

HowTo/DataTreeManipulation

From Comparative Phylogenetics in R

The commands referenced below are all part of special phylogenetic packages in R, not the basic R install. Be sure that you have installed and loaded the packages `ape` and `geiger`, which contain the commands referenced below before continuing. For example:

```
library(ape)
library(geiger)
```

This loads the packages `ape` and `geiger` and their required packages, `gee`, `nlme`, `lattice`, `MASS`, `mvtnorm`, `msm` and `ouch` into your R session.

To execute some of the worked examples below yourself, save the sample *Geospiza* phylogeny (`geospiza.nex`) and dataset (`geospiza.txt`) to your working directory and load them into memory using these commands

```
geotree <- read.nexus("geospiza.nex")
geodata <- read.table("geospiza.txt")
```

Contents

- 1 How do I designate a specific taxon to be the root of my phylogeny?
- 2 How can I resolve polytomies in my phylogeny?
- 3 How can I collapse very short branches into polytomies?
- 4 How can I see the length of the branches in my phylogeny?
- 5 How can I change the lengths of the branches in my phylogeny?
- 6 How can I see the list of taxa represented in my phylogeny?
- 7 How can I verify that the taxa listed in my data table match those at the tips of my phylogeny?
- 8 Is there a shorthand way to refer to a specific list of taxa (for example, all members of a particular clade)?
- 9 How can I remove taxa from my phylogeny?
-

10 How can I see a plot of my phylogeny?

- 11 How can I identify all the branches belonging to a particular subclade?
- 12 How can I identify the node representing the most recent common ancestor of a pair of taxa?
- 13 How do I calculate the patristic distance between two taxa?
- 14 How do I calculate the patristic distance between two internal nodes or an internal node and a tip?
- 15 How do I calculate the distance from an internal node to the tips of an ultrametric phylogeny?
- 16 References

How do I designate a specific taxon to be the root of my phylogeny?

The general syntax is

```
rootedphylogeny <- root(phylogeny, outgroup)
```

The *Geospiza* tree is already rooted at taxon "olivacea". This command will reroot the tree at taxon "fusca" and save the rerooted tree as a new phylo object "rerootedgeotree".

```
rerootedgeotree <- root(geotree, "fusca")
```

You can also just modify the existing phylo object.

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

Note that rerooting produces a basal trichotomy . . . essentially this command roots the tree at the node subtending taxon fusca, not the taxon itself.

The root command will also root the tree at a group of taxa so long as that group is monophyletic. Otherwise, an error is returned.

How can I resolve polytomies in my phylogeny?

The multi2di command will break polytomies in random order.

```
dichotomousphylogeny <- multi2di(phylogeny, random = TRUE)
```

If you want to collapse these zero length branches into 'true multichotomies' ou can also choose to break polytomies in the order that taxa appear in the list of tip names.

```
dichotomousphylogeny <- multi2di(phylogeny, random = FALSE)
```

NOTE: that if you have zero length branches in your tree, ape will recognize the tree as binary if you type:

```
is.binary.tree(phylogeny)
```

If you want your zero length branches into a 'true multichotomy' follow the info in the section below.

How can I collapse very short branches into polytomies?

Use the di2multi command. The basic syntax is

```
collapsedphylogeny <- di2multi(phylogeny, tolerancevalue)
```

Branches shorter than the tolerance value will be collapsed into polytomies. If unspecified, the tolerance value defaults to 10E-8. For example:

```
collapsedgeotree <- di2multi(geotree, 0.03)
```

This should produce two polytomies in the Geospiza phylogeny.

How can I see the length of the branches in my phylogeny?

The vector of branch lengths for a phylo object (say, the object "phylogeny") is contained in phylogeny\$edge.length. For the Geospiza dataset, you can view this vector by typing:

```
geotree$edge.length
```

How can I change the lengths of the branches in my phylogeny?

Use the compute.brln function. This is the basic syntax

```
phylogeny <- compute.brln(phylogeny, method, power)
```

Method can equal Grafen (the default), a vector, or a user defined function (for example, a function that generates random branch lengths). If a vector shorter than the number of branches is given as the value of the argument method, that vector is iterated over the branches in sequence. If power is provided, it exponentiates the branch lengths by the specified power.

The following syntax will ultrametricize the Geospiza phylogeny using Grafen's method (Grafen, 1989).

```
geotree <- compute.brlen(geotree, method="Grafen")
```

While this syntax will set all branch lengths equal to one

```
geotree <- compute.brlen(geotree, 1)
```

And this will alternate branch lengths of one and two throughout the phylogeny (if, for some reason, you wanted to do that ...).

```
geotree <- compute.brlen(geotree, c(1, 2))
```

How can I see the list of taxa represented in my phylogeny?

These are stored in phylogeny\$tip.label. For example, typing

```
geotree$tip.label
```

will show you a list of all the taxa in the Geospiza phylogeny.

How can I verify that the taxa listed in my data table match those at the tips of my phylogeny?

Geiger contains the useful function name.check that does exactly that. First, make sure that you have geiger loaded.

```
library(geiger)
```

The basic syntax for name.check is

```
name.check(phylogeny, data)
```

So for the Geospiza dataset, we'd type

```
name.check(geotree, geodata)
```

This will return the following data

```
$Tree.not.data  
[1] "olivacea"
```

```
$Data.not.tree  
character(0)
```

Which tells us that the taxon "olivacea" is missing from the character dataset. We can drop olivacea from the phylogeny like this.

```
geotree <- drop.tip(geotree, "olivacea")
```

Running name.check now should produce the output "OK", indicating that the taxa in the tree and character dataset match.

Is there a shorthand way to refer to a specific list of taxa (for example, all members of a particular clade)?

One approach is to concatenate all the taxon names into a named vector, which can then be used as an input argument in other functions. For example, using the Geospiza dataset.

```
cladeA = c("pauper", "psittacula", "parvulus")
```

Note that you need to enclose the taxon names in quotes, otherwise R will look for objects in memory named pauper, psittacula and parvulus.

The package geiger offers the function node.leaves, which facilitates subclade designation if you know the basal node of a clade. The basic syntax is:

```
node.leaves(phylogeny, node)
```

so, we could also designate cladeA above by typing

```
cladeA <- node.leaves(geotree, 26)
```

because node 26 subtends the clade composed of those three species. One can determine the number of the desired node with the `mrca` command, which yields the node number of the most recent common ancestor of a pair of taxa.

```
mrca(geotree)[ "pauper", "psittacula" ]
```

You can do the whole thing in one step like this:

```
cladeA <- node.leaves(geotree, mrca(geotree)[ "pauper", "psittacula" ])
```

NOTE: `node.leaves` will be public in the release of the new version of `geiger`

How can I remove taxa from my phylogeny?

Use the `drop.tip` command. The basic syntax is:

```
drop.tip(phylogeny, tip)
```

Where `tip` is a single taxon or a vector of taxon name.

The example below will remove taxa "pauper", "psittacula" and "parvulus" (previously designated as "CladeA") from the *Geospiza* phylogeny and save the resulting phylogeny in "culledtree".

```
culledtree <- drop.tip(geotree, cladeA)
```

How can I see a plot of my phylogeny?

A quick plot for the *Geospiza* tree is generated by

```
plot.phylo(geotree)
```

`Plot.phylo` is actually a very powerful command with many options, for example, it can label internal nodes, modify line widths, modify the plot style (cladogram, fan, radial), etc. A full list of the switches and syntax can be obtained by typing:

```
help(plot.phylo)
```

How can I identify all the branches belonging to a particular subclade?

The general syntax is:

```
branchlist <- which.edge(phylogeny, group)
```

In the case of the Geospiza example, the branches that unite the species "pauper", "psittacula", and "parvulus" (CladeA defined above) are given by:

```
branchlist <- which.edge(geotree, cladeA)
```

This returns a list of integers, which identify the rows in the edge matrix of the Geospiza phylogeny that belong to the specified clade, which we stored as "geotree".

You can see what the edge matrix looks like by typing:

```
geotree$edge
```

By doing so, we can see that the edge matrix is an N by 2 list of the N branches in the phylogeny. Each node in the phylogeny is assigned a number, with each branch being defined by the numbers of the nodes bracketing it.

You can extract just the portion of the edge matrix containing the branches in cladeA like this:

```
Abranches<-geotree$edge[branchlist, ]
```

Remember that "branchlist" in this worked example is a vector of integers returned by which.edge.

How can I identify the node representing the most recent common ancestor of a pair of taxa?

Use the mrca command. The basic syntax is

```
mrca(phylogeny)[ "taxon1", "taxon2" ]
```

In the Geospiza example, this will return the number of the node in geotree that represents the most recent common ancestor of taxa "pauper" and "parvulus".

```
mrca(geotree)[ "pauper", "parvulus" ]
```

You should get the answer "26". You can verify this by first setting the labels for all the nodes to integers representing the order in which they appear in the phylogeny, and then replotting the tree with the switch `show.node.label=TRUE`

```
geotree$node.label<-((length(geotree$tip)+1):((length(geotree$tip)*2)-1))
plot(geotree, show.tip.label=TRUE, show.node.label=TRUE)
```

How do I calculate the patristic distance between two taxa?

Use the `cophenetic` command. For example:

```
cophenetic(geotree)["pallida", "conirostris"]
```

will calculate the patristic distance between the taxa "pallida" and "conirostris" in the *Geospiza* phylogeny.

```
cophenetic(geotree)
```

yields a matrix of all distances between taxa.

How do I calculate the patristic distance between two internal nodes or an internal node and a tip?

Use the `dist.nodes` command. For example:

```
dist.nodes(geotree)
```

yields a matrix of all distances between all nodes (internal and external), and

```
dist.nodes(geotree)[15, 20]
```

gives the distance between internal node 15 and internal node 20.

If you want the distance between a node and a tip, you need to know the number of the tip in question. You can see the numbering of the tips by typing.

```
geotree$tip
```

For example, taxon "fuliginosa" is tip 1 in the *Geospiza* dataset, so


```
dist.nodes(geotree)[1, 15]
```

yields the distance between "fulginosa" and the internal node labeled 15.

How do I calculate the distance from an internal node to the tips of an ultrametric phylogeny?

Use the `branching.times` command.

```
branching.times(geotree)
```

This returns a numeric vector of the branching times (distances from the nodes to the tips) for all nodes. The names of the vector are drawn from `phylogeny$node.label` if node labels have been specified, otherwise they are numbered in the order in which they appear in the edge matrix of the phylogeny.

References

Grafen, A. (1989). The phylogenetic regression. *Phil. Trans. Roy. Soc. B.* 326: 119–156

Credit: Most of the information on this page is based on the book *Analysis of Phylogenetics and Evolution with R* (Paradis, 2006).

[Back to the HowTo/Table of Contents](#)

Retrieved from "<https://www.r-phylo.org/w/index.php?title=HowTo/DataTreeManipulation&oldid=255>"

Categories: [HowTo](#) | [R Help](#) | [Comparative Methods Help](#)

-
- This page was last modified on 14 March 2008, at 10:32.
 - Content is available under [Creative Commons Attribution 4.0 International](#) unless otherwise noted.