,"_",D1) and copy down. (be sure to copy and "paste special / values" if you don't want it to remain an equation)

Generating a phylogenetic tree with Phylomatic

- Copy and paste the concatenated list into Phylomatic
  - It is okay to have "sp." for unidentified species, it simply will match by genus.
- Output format: select "newick"
- Master phylogeny: select "maximally resolved seed plant tree"

What is a “newick” string?

- A newick string is a way of representing a phylogenetic tree in text format.
- The tree depicted below would be displayed in text like this:
  - ((sp1,(sp2,sp3)),sp4)



Getting started in Phylocom

- Download from: http://www.phylodiversity.net/phylocom
- There are useful tutorials at:
  - http://bodegaphylo.wikispot.org/Community_Phylogenetics
  - http://people.ucalgary.ca/~smvamosi/phylocom.htm
- Open the folder containing Phylocom on your computer. (*for Mac, see Appendix 1 “Troubleshooting Phylocom” at end of this tutorial)
  - For Windows: Go to folder “W32”
    - Double click “phylocom.bat”
    - Your command line console should open up with some Phylocom headings.
- *All commands in phylocom will now look for files in and deposit outputs to your W32 folder.
- **Phylocom is picky, so note **Appendix 1** in this tutorial “**Troubleshooting Phylocom**”.

Assigning branch lengths to your phylogeny in Phylocom

- Phylocom comes with a file called “wiskstrom.ages”, which contains known fossil ages of ancestral angiosperms.
- Phylocom can use these ages to assign branch lengths to your tree that are scaled with respect to time.
- Go to the “examples” folder, open “bladj_example” and **copy** the file “wikstrom.ages”
  - *NOTE: There is also a file just called “ages” in the “examples” folder. That file goes with the hypothetical example dataset and will not work to assign branch lengths to your tree.
- Paste “wikstrom.ages” into your W32 folder.
- Delete “wikstrom.” from the name so you are left with “ages”
- Copy your phylogeny text (newick) file, (for this tutorial, use file “phylo_B2TF_origin.txt”)
  - Paste it into W32 and call it “phylo” with no extension (i.e., delete “.txt” – you may need to adjust folder settings to “show extensions of known file types” before you can even see that the file still contains “.txt” then delete the extension. For Windows 7, click the ? Help button and search “show file extensions”)

Assigning branch lengths to your phylogeny in Phylocom

- With your “phylo” and “ages” files in the W32 folder, you can now assign branch lengths.
- In the command line terminal, type “phylocom bladj > phylo.ages.txt”
  - All commands begin with “phylocom …”
  - Bladj stands for the branch length adjust algorithm
  - The > symbol tells it to export your result to a file of the following name and type.
  - A text file will show up in W32 called “phylo.ages.txt” (but you could name it anything you want)
  - Open the result file to make sure it worked. You should see a newick string with lengths added, like “(genus_sp:75.000)”
- Save this somewhere; now you can copy this aged tree into W32 and make it your “phylo” file for other analyses.
- *Remember, **Appendix 1** in this tutorial contains hints for **troubleshooting Phylocom**.

Understanding community phylogenetic metrics NRI and NTI

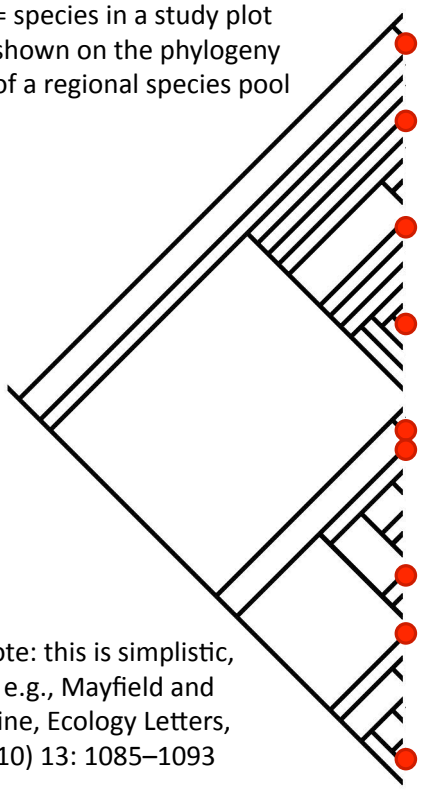
- Next we'll analyze Net Relatedness Index (NRI) and Nearest Taxon Index (NTI) for sample communities.
- These increasingly commonly used metrics reflect phylogenetic signal in the organization of your sample plant communities.
- Both compare your sample communities to random draws from the regional species pool
  - *Note: regional species pool = phylo file.
- How you define your regional pool is up to you. It is usually an area that contains all of the species that you think could colonize your study plot.
  - *Changing your regional species pool can give you different results; scale is important! (see Swenson *et al., Ecology, 87(10), 2006, pp. 2418–2424*)

Understanding community phylogenetic metrics NRI and NTI

- Both metrics tell you whether taxa in your community are more closely (clustered) or less closely (over-dispersed) related to each other than you would expect by chance.
  - The difference between them is in phylogenetic scale.
- NRI tells you about the whole phylogeny, and is a good indicator when the environment has excluded entire clades (groups of related species).
  - Abiotic filtering
- NTI tells you about the tips of the phylogeny, and is a good indicator when species exclusion is due to high competition between closely related species.
  - Limiting similarity

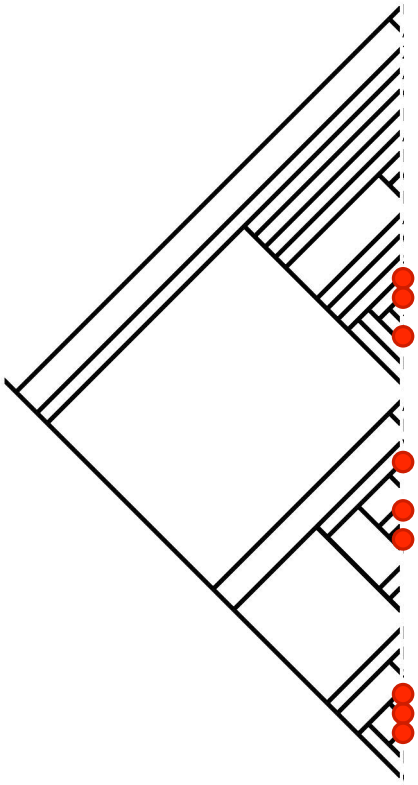
Limiting similarity

= species in a study plot shown on the phylogeny of a regional species pool



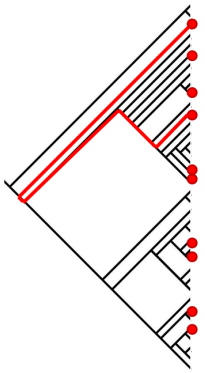
*Note: this is simplistic, see e.g., Mayfield and Levine, Ecology Letters, (2010) 13: 1085–1093

Abiotic filtering

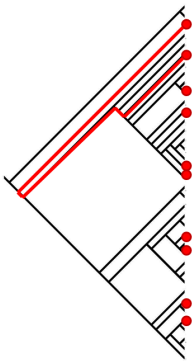


Understanding community phylogenetic metrics NRI and NTI

- The raw measure for NRI is mean pairwise phylogenetic distance (MPD).
  - The mean of the distances (length along phylogeny) between all possible pairs of species.
- The raw measure for NTI is mean nearest taxon distance (MNTD)
  - The mean of the distances between each species and its most closely related species.



Red lines indicate measured distances between taxa in study plot



Analyzing community phylogenetic metrics in Phylocom

- MPD.rnd and MNTD.rnd are the mean MPD and MNTD from your null communities generated by random draws from your regional species pool.
- NRI = -1 * (MPD.sample – MPD.rnd) / sd(MPD.rnd)
- NTI = -1 * (MNTD.sample – MNTD.rnd) / sd(MNTD.rnd)
  - Dividing by st.dev. scales the index by the variance among the randomized plots. This standardizes for regional phylogenies of different sizes (look at equations and think about why that's true).
- Statistical significance is determined by ranking the order of plots (rnd and sample) by MPD or MNTD.
  - If MPD.sample < MPD.rnd
    - p = 2*(#plots with MPD ≤ MPD.sample) / total # plots
    - Multiply by 2 because it is a two-tailed test.
    - So with 999 randomized plots, a MPD.sample ranking 15 would give p = 0.03.

Analyzing NRI and NTI in Phylocom

- You will need a phylogenetic tree with branch lengths assigned, and a "sample" file.
- Your sample file is a text file and contains three columns with no headers, separated by tabs:
  - 1) plot name (you can have multiple plots in this list)
  - 2) abundance of the species in your study plot
  - 3) species name exactly as it occurs in your phylogeny

b2_1991	7	aleurites_moluccana
b2_1991	1	anacardium_occidentale
b2_1991	12	annona_muricata
b2_1991	1	annona_sp.

- *Note: The R package "picante" can also measure these. But **Phylocom is useful** because its algorithms **can deal with polytomies** (multifurcations in phylogeny as opposed to bifurcations).

Analyzing NRI and NTI in Phylocom

- Copy your sample file into folder W32 and call it "sample" (remember, no extensions like ".txt").
  - Tutorial file: "sample_B2TF_1993.txt"
- Copy your phylogenetic tree file with branch lengths (previous section) into folder W32 and name it "phylo".
  - Tutorial file: "phylo.ages_B2TF_origin.txt".
- Go to your command line terminal with phylocom started (phylocom.bat in W32 folder)
  - Type: phylocom comstruct > result.xls
- The results suggest some phylogenetic clustering by NRI and NTI, but rank shows the result to be non-significant for both (see "Understanding NRI and NTI" in this tutorial).

Analyzing NRI and NTI in Phylocom

- Using switches in Phylocom
  - Switches are options, specified by the dash "-" symbol, that can be added after commands in any order.
  - The default number of randomized draws from the regional species pool is 2,999. You can specify 5,000 randomizations by typing:
    - phylocom comstruct –r 5000 > result.xls
  - Another comstruct default is that species are treated as present/absent. Use the "-a" switch weight the analyses by the abundances in your sample file:
    - phylocom comstruct –r 5000 –a > result.xls

Analyzing NRI and NTI in Phylocom

- Using your own null model
  - In my case, for analyzing community changes in the Biosphere 2 tropical forest mesocosm over time, I used my own null model (one more appropriate for a community that has experienced only extinction and no recruitment) to generate null communities.
  - So to analyze NRI and NTI, just create a sample list containing all of your null communities, and use:
    - phylocom comstruct –r 1 > result.xls
  - Limiting the analysis to one randomization ensures that it runs quickly across your null communities.
  - Then delete all entries in the result except plot name, MPD, and MNTD.
  - Put those raw results of your observed community at the top of the null communities in Excel.
  - Calculate average and st dev of MPD and MNTD of the null communities and calculate NRI and NTI using the equations in the previous section of this tutorial.
  - Determine significance by sorting MPD and MNTD ascending and numbering (ranking) them in order (see previous section of tutorial).

Part 2

Understanding and analyzing trait patterns in community assembly

Understanding community trait pattern metrics CWM and CWV

- Community weighted mean (CWM) and community weighted variance (CWV) are simple, but powerful metrics in the analysis of community assembly and responses to change.
- These are simply the first two moments in the abundance-distribution of a given trait in the community.
- These metrics are weighted by the relative abundance of each species.
  - The argument for abundance-weighting is that species are rarely evenly distributed in an environment, and the functioning of a forest is a result of the traits of each *plant*, not of each species.

Calculating community trait pattern metrics CWM and CWV

- Relative abundance (A)
  - A1 = #individuals in sp1 / total #individuals
- Community weighted mean
  - CWM = A1*sp1 trait value + A2*sp2 trait value + A3…
    - or the sum of abundance weighted trait values
- Community weighted variance
  - CWV = A1(value.of.trait1 – CWM)^2 + A2(v.trait2 – CWM)^2 +A3…
    - or the sum of abundance weighted squared deviations of species trait values from the community weighted mean

Calculating community trait pattern metrics CWM and CWV

- Calculating CWM and CWV with an R script
  - The equations are simple, but calculation is cumbersome across many samples or null communities of different sizes → use R!
  - An R script for calculating these across any number of plots of variable sizes is included with this tutorial, called "CWM_CWV.r"

Calculating community trait pattern metrics CWM and CWV

- CWM_CWV.r script for R
  - First, download the package "doBy" and all of its dependencies into R (if you don't know how to do this, Google "downloading packages in R")
  - Input data should be a .csv file (saved as such from Excel) called "rawdata.csv" that looks like this (with precisely these headings):

plot	species	abund	v.trait
plot1	aleurites_moluccana	6	12.2
plot1	anacardium_occidentale	1	24.8
plot1	annona_muricata	10	7

  - "v.trait" is the value of the trait (like max height), "abund" is the number of individuals of that species in the plot
  - It is okay if the file "rawdata.csv" has other columns with other information.
  - HOWEVER: rows with missing values (like no trait data) must be omitted!

Calculating community trait pattern metrics CWM and CWV

- CWM_CWV.r script for R
  - In R, set your working directory to the folder containing "rawdata.csv"
    - Tutorial file: "rawdata.csv" contains specific leaf area data from the Biosphere 2 tropical forest mesocosm at several time points.
  - Open up the file "CWM_CWV.r" either in R as a script, or in notepad. Copy and paste the entire code at once into R.
  - The script will output a file called "CWM_CWV_output.csv" into the folder specified as your working directory.
  - Opened in Excel, it will look like this:

plot	cwm	cwv
plot1	0.316327	0.213997
plot2	0.494444	0.244506
plot3	0.538462	0.248521

Part 3

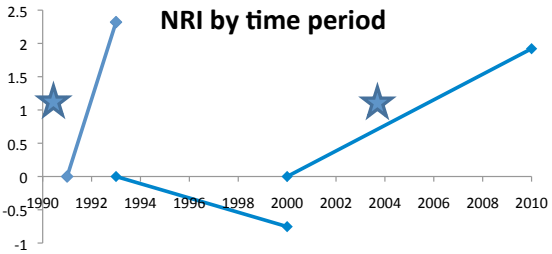
Understanding trends in community phylogenetics and traits

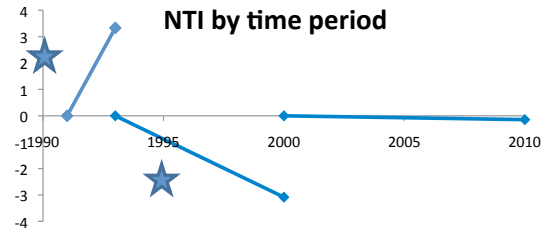
Understanding trends in community phylogenetics and traits

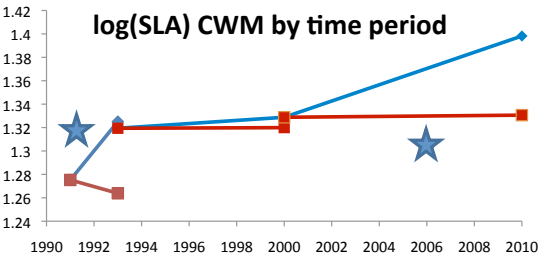
- You've learned how to analyze community phylogenetic trends (NRI and NTI) and community trait patterns (CWM and CWV).
  - Are they related? How do you expect them to change with respect to one another?
- All natural changes in a species pool are driven by species interactions with each other and with their environment.
  - **Those interactions are mediated by traits**.
- Let's take a look at an example of phylogenetic and trait patterns changing in tandem.
  - Data from analyses of the Biosphere 2 tropical forest biome (TC Taylor *et al.,* MS in prep)

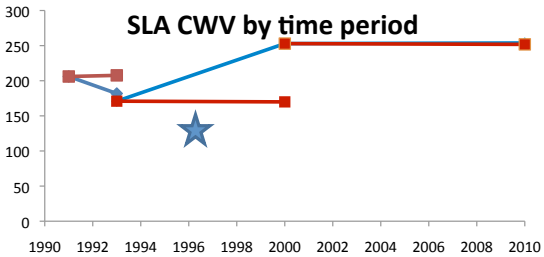
Understanding trends in community phylogenetics and traits

- Compare these trends in specific leaf area CWM and CWV with phylogenetic trends NRI and NTI
- When NRI goes up (environmental filtering), community trait means change.
- When NTI goes down (limiting similarity) trait variance increases!
- Again, this may be a bit simplified, read the literature and think about it!









Appendix 1

Troubleshooting Phylocom

Troubleshooting phylocom

- For all users
  - You must be able to see and delete extensions like ".txt" from files. Change settings to show extensions for known file types (next two slides).
  - If Phylocom crashes when you try to execute a command, it is usually because one of your files is not in the right format, examine them closely. Compare them to files in "examples" folder, or to the files included in this tutorial.

Troubleshooting Phylocom

- Mac users
  - Also see file accompanying this tutorial called "BJE_Phylocom_notes_for_mac.docx"
    - Open up program "Terminal" in Utilities folder in Applications
    - Type pwd to see your current directory (e.g., /Users/owner)
    - Now type cd../../ to move you up two directories.
    - Then type cd Applications/phylocom-4.2/
    - Type ls to see a list of directories
    - Type cd mac to change directory to the mac folder
    - The "mac" folder is where you should dump all of your working files like phylo, ages, sample, etc. Phylocom outputs will also go to mac folder.
  - Excecuting commands requires typing "./" first, e.g., ./phylocom comstruct > results.xls
  - Seeing and deleting extensions of known file types (like ".txt")
    - Hold control and right-click the file, choose 'get info', go to 'name and extensions'. If it says "phylo.txt", delete the ".txt" so it just says "phylo".
    - This must be done for phylocom to read files like "phylo", "sample", "ages", and "traits".

Troubleshooting Phylocom

- For Windows users
  - The simplest way to get started is to execute the .bat file in your W32 folder to avoid the hassle of setting your working directory, etc.
  - You must delete file extensions (like ".txt") for phylo, sample, ages, etc. files.
    - Windows hides file extensions for known file types. In XP go to folder options and find and deselect radio button for "hide extensions of known file types"
    - In Windows 7, click the ?Help button and search "show file extensions" to see access to options in control panel.