

## Exercise 4: Using trees

### Learning objectives

By the end of this exercise, you should better understand:

1. How to use R and RStudio with phylogenies and character data
2. Fitting evolutionary models to discrete and continuous characters
3. How to reconstruct ancestral states for discrete and continuous character data
4. Brownian motion and the rate parameter  $\sigma^2$

Today we will use RStudio and some ready-made datasets included there to practice using phylogenetic comparative methods.

1. Open RStudio on your computer.
2. Install some packages by typing the following commands into the prompt, each line followed by “enter”:

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

3. Now, load these packages so we can work with them:

```
> library(ape)
> library(phytools)
> library(geiger)
```

4. Let's first use a dataset for mammals, which includes data on their body mass and home range size, and a phylogenetic tree.

Download the data and phylogeny by typing:

```
> data("mammal.data")
> data("mammal.tree")
```

5. You can look at the mammal tree by typing:

```
> plot(mammal.tree)
```

6. Look at the mammal data by typing:

```
> mammal.data
```

### Reflect:

**What kinds of characters are included in this dataset? What should we do first to work with these data?**

## Phylogenetic Independent Contrasts

7. Transform your data and create variables for each data type by entering the following one at a time into the prompt:

```
> bodymass <- log(mammal.data$bodyMass)
> names(bodymass) <- row.names(mammal.data)

> homerange <- log(mammal.data$homeRange)
> names(homerange) <- row.names(mammal.data)
```

8. You can check the data class of your data by entering:

```
> class(mammal.tree)
```

9. See if you can calculate the phylogenetic independent contrasts (PIC) for each character on your own.

### HINTS:

- remember to save your variables (I recommend pic.homerange and pic.bodymass)
- start by typing:

```
> help(pic)
```

10. Once you have calculated the PICs for each character, plot your values and their linear regression line by typing the following. HINT: remember to use the variable names from your PIC calculations:

```
> fit.pic<-lm(pic.bodymass~pic.homerange+0)

> plot(pic.homerange,pic.bodymass)

> abline(fit.pic,lwd=2,lty="dashed",col="red")
```

### Reflect:

**Are your variables strongly correlated?**

## Ancestral Character State Reconstruction for continuous characters

11. Using your body mass data from the last exercise, test the fit of the Brownian Motion (BM), Ornstein-Uhlenbeck (OU), Early-Burst (EB) and BM with trend (Trend) models using the command **fitContinuous**

To fit the BM model, type:

```
> fitContinuous(mammal.tree,bodymass,model="BM")
```

use this command to find how to test the other models:

```
> help("fitContinuous")
```

**Reflect:**

**Which model was the best-fit to your data? How did you decide this?**

12. Using the best-fitting model, reconstruct the ancestral states for your body mass data using the command **ace**. Assign it a variable (I suggest `ace.bodymass`)

13. Type your variable name to see the results.

14. Plot your ancestral character reconstructions using `contMap`  
(remember to change the text if you didn't give it the variable name `ace.bodymass`)

```
> contMap(mammal.tree,bodymass,method="user",anc.states=ace.bodymass$ace)
```

**Reflect:**

**What is the reconstructed body mass for the last common ancestor of mammals?**

**How can you find this information?**

15. Match the node numbers of your tree to those listed in your ace results by entering:

```
> plot(mammal.tree)
> nodelabels()
```

**Reflect:**

**What is the confidence interval of your root estimate?**

## Ancestral Character State Reconstruction for discrete characters

16. Load a new dataset called “eel.data” and “eel.tree” (see step 4). Plot the tree and look at the data. What kinds of characters are included?

17. Isolate the binary feeding mode character by entering:

```
> feed.mode<-setNames(eel.data[,1],rownames(eel.data))
```

18. Find the best-fitting model using the command **fitDiscrete**

**Reflect:**

**Which was the best-fitting model? Was there more than one?**

19. Now, using **ace** again, specify your best-fitting model and character type to calculate joint reconstructions of the ancestral states. Save the results in a variable called **fit.ace**

20. Plot your results by entering the following one at a time:

```
> plotTree(eel.tree, fsize=0.7, ftype="i", lwd=1, offset=0.5)
> nodelabels(node=1:eel.tree$Nnode+Ntip(eel.tree), pie=fit.ace$lik.anc, piecol=cols, cex=0.4)
> tiplabels(pie=to.matrix(feed.mode[eel.tree$tip.label], levels(feed.mode)), piecol=cols, cex=0.3)
> add.simmap.legend(colors=cols, prompt=FALSE, x=0.9*par()$usr[1], y=0.8*par()$usr[3], fsiz
e=0.8)
```

**Reflect:**

**What was the reconstructed ancestral state for eel feeding mode? How sure can you be?**

## Brownian motion simulations

21. Use the command **simBMphylo** to visualize how the value of the rate parameter  $\sigma^2$ , the number of taxa, and time influences the shape of the plot and simulated tree.