

Phylogenetic Independent Contrasts in R

Casey Dunn

April 12, 2013

Contents

1	Introduction	1
2	Simulations	1
2.1	Simulating trees	1
2.2	Simulating phenotypic data on trees	3
3	Independent contrasts	5
4	Phylogenetic signal	6
5	How this document was made	7

1 Introduction

This document introduces you to a variety of tools and approaches to analyzing phylogenetic data with R.

There is detailed information on phylogenetic analysis with R at both the R phylo wiki and the CRAN task page for phylogenetics. Also consult the excellent book by Paradis (2012).

Make sure that you have the most recent version of R installed. For the analyses below, you will need to install the following libraries in R:

- ape
- geiger
- phylobase

Update them if they are not the latest version.

2 Simulations

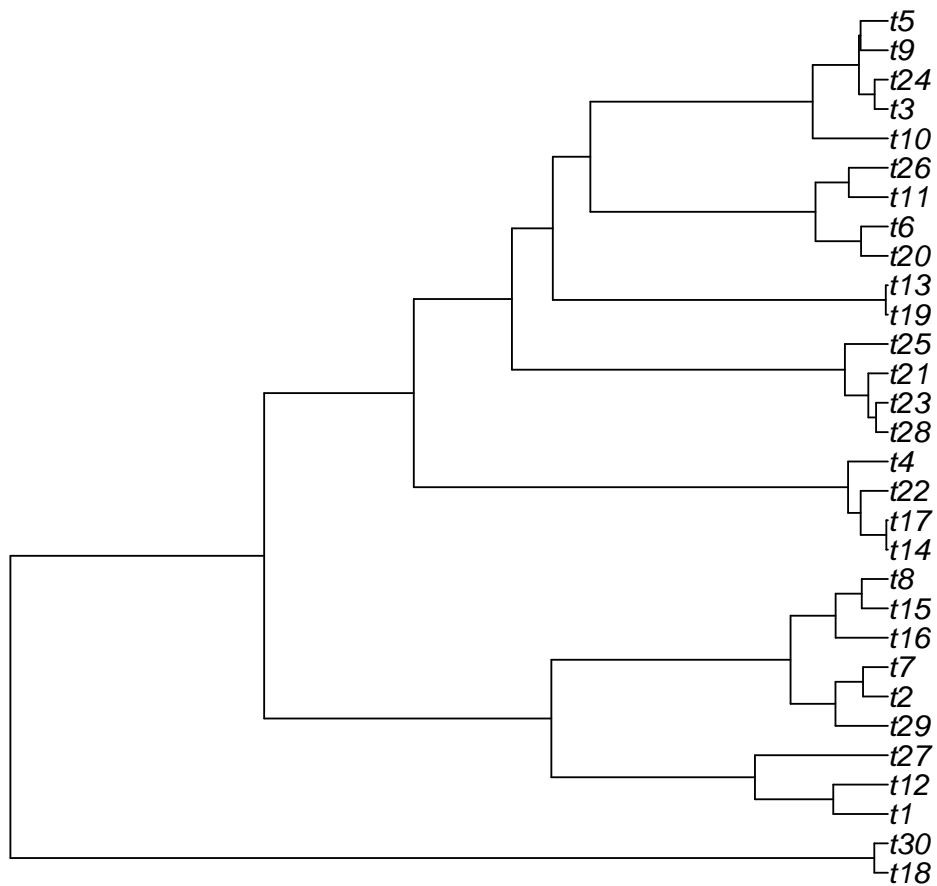
2.1 Simulating trees

First, load ape:

```
library(ape)
```

The simplest way to generate a random tree is with the `rcoal()` function:

```
t <- rcoal(30)  
plot(t)
```



The only required argument is the number of taxa. `rcoal()` generates ultrametric trees, the related function `rtree()` can be used to generate trees whose tips are not contemporaneous.

2.2 Simulating phenotypic data on trees

The `sim.char()` function from the `geiger` library is a convenient tool for simulating character evolution.

We'll simulate three characters, each with a variance of 1. The first two characters will have a covariance of 0.7 with each other. The third character will have a covariance of 0 with the other characters. First, make a variance-covariance matrix describes this relationship:

```
vcv <- diag(3)
vcv[1, 2] <- 0.7
vcv[2, 1] <- 0.7
vcv

##      [,1] [,2] [,3]
## [1,]  1.0  0.7  0.0
## [2,]  0.7  1.0  0.0
## [3,]  0.0  0.0  1.0
```

Now, generate multiple simulated datasets on tree `t` according to `vcv`:

```
library(geiger)
D <- sim.char(t, vcv, nsim = 100, model = "BM", root = 0)
D[, , 1]

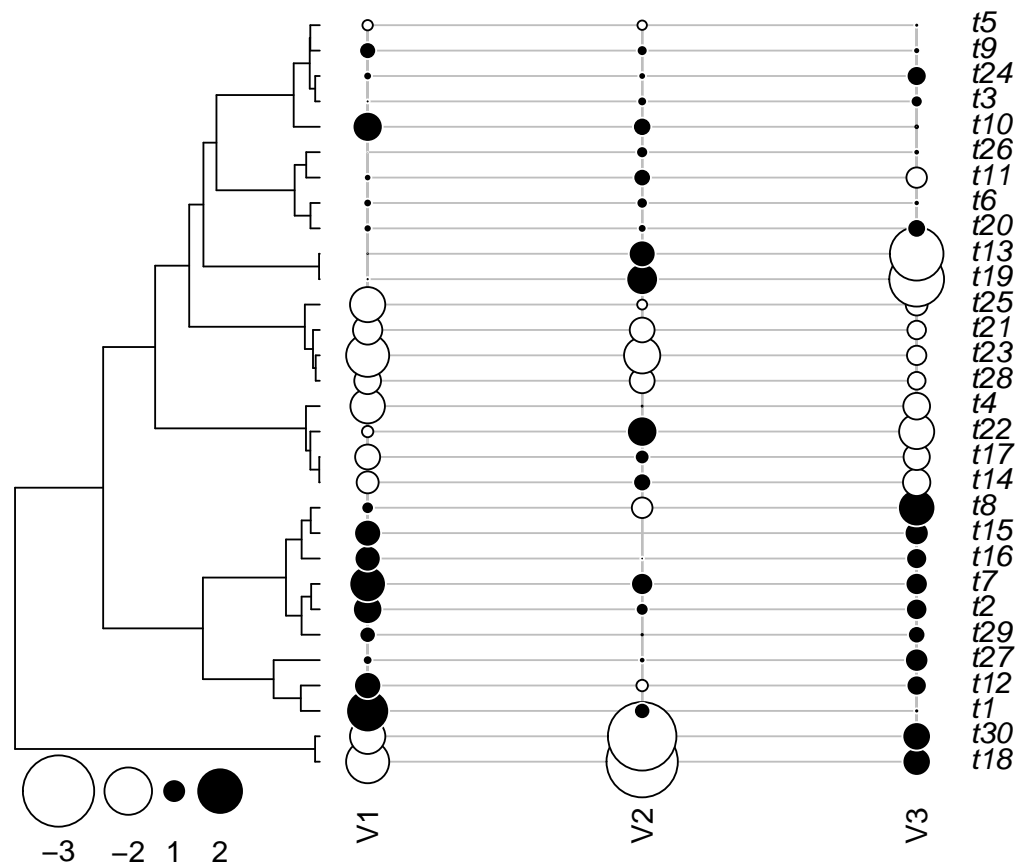
##      [,1]      [,2]      [,3]
## t18 -1.628802 -2.12330  0.07595
## t30 -1.453593 -2.04430  0.08165
## t1   0.313665  0.33917 -0.63395
## t12 -0.041795 -0.44340 -0.17259
## t27 -0.438492  0.06447 -0.07656
## t29 -0.286230 -0.16080 -0.25686
## t2   0.006977  0.24091 -0.13656
## t7   0.168436  0.50268 -0.11114
## t16 -0.066554 -0.01374 -0.14852
## t15 -0.047387 -0.11963 -0.05433
## t8   -0.372157 -0.69948  0.35210
## t14 -1.152393  0.38701 -1.62618
## t17 -1.220506  0.29282 -1.59149
## t22 -0.912792  0.73482 -1.86247
## t4   -1.435325 -0.14869 -1.60825
## t28 -1.259658 -0.81922 -1.33251
## t23 -1.626998 -1.13692 -1.38240
## t21 -1.317739 -0.81668 -1.35789
## t25 -1.451711 -0.39933 -1.45947
## t19 -0.560827  0.76932 -2.46519
## t13 -0.670083  0.63438 -2.42879
```

```
## t20 -0.471474  0.12701 -0.21873
## t6  -0.460982  0.20034 -0.59804
## t11 -0.487141  0.37389 -1.41629
## t26 -0.642711  0.21862 -0.58845
## t10  0.024133  0.39613 -0.87602
## t3   -0.564213  0.15545 -0.42273
## t24 -0.459552  0.09626 -0.17927
## t9   -0.273272  0.18664 -0.57666
## t5   -0.891110 -0.38471 -0.63923
```

This shows just the first simulation.

You can plot the data right onto the tree. The `adephylo` library has some nice tools for plotting multivariate data, but we first need to combine the tree and data into a `phylo4d` object:

```
library(adephylo)
D4 <- phylo4d(t, D[, , 1])
table.phylo4d(D4, box = FALSE)
```



3 Independent contrasts

Now construct independent contrasts for each of the three variables:

```
ic1 <- pic(D[, 1, 1], t)
ic2 <- pic(D[, 2, 1], t)
ic3 <- pic(D[, 3, 1], t)

cov(cbind(ic1, ic2, ic3))
```

```
##          ic1      ic2      ic3
## ic1  0.8590  0.9025 -0.1063
## ic2  0.9025  1.4614 -0.4364
## ic3 -0.1063 -0.4364  0.9614
```

This matrix is based on a single simulation, and will differ by chance from the original vcv matrix that we defined above.

4 Phylogenetic signal

Many studies use K (Blomberg et al., 2003) to assess the “phylogenetic signal” of a character. Under a Brownian motion model on a phylogenetic tree, a K of 1 is expected. For $K < 1$, relatives resemble each other less than expected. For $K > 1$, relatives resemble each other more than expected.

This measure has been implemented in the `picante` library.

```
library(picante)
Kcalc(D[, 1, 1], t)

##          [,1]
## [1,] 0.501
```

Now take a look at K for a variable that has been simulated without consideration of the tree:

```
ntips <- length(t$tip.label)
x <- rnorm(ntips, mean = 0, sd = 1)
names(x) <- names(D[, 1, 1])
Kcalc(x, t)

##          [,1]
## [1,] 0.01328
```

Neither of these values are exactly 1, but it is not clear if the difference is significant. This can be addressed via randomization tests as implemented by the `phylosignal()` function

```
phylosignal(D[, 1, 1], t)

##          K PIC.variance.obs PIC.variance.rnd.mean PIC.variance.P
## 1 0.501          0.9278          14.49          0.001
## PIC.variance.Z
## 1          -1.828

phylosignal(x, t)

##          K PIC.variance.obs PIC.variance.rnd.mean PIC.variance.P
## 1 0.01328          69.64          34.82          0.925
## PIC.variance.Z
## 1          1.631
```

5 How this document was made

This document is a computable data report compiled directly from the data. To recreate this file from the data, you will need to install:

- R (<http://www.r-project.org>).
- The R package knitr (<http://yihui.name/knitr/>), which can be installed from within R.
- pdflatex, which comes with LaTeX distributions (<http://www.latex-project.org/ftp.html>).

From within the knitr directory, launch R and run:

```
library(knitr)
knit("comparative_r.Rnw")
quit()
```

This will generate a new tex file. To compile this tex file into a pdf file, run the following at the shell command line:

```
pdflatex comparative_r.tex
```

In addition to recreating this document as-is, you can directly edit and add to the analyses in the .Rnw source file. You can also copy the R source code from the .Rnw file.

The following shows the versions of R software that were used to generate this document:

```
sessionInfo()

## R version 3.0.0 (2013-04-03)
## Platform: x86_64-apple-darwin10.8.0 (64-bit)
##
## locale:
## [1] en_US.UTF-8/en_US.UTF-8/en_US.UTF-8/C/en_US.UTF-8/en_US.UTF-8
##
## attached base packages:
## [1] grid      stats      graphics  grDevices  utils      datasets
## [7] methods   base
##
## other attached packages:
## [1] picante_1.5-2    nlme_3.1-109     vegan_2.0-7
## [4] permute_0.7-0    adephylo_1.1-3   ade4_1.5-2
## [7] phylobase_0.6.5  geiger_1.99-2    ncbi_2013.03.29
## [10] coda_0.16-1      lattice_0.20-15  Rcpp_0.10.3
## [13] digest_0.6.3     colorspace_1.2-2 deSolve_1.10-4
## [16] subplex_1.1-3    msm_1.1.4        mvtnorm_0.9-9994
## [19] MASS_7.3-26      ape_3.0-8        knitr_1.2
```

```
##  
## loaded via a namespace (and not attached):  
## [1] evaluate_0.4.3  formatR_0.7      splines_3.0.0    stringr_0.6.2  
## [5] survival_2.37-4 tools_3.0.0
```

References

- Blomberg, S. P., T. Garland, and A. R. Ives. 2003. Testing for phylogenetic signal in comparative data: behavioral traits are more labile. *Evolution* 57:717–745.
- Paradis, E. 2012. *Analysis of Phylogenetics and Evolution With R*. Springer Science+Business Media.