

Phylogenetic Independent Contrasts in R

Casey Dunn

April 1, 2013

Contents

1	Introduction	1
2	Getting oriented with R	1
3	Getting set up and loading the data	2
4	Calculating independent contrasts	6
5	Investigating the relationships between the variables	7
6	What if we didn't consider the phylogeny?	9
7	Understanding how the contrasts are calculated	11
7.1	Phylogenetic trees in R	12
7.2	Preparing the character data	15
7.3	Preparing the variances	17
7.4	Calculating the contrasts	17
8	How this document was made	18

1 Introduction

This example analysis presents the basics phylogenetic independent contrasts (Felsenstein, 1985). It expands on the example presented at the R phylogenetics wiki (http://www.r-phylo.org/wiki/HowTo/Phylogenetic_Independent_Contrasts).

This analysis uses the programming language R. In the past decade, a large community of phylogenetic biologists interested in character evolution have written many analysis tools in R. This set of tools provides a rich environment for the study of character evolution.

2 Getting oriented with R

R, like python and many other languages, can be used interactively, where you enter a few commands at a time, or in batch mode, where a series of commands are placed in a file and executed

all at once. We will use R interactively, and load R code that others have already written for phylogenetic analyses as we go.

There are already *many* general introductions, tutorials, and quick-references for R. I therefore won't provide a background on R itself, we will dive right into some analyses.

There is detailed information on phylogenetic analysis with R at both the R phylo wiki and the CRAN task page for phylogenetics.

For the analyses below, you will need to install the 'ape' library in R.

3 Getting set up and loading the data

Copy the data files ('Geospiza.txt' and 'Geospiza.nex') to a directory on your computer. Open the R interface (which was installed alongside the rest of the R components) or open a terminal window and launch an interactive R session with the command 'R'.

Now, change to the directory where your data files are:

```
setwd("DATADIR")
```

Where 'DATADIR' is the directory where you put the data files. The R command 'setwd()' is much like the shell command 'cd' - it controls your working directory.

Load the ape library, which has the functions we will use for phylogenetic independent contrasts:

```
library(ape)
```

Load and prepare the data:

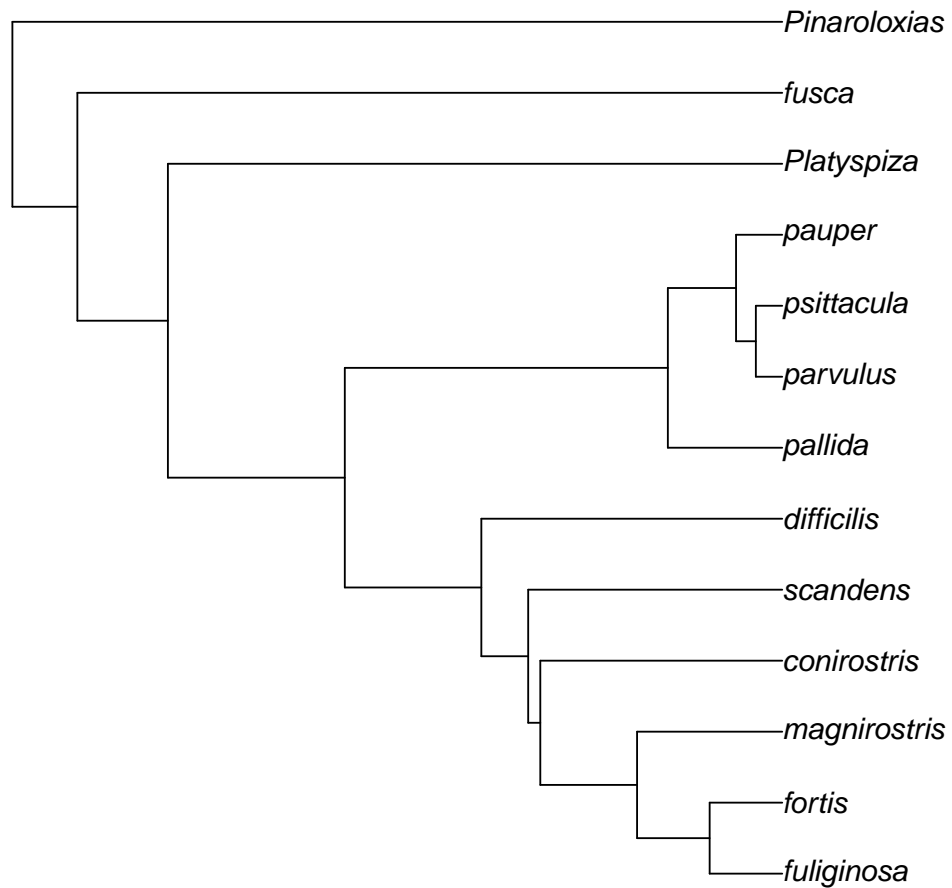
```
geodata <- read.table("Geospiza.txt")  
geotree <- read.nexus("Geospiza.nex")
```

We now need to remove a taxon in the tree that isn't in the data file.

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

Plot the tree:

```
plot(geotree)
```



And take a look at the character data:

```

geodata
#      wingL  tarsusL  culmenL  beakD  gonysW
# magnirostris 4.404    3.039    2.725  2.824  2.676
# conirostris  4.350    2.984    2.654  2.514  2.360
# difficilis   4.224    2.899    2.277  2.011  1.930
# scandens     4.261    2.929    2.622  2.145  2.037
# fortis       4.244    2.895    2.407  2.363  2.222
# fuliginosa   4.133    2.807    2.095  1.941  1.845

```

Independent contrasts

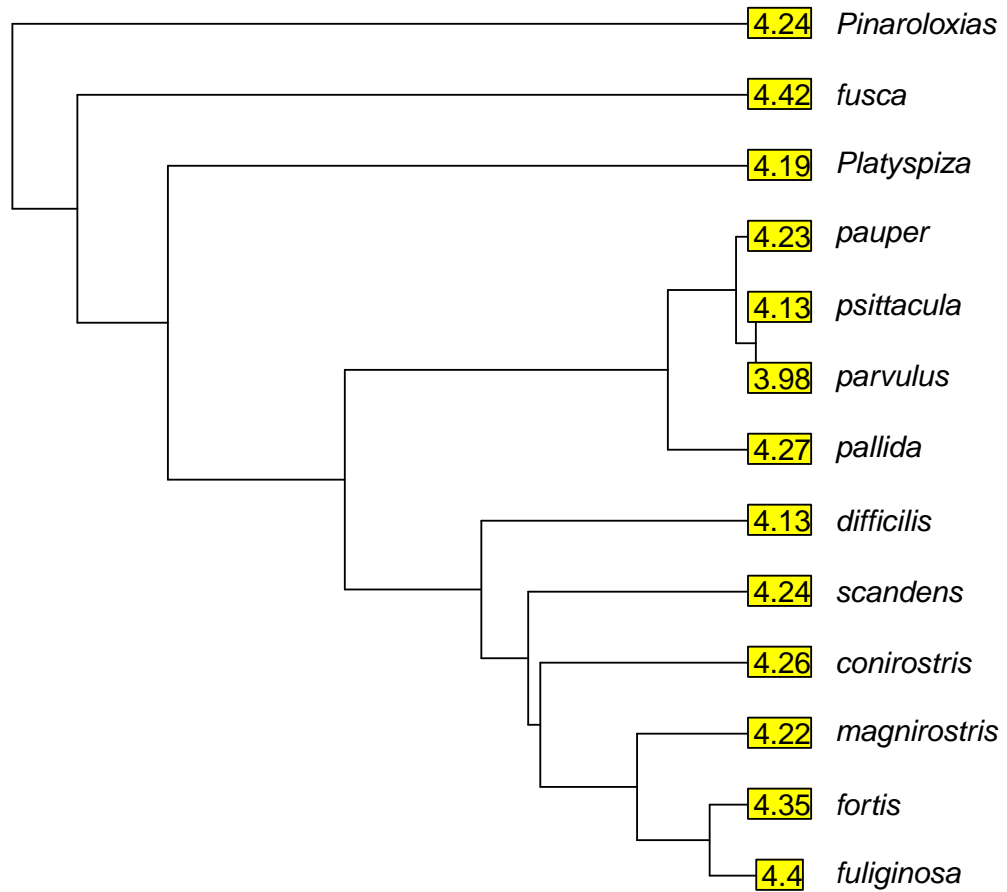
```
# pallida      4.265    3.089    2.430 2.016  1.949
# fusca        3.975    2.937    2.052 1.191  1.401
# parvulus     4.132    2.973    1.974 1.874  1.813
# pauper       4.232    3.036    2.187 2.073  1.962
# Pinaroloxias 4.189    2.980    2.311 1.548  1.630
# Platyspiza   4.420    3.271    2.331 2.347  2.282
# psittacula   4.235    3.049    2.260 2.230  2.074
```

Data are available for five different characters for all the taxa in the tree. We'll look at two of these characters, wing length and tarsus length. Extract these columns as their own variables to simplify later commands:

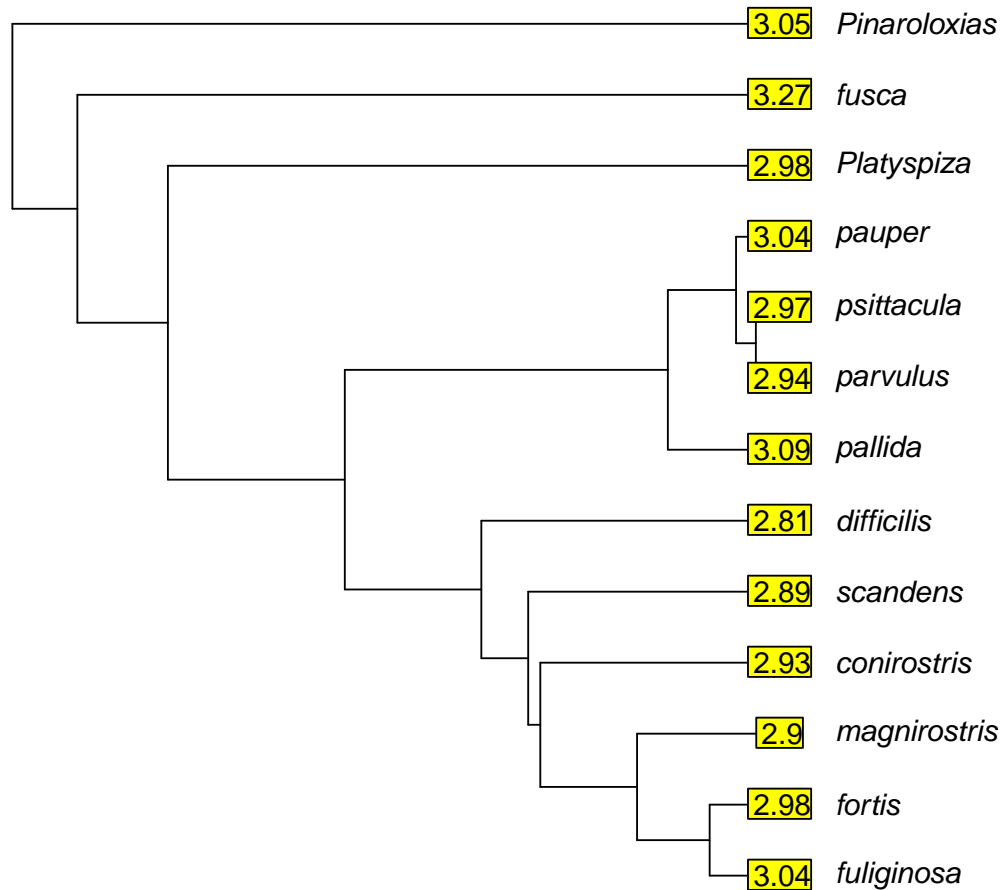
```
wingL <- geodata$wingL
tarsusL <- geodata$tarsusL
names(wingL) <- row.names(geodata)
names(tarsusL) <- row.names(geodata)
```

We can now plot these data directly onto the tips of the tree:

```
plot(geotree, label.offset = 0.04)
tiplabels(round(wingL, 2))
```



```
plot(geotree, label.offset = 0.04)
tiplabels(round(tarsusL, 2))
```



4 Calculating independent contrasts

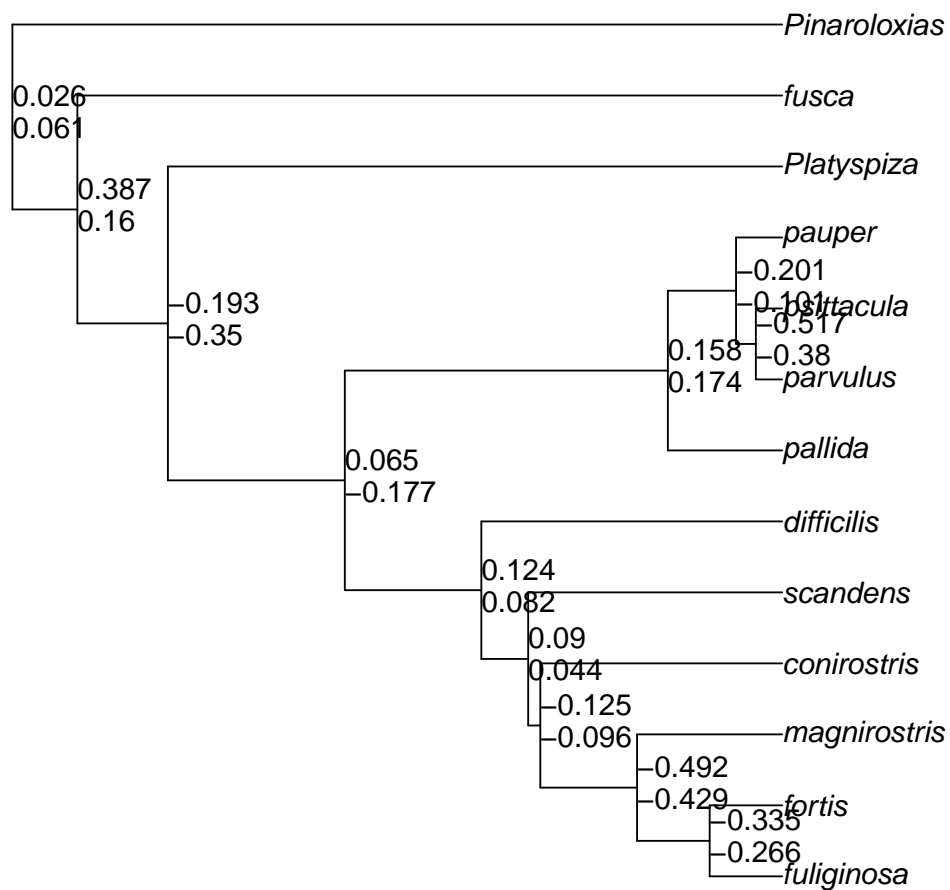
Calculate the independent contrasts for each of the variables:

```
ContrastwingL <- pic(wingL, geotree)
ContrasttarsusL <- pic(tarsusL, geotree)
```

And plot them on the tree:

Independent contrasts

```
plot(geotree)
nodelabels(round(ContrastwingL, 3), adj = c(0, -0.5), frame = "n")
nodelabels(round(ContrasttarsusL, 3), adj = c(0, 1), frame = "n")
```

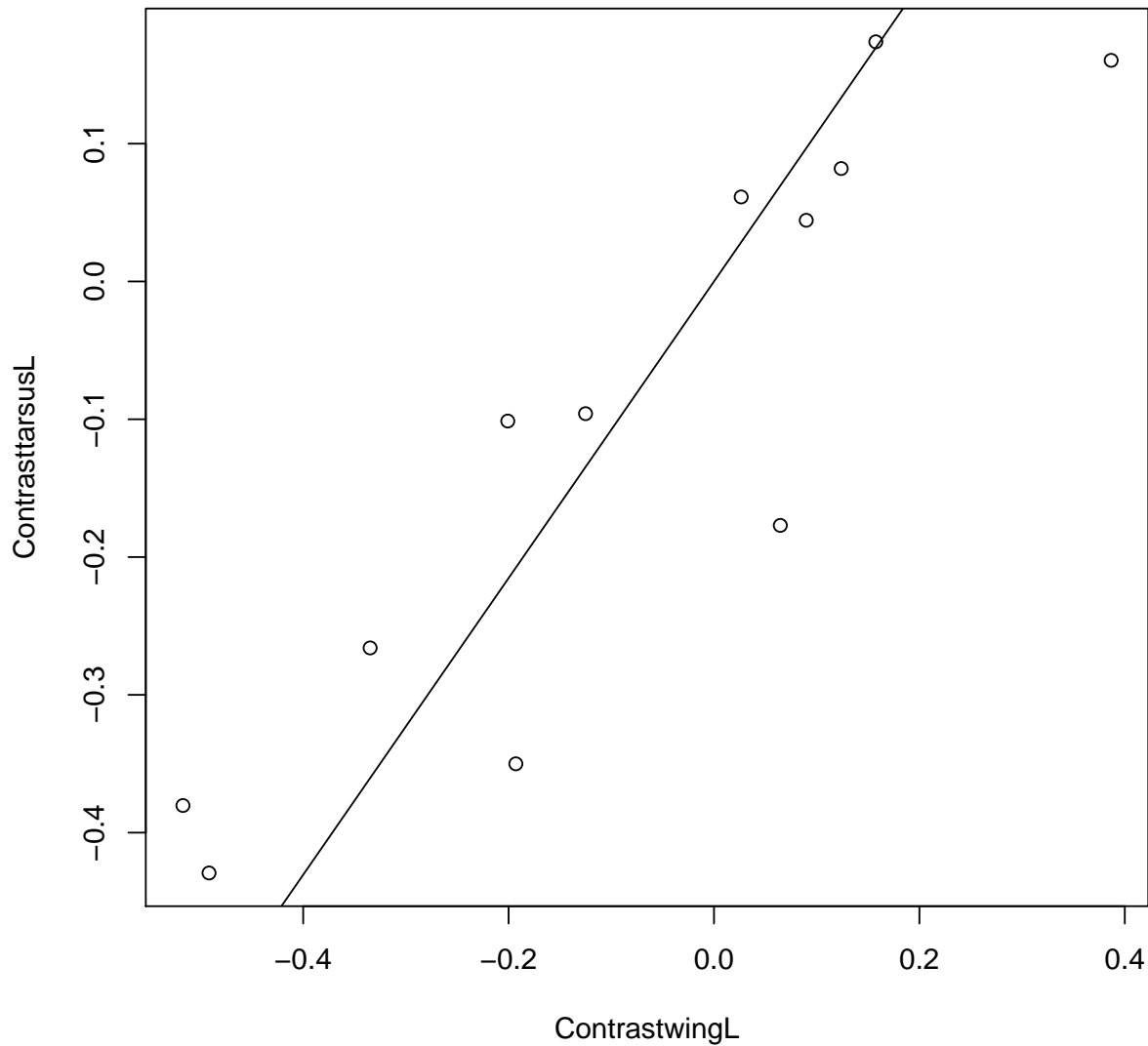


5 Investigating the relationships between the variables

```
RegressTarsusWing <- lm(ContrastwingL ~ ContrasttarsusL - 1)
summary.lm(RegressTarsusWing)
```

```
#  
# Call:  
# lm(formula = ContrastwingL ~ ContrasttarsusL - 1)  
#  
# Residuals:  
#      Min       1Q   Median       3Q      Max   
# -0.1077 -0.0418 -0.0257  0.0775  0.2551   
#  
# Coefficients:  
#              Estimate Std. Error t value Pr(>|t|)      
# ContrasttarsusL    1.076      0.157    6.86 2.7e-05 ***  
# ---  
# Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1  
#  
# Residual standard error: 0.126 on 11 degrees of freedom  
# Multiple R-squared:  0.81, Adjusted R-squared:  0.793   
# F-statistic:    47 on 1 and 11 DF,  p-value: 2.74e-05
```

```
plot(ContrastwingL, ContrasttarsusL)  
abline(RegressTarsusWing)
```

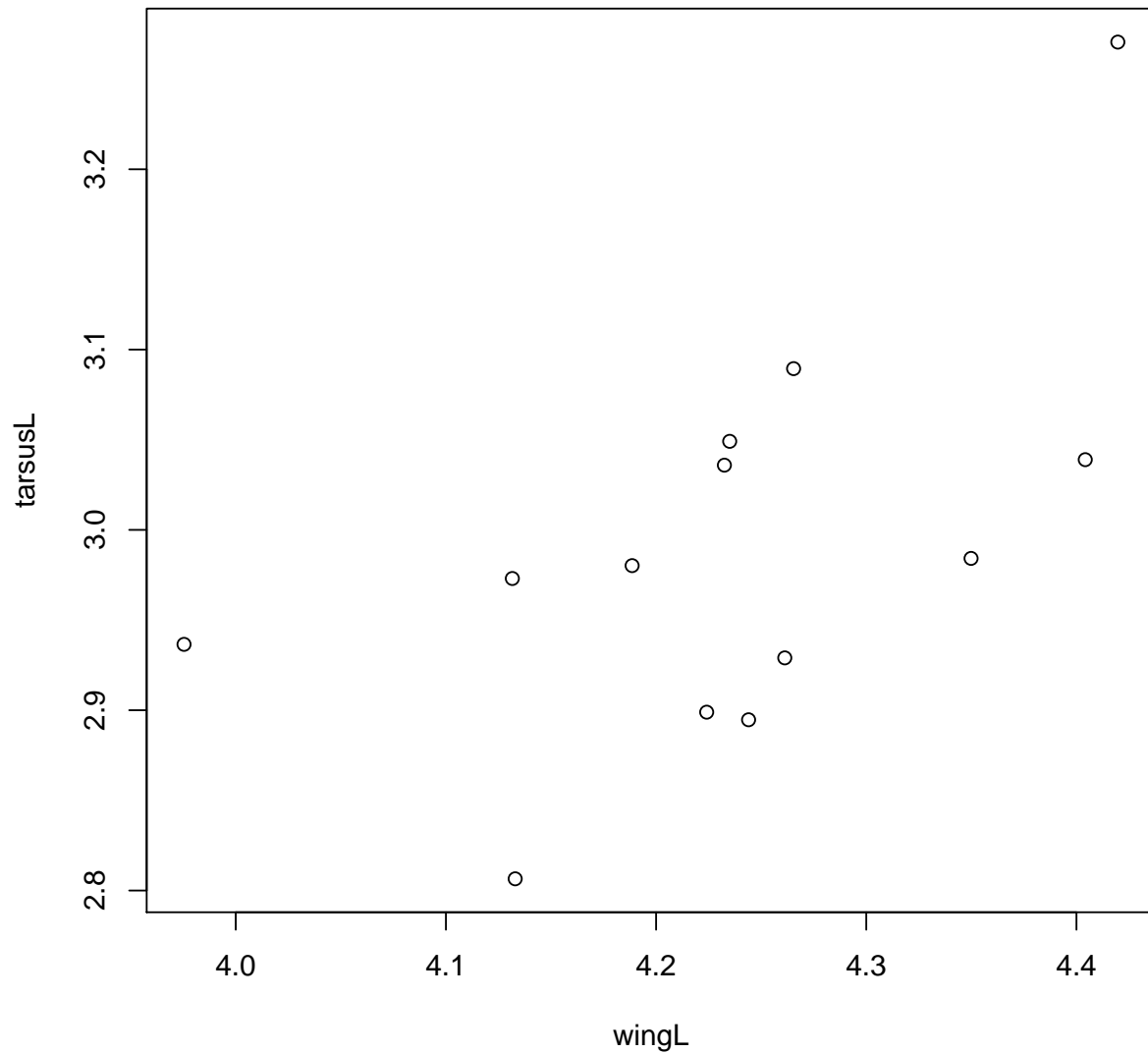
6 What if we didn't consider the phylogeny?

```
RegressTarsusWingNaive <- lm(wingL ~ tarsusL)
summary.lm(RegressTarsusWingNaive)

#
# Call:
# lm(formula = wingL ~ tarsusL)
#
```

```
# Residuals:
#      Min       1Q   Median       3Q      Max
# -0.2264 -0.0365  0.0112  0.0640  0.1390
#
# Coefficients:
#              Estimate Std. Error t value Pr(>|t|)
# (Intercept)    2.385      0.758     3.14  0.0093 **
# tarsusL        0.619      0.253     2.44  0.0327 *
# ---
# Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
#
# Residual standard error: 0.0996 on 11 degrees of freedom
# Multiple R-squared:  0.352, Adjusted R-squared:  0.293
# F-statistic: 5.96 on 1 and 11 DF,  p-value: 0.0327
```

```
plot(wingL, tarsusL)
```



7 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>). This document was generated with version 2.15.2.
- The R package knitr (<http://yihui.name/knitr/>), which can be installed from within R. This document was generated with version 0.9.

- pdflatex, which comes with LaTeX distributions (<http://www.latex-project.org/ftp.html>). This document was generated with version 3.1415926-2.4-1.40.13.

From within the knitr directory, launch R and run:

```
library(knitr)
knit("independent_contrasts.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 independent_contrasts.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.

References

Felsenstein, J. 1985. Phylogenies and the Comparative Method. *American Naturalist* 125:1–15.