

Lab 11: Introduction to R

Introduction

During lecture today you were given a brief introduction to R. We will use R for several of the labs for the remainder of the semester. As mentioned in lecture, the major impediment of R is the steep learning curve. So, *I do not expect* you to become an R phylogenetics expert this semester. *I do not expect* you to be creating your own R packages (though some of you with programming language experience may be inclined to do so, which is great). That being said, R is an incredibly powerful tool for phylogenetic and comparative analyses and something you should utilize for your research.

The goals of the R labs this semester are to:

- I. Have you become comfortable loading packages and working with the R interface
- II. Illustrate the power and utility of a few R commands
- III. Provide you with resources (tutorials, websites, listservs) so that you can continue your R training after this semester

Today we'll discuss some very basic functions in R to get you up and running, use some phylogenetics packages, and simulate some data sets. This represents a lot of information, so if you are new to R, you will probably not get to part III right away.

The goals of the lab today are to:

- I. Learn Some R Basics (if you are very familiar with R, you can skip this)
- II. Introduce Some Phylogenetic Functions
- III. Simulate Data Sets in R (advanced)

Exercise 1: Learn Some R Basics

When you first open R or R studio, you will be greeted with a command line prompt ("**>**"). You have seen similar prompts before when you used PAUP* ("paup**>**") and MrBayes ("MrBayes**>**"). Just as with these previous programs, the first thing you generally want to do is set your working directory for the day.

On a Mac: `> setwd ("/Users/mala/Desktop/IB200/Lab11")`

On a PC: `> setwd ("c:\\Users\\mala\\Desktop\\IB200\\Lab11")`

You can make sure this worked by asking R what your working directory is:

```
> getwd()
```

Packages

All R functions and datasets are stored in packages. Only when a package is loaded are its contents available. You can see which packages are currently installed using the command:

```
> library ()
```

You can individually install packages using:

```
> install.packages ("ape")
```

Yes, include the quotes (") here. The package here is called APE: Analyses of Phylogenetics and Evolution. There are many, many packages for phylogenetic comparative analyses, which you can read about here:

<http://cran.r-project.org/web/views/Phylogenetics.html>

Wouldn't it be great if we could just install all the packages at once? Actually we can do that.

```
> install.packages("ctv")
```

```
> library("ctv")
```

```
> install.views("Phylogenetics")
```

This might take a minute or two, but now you most likely won't have to install any more packages in R this semester. New versions of packages are made available on a fairly regular basis, so you can update your packages using:

```
> update.packages()
```

Once you have packages installed, you need to load them into R. You need to do this each time you open R and want to use a specific package. Do NOT include quotes (") here. To load a package:

```
> library (ape)
```

Help and Comments

Comments in R are indicated by the # symbol. Basically anything following these symbols is not read by R.

You can ask for help about a package in general by typing:

```
> library (help=nameofpackage)
```

You can ask for help regarding a specific function in a package by typing:

```
> help (read.tree)
```

Saving Your Work

You can save the objects created during an R session by typing:

```
> save.image ("mywork.Rdata") ##mywork can be changed to whatever you want
```

This will save your work in the current directory. When you re-open R, you can load these objects by typing:

```
> load ("mywork.Rdata")
```

Now that you have learned the basics of how R operates, you should now work through **Lesson1.R**

Question #1: Send me the answers to the questions in Lesson1.R

Exercise 2: Basic Phylogenetic Functions in R

Now that you have worked through some basic functions of R in the previous lesson, let's examine some specific phylogenetic functions. The APE package can read phylogenetic trees in Newick or NEXUS format using the `read.tree` or `read.nexus` commands. For today's lesson, let's use the `Geospiza.nex` file, which contains a phylogeny for a clade of birds. We'll also load a data file called `Geospizameasure.txt` that contains some measurements for those birds. Make sure these files are in your working directory. Load the packages APE, PHYTOOLS, and GEIGER.

```
> library (ape)
> library (phytools)
> library (geiger)
> geotree <- read.nexus("geospiza.nex")
> geodata <- read.table ("geospizameasure.txt")
```

Now let's plot this tree in several different ways.

```
> plot (geotree, edge.width = 2)
> plot (unroot (geotree), type = "unrooted", edge.width = 2)
> plot (geotree, type = "fan", edge.width = 2)
```

What does the `edge.width` function do?

Let's reroot the tree using the taxon `fusca`.

```
> georeroot <- root (geotree, "fusca")
```

Plot these two to see how they have changed

```
> plot (geotree)
> plot (georeroot)
```

You can examine the list of taxa in your phylogeny using:

```
> geotree$tip.label
```

You can examine the branch lengths of your phylogeny using:

```
> geotree$edge.length
```

We also loaded a data file containing measurement information. One thing we probably first want to do is make sure that the taxa in our phylogenetic tree match the taxa that we have measurement data for. We can do this using:

```
> treedata(geotree, geodata)
```

You should get a warning message indicating that one of the taxa in your tree was actually not included in your data table. The Geiger package automatically removes this taxon from your

analyses. Alternately, if you know that there is a specific taxon you want to remove from the tree you can use the APE function `drop.tip`

```
> geotreenofusca <- drop.tip(geotree, "fusca")
```

Exercise 3: Simulate Data Sets in R (Advanced)

If you are already familiar with the R interface and functions, you can simulate some data in R today. Work through the **regressionNull.R** script.

Additional Resources

Basic Introduction to R	http://cran.r-project.org/doc/manuals/R-intro.html
R Phylogenetics listserv	https://stat.ethz.ch/mailman/listinfo/r-sig-phylo
R Phylogenetics tutorials	http://www.r-phylo.org/wiki/HowTo_Table_of_Contents
Bodega Bay Tutorials	http://treethinkers.org/tutorials/
Phytools blog by Liam Revell	http://blog.phytools.org/