APPLICATION

phytools: an R package for phylogenetic comparative biology (and other things)

Liam J. Revell*

Department of Biology, University of Massachusetts Boston, Boston, MA 02125, USA

Summary

1. Here, I present a new, multifunctional phylogenetics package, phytools, for the R statistical computing environment.

2. The focus of the package is on methods for phylogenetic comparative biology; however, it also includes tools for tree inference, phylogeny input/output, plotting, manipulation and several other tasks.

3. I describe and tabulate the major methods implemented in phytools, and in addition provide some demonstration of its use in the form of two illustrative examples.

4. Finally, I conclude by briefly describing an active web-log that I use to document present and future developments for phytools. I also note other web resources for phylogenetics in the R computational environment.

Key-words: blogging, computational biology, evolution, phylogeny, statistics

Introduction

In recent decades, phylogenies have assumed a central role in evolutionary biology (Felsenstein 1985, 2004; Harvey & Pagel 1991; Losos 2011). Among phylogeneticists, the scientific computing environment R (R Development Core Team 2011) has grown by leaps and bounds in popularity, particularly since the development of the multifunctional 'ape' (Analysis of Phylogenetics and Evolution) R package (Paradis, Claude & Strimmer 2004) and since the publication of Paradis's 'UseR!' phylogenetics book (Paradis 2006). Recent years have witnessed a rapid expansion of the phylogenetic capabilities of R in the form of numerous contributed packages. Most, such as the popular packages 'geiger' (Harmon *et al.* 2008) and 'phangorn' (Schliep 2011), work by building off the functionality and data structures developed in ape.

In this note, I will describe a new phylogenetics package written in the R language. This library is called 'phytools' (phylogenetic tools for comparative biology – and other things) and can be installed from the Comprehensive R Archive Network, CRAN. In this package, I have focused primarily on implementing several phylogenetic comparative methods from my own work (Revell 2008, 2009, 2010; Revell, Harmon & Collar 2008; Revell & Harrison 2008; Revell &

Collar 2009; Lindenfors, Revell & Nunn 2010; Mahler *et al.* 2010; Revell *et al.* in press); however, I have also included several new functions of interest and use to the phylogenetics community that were not previously implemented in R (e.g. Cavalli-Sforza & Edwards 1967; Baum 1992; Ragan 1992; Nielsen 2002; Huelsenbeck, Nielsen & Bollback 2003; Bollback 2006; O'Meara *et al.* 2006; Ives, Midford & Garland 2007; Sidlauskas 2008), as well as a number of simple utility functions for reading, writing, plotting and manipulating special types of phylogenetic trees.

Invariably, I have strived to maximize the interactivity of phytools with the ape package. For instance, one of the new functionalities of phytools is the capacity to generate, plot, read, and write stochastic character mapped trees (Nielsen 2002). Rather than create a new type of R object to store stochastically mapped phylogenies, I instead build directly on the existing 'phylo' structure developed for ape and employed in many other R phylogenetics packages. At present, phytools is not interoperable with the 'phylobase' package (R Hackathon *et al.* 2011), although this capability will be added in the future.

In the sections that follow, I will describe the major functionality of the phytools library; I will provide two illustrative examples that demonstrate some of the functionality of phytools; and, finally, because phytools is a work in progress, I will describe a web-log that I will use to keep phytools users up to date on bugs, updates, and future software development for the package.

*Correspondence author. E-mail: liam.revell@umb.edu
Correspondence site: http://www.respond2articles.com/MEE/

Description

The phytools library is written entirely in the scientific computing language, R (R Development Core Team 2011). It takes advantage of functionality developed in other packages, particularly the core phylogenetics package ape (Paradis, Claude & Strimmer 2004), for many types of phylogenetic tree input and manipulation. It also uses the phylogenetic inference package phangorn (Schliep 2011) for inference and certain other calculations. In addition, the phytools package depends, imports, or suggests several other R libraries either directly, or via its dependencies. These include the following packages: animation: Xie (2011); calibrate: Graffelman (2010); igraph: Csardi & Nepusz (2006); mnormt: Genz & Azzalini (2011); msm: Jackson (2011); numDeriv: Gilbert (2011); quadprog: Turlach & Weingessel (2010).

So far, I have implemented numerous functions for the phytools package; however, I should also note that phytools is a work in progress and I expect the capabilities of phytools to expand considerably in the coming years. In

Table 1. Major functions of the phytools package

Function name	Description
add.everywhere	Adds a tip to all edges of a tree
allFurcTrees	Generates all possible bi- and multifurcating trees for a set of taxa
anc.trend	Performs ancestral character estimation with a trend using likelihood
branching.diffusion	Creates an animation of Brownian motion evolution with speciation
brownie.lite	Fits a multiple rate Brownian evolution model using likelihood (O'Meara *et al.* 2006)
drop.clade	Removes a specific clade from the tree
estDiversity	Estimates the lineage density at each node of the tree based on a model (Mahler *et al.* 2010)
evol.rate.mcmc	Locates the position of a shift in the evolutionary rate using Bayesian MCMC (Revell *et al.* in press)
evol.vcv	Fits multiple evolutionary variance-covariance matrices to the tree using likelihood (Revell & Collar 2009)
exhaustiveMP	Performs exhaustive and branch & bound maximum parsimony inference
fastBM	Conducts fast Brownian motion simulation using multiple models (standard, with a trend, with bounds)
fitDiversityModel	Fits a diversity dependent phenotypic evolution model (Mahler *et al.* 2010)
gammatest	Conducts γ-test of Pybus & Harvey (2000)
ls.tree	Computes the least-squares branch lengths for a tree topology and distance matrix (Cavalli-Sforza & Edwards 1967)
ltt	Creates lineage-through-time plot, with extinct branches
make.era.map	Maps temporal intervals on a phylogenetic tree
make.simmap	Generates stochastic character map for discrete character data and a model of evolution (Nielsen 2002; Huelsenbeck, Nielsen & Bollback 2003)
map.overlap	Computes the similarity of two different character histories
mrp.supertree	Estimates the matrix representation parsimony supertree from a set of phylogenies (Baum 1992; Ragan 1992; Bininda-Emonds 2004)
optim.phylo.ls	Performs phylogeny inference under the least-squares criterion (Cavalli-Sforza & Edwards 1967; Felsenstein 2004)
paste.tree	Pastes two trees together
phyl.cca	Conducts phylogenetic canonical correlation analysis (Revell & Harrison 2008).
phyl.pairedttest	Performs a phylogenetic paired *t*-test (Lindenfors, Revell & Nunn 2010).
phyl.pca	Performs phylogenetic principal components analysis (Revell 2009)
phyl.resid	Conducts phylogenetic size-correction, using least-squares regression (Revell 2009)
phyl.RMA	Performs phylogenetic reduced major axis regression
phylANOVA	Performs the phylogenetic ANOVA (Garland *et al.* 1993) with *posthoc* comparison of group means
phylomorphospace	Projects a phylogenetic tree into bivariate morphospace (Sidlauskas 2008; e.g. Fig. 1).
phylosig	Computes phylogenetic signal using two different methods (Pagel 1999; Freckleton, Harvey & Pagel 2002; Blomberg, Garland & Ives 2003) and incorporating sampling error (Ives, Midford & Garland 2007)
plotSimmap	Plots a stochastic character map format tree (Fig. 3)
plotTree	Plots a phylogenetic tree with several options (Fig. 2)
read.simmap	Reads one or multiple stochastic character map format trees from file (Bollback 2006)
reorderSimmap	Reorders the edges of a stochastic map format tree
reroot	Re-roots a phylogenetic tree at an arbitrary position along an edge
sim.history	Simulates a stochastic history for a discretely valued character trait on the tree
sim.rates	Simulates multiple evolutionary rates on the tree using a Brownian evolution model
treeSlice	Cuts a tree and returns all subtrees
write.simmap	Function writes stochastic map style trees to file

Table 1, I provide an annotated list of the major functions thus far implemented in the phytools library. These functions cover methods in a few different areas of phylogenetic biology, described later.

COMPARATIVE BIOLOGY

Several methods in phylogenetic comparative biology have been implemented in phytools. These cover a wide range of areas including ancestral character estimation (e.g. `anc.trend`), likelihood-based methods for studying the evolution of character traits over time (e.g. `brownie.lite`, `evol.vcv`, `fitDiversityModel` and `phylosig`), a Bayesian method for detecting the location of a rate shift in the tree (`evol.rate.mcmc`), estimation of phylogenetic signal, including with sampling error (`phylosig`), and various methods for statistical hypothesis testing in a phylogenetic context (e.g. `phyl.cca`, `phyl.pairedttest`, `phyl.pca` and `phyl.resid`).

SIMULATION

Several simulation methods are implemented in phytools. These include Brownian motion simulation under various conditions (`fastBM`), simulation of discrete character evolution (`sim.history`), simulation of stochastic character maps (`make.simmap`) and simulation of multiple evolutionary rates (`sim.rates`), among other functionality (Table 1).
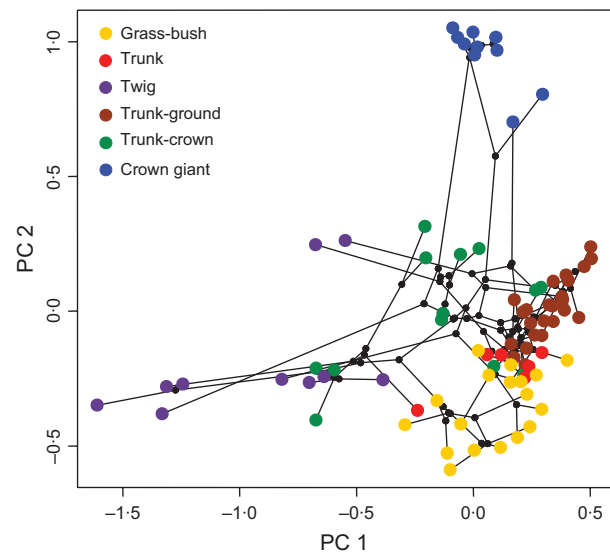
PHYLOGENETIC INFERENCE

A few different phylogenetic inference procedures are implemented in the phytools package. These functions are, in general, highly dependent on calculations and algorithms in the phangorn library of Schliep (2011). Some of the functionality includes matrix representation parsimony supertree estimation (`mrp.supertree`) and least-squares phylogeny inference (`optim.phylo.ls`; Table 1).

GRAPHICAL METHODS

Several graphical methods are implemented in phytools. Among these are projection of a tree into bivariate morphospace (`phylomorphospace`; Fig. 1), plotting stochastic character maps and histories (`plotSimmap`), lineage through time plotting with extinct lineages (`ltt`), animation of Brownian motion and speciation (`branching.diffusion`), and other functions (Table 1).

UTILITY FUNCTIONS

In addition to the aforesaid scientific functions, phytools also includes a number of utility functions for phylogeny input, output and manipulation. These are meant to supplement and complement the existing diverse array of utility functions in the ape and phangorn packages. Several of these functions are listed in Table 1.

Fig. 1. Phylomorphospace plot of two principal component axes obtained from the data of Mahler *et al.* (2010) for Caribbean *Anolis* lizards. Each point represents a species phenotypic value (coloured) or a hypothesized ancestral phenotype (black). Lines connect related species through hypothetical ancestors (Sidlauskas 2008). Different colours represent species in different ecomorphological categories that have been described based on microhabitat utilization, ecology, and gross phenotypic similarity (Losos 2009; colour code indicated in legend inset). The phylogeny has been pruned to include only these so-called 'ecomorph' species.

Examples

To demonstrate the use of phytools, I have created two short illustrative examples which can be easily reproduced by the reader. In the first, I use simulated data and the phytools function `evol.rate.mcmc` to identify the location of a shift in the evolutionary rate over time (Revell *et al.* in press). In the second, I simulate a stochastic discrete character history and a continuous character with different rate conditioned on the discrete character state, and then I fit a multi-rate Brownian character evolution model using the phytools function `brownie.lite` (O'Meara *et al.* 2006).

EXAMPLE 1: DETECTING THE LOCATION OF A RATE SHIFT

In this example, I first simulate a stochastic pure-birth phylogeny; next, I simulate evolutionary change for a single continuously valued character on the phylogeny under two different evolutionary rates in different parts of the tree; I analyse the tree and data using the Bayesian MCMC method for identifying the location of a shift in the evolutionary rate over time (Revell *et al.* in press); finally, I analyse the MCMC results to estimate the location of the shift and the evolutionary rates tipward and rootward of this point.

First, I loaded the phytools package. This will also load ape and other required packages on first instantiation:

```
> # load the phytools package (and ape)
> require(phytools)
```

```
Loading required package: phytools
Loading required package: ape....
```

Next, I set the random number seed for reproducibility (here, it is just set to 1):

```
> set.seed(1) # set the seed
```

I use the ape function `rbdtree` to simulate a stochastic pure-birth tree. In this instance, the tree has 91 taxa.

```
> # simulate a tree (using ape)
> tree<-rbdtree(b=log(50),d = 0,Tmax=1)
```
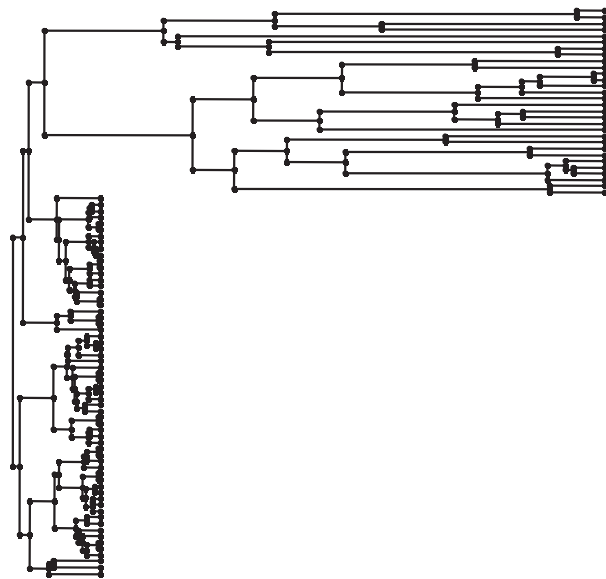
Now, for the purposes of simulation, I split the tree at a pre-determined position – specified here by the number of the descendant node and the distance along the edge from the root. To do this, I use the phytools function `splitTree`. It should be noted that the node and edge position used below are only guaranteed to work conditioned on having set the random number seed at 1 (see above), otherwise a different split point should be chosen.

```
> trees<-splitTree(tree,list(node = 153,bp =
0·09))
```

Now, I stretch the branches in one part of the tree and reattach the subtrees for simulation. To reattach the two parts of the tree, I use the phytools function `paste.tree` as follows:

```
> # stretch the branches in one part of the tree
> trees[[2]]$edge.length<-
trees[[2]]$edge.length*10
> # reattach the two subtrees
> sim.tree<-paste.tree(trees[[1]],trees[[2]])
```

I can then plot the generating tree for simulation (which has its branches stretched to be proportional to the evolutionary rate multiplied by time; Fig. 2), using the phytools function `plotTree`, and simulate on this stretched tree using phytools function `fastBM`:

Fig. 2. Phylogenetic tree simulated in illustrative example 1. The branches of the tree have been rescaled to represent evolutionary rate multiplied by elapsed time (thus, longer branches have higher evolutionary rates).

```
> # plot the generating tree for simulation
> plotTree(sim.tree,fsize = 0·5)
> x<-fastBM(sim.tree) # simulate on the tree
```

Now, I perform Bayesian MCMC analysis using the phytools function `evol.rate.mcmc` (Revell *et al.* in press). This analysis required about 20 min on a Dell i5 650 CPU running at 3·20 GHz.

```
> # perform MCMC > res<-evol.rate.mcmc(tree,
x,ngen = 100000)
Control parameters (set by user or default):
List of 11
$ sig1 : num 4·34
$ sig2 : num 4·34
$ a : num 0·0222
$ sd1 : num 0·867
$ sd2 : num 0·867
$ sda : num 0·00444
$ kloc : num 0·2
$ sdlnr : num 1
$ rand.shift: num 0·05
$ print : num 100
$ sample : num 100
Starting MCMC run:
state sig1 sig2 a node bp likelihood
04·3352 4·3352 0·0222 35 0·345 −130·7022
state sig1 sig2 a node bp likelihood
1004·8517 0·9447 −0·0248 93 0·044 −116·3554
state sig1 sig2 a node bp likelihood
2006·7915 1·2872 −0·0205 153 0·057 −101·0501
....
Done MCMC run.
```

The MCMC function first prints the control parameters (which can be set by the user, although above they have been given their default values, see below), and then prints the state of the MCMC chain at a frequency given by the control parameter `print` (here, every 100 generations; generations after 200 not shown above).

Next, I can estimate the location of the shift point by finding the split in the posterior sample with the smallest summed distance to all the other samples (this is one of multiple possible criteria; see Revell *et al.* in press). For this analysis, I use the phytools function `minSplit` and exclude the first 20 000 generations as burn-in:

```
>est.split<-minSplit(tree,
res$mcmc[201:nrow(res$mcmc),])
```

This analysis took about 6 s to run on the same hardware as described earlier.

Finally, I need to pre-process the posterior sample to get the sampled rates tipward and rootward of the average shift, for each sample (see Revell *et al.* in press). I do this using the phytools function `posterior.evolrate`. I can then print the results (estimated shift point and evolutionary rates) to screen:

```
> pp.result<-posterior.evolrate(tree,
est.split,res$mcmc[201:nrow(res$mcmc),],
res$tips[201:nrow(res$mcmc)])
> est.sig1<-mean(pp.result[,"sig1"])
```

```
> est.sig2 <-mean(pp.result[,"sig2"])
> est.split
$node
[ 1] 153
$bp
[ 1] 0·1003489
> est.sig1
[ 1] 1·010709
> est.sig2
[ 1] 10·24914
```

This analysis took about 9 s to complete.

Here, the parameter estimates are very close to the generating shift point of [153, 0·09] and the generating evolutionary rates of $\sigma_1^2 = 1·0$ and $\sigma_2^2 = 10·0$.

It should be noted that in actual practice, the authors should pay much closer attention to the control parameters of the MCMC than is given here, and in particular, to the proposal distribution for the model parameters. More information about function control can be obtained by calling the help file of evol.rate.mcmc:

```
> ? evol.rate.mcmc
```

or by referring to Revell *et al.* (in press). In addition, users should assess convergence and compute effective sample sizes for their samples from the posterior distribution. This can be carried out using the MCMC diagnostics package 'coda' (Plummer *et al.* 2006). Please refer to Revell *et al.* (in press) for more information about this method.

EXAMPLE 2: SIMULATE AND ANALYSE MULTI-RATE BROWNIAN EVOLUTION

In this example, I first simulate the character history of a discretely valued character trait with three states evolving on a phylogeny. I then simulate the evolution of a continuous trait with a rate that depends on the value of the discrete trait. Finally, I fit single and multiple rate evolutionary models to the data and tree using the likelihood method of O'Meara *et al.* (2006).

After loading phytools, I first set the seed (arbitrarily to 10; done here for reproducibility only):

```
> set.seed(10)
```

Now, I simulate a stochastic pure-birth tree using ape:

```
> tree<-rbdtree(b = log(50),d = 0,Tmax = 1)
```

This tree contains 129 taxa. Next, I simulate a stochastic character history on the tree for a character with three states, *A*, *B*, and *C*, using the phytools function sim.history as follows:

```
> # this is our transition matrix
> Q<-matrix(c(-2,1,1,1,-2,1,1,1,-2),3,3)
> rownames(Q)<-colnames(Q)<-
c("A","B","C")
> mtree<-sim.history(tree,Q)
```

I can plot the simulated history using the phytools function plotSimmap to see what it looks like:

```
> # set colors
> cols<-c("red","blue","green");
> names(cols)<-rownames(Q)
```
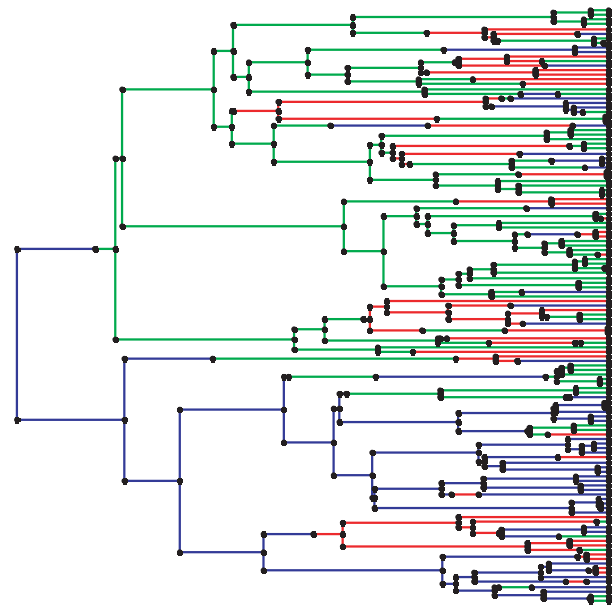

Fig. 3. Simulated stochastic history for illustrative Example 2. Here, different coloured branches indicate a different simulated state for a discretely valued character trait evolving on the tree. The colour code is A: red, B: blue, and C: green.

```
> # plot tree with labels off
> plotSimmap(mtree,cols,ftype="off")
```

This visualization is shown in Fig. 3.

Next, I simulate continuous character evolution using three different rates using the phytools function sim.rates:

```
> # set rates
> sig2 <-c(1,10,100)
> names(sig2)<-rownames(Q)
> X<-sim.rates(mtree,sig2) # simulate
```

Finally, I fit a multi-rate Brownian model using the likelihood method of O'Meara *et al.* (2006) with the phytools function brownie.lite. This likelihood optimization took about 3 s to run on the same hardware described earlier.

```
> fit.bm<-brownie.lite(mtree,X,maxit=4000)
> fit.bm
$sig2.single
[ 1] 47·15693
$a.single
[ 1] −5·242448
$var.single
[ 1] 34·47709
$logL1
[ 1] −329·7582
$k1
[ 1] 2
$sig2.multiple
B C A
10·4646399 99·9563314 0·8247106
...
$logL.multiple
[ 1] −287·4271
$k2
[ 1] 4
```

```
$P.chisq
[ 1] 4·129091e-19
$convergence
[ 1] "Optimization has converged."
```

It should be noted that the order of the three rate regimes in the fitted model is the order in which they are encountered in the tree (Fig. 3), rather than in alphabetical or numerical order. In this case, the fitted parameter estimates (0·82, 10·46, 99·96) are very close to their generating values (1, 10, and 100). For more details on this likelihood method, please refer to O'Meara *et al.* (2006) or Revell (2008).

phytools development web-log and other resources

This package so far implements a number of methods for phylogenetic comparative biology, phylogeny inference, tree manipulation and graphing. However, the phytools project is one in progress. To keep users of phytools up to date on bugs, improvements, and new functionality, I maintain an active web-log (i.e. 'blog'; http://phytools.blogspot.com). This blog acts as both a conduit between the developer (presently myself) and users of the phytools package, as well as a sort of open lab notebook (Butler 2005; Bradley *et al.* 2011) in which I document the details of bug fixes, software implementation, and use. Most of the functions listed earlier have already been featured on the blog (in the course of their development and refinement). Future work on phytools will also be documented here.

Finally, in addition to my blog, there are a number of other helpful web and email forums for phytools and phylogenetics in the R language generally. The phylogenetics CRAN Task View (http://cran.r-project.org/web/views/Phylogenetics.html) and R-sig-phylo email mailing list (https://stat.ethz.ch/mailman/listinfo/r-sig-phylo) are two of the most important such resources.

Citation of phytools

Scientists using phytools in a published paper should cite this article. Users can also cite the phytools package directly if they are so inclined. Citation information can be obtained by typing:

```
> citation("phytools")
```
at the command prompt.

Acknowledgements

References

Baum, B.R. (1992) Combining trees as a way of combining data sets for phylogenetic inference, and the desirability of combining gene trees. *Taxon*, **41**, 3–10.

Bininda-Emonds, O.R.P. (2004) The evolution of supertrees. *Trends in Ecology and Evolution*, **19**, 315–322.

Blomberg, S.P., Garland Jr., T. & Ives, A.R. (2003) Testing for phylogenetic signal in comparative data: behavioral traits are more labile. *Evolution*, **57**, 717–745.

Bollback, J.P. (2006) SIMMAP: stochastic character mapping of discrete traits on phylogenies. *BMC Bioinformatics*, **7**, 88.

Bradley, J.-C., Lang, A.S.I.D., Koch, S. & Neylon, C. (2011) Collaboration using open notebook science in academia. *Collaborative Computational Technologies for Biomedical Research* (eds S. Ekins, M.A.Z. Hupcey & A.J. Williams), pp. 425–452. John Wiley & Sons, Hoboken, NJ.

Butler, D. (2005) Elecronic notebooks: a new leaf. *Nature*, **436**, 20–21.

Cavalli-Sforza, L.L. & Edwards, A.W.F. (1967) Phylogenetic analysis: models and estimation procedures. *American Journal of Human Genetics*, **19**, 233–257.

Csardi, G. & Nepusz, T. (2006) The igraph software package for complex network research. *InterJournal, Complex Systems*, **????**, 1695.

Felsenstein, J. (1985) Phylogenies and the comparative method. *American Naturalist*, **125**, 1–15.

Felsenstein, J. (2004) *Inferring Phylogenies*. Sinauer, Sunderland, MA.

Freckleton, R.P., Harvey, P.H. & Pagel, M. (2002) Phylogenetic analysis and comparative data: a test and review of evidence. *American Naturalist*, **160**, 712–726.

Garland Jr., T., Dickerman, A.W., Janis, C.M. & Jones, J.A. (1993) Phylogenetic analysis of covariance by computer simulation. *Systematic Biology*, **42**, 265–292.

Genz, A. & Azzalini, A. (2011) *mnormt: The Multivariate Normal and t Distributions. R Package Version 1.4-0.* http://cran.r-project.org/package=mnormt.

Gilbert, P. (2011) *numDeriv: Accurate Numerical Derivatives. R package Version 2010.11-1.* http://cran.r-project.org/package=numDeriv.

Graffelman, J. (2010) *calibrate: Calibration of Scatterplot and Biplot Axes. R Package Version 1.7.* http://cran.r-project.org/package=calibrate.

Harmon, L.J., Weir, J.T., Brock, C.D., Glor, R.E. & Challenger, W. (2008) GEIGER: investigating evolutionary radiations. *Bioinformatics*, **24**, 129–131.

Harvey, P.H. & Pagel, M.D. (1991) *The Comparative Method in Evolutionary Biology*. Oxford University Press, Oxford UK.

Huelsenbeck, J.P., Nielsen, R. & Bollback, J.P. (2003) Stochastic mapping of morphological characters. *Systematic Biology*, **52**, 131–158.

Ives, A.R., Midford, P.E. & Garland Jr., T. (2007) Within-species variation and measurement error in phylogenetic comparative methods. *Systematic Biology*, **56**, 252–270.

Jackson, C.H. (2011) Multi-state models for panel data: the msm package for R. *Journal of Statistical Software*, **38**, 1–29.

Lindenfors, P., Revell, L.J. & Nunn, C.L. (2010) Sexual dimorphism in primate aerobic capacity: a phylogenetic test. *Journal of Evolutionary Biology*, **23**, 1183–1194.

Losos, J.B. (2009) *Lizards in an Evolutionary Tree: Ecology and Adaptive Radiation of Anoles*. University of California Press, Berkeley, CA.

Losos, J.B. (2011) Seeing the forest for the trees: the limitations of phylogenies in comparative biology. *American Naturalist*, **177**, 709–727.

Mahler, D.L., Revell, L.J., Glor, R.E. & Losos, J.B. (2010) Ecological opportunity and the rate of morphological evolution in the diversification of Greater Antillean anoles. *Evolution*, **64**, 2731–2745.

Nielsen, R. (2002) Mapping mutations on phylogenies. *Systematic Biology*, **51**, 729–739.

O'Meara, B.C., Ané, C., Sanderson, M.J. & Wainwright, P.C. (2006) Testing for different rates of continuous trait evolution using likelihood. *Evolution*, **60**, 922–933.

Pagel, M. (1999) Inferring the historical patterns of biological evolution. *Nature*, **401**, 877–884.

Paradis, E. (2006) *Analysis of Phylogenetics and Evolution with R*. Springer, New York, NY.

Paradis, E., Claude, J. & Strimmer, K. (2004) APE: analyses of phylogenetics and evolution in R language. *Bioinformatics*, **20**, 289–290.

Plummer, M., Best, N., Cowles, K. & Vines, K. (2006) CODA: convergence diagnosis and output analysis for MCMC. *R News*, **6**, 7–11.

Pybus, O.G. & Harvey, P.H. (2000) Testing macro-evolutionary models using incomplete molecular phylogenies. *Proceedings of the Royal Society, B*, **267**, 2267–2272.

R Development Core Team (2011) *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

R Hackathon (alphabetically: Bolker, B., Butler, M., Cowan, P., de Vienne, D., Eddelbuettel, D., Holder, M., Jombart, T., Kembel, S., Michonneau, F., Orme, D., O'Meara, B., Paradis, E., Regetz, J., Zwickl, D.) (2011) *phylobase: Base Package for Phylogenetic Structures and Comparative Data. R Package Version 0.6.3*. http://cran.r-project.org/package = phylobase.

Ragan, M.A. (1992) Phylogenetic inference based on matrix representation of trees. *Molecular Phylogenetics and Evolution*, **1**, 53–58.

Revell, L.J. (2008) On the analysis of evolutionary change along single branches in a phylogeny. *American Naturalist*, **172**, 140–147.

Revell, L.J. (2009) Size-correction and principal components for interspecific comparative studies. *Evolution*, **63**, 3258–3268.

Revell, L.J. (2010) Phylogenetic signal and linear regression on species data. *Methods in Ecology and Evolution*, **1**, 319–329.

Revell, L.J. & Collar, D.C. (2009) Phylogenetic analysis of the evolutionary correlation using likelihood. *Evolution*, **63**, 1090–1100.

Revell, L.J., Harmon, L.J. & Collar, D.C. (2008) Phylogenetic signal, evolutionary process, and rate. *Systematic Biology*, **57**, 591–601.

Revell, L.J. & Harrison, A.S. (2008) PCCA: a program for phylogenetic canonical correlation analysis. *Bioinformatics*, **24**, 1018–1020.

Revell, L.J., Mahler, D.L., Peres-Neto, P.R. & Redelings, B.D. (in press) A new method for identifying exceptional phenotypic diversification. *Evolution*, doi: 10.1111/j.1558-5646.2011.01435.x

Schliep, K.P. (2011) phangorn: phylogenetic analysis in R. *Bioinformatics*, **27**, 592–593.

Sidlauskas, B. (2008) Continuous and arrested morphological diversification in sister clades of characiform fishes: a phylomorphospace approach. *Evolution*, **62**, 3135–3156.

Turlach, B.A. & Weingessel, A. (2010) *quadprog: Functions to Solve Quadratic Programming Problems. R Package Version 1.5-3*. http://cran.r-project.org/package = quadprog.

Xie, Y. (2011) *animation: A Gallery of Animations in Statistics and Utilities to Create Animations. R Package Version 2.0-4*. http://cran.r-project.org/package = animation.