EEB 174

Project overview

Your final project is a chance to show off your programming skills in a biological context. Just about any project will work as long as it satisfies the following criteria:

- is grounded in a biological discipline
- requires you to perform operations in the shell
- uses python, R, or shell to manipulate text data
- uses python to read/write text files
- requires three or more novel functions
- includes a visualization of data in R
- includes a final report in markdown or latex with commented code and embedded figures
- uses a history of version control.

In addition, you will be required to give a short presentation on your project.

If you have no idea about possible directions I offer the following suggestions. Find a clade with a good fossil record and examine its evolutionary history. You can download occurrence data from the Paleo Biology Data Base. It is fairly straightforward to download a big csv file from this site and once you have that in hand you can go to work. Here are some ideas suggestions (you must use your code to generate plots):

1. Extract the number of unique species.
2. Plot diversity of your group though time (something like this). Here is one library for plotting geologic timescales in R
3. Plot the stratigraphic record of genera or families in your data set (something like this
4. Plot the occurrences of fossil species (map plotting examples in python here
5. Use your occurrence data to calculate speciation and extinction rats through time. You can do this using the program Pyrate but you will need to format your data appropriately (another chance to show off your programming skills). See this for a description of the general approach: Silvestro D., Schnitzler J., Liow L.H., Antonelli A. and Salamin N. (2014) Bayesian Estimation of Speciation and Extinction from Incomplete Fossil Occurrence Data. Systematic Biology, 63, 349-367. The link to pyrate is here; you will need to do the tutorials here
6. The R packages paleotree, fossil and strap may also be of interest but may also require you to find or create a phylogenetic tree for your taxa.

Note that this suggested project will require you to grapple with unfamiliar programs, search through literature, and manipulate novel data sets. In other words, it may feel frustrating and hard at times. This is normal and a necessary part of growing your programming skills.

If you are thinking of pursuing this idea for the Friday deadline, download occurrence data for three (or more) groups (we can't all work on dinosaurs...) and use your `tail`, `cut`, `sort` and `uniq` commands to find the number of unique species in the group. Write a text file called **"paleo-project-idea.txt"** and rank your preference for the three groups. Report the shell command and the result (species count) for each group and commit this to a new repo called **eeb-174-final-project**.