

Half-day Workshop on Phylogenetic Comparative Methods

Simon Joly

2023-02-10

Contents

1	Preface	5
2	Other useful information	7
3	Disclaimer	9
4	Ahead of the workshop	11
4.1	Install R and the required packages	11
4.2	Working with github	12
5	An introduction to Phylogenetic Comparative Methods	13

Chapter 1

Preface

This document consist in an introduction to the comparative methods. It contains theory as well as practical examples in R on Phylogenetic Generalized Least Squares (PGLS). It was developed for a half-day workshop that consists in short presentations followed by R exercises. Note that the present document should pretty much stand by itself because most of the theory given in the presentations are incorporated into the theory sections. Therefore, this document should contain all the necessary information to understand the examples.

I assume that the readers are “reasonably” familiar with R as well as with linear regression and its assumptions. There are a lot of good R introductory tutorials on the web and for linear models. Zuur et al. (2007) provide a good introduction to linear models, mixed-effects models and model comparison. Good introductions to model fitting in R can also be found on Dolph Schluter’s webpage and among the QCBS workshops.

Chapter 2

Other useful information

The links below might provide additional information of interest.

Luke Harmon's book - Phylogenetic Comparative Methods

Course on Phylogenetic Comparative Methods

Introduction to phylogenies in R - pdf

Reading and making phylogenetic trees in R - pdf

Chapter 3

Disclaimer

This tutorial is provided as is, without any guaranty that it will work or that the analyses will be up to date.

Chapter 4

Ahead of the workshop

Here are a few things you should know and you should do ahead of the workshop.

4.1 Install R and the required packages

To perform the examples of this document, you will need to have the R software installed on your computer. I also recommend that you install RStudio. If R Studio is not required, it facilitates interactions between scripts and the R console and provides many great tools.

After installing R, you will have to install some packages. For this specific tutorial, we will need to load the following R packages.

```
library(nlme)
library(ape)
library(RColorBrewer)
library(ggplot2)
```

To execute the code of this tutorial in R, you can just copy and paste the code in the boxes in your R console. This will replicate the analyses presented in the tutorial.

If some of the packages above are not yet installed on your computer, you get error messages when trying to load them. If this is the case, you will have to install them using the function `install.packages()`. You only have to install them only once.

```
install.packages('nlme')
install.packages('ape')
install.packages('RColorBrewer')
install.packages('ggplot2')
```

Once the packages are installed, you can load the packages using the `library()` function. Also note that if you are using both the packages `nlme` and `ape`, `nlme` should be loaded first. If you don't do this, you might get errors; you could then restart R and start over.

4.2 Working with github

All the data of the workshop are deposited in a github repository. The easiest way to access them is to go to the Workshop repository and clone it on your computer. The option to clone the directory is in a green box called “code”.

Once you have cloned the repository, you need to open R (R studio) and set the working directory to the repository folder (“/ComparativeMethods-HalfDayWorkshop”). In RStudio, to change the working directory, you choose `>Session>Set Working directory>Choose Directory`. You are now ready for the workshop.

Chapter 5

An introduction to Phylogenetic Comparative Methods

Phylogenetic comparative methods were introduced by Joseph Felsenstein in 1985. The idea of phylogenetic comparative methods was to correct for the non-independence of species in statistical tests because of their shared evolutionary histories. Indeed, two species may look similar not because they live in the same environment but because they are closely related. For instance, considering the following angiosperm phylogeny.

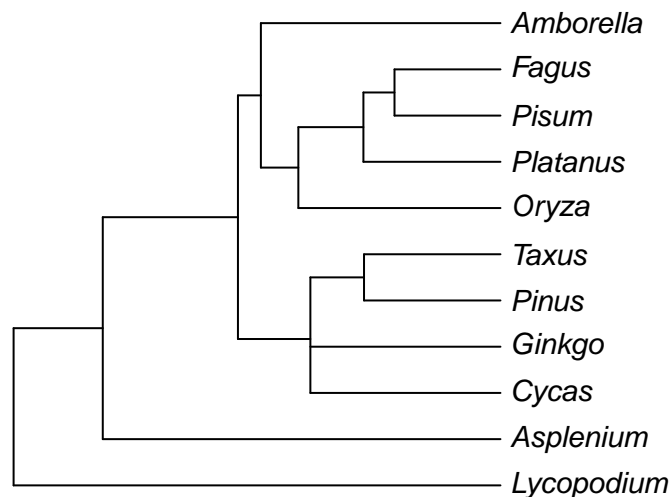
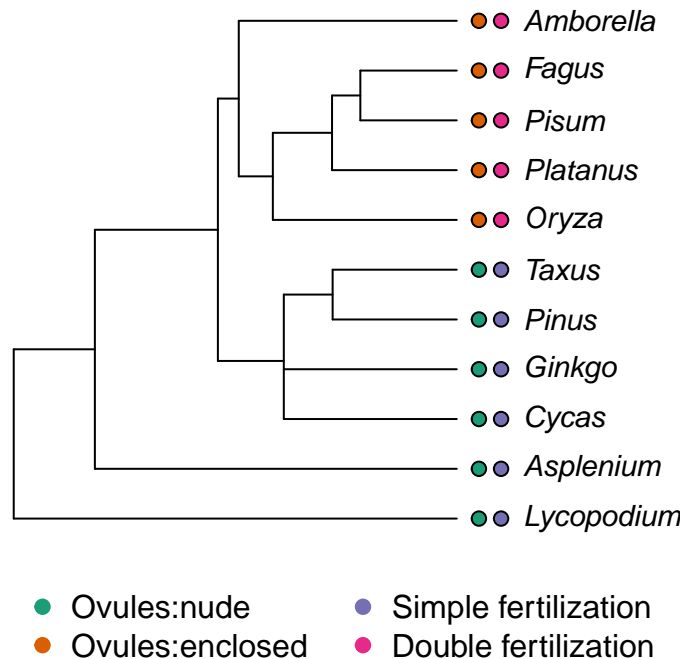


Figure 5.1: land plant phylogeny

It is clear that *Fagus* (beech) and *Pisum* (pea) are more likely to share similar characteristics compared to *Asplenium* (a fern), because they share a more recent common ancestor. In other words, their evolutionary histories are shared over a longer period than with *Asplenium*. As such, they have more chance to have more similar traits (and in fact they do). For instance, take two characters, ovule and fertilization type, within this group.



Ignoring the phylogeny, we might be tempted to see a strong correlation between these two characters. Indeed, the states between the two characters show a perfect correspondence. Using standard contingency table statistics, we could do a Fisher exact test:

```
fisher.test(matrix(c(5,0,0,6),ncol=2))
```

```
##
## Fisher's Exact Test for Count Data
##
## data: matrix(c(5, 0, 0, 6), ncol = 2)
## p-value = 0.002165
## alternative hypothesis: true odds ratio is not equal to 1
## 95 percent confidence interval:
## 2.842809 Inf
## sample estimates:
## odds ratio
## Inf
```

The test suggests that the association is highly significant. However, we know that the comparisons made are not completely independent. Actually, both characters evolved only once, and this along the same branch.

A more appropriate question would be “what is the probability that two characters evolved along the same branch?”. This can be calculated using a contingency table, but this time taking the observations along the branches of the phylogeny. In the example, there are 18 branches and the two characters evolved only once and on the same branch. The contingency table when considering the changes along the branches looks like this:

	Change in trait 2	No change in trait 2
Change in trait 1	1	0
No change in trait 1	0	17

With this table, Fisher’s exact test will give the following result:

```
fisher.test(matrix(c(1,0,0,17),ncol=2))

##
## Fisher's Exact Test for Count Data
##
## data: matrix(c(1, 0, 0, 17), ncol = 2)
## p-value = 0.05556
## alternative hypothesis: true odds ratio is not equal to 1
## 95 percent confidence interval:
##  0.4358974      Inf
## sample estimates:
## odds ratio
##      Inf
```

You can see that the result is no longer significant. While this approach is correct, more powerful comparative methods have been developed. One useful and powerful approach is the Phylogenetic Generalized Least Squares (PGLS) and it is the one that will be introduced next. But first, let’s do some revision and look briefly at the standard regression to better understand PGLS.