# Two new graphical methods for mapping trait evolution on phylogenies

## Liam J. Revell*

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

## Summary

**1.** Modern phylogenetic comparative biology uses data from the relationships between species (phylogeny) combined with comparative information for phenotypic traits to draw model-based statistical inferences about the evolutionary past. Recent years have seen phylogeny methods for evolutionary inference become central in the study of organic evolution.

**2.** Here, I present two different graphical methods for visualizing phenotypic evolution on the tree. Method 1 is a new approach for plotting the posterior density of stochastically mapped character histories for a binary (two-state) phenotypic trait on a phylogeny. Method 2 is a closely related technique that uses ancestral character estimation to visualize historical character states for a continuous trait along the branches of a tree.

**3.** One shortcoming of Method 2 is that by mapping the point estimates of ancestral states along the branches of the tree, we have effectively ignored the uncertainty associated with ancestral character estimation of continuous traits. To alleviate this issue, I propose a new method for visualizing ancestral state uncertainty using a type of projection of the tree into morphospace called a 'traitgram.'

**4.** All of these approaches should prove useful in summarizing complex comparative inferences about ancestral character reconstruction. They are implemented in the freely available and open-source R phylogenetics package 'phytools.'

**Key-words:** ancestral states, comparative method, interspecific data, stochastic character mapping, traitgram

## Introduction

Evolutionary biology is a historical discipline: large evolutionary changes take place over long time-scales and thus cannot be studied directly. One of the most important tools in the arsenal of scientists studying the evolutionary past is phylogenetic comparative biology (Felsenstein 1985, 2004; Harvey & Pagel 1991). Phylogenetic comparative biologists combine the phylogeny (that is, one or multiple estimates of the relationships and divergence times between taxa) with data for the phenotypic traits of species to make inferences about evolutionary processes in deep time (Harmon *et al.* 2010; Mahler *et al.* 2010; O'Meara 2012). In recent years, phylogenetic comparative approaches have become central in evolutionary biology (Miles & Dunham 1993; Freckleton, Harvey & Pagel 2002; Losos 2011).

One component of studying the evolutionary past is reconstructing the past phenotypes of extinct ancestral species from the trait values of their extant descendants (Huey & Bennett 1987; Maddison 1991; Schluter *et al.* 1997; Pagel 1999). Modern approaches for ancestral character reconstruction treat the states at ancestral nodes as parameters or variables in a statistical model and use techniques such as maximum likelihood or Bayesian MCMC to fit the model of evolution and estimate

ancestral character values. For example, using a method called stochastic character mapping, developed by Neilsen (2002; Huelsenbeck, Nielsen & Bollback 2003), we can sample possible histories of a discretely valued character from their Bayesian posterior probability distribution. A single such character history is meaningless; however, in aggregate, a large number of discrete stochastic character maps (say, 100 to 1000) can be used to estimate the history of a discretely valued character trait evolving on the tree. In addition, the variability among sampled histories provides a measure of the uncertainty we have about trait evolution in our clade of interest (e.g. Collar *et al.* 2009; Kelly, Near & Alonzo 2012).

A stochastic character mapping analysis will result in a large number of discrete character histories on a phylogeny. These data are difficult to visualize in a figure for publication. For instance, one might display the result of one or a small number of representative stochastic histories (e.g. Collar *et al.* 2009; Price *et al.* 2013); but this ignores the fact that the stochastic maps should be considered in aggregate. Alternatively, one might compute the marginal probability of the character being in each state at all internal nodes of the tree (e.g. Kelly, Near & Alonzo 2012); but this ignores the trait values for the character along internal edges, which represent an important component of stochastic maps. Both of these approaches have been taken in recent empirical studies.

In this note, I present a new approach for visualizing the posterior density of a binary character state on a phylogenetic

*Correspondence author. E-mail: liam.revell@umb.edu

tree from stochastic mapping. In addition, I present a closely related, visually similar technique for mapping the estimated state of a continuous character on the tree. I have illustrated the two methods with an empirical analysis of published data. R is a programming language and scientific computing environment with a large and growing community of users in phylogenetic biology (R Development Core Team 2013). Both of these methods are implemented in the freely available, multi-functional, open-source R phylogenetics package, 'phytools' (Revell 2012). phytools is available from the Comprehensive R Archive Network, CRAN.

## Methods and results

In this note, I present two closely related approaches for visualizing trait evolution on a phylogeny. Method 1 is a technique for visualizing the posterior density of character histories obtained by aggregating the results from a set of stochastic character maps. Method 2 is a related approach to graph the reconstructed states for ancestral species in a continuous character along the edges and nodes of a phylogeny. I will describe each method and illustrate its use with an empirical example from published studies, in the sections below.

*Visualizing the aggregate result of stochastic mapping* – Method 1 involves aggregating the results of a large number of stochastic character maps. For example, Fig. 1 shows three of 1000 stochastic character maps for the binary character of

piscivory (red) vs. non-piscivory (blue) in the phylogeny of sunfishes (Centrarchidae; Near, Bolnick & Wainwright 2005; Collar *et al.* 2009; Revell & Collar 2009).

Stochastic character mapping is a technique whereby possible character histories are sampled in proportion to their probability. Each sample contains a specific, unique and subtly or markedly different character history – for instance, the first two trees of Fig. 1 imply that piscivory was the ancestral feeding mode in centrarchids; whereas the third stochastic map implies non-piscivory at the root node of the tree. Consequently, stochastic character maps are only meaningful when considered in aggregate.

Here, I propose a new approach for visualizing the aggregate result from stochastic character mapping of a binary trait. According to this procedure, we finely fraction the branches of the tree and then compute the posterior probability of the state being in either condition '*0*' or condition '*1*' for each fraction as the relative frequency across all stochastic maps. When a fraction is bisected by a state change, the two states both contribute to the cumulative probability – weighted by the relative branch length spent in each state. We can then plot these probabilities on the branches of the tree using a colour map to translate the probability to a plotted colour. The result (so long as a sufficient number of stochastic histories are generated) is that the probability density appears to change continuously along the branches of the tree. This method is implemented in the function `densityMap` of the R phylogenetics package, phytools
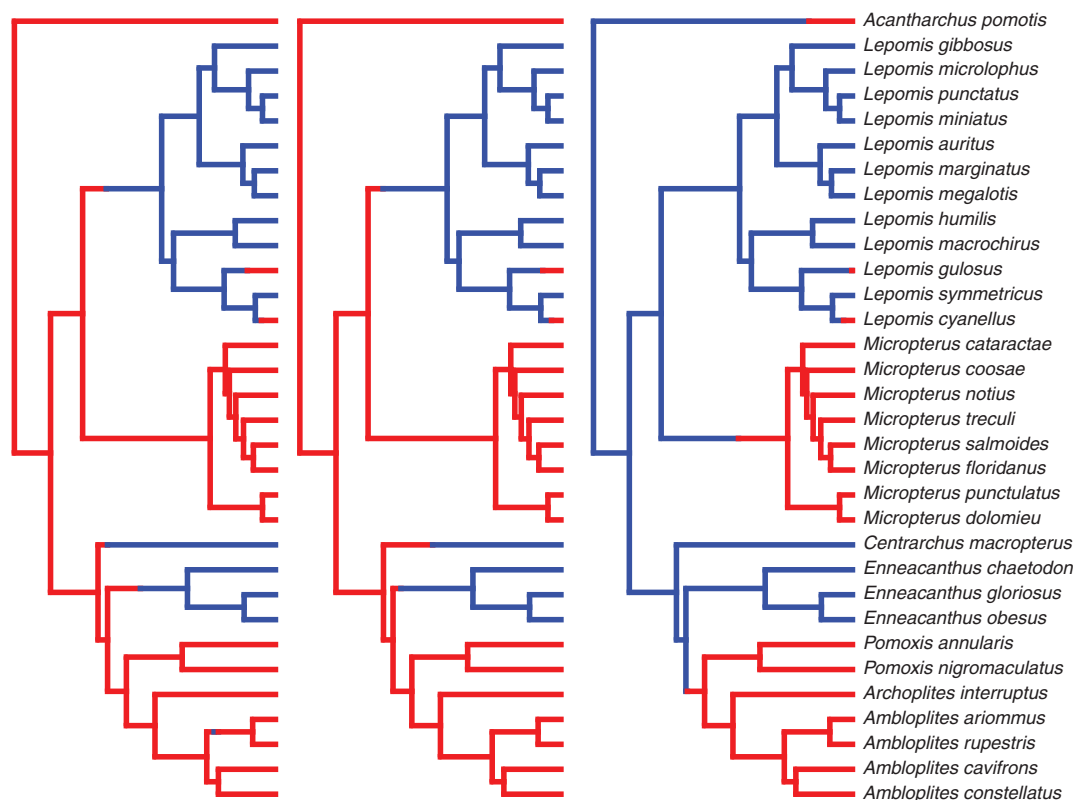


**Fig. 1.** A sample of three stochastic character maps for the evolution of feeding mode in Centrarchidae. Phylogeny and data are from Near, Bolnick & Wainwright (2005) and Collar *et al.* (2009).

(Revell 2012). I have applied the method to the full sample of 1000 stochastically mapped trees for feeding mode in Centrarchidae, of which just three stochastic maps are given in Fig. 1. The result is shown in Fig. 2. The horizontal bar at the bottom of the graph in Fig. 2 is at the same time both a legend and scale: the colour map translates the heat colours (blue through red) into the posterior probability of feeding mode being in the condition 'piscivory'; and the total length of the bar also provides a scale for the branch lengths of the tree. Figure 2 suggests that the ancestral feeding mode should be considered highly uncertain for about the first 1/3 of the tree; but this was followed by multiple acquisitions of piscivory or non-piscivory or both. This visualization, particularly the uncertain condition of many ancestral branches in the tree, is more revealing that any single stochastic map considered in isolation. Stochastic character mapping can be conducted in R using the phytools function `make.simmap` (Revell 2012); or with the stand-alone Mac OS program SIMMAP (Bollback 2006). (For a description of stochastic mapping in phytools, please refer to the phytools documentation.) Stochastically mapped phylogenies simulated using SIMMAP v1.0 or v1.5 can easily be read into the R environment using the phytools function `read.simmap` (Revell 2012).

*Visualizing ancestral estimates for continuous traits –* Whereas Method 1 is designed to aggregate the results of many discrete character hist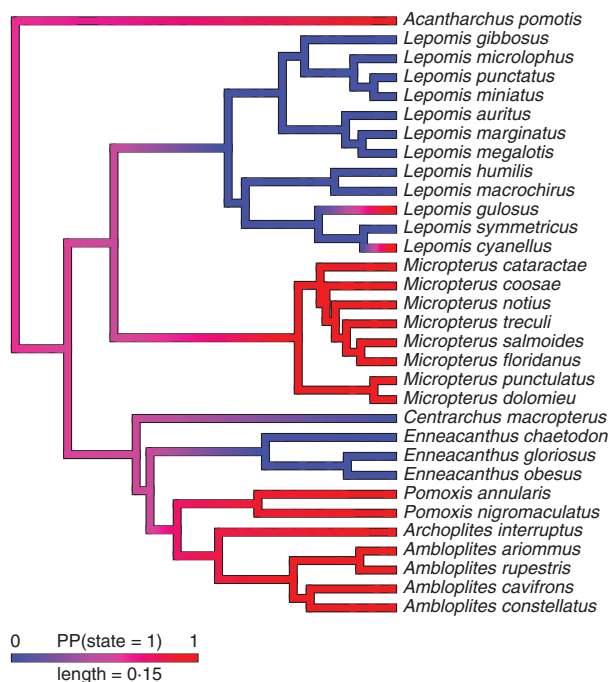ories, Method 2 uses a visually similar approach to plot the estimated history of a continuously valued trait on the tree. Here, we first estimate the ancestral characters at internal nodes using, for instance, maximum likelihood (ML; Schluter *et al.* 1997). For simplicity, I assume Brownian motion as a model for trait evolution (Cavalli-Sforza & Edwards 1967; Felsenstein 1973, 1985); however, the method could be adapted to other evolutionary models such as the Ornstein–Uhlenbeck model (Hansen 1997; Butler & King 2004). Next, we finely fraction all edges (as in Method 1), and then compute the state estimates at the midpoint of each fraction via interpolation. Again, this creates the visual appearance of continuous colour change along the edges of the tree. I interpolated the internal states along branches using equation (3) of Felsenstein (1985). Here, the target ancestral value at some arbitrary point along an edge in the tree, $\hat{a}$, is a function of the observed or ML estimated states at the subtending nodes or tips $i$ and $j$ ($x_i$ and $x_j$), as well as the phylogenetic distances to nodes $i$ and $j$ ($v_i$ and $v_j$):

$$\hat{a} = \frac{x_i/v_i + x_j/v_j}{1/v_i + 1/v_j}.$$

Figure 3 shows an example analysis of the evolution of body size (log-transformed snout-to-vent length, SVL) in 100 species of Greater Antillean *Anolis* lizards (Mahler *et al.* 2010). As in Fig. 2, the horizontal bar is both a legend and scale. The colour map represents observed and reconstructed values for SVL on the tree. Red colours correspond with relatively low values for SVL; whereas green through blue colours represent larger observed and reconstructed trait values. This visualization can be reproduced using the phytools function `contMap`.

One obvious shortcoming of this approach is that by plotting only the point estimate of phenotype at the nodes and along the branches of the tree, we are ignoring the considerable uncertainty that can be inherent in ancestral state estimation (Schluter *et al.* 1997; Losos 2011). Figure 4 shows a different kind of visualization – a type of projection of the phylogenetic tree into trait space, called a 'traitgram' by Ackerly (2009), here performed using a simulated phylogeny and dataset. In a traitgram, the vertical position of nodes and edges give known or estimated phenotypic trait values, while the horizontal position gives time or patristic distance from the root. Traitgrams have been used in a number of prior publications (e.g. Ackerly 2009; Evans *et al.* 2009). The innovative aspect of Fig. 4 is that in addition to showing the estimated ancestral values for the phenotypic trait, it also shows uncertainty about ancestral states along branches and at nodes. To do this, I computed the confidence intervals at internal nodes and then linearly interpolated along branches. Increasing quantiles away from the point estimate are shown via increased transparency of the plotted density. This visualization is implemented as a method of the phytools function `fancyTree`, using `type = 'phenogram95'`. `fancyTree` calls the phytools function `phenogram` internally. One nice aspect of `phenogram` over other implementations of traitgram plotting (e.g. Kembel *et al.* 2010) is that `phenogram` can accept user-supplied penalties (or 'costs') for label overlap and then uses numerical methods



**Fig. 2.** The results from 1000 stochastic character maps (the three shown in Fig. 1, plus 997 others) displayed in aggregate. The colour of edges in the tree gives the posterior probability (computed as the relative frequency across stochastic maps) of each feeding mode through the history of the clade. Red indicates high posterior probability of piscivorous feeding mode. The length of the legend also gives a scale for the branch lengths of tree, in this case in units of substitution per site for five mitochondrial and four nuclear genes (Near, Bolnick & Wainwright 2005).
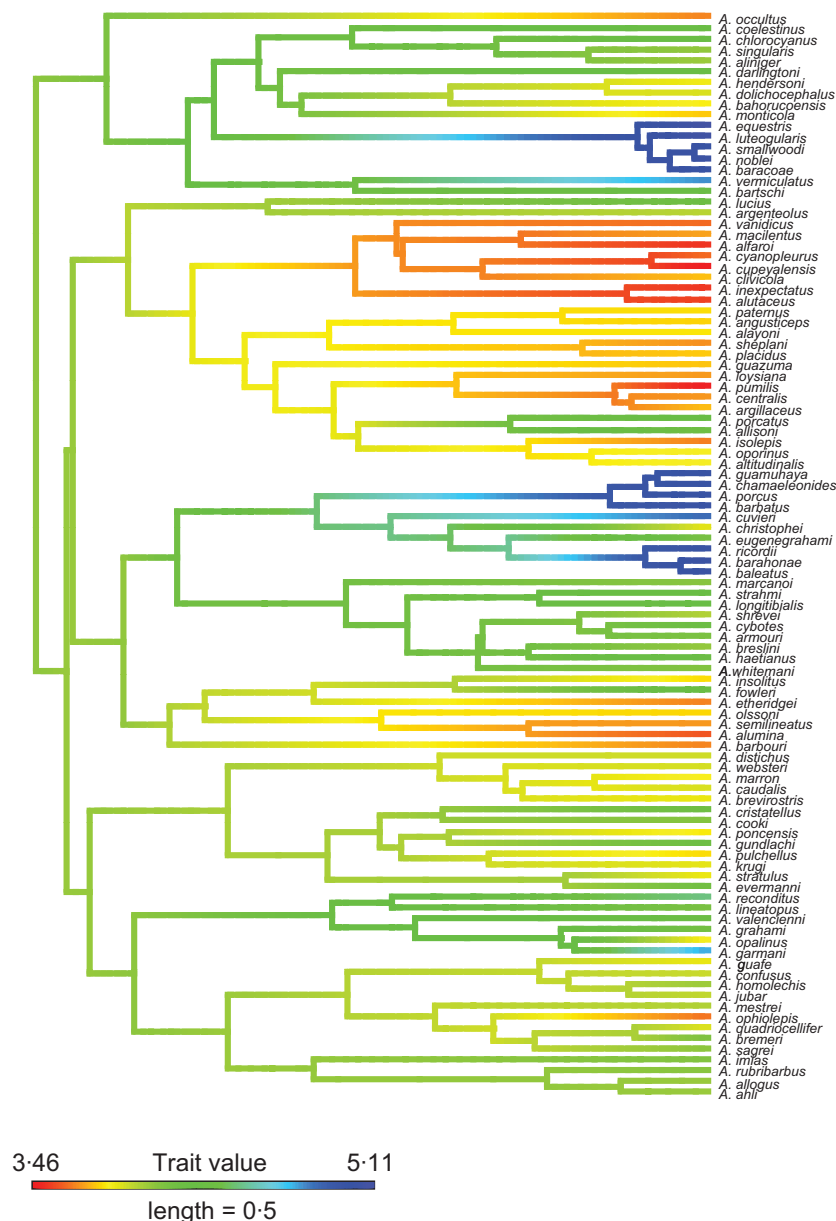
**Fig. 3.** Ancestral character estimation of body size (log-transformed snout-to-vent length) along the branches and nodes of the tree for Greater Antillean *Anolis* lizards. The total tree length was arbitrarily rescaled to unit length in this analysis, following the original authors. Phylogeny and data are from Mahler *et al.* (2010).

to optimize the spacing of tip labels on the trait axis. This can be very helpful because multiple species often have similar trait values (Fig. 4), which can make their tip labels difficult to read if plotted conventionally (i.e. directly adjacent to the tips).

## Discussion

Herein, I present two new methods, implemented in computer software, to visualize trait evolution on a phylogeny. The first of these, Method 1, is an approach to aggregate the results from a procedure called stochastic character mapping. In stochastic character mapping, we sample character histories (such as those shown in Fig. 1) in proportion to their probability conditioned on a model (Neilsen 2002; Huelsenbeck, Nielsen

& Bollback 2003). These character-mapped phylogenies can then be used in subsequent analyses, such as in model-fitting for comparative biology (O'Meara *et al.* 2006) or in computing the mean and distribution of trait changes on the tree (e.g. Kelly, Near & Alonzo 2012). Thus, for instance, one might fit a model of evolution to each tree in the set of mappings, and then compute the mean result across trees (Collar *et al.* 2010; Revell 2013). One difficulty with these mappings is that they are relatively meaningless in isolation and should be considered in aggregate. There is no difficulty for analytic techniques such as those described previously, but presents a challenge for visualization. The typical response in prior studies has been to plot one or a small number of stochastic maps; or to compute and plot the posterior probability at nodes, but ignore state
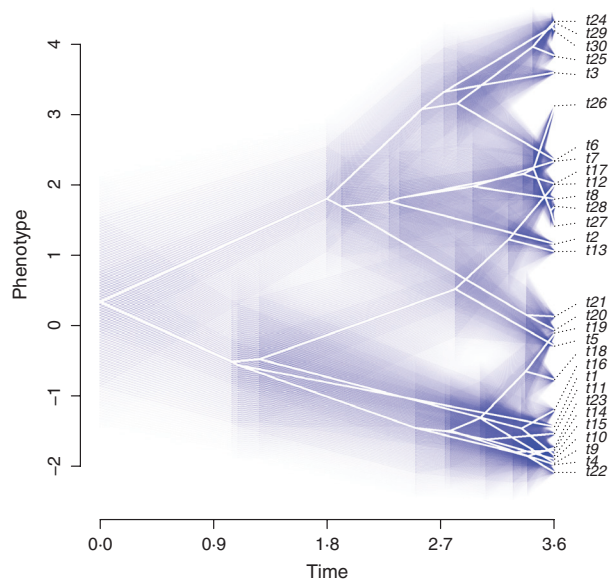
**Fig. 4.** A projection of the phylogeny into a space defined by pheno-type (on *y*) and time since the root. The vertical position of nodes and branches are computed via ancestral character estimation using likeli-hood (e.g. Schluter *et al.* 1997). Uncertainty is shown via increasing transparency of the plotted blue lines around the point estimates with the entire range showing the 95% confidence interval.

changes that occur along the branches of the tree (e.g. Collar *et al.* 2009, 2010; Kelly, Near & Alonzo 2012; Price *et al.* 2013). In this article, I illustrate a new method for visualizing the aggregate result of stochastic mapping of a binary character in which we map posterior density on the tree using a continuous colour gradient.

An inherent shortcoming of this method is that it is only sen-sible for a binary trait – in which case the plotted colour is a proxy for the probability of being in an identified condition (such as the feeding mode 'piscivory'). For multistate character traits, options for visualizations of this type are less obvious; however, it would be no trouble to plot the posterior probabil-ity of being in any specific discrete character state. For instance, given nucleotide data, we could use this method to plot the 'probability of *T*' vs. 'the probability of not-*T*' on the branches of the tree, even though there are actually four dis-tinct nucleotide states. One option for plotting the probability of three states is to use saturation in three dimensions on an RGB ('red-green-blue') colour model. This may be incorpo-rated into a future version of 'phytools,' but extension to four or more trait values is not obvious.

Method 2 is a closely related visualization technique, but for continuous trait evolution. Here, I computed the ML estimates at internal nodes, and then I interpolated ancestral condition along the branches of the tree using equation (3) of Felsenstein (1985). Visualizing quantitative trait evolution for a single character can alternatively be performed using the 'traitgram' approach in which the vertical dimension of a phylogeny is projected onto a phenotypic trait axis (e.g. Fig. 4). The horizontal dimension is time, and edges connect species trait values to the trait values of their hypothesized ancestors (Ack-erly 2009; Kembel *et al.* 2010; Revell 2012). The trouble with

this method is that the phylogenetic relationships of species in a traitgram can be difficult to discern, particularly when species in different parts of the phylogeny have similar phenotypes. The new visualization technique presented here alleviates this difficulty by simultaneously portraying trait evolution and representing the phylogeny in conventional way.

As noted earlier, a significant shortcoming of this approach is that (by mapping the point estimate of ancestral traits) we are ignoring the considerable uncertainty that invariably accompanies said estimates. It is not clear how to incorporate measures of uncertainty without totally obfuscating the clarity of visualizations such as that of Fig. 3. To address this concern, I propose a new method to graph uncertainty in traitgrams. This is illustrated in Fig. 4 with simulated data. Unfortunately, this visualization is still limited by the difficulty traitgrams pose in discerning phylogenetic relationships, as described above.

All the plotting techniques described in this article are imple-mented in the open-source R phylogenetics package 'phytools' (Revell 2012). Phylogenetic comparative biology in R has exploded in recent years, with many new packages and func-tions (e.g. Harmon *et al.* 2008; Kembel *et al.* 2010; Boettiger, Coop & Ralph 2012; Fitzjohn 2012). Most, including phytools, rely extensively on the important core phylogenetics package, 'ape' (Paradis, Claude & Strimmer 2004; Paradis 2012).

Phylogenetic comparative biology has become central in the study of evolution over the past 25 years. Many methodologi-cal advances have been made, but these methods also pose new and interesting visualization challenges. Here, I describe sev-eral methods for visualizing ancestral states on the branches and nodes of a phylogeny. These approaches add considerably to our set of plotting techniques for ancestral character recon-struction on phylogenies.

## References

Ackerly, D. (2009) Conservatism and diversification of plant functional traits: evolutionary rates versus phylogenetic signal. *Proceedings of the National Academy of Sciences*, **106**, 19699–19706.

Boettiger, C., Coop, G. & Ralph, P. (2012) Is your phylogeny informative? Mea-suring the power of comparative methods. *Evolution*, **66**, 2240–2251.

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

Butler, M.A. & King, A.A. (2004) Phylogenetic comparative analysis: a modeling approach for adaptive evolution. *American Naturalist*, **164**, 683–695.

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

Collar, D.C., O'Meara, B.C., Wainwright, P.C. & Near, T.J. (2009) Piscivory lim-its diversification of feeding morphology in centrarchid fishes. *Evolution*, **63**, 1557–1573.

Collar, D.C., Schulte, J.A. II, O'Meara, B.C. & Losos, J.B. (2010) Habitat use affects morphological diversification in dragon lizards. *Journal of Evolutionary Biology*, **23**, 1033–1049.

Evans, M.E.K., Smith, S.A., Flynn, R.S. & Donoghue, M.J. (2009) Climate, niche evolution, and diversification of the "bird-cage" evening primroses (Oenothera, Sections Anogra and Kleinia). *American Naturalist*, **173**, 224–240.

Felsenstein, J. (1973) Maximum likelihood estimation of evolutionary trees from continuous characters. *American Journal of Human Genetics*, **25**, 471–492.

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

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

Fitzjohn, R.G. (2012) Diversitree: comparative phylogenetic analyses of diversification in R. *Methods in Ecology and Evolution*, **3**, 1084–1092.

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.

Hansen, T.F. (1997) Stabilizing selection and the comparative analysis of adaptation. *Evolution*, **51**, 1341–1351.

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

Harmon, L.J., Losos, J.B., Davies, T.J., Gillespie, R.G., Gittleman, J.L., Jennings, W.B. *et al.* (2010) Early bursts of body size and shape evolution are rare in comparative data. *Evolution*, **64**, 2385–2396.

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.

Huey, R.B. & Bennett, A.F. (1987) Phylogenetic studies of coadaptation: preferred temperatures versus optimal performance temperatures of lizards. *Evolution*, **41**, 1098–1115.

Kelly, N.B., Near, T.J. & Alonzo, S.H. (2012) Diversification of egg-deposition behaviours and the evolution of male parental care in darters (Teleostei: Percidae: Etheostomatinae). *Journal of Evolutionary Biology*, **25**, 836–846.

Kembel, S.W., Cowan, P.D., Helmus, M.R., Cornwell, W.K., Morlon, H., Ackerly, D.D., Blomberg, S.P. & Webb, C.O. (2010) Picante: R tools for integrating phylogenies and ecology. *Bioinformatics*, **26**, 1463–1464.

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

Maddison, W.P. (1991) Squared-change parsimony reconstructions of ancestral states for continuous-valued characters on a phylogenetic tree. *Systematic Biology*, **56**, 701–710.

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.

Miles, D.B. & Dunham, A.E. (1993) Historical perspectives in ecology and evolutionary biology: the use of phylogenetic comparative analyses. *Annual Reviews of Ecology and Systematics*, **24**, 587–619.

Near, T.J., Bolnick, D.I. & Wainwright, P.C. (2005) Fossil calibrations and molecular divergence time estimates in centrarchid fishes (Teleostei: Centrarchidae). *Evolution*, **59**, 1768–1782.

Neilsen, 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.

O'Meara, B.C. (2012) Evolutionary inferences from phylogenies: a review of methods. *Annual Review of Ecology, Evolution and Systematics*, **43**, 267–285.

Pagel, M. (1999) The maximum likelihood approach to reconstructing ancestral character states of discrete characters on phylogenies. *Systematic Biology*, **48**, 612–622.

Paradis, E. (2012) *Analysis of Phylogenetics and Evolution with R*, 2nd edn. Springer, New York, NY.

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

Price, S.A., Tavera, J.J., Near, T.J. & Wainwright, P.C. (2013) Elevated rates of morphological and functional diversification in reef-dwelling haemulid fishes. *Evolution*, **67**, 417–428.

R Development Core Team. (2013) *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna Austria.

Revell, L.J. (2012) phytools: an R package for phylogenetic comparative biology (and other things). *Methods in Ecology and Evolution*, **3**, 217–223.

Revell, L.J. (2013) A comment on the use of stochastic character maps to estimate evolutionary rate variation in a continuously valued trait. *Systematic Biology*, **62**, 339–345.

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

Schluter, D., Price, T., Mooers, A.Ø. & Ludwig, D. (1997) Likelihood of ancestor states in adaptive radiation. *Evolution*, **51**, 1699–1711.