Syst. Biol. 63(3):397–408, 2014

DOI:10.1093/sysbio/syu005
Advance Access publication February 4, 2014

A Linear-Time Algorithm for Gaussian and Non-Gaussian Trait Evolution Models

LAM SI TUNG HO[1] and CÉCILE ANÉ[1,2,*]

[1]*Department of Statistics, University of Wisconsin - Madison, Madison, WI 53706, USA*
[2]*Department of Botany, University of Wisconsin - Madison, Madison, WI 53706, USA*
**Correspondence to be sent to Department of Statistics, 1300 University Ave., Madison WI 53706, USA; E-mail: ane@stat.wisc.edu*

Received 12 January 2013; reviews returned 16 May 2013; accepted 24 January 2014
Associate Editor: Luke Harmon

Abstract.—We developed a linear-time algorithm applicable to a large class of trait evolution models, for efficient likelihood calculations and parameter inference on very large trees. Our algorithm solves the traditional computational burden associated with two key terms, namely the determinant of the phylogenetic covariance matrix **V** and quadratic products involving the inverse of **V**. Applications include Gaussian models such as Brownian motion-derived models like Pagel's lambda, kappa, delta, and the early-burst model; Ornstein-Uhlenbeck models to account for natural selection with possibly varying selection parameters along the tree; as well as non-Gaussian models such as phylogenetic logistic regression, phylogenetic Poisson regression, and phylogenetic generalized linear mixed models. Outside of phylogenetic regression, our algorithm also applies to phylogenetic principal component analysis, phylogenetic discriminant analysis or phylogenetic prediction. The computational gain opens up new avenues for complex models or extensive resampling procedures on very large trees. We identify the class of models that our algorithm can handle as all models whose covariance matrix has a 3-point structure. We further show that this structure uniquely identifies a rooted tree whose branch lengths parametrize the trait covariance matrix, which acts as a similarity matrix. The new algorithm is implemented in the `R` package `phylolm`, including functions for phylogenetic linear regression and phylogenetic logistic regression.

Statistical methods are well developed now to account for phylogenetic correlation in the analysis of quantitative traits across species. Most of these methods use models derived from a Brownian motion (BM) or an Ornstein-Uhlenbeck (OU) process along the tree, where the tree is informed by molecular data. These models are now used on increasingly large data sets from tree-of-life projects with large numbers of taxa (e.g., Smith and Donoghue 2008; Cooper and Purvis 2010; Jetz et al. 2012) or on a large number of traits such as from expression profiles (e.g., Brawand et al. 2011). The traditional likelihood calculations use generic methods and do not raise computational issues for a few traits and a dozen taxa. However, these methods are challenged by large phylogenetic covariance matrices on large numbers of taxa, causing very long computing times at best, and possibly failure or inaccurate results (Hadfield and Nakagawa 2010). For instance, Venditti et al. (2011) studied body mass across 3185 mammal species. Using generic methods, one single analysis of their data took us 1 hour 20 minutes (using the BM model in `fitContinuous` from the R package `geiger` v1.3-1). To determine lineages where body mass evolution experienced a change in its evolutionary rate, their complex Bayesian model required millions of likelihood calculations. Therefore, traditional likelihood calculations were prohibitively slow and a fast algorithm was absolutely needed for their study. Recently, Freckleton (2012) showed how to calculate the likelihood of BM models in linear time based on an algorithm proposed by Felsenstein (1973), which Venditti et al. (2011) used (personal communication). Independently, FitzJohn (2012) recently proposed a fast Gaussian elimination method to calculate the likelihood of BM and OU models, philosophically similar to the pruning algorithm for molecular sequence data. However, the applicability of these methods to non-Gaussian traits is not clear and remains open (Freckleton 2012).

We present here a linear-time algorithm that applies to a broader class of phylogenetic trait models than the fast algorithms cited above. We separately calculate two computationally difficult terms involving the phylogenetic covariance matrix. These are necessary for many phylogenetic methods, making our algorithm applicable to a number of non-Gaussian models for which these two terms are used separately and in various ways. Our algorithm can be used to speed up parameter estimation and likelihood calculation of trait data under modified BM models like Pagel's lambda, kappa, delta, and the early-burst model (as can also be done with the methods of Felsenstein 1973; Freckleton 2012). These models can be used to quantify the level of phylogenetic signal in single traits, or they can quantify and account for this phylogenetic signal while testing the correlation between several traits. Our algorithm can also be used to fit OU models with possibly varying selection parameters along the tree (Butler and King 2004; Beaulieu et al. 2012), as does FitzJohn (2012). Additionally, our algorithm applies to several non-Gaussian models like phylogenetic logistic regression, phylogenetic Poisson regression, phylogenetic generalized linear model (Paradis and Claude 2002; Ives and Garland 2010), and phylogenetic generalized linear mixed models (Ives and Helmus 2011), thus allowing these methods to be used on very large data sets for the first time. Other applications include phylogenetic principal component analysis (PCA; Revell 2009), phylogenetic discriminant analysis (Motani and Schmitz 2011) and phylogenetic prediction (Garland and Ives 2000).

Our method is implemented in the `R` package `phylolm`, with functions for phylogenetic linear regression and phylogenetic logistic regression. Our algorithm requires that the underlying covariance matrix $\mathbf{V}$ for which inversion is to be avoided belong to a class of generalized 3-point structured matrices. In what follows, we first describe the terms responsible for the computational bottleneck, our algorithm for calculating them efficiently, and computation time comparisons. We then describe the class of 3-point structured matrices that our algorithm applies to, with its connection to a rooted tree structure. Finally, we show how our algorithm applies to a number of tools for trait evolution, including methods for which no other fast algorithm was available before.

The Algorithm

Calculating the Likelihood for Gaussian Traits

To motivate the method, we start by examining the likelihood for Gaussian trait models as it is used for both maximum likelihood and Bayesian inference. In this context, we explain what terms form the computational bottleneck. Consider a single trait, with data Y at n tips of a tree and an evolution model under which Y has a multivariate normal distribution with mean μ_θ and phylogenetic covariance matrix $\mathbf{V}_\theta$, for some covariance parameter θ. This covariance takes a specific form depending on the underlying trait evolution model. The simplest example is the BM model for the error term, which corresponds to $\theta=\sigma^2$ and $\mathbf{V}_\theta(i,j)=\sigma^2 t_{ij}$, where t_{ij} is the tree length from the root to the last common ancestor of tips i and j. In a regression framework, the mean value is $\mu_\theta=\mathbf{X}\beta$ where the matrix $\mathbf{X}$ contains predictors in its columns including a column vector of ones, $\mathbf{1}$. If there are no predictors then $\mu_\theta=\beta_1\mathbf{1}$ is constant and the intercept β_1 typically represents the ancestral trait value at the root. The likelihood $f(Y|\theta)$ is then given by

$$-2\log f(Y|\theta)=n\log(2\pi)+\log|\mathbf{V}_\theta| +(Y-\mathbf{X}\beta)'\mathbf{V}_\theta{}^{-1}(Y-\mathbf{X}\beta).$$

Given θ, the estimated regression coefficients are $\hat{\beta}=(\mathbf{X}'\mathbf{V}_\theta^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}_\theta^{-1}Y$. If there are no predictors, then this is simply the estimated mean (or ancestral state at the root) and simplifies to $\hat{\beta}_1=(\mathbf{1}'\mathbf{V}_\theta^{-1}\mathbf{1})^{-1}\mathbf{1}'\mathbf{V}_\theta^{-1}Y$.

Traditional calculations use generic methods for calculating the determinant $|\mathbf{V}_\theta|$ and the matrix inverse $\mathbf{V}_\theta^{-1}$, such as the Cholesky or QR decomposition, or singular value decomposition (Lawson and Hanson 1995). However, the full matrix inverse $\mathbf{V}_\theta^{-1}$ is not needed. Our approach consists in calculating the following key terms only: $|\mathbf{V}_\theta|$ and the quadratic products $Y'\mathbf{V}_\theta^{-1}Y$ and $\mathbf{X}'\mathbf{V}_\theta^{-1}Y$, without actually inverting $\mathbf{V}_\theta$. The general algorithm to calculate these quadratic products can then also be used to obtain the precision matrix $\mathbf{X}'\mathbf{V}_\theta^{-1}\mathbf{X}$ whose inverse gives the variances of the β regression coefficients. These provide the basis to get standard errors, p-values and test hypotheses about coefficients. Note that alternative fast algorithms (Felsenstein 1973; Freckleton 2012; FitzJohn 2012) focus on the likelihood only, not on the standard errors of coefficients as we do here. The gain in computation efficiency is all the more important when the phylogenetic signal parameters in θ are unknown, requiring repeated iterations to estimate θ by maximum likelihood or Markov Chain Monte Carlo (MCMC).

Hadfield and Nakagawa (2010) already discussed the computational burden of inverting the covariance matrix. They report attempts to calculate $\mathbf{V}^{-1}$ for the 4510 mammal species supertree from Bininda-Emonds et al. (2007) that either failed completely or took up to one month of computing time. They discussed similarities between the phylogenetic linear mixed model and animal models in quantitative genetics, for which efficient algorithms have been developed early on. They noted that an expanded covariance matrix can be inverted analytically if the traits at all internal nodes are considered. Therefore, they argued for algorithms that track the unknown trait values at internal nodes, such as expectation-maximization (EM; Dempster et al. 1977) or MCMC for Bayesian analyses. Here, our algorithm removes the constraint of inverting a large matrix without relying on an expanded state space or on the convergence of an EM algorithm.

The Algorithm

This algorithm is designed to calculate the determinant $|\mathbf{V}|$ and quadratic quantities of the form $\mathbf{Q}=\mathbf{X}'\mathbf{V}^{-1}\mathbf{Y}$, for any covariance matrix $\mathbf{V}$ within the class of "3-point structured" matrices. This concept is somewhat technical and postponed to the next section, where we show a connection between a 3-point condition on similarity matrices and a rooted tree structure. $\mathbf{X}$ and $\mathbf{Y}$ must have the same number of rows as $\mathbf{V}$, but can have any number of columns. The algorithm also needs to recursively calculate the following quantities: $p=\mathbf{1}'\mathbf{V}^{-1}\mathbf{1}$, $\hat{\mu}_Y=\mathbf{1}'\mathbf{V}^{-1}\mathbf{Y}/p$, and $\tilde{\mu}_X'=\mathbf{X}'\mathbf{V}^{-1}\mathbf{1}/p$. For the BM phylogenetic regression above, the algorithm needs to be applied to the response trait $\mathbf{Y}$ (one column) and to $\mathbf{X}$ containing one column per predictor. Then $\hat{\mu}_Y$ and $\tilde{\mu}_X$ give the ancestral state estimates of each trait at the root and p is proportional to the precision of these estimates. To obtain the standard errors of regression coefficients, the algorithm also needs to be applied with $\mathbf{Y}$ replaced by $\mathbf{X}$ to get $\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}$. The algorithm runs in linear time without matrix inversion using a tree-traversal algorithm, as follows. Based on its 3-point structure, we can write $\mathbf{V}$ as

$$\mathbf{V}=t\mathbf{1}\mathbf{1}'+\mathbf{A} \quad \text{with } \mathbf{A}=\begin{pmatrix}\mathbf{V}_1 & & \mathbf{0}\\ & \ddots & \\ \mathbf{0} & & \mathbf{V}_k\end{pmatrix} \qquad (1)$$

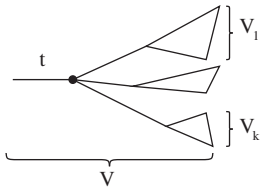

FIGURE 1. Recursion for the tree traversal: at each internal node, $\mathbf{V}$ is of the form $t\mathbf{1}\mathbf{1}'+\mathbf{A}$ with $\mathbf{A}$ block diagonal. Each block $\mathbf{V}_s$ of $\mathbf{A}$ is of this form too.

where $t \geq 0$ and $\mathbf{A}$ is a diagonal matrix with one block for each subtree stemming from the root. Each block $\mathbf{V}_s$ in $\mathbf{A}$ is of the form (1) for the corresponding subtree rooted by its parent edge (Fig. 1). The algorithm uses a post-order tree traversal, starting from the external edges and working toward the root.

1. Initialization: for a tree with a single tip, $\mathbf{V}=t>0$. Set $\log|\mathbf{V}|=\log t$, $p=1/t$, $\hat{\mu}_Y=\mathbf{y}$, $\tilde{\mu}'_X=\mathbf{x}'$ and $\mathbf{Q}=\mathbf{x}'\mathbf{y}/t$, where $\mathbf{x}$ and $\mathbf{y}$ are the rows of $\mathbf{X}$ and $\mathbf{Y}$ corresponding to the current tip.
2. For a tree with two or more tips, consider (1) for $\mathbf{V}$ on this tree. Denote p_s, $\tilde{\mu}'_{X,s}$, $\hat{\mu}_{Y,s}$ and $\mathbf{Q}_s$ the quantities obtained recursively from subtree s. Define $p_A=\sum_{s=1}^k p_s$ and weights $w_s=p_s/p_A$. Then $p=p_A/(1+tp_A)$, $\hat{\mu}_Y=\sum_{s=1}^k w_s\hat{\mu}_{Y,s}$, $\tilde{\mu}'_X=\sum_{s=1}^k w_s\tilde{\mu}'_{X,s}$, $\mathbf{Q}=\sum_{s=1}^k \mathbf{Q}_s-\frac{tp_A^2}{1+tp_A}\tilde{\mu}'_X\hat{\mu}_Y$ and $\log|\mathbf{V}|=\sum_{s=1}^k\log|\mathbf{V}_s|+\log(1+tp_A)$.
3. At the root of the full tree: return $\log|\mathbf{V}|$ and $\mathbf{Q}$ then stop.

For a growing number of taxa n, our algorithm is linear in time because it uses a single tree traversal. It is also linear in memory requirement as it does not seek to calculate the large $\mathbf{V}^{-1}$ matrix. If one actually needed the full inverse matrix, then the Woodbury formula could be used to obtain $\mathbf{V}^{-1}$ in $O(n)$ steps, as detailed in the proof below.

Proof of formulas for the algorithm. We combine the decomposition (1) for $\mathbf{V}=\mathbf{A}+t\mathbf{1}'\mathbf{1}$ with the Woodbury formula for the matrix inverse and Sylvester's determinant (Hager 1989) for matrices of the form $\mathbf{M}=\mathbf{A}+\mathbf{UCV}$:

$$\mathbf{M}^{-1}=\mathbf{A}^{-1}-\mathbf{A}^{-1}\mathbf{U}\left(\mathbf{C}^{-1}+\mathbf{V}\mathbf{A}^{-1}\mathbf{U}\right)^{-1}\mathbf{V}\mathbf{A}^{-1},$$

$$\det\mathbf{M}=\det\mathbf{A}\det\mathbf{C}\det\left(\mathbf{C}^{-1}+\mathbf{V}\mathbf{A}^{-1}\mathbf{U}\right).$$

We use these formulas with the block diagonal matrix $\mathbf{A}$ for which each block $\mathbf{V}_i$ is treated recursively, and obtain:

$$\mathbf{V}^{-1}=\mathbf{A}^{-1}-\frac{t}{1+tp_A}\mathbf{A}^{-1}\mathbf{1}\mathbf{1}'\mathbf{A}^{-1}$$

where $p_A=\mathbf{1}'\mathbf{A}^{-1}\mathbf{1}=\sum_{i=1}^k\mathbf{1}'\mathbf{V}_i^{-1}\mathbf{1}$, and $\det\mathbf{V}=(1+tp_A)\det\mathbf{A}$. The recursion for the determinant is immediate then, because $|\mathbf{A}|=\prod_{s=1}^k|\mathbf{V}_s|$. The recursion for the other components uses the facts that $p_A\hat{\mu}'_X=\mathbf{X}'\mathbf{A}^{-1}\mathbf{1}$ and $p_A\hat{\mu}_Y=\mathbf{1}'\mathbf{A}^{-1}\mathbf{Y}$. ■

Link to other Fast Methods

Our algorithm is based on a tree traversal similar to that used for independent contrasts, which are computed in linear time for each trait separatelys (Felsenstein 1973; Freckleton 2012). These authors use both the unnormalized contrasts and the contrast variances to calculate the trait likelihood. Our algorithm for $\hat{\mu}_Y$ and its precision p parallels the calculation of independent contrasts. The Gaussian elimination method by FitzJohn (2012) is also a pruning algorithm, specifically designed to calculate the likelihood under BM and OU models. Our algorithm calculates key terms separately using a rank-one perturbation of the covariance matrix, for which the 3-point condition is required. The advantage of this method is its application to a broad class of models including non-Gaussian models, and its application to methods using some components separately like phylogenetic PCA or phylogenetic prediction (as detailed later).

Implementation and Computing Gain

The algorithm was implemented in the `R` package `phylolm`, with a function to perform phylogenetic regression for quantitative responses under various models for the residual error: BM, BM with trend if the tree is not ultrametric, Pagel's λ, κ, and δ models, and OU models with constant selection strength α and variance rate σ^2. These covariance parameters are estimated jointly with coefficient parameters in the linear model using maximum likelihood. The linear model can combine quantitative and categorical predictors. The package is available at http://cran.r-project.org/web/packages/phylolm/ .

We quantified the gain in computing time for 10 to 4507 taxa. The tree used for simulation was obtained by randomly sampling the desired number of taxa from the 4507-species ultrametric tree in Bininda-Emonds et al. (2007), rescaled to a total tree height of 1. For various models of phylogenetic trait evolution, 100 data sets were simulated with the following parameters: ancestral state $\mu=0$, variance rate $\sigma^2=1$, optimal value of 1 and $\alpha=1$ for the OU model, $\lambda=0.5$, $\kappa=0.5$, $\delta=0.5$, or $r=-1$ for the EB model. The parameters were then estimated on 3.0 GHz processing units using 3 implementations: our fast algorithm in `phylolm`, matrix inversion as used in `fitContinuous` in package `geiger` version 1.3-1, and the newer version of `fitContinuous` in `geiger` version 1.99-3, which uses FitzJohn's fast pruning method (2012). Estimation with `fitContinuous` v1.3-1 was performed on fewer than 100 replicates for 1000 or more taxa, due to very long computing times (3 on 4507 taxa for models other

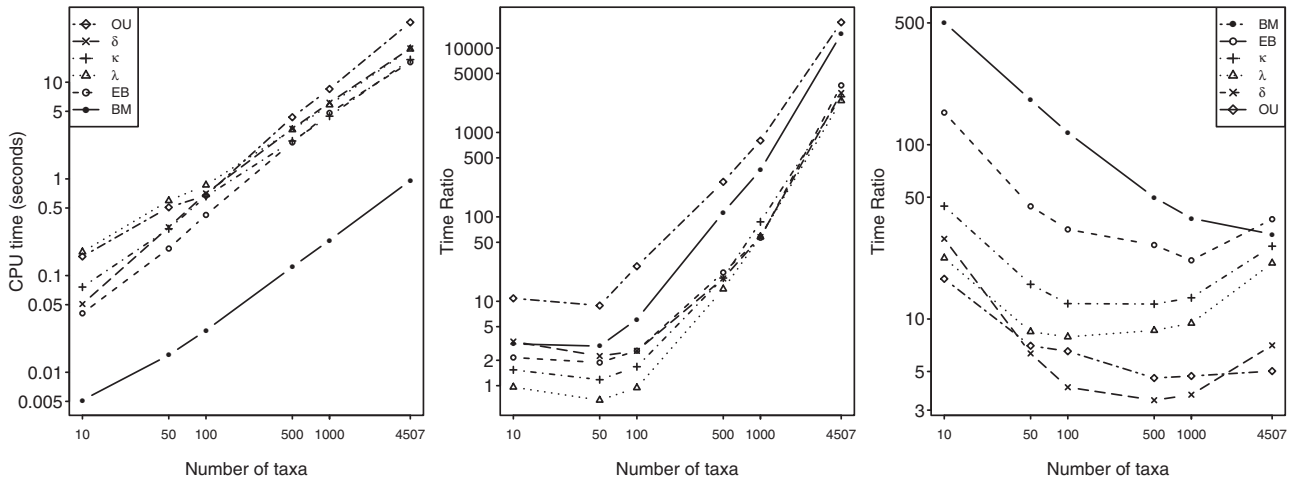

FIGURE 2. Running time for fitting phylogenetic regression models with maximum likelihood under BM, early-burst (EB), Pagel's λ, κ, and δ models, and the OU model. Left: average computing time of the proposed algorithm as implemented in `phylolm` showing a linear increase with tree size. Middle: Ratio of average computing time using `fitContinuous` v1.3-1 to `phylolm`'s average computing time. The use of matrix inversion causes a time requirement that explodes with tree size. Right: Ratio of average computing time using `fitContinuous` v1.99-3 to `phylolm`'s average computing time. The use of FitzJohn's pruning algorithm restores a linear computing time with increasing tree size.

than BM, 10 otherwise). The average computing time using `phylolm` is displayed in figure 2 (left), showing a linear increase with the number of taxa. Compared to `phylolm`, `fitContinuous` v1.3-1 is much slower (Fig. 2, center) due to the use of matrix inversion. For instance, our fast algorithm provides over a 100-fold speed increase on 500 taxa under BM, and over a 14,000-fold speed increase for the OU model on 4507 taxa. Freckleton (2012) reported similar computing gains (although for different implementations) with his linear-time method being 300 to 900 times faster than the direct matrix inversion method on a tree with 1000 tips. Thanks to its use of FitzJohn's pruning algorithm, `fitContinuous` v1.99-3 does not suffer the exploding computing burden with increasing tree size (Fig. 2, right). Its computing time was no more than 50 times that of `phylolm` on the 4507-taxon tree. `phylolm` still showed a computing gain in most situations. Note that `phylolm` incurs an overhead compared to `fitContinuous` for calculating the covariance and standard errors of coefficients for instance. Differences in the optimization procedures used by both functions might also partly explain differences in computing time.

To assess the computing gain for a non-Gaussian evolutionary model, phylogenetic logistic regression was performed on the 3185-taxon tree used by Venditti et al. (2011). A simulation was inspired by Ives and Garland (2010), who explained antipredator behavior (either flee-fight or hide) by the species' body size and group size. We simulated antipredator behavior of 3185 mammals according to their standardized log-transformed body sizes (from Venditti et al. 2011) using the binary trait model and parameters estimated in Ives and Garland (2010): $\alpha = 0.5$ (phylogenetic signal), $b_0 = -0.82$ (intercept), and $b_1 = 0.096$ (body size effect). This simulated data set was analyzed on a 2.2 GHz processing unit using our algorithm implemented in `phyloglm` in our R package `phylolm`, as well as a similar R function using matrix inversion instead. Estimation with `phyloglm` took 11.75 minutes, more than 70 times faster than when using matrix inversion, which took 862.8 minutes. The estimated coefficients were close to the true values ($\hat{b}_0 = -0.80$ and $\hat{b}_1 = 0.07$ with `phyloglm` and $\hat{b}_0 = -0.79$ and $\hat{b}_1 = 0.08$ with matrix inversion). Phylogenetic signal was better estimated by `phyloglm` ($\hat{\alpha} = 0.32$) while matrix inversion led to a convergence failure (due to $\hat{\alpha} = 0.006$ being close to 0). Ives and Garland (2010) recommend running 2000 bootstrap replicates to obtain confidence intervals on the estimated parameters, which becomes possible with the computing gain offered by our algorithm. Note also that further computing gain could be achieved by coding the algorithm in C internally, as our current implementation uses native R code in version 2.0 of `phylolm`.

THE (GENERALIZED) 3-POINT STRUCTURE

Definition 1 A matrix $\mathbf{V}$ has a 3-point structure if it is symmetric, with nonnegative entries ($V_{ij} \geq 0$) and satisfies the following 3-point condition: for any i, j, k (not necessarily distinct), the two smallest of V_{ij}, V_{ik}, and V_{jk} are equal.

These conditions are easy to check, by going over each triple i, j, k of indices and comparing the values V_{ij}, V_{ik}, V_{jk}. When $i = k$, the condition implies that $V_{ij} \leq V_{ii}$, so that $\mathbf{V}$ can be interpreted as measuring similarities. Note that our 3-point condition differs from the ultrametric condition in Semple and Steel (2003) in that it is defined for similarity maps instead of dissimilarity maps, and is required to hold even when 2 of the 3 indices (i, j, k)

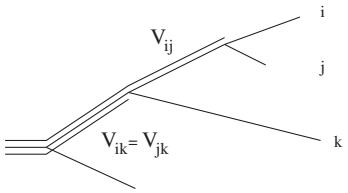

FIGURE 3. A matrix $\mathbf{V}$ is 3-point structured if it is symmetric with $V_{ij} \geq 0$ and if it satisfies the 3-point condition: for any i, j and k, the two smallest of V_{ij}, V_{ik} and V_{jk} are equal. Note that this condition can be verified without knowledge of the underlying tree.

are identical. As we show below, the 3-point structure implies a tree structure:

Theorem 2 $\mathbf{V}$ *is 3-point structured if and only if it is the covariance matrix of a random variable at the tips of* some *rooted tree under a BM model. (see Fig. 3).*

Proof. The covariance matrix of a random variable under BM evolution is obviously symmetric with nonnegative entries and satisfies the 3-point condition. However, the reverse is far from obvious. Our proof of the reverse shares some common ideas with the proof of the fundamental theorem identifying a nonrooted tree from the 4-point condition in Semple and Steel (2003, theorem 7.2.6). Consider a symmetric matrix $\mathbf{V}$ satisfying the 3-point condition, with nonnegative entries. Note that $\mathbf{V}$ is not assumed to be a covariance matrix. We define a dissimilarity map $\delta_{i,j} = -2V_{i,j}$ if $i \neq j$ and $\delta_{i,i} = 0$ for all i. δ has many negative terms, thus does not immediately identify a tree or positive edge weights. However the 3-point condition for $\mathbf{V}$ ensures that δ is an ultrametric as defined in Semple and Steel (2003, p.149). So there exists an equidistant representation of δ (theorems 7.2.5 and 7.2.8). In other words, there exists a rooted tree and a set of edge weights such that $\delta_{i,j}$ is the sum of edge weights along the path from i to j. Also, all internal edge weights are positive and the distance d_0 from the root to every tip is the same. We extend the tree with a root edge of weight $-d_0$ and add $V_{i,i}$ to the weight of the external edge connected to i. Then all edge weights of the new tree are positive and the covariance matrix of a BM process evolving along this new tree is exactly $\mathbf{V}$. ■

Note that our algorithm is readily applicable to matrices of the form $\mathbf{V} = \mathbf{D}_1 \widetilde{\mathbf{V}} \mathbf{D}_2$, when $\widetilde{\mathbf{V}}$ is 3-point structured and $\mathbf{D}_1$, $\mathbf{D}_2$ are diagonal matrices with nonzero diagonal elements. Indeed, the quadratic term $\mathbf{X}'\mathbf{V}^{-1}\mathbf{Y}$ can be obtained with our algorithm using $\widetilde{\mathbf{V}}$ and the transformed data $\mathbf{D}_2^{-1}\mathbf{X}$ and $\mathbf{D}_1^{-1}\mathbf{Y}$, which are straightforward to calculate when $\mathbf{D}_1$ and $\mathbf{D}_2$ are diagonal. The determinant is also easy to calculate from $|\mathbf{V}| = |\mathbf{D}_1||\widetilde{\mathbf{V}}||\mathbf{D}_2|$. We call such matrices $\mathbf{V}$ *generalized 3-point structured*. Also, $\mathbf{V} + \mathbf{D}$ is again generalized 3-point structured whenever $\mathbf{V}$ is and for every nonnegative diagonal matrix $\mathbf{D}$. This fact can be used when $\mathbf{D}$ represents measurement error or variation among individuals of the same species, for instance.

To check whether a matrix $\mathbf{V}$ satisfies the 3-point condition and leads to a particular tree topology T, we need to check that for any 2 tips i and j, $V_{ij} \leq V_{ii}$. If $\mathbf{V}$ is known to be a covariance matrix, then this is automatically satisfied. Next, we also need to check that for any 3 distinct tips, if i and j are more closely related to each other than to k on T, then $V_{ik} = V_{jk} \leq V_{ij}$. The equality $V_{ik} = V_{jk}$ will automatically hold if V_{ab} can generally be expressed as a feature of the last common ancestor of a and b, because the pairs i, k, and j, k share the same common ancestor. In what follows, we use this criterion to check the 3-point condition for models arising from a particular rooted tree topology.

Weak 3-point Condition and Negative Correlations

Our algorithm does not strictly require that all branch lengths (ts) are nonnegative on the rooted tree identified from $\mathbf{V}$ by the Theorem. It only requires nonnegative branch lengths at the tips, and $1 + tp_A > 0$ at each step of the algorithm. Therefore, our algorithm would apply to a broader class of covariance matrices than to just 3-point structured matrices, with some potentially negative correlations. This class of models would still require the following weak 3-point condition: for each triple i, j, k, at least two of V_{ij}, V_{jk}, and V_{ki} are equal.

One example is the PGLS method proposed by Martins and Hansen (1997). In their models the error variance depends on the mean, using the estimated mean μ_a at the most recent common ancestor of tips i and j. For instance, they propose $V_{ij} = \mu_a(1-\mu_a)\exp(-\alpha t_{ij})$ for a binary response. This formula ensures that $V_{ik} = V_{jk}$ whenever i and j are more closely related to each other than to k, so the weak 3-point condition is satisfied. The shared value $V_{ik} = V_{jk}$ might not necessarily be smaller than V_{ij}, however.

For a dissimilarity matrix, the weak 4-point condition identifies a unique unrooted tree with possibly negative branch lengths (Semple and Steel 2003). However, the weak 3-point condition on similarity matrices does not necessarily identify a unique rooted tree. As a counter-example, the similarity matrix below

$$\mathbf{V} = \begin{pmatrix} 1 & a & a & b \\ a & 1 & b & a \\ a & b & 1 & b \\ b & a & b & 1 \end{pmatrix}$$

satisfies the weak 3-point condition for any $a, b \geq 0$. However, if $a \neq b$, we can see that $\mathbf{V}$ does not correspond to any of the 15 possible 4-taxon rooted trees (with possibly negative branch lengths), in the sense that it cannot be written as a matrix of shared path lengths $(t_{ij})_{i,j}$.

Gaussian Trait Models with the 3-point Structure

Modified Brownian Motion Models

Evolution models under BM retain their 3-point structure when branch lengths are transformed. Typically, branch lengths are first inferred as divergence times based on molecular data, but can be adaptively stretched or shrunk to better fit the trait data. The goal of these transformations is to use branch lengths as parameters for the phylogenetic correlation between species, as expected from their shared history. Such transformations include Pagel's lambda (Pagel 1999) to account for a mixture of BM and independent evolution, which is equivalent to the phylogenetic mixed model (Lynch 1991). All internal branches are multiplied by λ in $[0,1]$ while the tips are maintained at their original distance from the root. Also included is the kappa transformation (Pagel 1997) where all branch lengths are raised to the power κ in $[0,1]$, the delta transformation (Pagel 1999) where the distance from the root to all nodes is raised to the power $\delta > 0$, the similar ρ transformation by Grafen (1989), and the acceleration/deceleration (ACDC) model (Blomberg et al. 2003). Models with BM evolution and a variable rate of variance accumulation also belong to the class of 3-point structured models, including models with a number of different rates $\sigma_1^2, \ldots, \sigma_k^2$ in different parts of the tree (O'Meara et al. 2006; Thomas et al. 2006; Revell and Harmon 2008; Venditti et al. 2011; Eastman et al. 2011; Revell et al. 2012), or models with a continuously varying rate along the tree such as the early-burst (EB) model (Harmon et al. 2010), which is equivalent to the ACDC transformation. The 3-point structure is maintained under a directional trend (estimable when taxa do not all have the same age), because the covariance matrix under BM evolution depends on the transformed branch lengths only, not the mean value.

Multivariate Response

If a multivariate response $\mathbf{Y}$ is considered with m different quantitative traits, the most common model of evolution assumes that the error structure has both a BM-modified phylogenetic covariance $\mathbf{V}$ as considered above, and a between-trait covariance structure $\mathbf{C}$. In other words, the covariance between traits k and l at taxa i and j is taken to be $\text{cov}(Y_{i,k}, Y_{j,\ell}) = C_{k\ell} V_{ij}$. The full covariance matrix of the vectorized $Y_{k,i}$ values is then the Kronecker product $\mathbf{C} \otimes \mathbf{V}$. This matrix is not 3-point structured, but its inverse is simply $\mathbf{C}^{-1} \otimes \mathbf{V}^{-1}$ and its determinant is $|\mathbf{C}|^n |\mathbf{V}|^m$. With a small number of traits m, $\mathbf{C}^{-1}$ and $|\mathbf{C}|$ are easy to calculate, and our algorithm can be used for the parts involving $\mathbf{V}^{-1}$ and $|\mathbf{V}|$. More complicated models have been proposed to allow for different phylogenetic signals for different traits. Freckleton (2012) considered Pagel's λ transformation with parameter λ_k for trait k, then using $\text{cov}(Y_{i,k}, Y_{j,\ell}) = C_{k\ell} V_{ij}(\sqrt{\lambda_k \lambda_\ell})$ and independent contrasts to do the computations in linear time. Here again, the covariance matrix of the full response data is not 3-point structured, but our method can still be used for the particular model above, because each matrix $\mathbf{V}(\sqrt{\lambda_k \lambda_\ell})$ is proportional to the same 3-point structured $\mathbf{V}(1)$ up to its diagonal (in Appendix).

Within-Species Variation

Individual values can be used at the tips instead of species averages as in Ives et al. (2007); Felsenstein (2008) or Revell and Reynolds (2012). This can greatly contribute to increasing tree size and computational burden. Each species is then represented by a star tree of the k individuals from that species. For a univariate response, the branches is this star tree are given length t_{within} to represent and estimate the within-species variance. At this star tree, $p_A = k/t_{\text{within}}$ is the precision of the species average at the root of the star tree, $p = (t + t_{\text{within}}/k)^{-1}$ is the precision at the root of the parent edge, $w_s = 1/k$ equally weights the k individuals and $\hat{\mu}_Y = \bar{y}$ is the usual species average. For a multivariate response with m different traits in $\mathbf{Y}$, our algorithm can be applied with no modification if the within-species phenotypic covariance between traits $\mathbf{C}_w$ is assumed to be the same as the between-species phylogenetic covariance between traits $\mathbf{C}_b$. In that case, the full covariance matrix for all the Y values stacked as a single vector is $\mathbf{C}_b \otimes \mathbf{V}$ and easy to handle. In most cases however, the assumption of equal within-species and between-species trait correlations is unreasonable. The general model considered by Ives et al. (2007) and Felsenstein (2008) leads to the covariance matrix $\mathbf{C}_b \otimes \mathbf{V} + \mathbf{C}_w \otimes \mathbf{I}$ for the multivariate response $\mathbf{Y}$. This matrix is not 3-point structured, but again our method can be adapted to recursively calculate the key terms involving its inverse and its determinant (Appendix).

Ornstein-Uhlenbeck Models

The covariance matrix for the univariate OU model also has a 3-point structure provided that the underlying tree is ultrametric. Indeed, it is given by

$$V_{ij} = \frac{\sigma^2}{2\alpha}(1 - e^{-2\alpha t_{ij}})e^{-\alpha d_{ij}} \quad \text{or} \quad V_{ij} = \frac{\sigma^2}{2\alpha} e^{-\alpha d_{ij}}$$

depending on whether the value at the root is considered to be a fixed parameter or is integrated out as a random variable with the stationary distribution (Hansen 1997; Ho and Ané 2013). Here t_{ij} is the time of shared ancestry between taxa i and j and d_{ij} is the tree distance between the two taxa. The part involving t_{ij} satisfies the 3-point condition but the part with d_{ij} does not necessarily. It does if the tree is ultrametric because then d_{ij} only depends on the age of the last common ancestor of i and j, as we can write $d_{ij} = 2(T - t_{ij})$ where T is the tree height.

The more general model of adaptive evolution considered by Butler and King (2004) includes variable optima, to allow for different constraints on different lineages. The 3-point structure is retained under this model because changes in the selection optimum do not affect the covariance matrix.

When the tree is not ultrametric, the OU covariance matrix has a generalized 3-point structure. To see this, consider a modified tree where the external branch to taxon i is extended (or shortened) by u_i, in such a way that the modified tree is ultrametric. The modified covariance matrix $\widetilde{\mathbf{V}}$ is then 3-point structured. It is a simple modification of the original covariance matrix: $\mathbf{V}=\mathbf{D}_u\widetilde{\mathbf{V}}\mathbf{D}_u$ where $\mathbf{D}_u$ has entries $e^{\alpha u_i}$ on the diagonal and zeros elsewhere. Therefore, the covariance of the OU model on a nonultrametric tree is indeed generalized 3-point structured. Beaulieu et al. (2012) further extended this adaptive evolution model by allowing different variance rates and/or different selection strengths on different parts of the tree. The covariance matrix under this model also has a generalized 3-point structure. To see this, we consider "epochs" or selection regimes, where epochs are separated by speciation events or changes in the selection regime. Beaulieu et al. (2012) showed that the covariance matrix is then

$$V_{ij}=\exp\left(-\sum_{\gamma=1}^{\kappa(i)}\alpha_{i,\gamma}(s_{i,\gamma}-s_{i,\gamma-1})-\sum_{\gamma=1}^{\kappa(j)}\alpha_{j,\gamma}(s_{j,\gamma}-s_{j,\gamma-1})\right)\times\left(\sum_{\gamma=1}^{\kappa(i,j)}\frac{\sigma^2_{ij,\gamma}}{2\alpha_{ij,\gamma}}\left(e^{2\alpha_{ij,\gamma}s_{ij,\gamma}}-e^{2\alpha_{ij,\gamma}s_{ij,\gamma-1}}\right)\right)$$

where the s values are transition times between successive epochs going forward in time with $s=0$ at the root of the tree, $\kappa(i)$ is the number of epochs on the path from the root to taxon i, $\kappa(i,j)$ is the number of epochs on the path from taxon i to taxon j, and the α and σ^2 values are the selection strengths and variance rates during each epoch. If the selection strength is constant then the first term simplifies to $e^{-\alpha t_{ii}-\alpha t_{jj}}=e^{-2\alpha t_{ij}-\alpha d_{ij}}$. If the variance rate is also constant throughout the tree, then the second term simplifies to $\frac{\sigma^2}{2\alpha}(e^{2\alpha t_{ij}}-1)$, so that the formula above generalizes the one-regime covariance matrix when the state at the root is a fixed parameter of the model. We can rewrite $\mathbf{V}=\mathbf{D}\widetilde{\mathbf{V}}\mathbf{D}$ where $\mathbf{D}$ has diagonal element $\exp\left(-\sum_{\gamma=1}^{\kappa(i)}\alpha_{i,\gamma}(s_{i,\gamma}-s_{i,\gamma-1})\right)$ for taxon i, and

$$\widetilde{V}_{ij}=\sum_{\gamma=1}^{\kappa(i,j)}\frac{\sigma^2_{ij,\gamma}}{2\alpha_{ij,\gamma}}\left(e^{2\alpha_{ij,\gamma}s_{ij,\gamma}}-e^{2\alpha_{ij,\gamma}s_{ij,\gamma-1}}\right),$$

which only depends on the regimes affecting the root to the last common ancestor of i and j. Then $\widetilde{\mathbf{V}}$ is 3-point structured, so $\mathbf{V}$ is generalized 3-point structured and our algorithm can be applied to the OU model with variable selection parameters.

More recently, the multivariate OU model has been used to analyze a response $\mathbf{Y}$ with m traits. For this model, the selection strength $\boldsymbol{\alpha}$ and the variance accumulation rate $\boldsymbol{\sigma}^2$ are now $m\times m$ matrices. The covariance between the m traits at tips i and j is then (King and Butler 2009; Bartoszek et al. 2012)

$$\begin{aligned}\operatorname{cov}(\mathbf{Y}_{i\cdot},\mathbf{Y}_{j\cdot})&=e^{-\boldsymbol{\alpha}t_{ii}}\left(\int_0^{t_{ij}}e^{\boldsymbol{\alpha}v}\boldsymbol{\sigma}^2e^{\boldsymbol{\alpha}v}dv\right)e^{-\boldsymbol{\alpha}t_{jj}}\\&=e^{-\boldsymbol{\alpha}t_{ii}}\left(e^{\boldsymbol{\alpha}t_{ij}}\boldsymbol{\gamma}e^{\boldsymbol{\alpha}t_{ij}}-\boldsymbol{\gamma}\right)e^{-\boldsymbol{\alpha}t_{jj}}\end{aligned}$$

where the asymptotic covariance $\boldsymbol{\gamma}$ can be found by solving $\boldsymbol{\gamma}\boldsymbol{\alpha}'+\boldsymbol{\alpha}\boldsymbol{\gamma}=\boldsymbol{\sigma}^2$. If selection acts independently on all traits, i.e., if $\boldsymbol{\alpha}$ is diagonal (which can be obtained by pre-transforming the traits using the eigenvectors of $\boldsymbol{\alpha}$), then the phylogenetic covariance between taxa is generalized 3-point structured for each pair of traits k and ℓ: $\operatorname{cov}(Y_{i,k},Y_{j,\ell})=\sigma^2_{k\ell}e^{-\alpha_k t_{ii}-\alpha_\ell t_{jj}}(e^{(\alpha_k+\alpha_\ell)t_{ij}}-1)/(\alpha_k+\alpha_\ell)$. If the traits are under different selection strengths α_k, then different 3-point structured matrices apply to different traits and there is no obvious adaptation of our method to speed up the calculations. Under the special case that $\alpha_k=\alpha$ are all equal, then the same phylogenetic signal applies to all traits. The covariance of the full vectorized $\mathbf{Y}$ is then $\boldsymbol{\sigma}^2\otimes\mathbf{V}_\alpha$ with the same phylogenetic covariance $\mathbf{V}_\alpha$ as in the univariate case. So again, terms involving its inverse and determinant can be calculated efficiently in that case.

Applications to Non-Gaussian Models

Our algorithm can be applied to non-Gaussian traits, as long as the computational bottleneck lies in the calculation of quadratic forms involving the inverse of a generalized 3-point structured covariance matrix, and/or its determinant. It is the case for phylogenetic logistic regression, phylogenetic Poisson regression and phylogenetic generalized linear mixed models, as we explain in this section.

Phylogenetic Logistic Regression

Ives and Garland (2010) proposed statistical methods for phylogenetic logistic regression when the trait of interest is binary. For instance, they showed that group size influences antipredator behavior in African antelopes, with solitary species being more likely to hide from predators ($y=0$) than to flee or fight ($y=1$). Their method accounted for phylogenetic correlation and a possible body size effect. In their model, the binary trait evolves according to a two-step process, leading to both phylogenetic correlation and correlation with predictors. In the first step, the binary trait is assumed to evolve along a phylogenetic tree according to a two-state continuous time Markov process with transition rate α and asymptotic probability $\bar{\mu}$ of being in state 1. In the second step, the predictors $\mathbf{X}$ influence the binary trait

values $\mathbf{Y}$ at each tip in such a way that the expectation of $\mathbf{Y}$ satisfies

$$\boldsymbol{\mu} = \frac{\exp(\mathbf{X}\beta)}{1+\exp(\mathbf{X}\beta)}$$

(details in Ives and Garland 2010). Under this model, the covariance of $\mathbf{Y}$ is

$$V_{ij} = \begin{cases} \mu_i(1-\mu_i) & \text{if } i=j \\ s_i s_j \exp(-\alpha d_{ij}) & \text{if } i \neq j \end{cases}$$

where d_{ij} is the tree distance between tips i and j and

$$s_i = \begin{cases} \mu_i\sqrt{1-\bar{\mu}}/\sqrt{\bar{\mu}} & \text{if } \mu_i < \bar{\mu} \\ (1-\mu_i)\sqrt{\bar{\mu}}/\sqrt{1-\bar{\mu}} & \text{if } \mu_i > \bar{\mu}. \end{cases}$$

This covariance matrix is generalized 3-point structured (using the s_i values to form the diagonal matrix $\mathbf{D}$). The parameters are estimated by an iterative procedure. Given the phylogenetic correlation α, $\hat{\beta}$ is updated using quasi-log-likelihood by solving the following equation

$$\mathbf{U}(\hat{\beta}(\alpha)|\alpha) = (\mathbf{AX})'\mathbf{V}(\alpha)^{-1}(\mathbf{Y}-\boldsymbol{\mu}) = 0 \qquad (2)$$

where $\mathbf{A}$ is the diagonal matrix with diagonal terms $\mu_i(1-\mu_i)$. To reduce bias, the penalized quasi-log-likelihood score adds a term involving the Fisher information matrix $\mathbf{I}(\beta)$:

$$U_j^*(\hat{\beta}(\alpha)|\alpha) = U_j(\hat{\beta}(\alpha)|\alpha) + 1/2\,\text{trace}\{\mathbf{I}(\beta)^{-1}[\partial\mathbf{I}(\beta)/\partial\beta_j]\} = 0 \text{ for all } j.$$

where $\mathbf{I}(\beta) = (\mathbf{AX})'\mathbf{V}(\alpha)^{-1}(\mathbf{AX})$. Here again, $\mathbf{U}(\hat{\beta}(\alpha)|\alpha)$ and $\mathbf{I}(\beta)$ take the form of a quadratic term that can be calculated efficiently. Based on the new estimated value $\hat{\beta}$ (and thus new estimates of $\bar{\mu}$ and $\boldsymbol{\mu}$), $\hat{\alpha}$ is updated by minimizing

$$\text{SS}(\hat{\alpha}(\boldsymbol{\mu})|\boldsymbol{\mu}) = -\frac{1}{2}\left(\log|\mathbf{V}(\alpha)| + (\mathbf{Y}-\boldsymbol{\mu})'\mathbf{V}(\alpha)^{-1}(\mathbf{Y}-\boldsymbol{\mu})\right),$$

which contains both the determinant of the covariance matrix and a quadratic term. Finally, the estimated standard error of coefficients $\hat{\beta}$ given $\hat{\alpha}$, $\hat{\sigma}^2(\hat{\beta}|\hat{\alpha}) = \mathbf{I}(\beta)^{-1}$, is again evaluated using a quadratic product. Therefore, all these steps can be carried out efficiently using our algorithm.

Phylogenetic Logistic Mixed Models

Phylogenetic mixed models use a regression formula that involves a random effect with phylogenetic correlation. For example, Ives and Helmus (2011) investigate the presence/absence of n species at m sites to determine whether closely related species are more likely to co-occur at the same sites, and to quantify the phylogenetic signal in species co-occurrence. The observations $\mathbf{Y}$ form a long nm vector of all binary outcomes, for each species at each site. The authors consider the following phylogenetic logistic mixed model with binomial distribution $\text{IP}\{Y_i = 1\} = \mu_i$ and

$$\log\frac{\mu_i}{1-\mu_i} = \alpha_{\text{spp}[i]} + b_i + c_{\text{site}[i]}$$

$$\mathbf{b} \sim \mathcal{N}(0, \mathbf{I}_m \otimes \sigma^2_{\text{spp}}\Sigma_{\text{spp}}), \quad \mathbf{c} \sim \mathcal{N}(0, \sigma^2_{\text{site}}\mathbf{I}_m)$$

where $\alpha_{\text{spp}[i]}$ is an unobserved (fixed) species effect and the random effects b_i carry the phylogenetic correlation. The phylogenetic covariance matrix among species Σ_{spp} is derived either from a BM model or an OU model and hence is generalized 3-point structured. The full covariance matrix of the random effects is then

$$\mathbf{C} = \mathbf{I}_m \otimes (\sigma^2_{\text{spp}}\Sigma_{\text{spp}} + \sigma^2_{\text{site}}\mathbf{I}_n).$$

To estimate the parameters using approximate maximum likelihood, Ives and Helmus (2011) iteratively update $\boldsymbol{\alpha}$, $\mathbf{b}$, and $\mathbf{c}$ given the current estimates of σ^2_{spp} and σ^2_{site} using penalized quasi-likelihood (equations (B.3) and (B.4) in Ives and Helmus 2011). This step requires the calculation of quadratic terms, where the working covariance matrix is $\mathbf{V} = \mathbf{C} + \mathbf{W}^{-1}$ and $\mathbf{W}$ is the $nm \times nm$ diagonal matrix with diagonal elements $\mu_i(1-\mu_i)$. Our algorithm is applicable to $\mathbf{V}$ because it is block diagonal, and each block has a generalized 3-point structure. Second, the estimates of σ^2_{spp} and σ^2_{site} are updated based on the new estimated values of $\boldsymbol{\alpha}$, $\mathbf{b}$ and $\mathbf{c}$ using maximum likelihood (ML) or restricted maximum likelihood (REML). For this, the log-likelihood function being optimized (eq. (B.5) in Ives and Helmus 2011) involves both the determinant of $\mathbf{V}$ and quadratic terms, which can again be calculated in linear time using our algorithm. This likelihood criterion is further used to build confidence intervals for the variances σ^2_{spp} and σ^2_{site}, one further step that can benefit from computational efficiency.

Phylogenetic Poisson Regression

Lavergne et al. (2013) used phylogenetic Poisson regression to investigate whether the local diversification rate of angiosperm families in the western Mediterranean basin was affected by the family's migration rate across the strait of Gibraltar. To do so, the number $\mathbf{Y}$ of local endemic species per family was explained by the family migration rate $\mathbf{X}$ using the phylogenetic GEE method (Paradis and Claude 2002), assuming that the counts $\mathbf{Y}$ follow a Poisson distribution with mean $\boldsymbol{\mu} = \exp(\mathbf{X}\beta)$ and covariance $\mathbf{V} = \phi\mathbf{A}^{1/2}\mathbf{C}\mathbf{A}^{1/2}$ across species. Here ϕ is a dispersion parameter, $\mathbf{A}$ is a diagonal matrix with entries $\boldsymbol{\mu}$ on the diagonal, and $\mathbf{C}$ is a phylogenetic correlation matrix. The coefficients β can be estimated from the generalized estimating equations (similar to equation (2) for logistic regression)

$$(\mathbf{AX})'\mathbf{V}^{-1}(\mathbf{Y}-\boldsymbol{\mu}) = 0.$$

These equations can be solved through an iterative process in which β is updated at each step t by:

$$\beta_{t+1} = \beta_t - \mathbf{I}(\beta)^{-1}[(\mathbf{AX})'\mathbf{V}^{-1}(\mathbf{Y}-\boldsymbol{\mu})]$$

where $\mathbf{I}(\beta) = (\mathbf{AX})'\mathbf{V}(\alpha)^{-1}(\mathbf{AX})$. Finally, the standard error of $\hat{\beta}$ can be estimated from $\hat{\sigma}^2(\hat{\beta}) = \mathbf{I}(\beta)^{-1}$. Because the phylogenetic correlation $\mathbf{C}$ is derived from a BM model in Paradis and Claude (2002), $\mathbf{V}$ is generalized 3-point structured and the quadratic terms above can be calculated efficiently by our algorithm.

Hadfield and Nakagawa (2010) described a somewhat different model (in the class of phylogenetic generalized linear mixed models) where the response trait follows a Poisson distribution with mean $\boldsymbol{\mu} = \exp(\mathbf{X}\beta + \mathbf{b} + \mathbf{e})$. Here taxon i has random effects b_i and e_i, where $\mathbf{b}$ is assumed to follow a Brownian motion along the tree and $\mathbf{e}$ is an independent Gaussian noise. This model is equivalent to Pagel's lambda model for the random effects $\mathbf{b}+\mathbf{e}$. Hadfield (2010) uses MCMC for this model in a Bayesian framework. To avoid numerical matrix inversion, the state space is increased to track all random effects at the $n-2$ non-root internal nodes. Our algorithm could also be used for this model, except that Hadfield (2010) makes further use of the sparsity of $\mathbf{V}^{-1}$ when this matrix is expanded to include internal nodes, but $\mathbf{V}^{-1}$ is no longer sparse when restricted to the tips of the tree. Therefore our algorithm might not bring much computational gain in this MCMC implementation. However, computational benefits would be substantial in a frequentist approach, using approximate ML as in Ives and Helmus (2011). A more general model than that considered by Hadfield (2010) would allow the random effect to follow any evolution trait model, such as OU models or variable rate BM models. For a Poisson-distributed trait $\mathbf{Y}$, the mean μ_i at taxon i is assumed to satisfy

$$\boldsymbol{\mu} = \exp(\mathbf{X}\beta + \mathbf{b})$$

where $\mathbf{b} \sim \mathcal{N}(0, \mathbf{V})$ and $\mathbf{V}$ is any generalized 3-point structured phylogenetic covariance matrix. Penalized quasi-likelihood estimation and REML estimation can thus be combined iteratively, as done in Ives and Helmus (2011). For a Poisson-distributed response, the working covariance matrix is $\mathbf{V}+\mathbf{W}^{-1}$ with $\mathbf{W}$ diagonal and $W_{ii} = \mu_i$ (Liang 2007). This matrix can be handled by the algorithm presented here because it is still generalized 3-point structured.

Further Applications

The components we derive are the key building blocks for other tools, such as fitting models with REML, reduced major axis (RMA) regression (Ives et al. 2007; Hansen and Bartoszek 2012), phylogenetic PCA or phylogenetic discriminant analysis as in Motani and Schmitz (2011). For instance, the PCA approach by Revell (2009) is based on the phylogenetically corrected means $\hat{\mu}_X$ and the evolutionary covariance among traits $\mathbf{C} = (\mathbf{X} - \mathbf{1}'\hat{\mu}_X)'\mathbf{V}^{-1}(\mathbf{X} - \mathbf{1}'\hat{\mu}_X)$. Here $\mathbf{X}$ is an $n \times m$ matrix for multivariate data across n taxa and m traits. If the underlying phylogenetic covariance matrix is generalized 3-point structured, then the algorithm presented here can be used in these applications as well.

Prediction intervals were used in Garland and Ives (2000) to determine if certain species deviate significantly from an allometric prediction. More recently, there has been renewed interest in phylogenetic prediction (e.g., Cai et al. 2011) to predict traits of interest across wild species to help plant breeding of cultivated species. Let s_0 be a new taxon that can be placed on the current tree. Given the phylogenetic covariance parameter θ, the following prediction of the response trait $Y(s_0)$ at the new taxon is the best linear unbiased prediction (Goldberger 1962):

$$\hat{Y}(s_0) = \mathbf{c}_\theta'\mathbf{V}_\theta^{-1}\mathbf{Y} + (\mathbf{x} - \mathbf{X}'\mathbf{V}_\theta^{-1}\mathbf{c}_\theta)'\hat{\beta}$$

where $\mathbf{x}$ are the known covariate(s) at the new phylogenetic location, $\mathbf{c}_\theta$ is the vector of covariances between the new taxon and the sampled taxa, $\hat{\beta}$ contains the estimated regression coefficients from the sampled taxa with observed response data $\mathbf{Y}$ and covariates $\mathbf{X}$, and $\mathbf{V}_\theta$ is the covariance matrix among the sampled taxa as before. This is also called "universal kriging" in spatial statistics (Cressie 1993). Prediction intervals can then be obtained based on the prediction variance:

$$\mathrm{IE}(\hat{Y}(s_0) - Y(s_0))^2 = V_\theta(s_0, s_0) - \mathbf{c}_\theta'\mathbf{V}_\theta^{-1}\mathbf{c}_\theta + (\mathbf{x} - \mathbf{X}'\mathbf{V}_\theta^{-1}\mathbf{c}_\theta)'(\mathbf{X}'\mathbf{V}_\theta^{-1}\mathbf{X})^{-1}(\mathbf{x} - \mathbf{X}'\mathbf{V}_\theta^{-1}\mathbf{c}_\theta).$$

All these quantities can be calculated easily with our algorithm for quadratic products, without the need to invert the large matrix $\mathbf{V}_\theta$. These quantities were already used in Garland and Ives (2000) for the case of simple linear regression, where the tree needed to be re-rooted for each new taxonomic position where prediction was desired. Our approach is general to any regression model, fast enough for very large trees, and does not require any re-rooting.

Discussion

Speed-up for Large Scale and Iterative Analyses

The utility of this algorithm comes in when the likelihood of trait data or other score functions need to be evaluated on large data sets or for many iterations. The size of data sets continues to increase as phylogenetic trees continue to become more and more accurate and more taxon-rich. For instance, the goal of the iPlant Tree of Life project (Goff et al. 2011) is to scale up tree inference and post-tree analyses, to hundreds of thousands of taxa. Trait analyses on such large trees are possible with the proposed algorithm. Also, the important sample size here is the number of terminals with trait data, which can be many times larger than the number of species

when individual organisms are used instead of species averages (e.g., Revell and Reynolds 2012).

The need for large numbers of iterations comes from various sources. First, iterations are needed to optimize model parameters and maximize the likelihood, or to search for the area of high posterior probability in MCMC algorithms. Such iterations are needed to optimize Pagel's λ or κ, the rate of variance decrease in an early-burst model, the selection strength α in an OU models, or the estimated ratio of within-species variance versus between-species variance. The number of iterations or MCMC generations that are needed can be substantial as models become very complex, such as with an unbounded number of rate shifts at unknown locations in the tree (Eastman et al. 2011) or an unknown number of selective regime changes at unknown locations Beaulieu et al. (2012). Hadfield (2010) noted that each MCMC step can be sped up considerably by tracking all random effects at nonroot internal nodes, at the expense of more iterations for convergence. Second, many iterations are used when tree uncertainty is taken into account, either in an integrated Bayesian framework (as in BayesTraits, Pagel 1999), or when repeating the analysis on a large sample of phylogenetic trees. Finally, bootstrap procedures are now increasingly used to test hypotheses or assess uncertainty in the resulting parameters (Garland et al. 1993; Freckleton and Harvey 2006). For instance, Boettiger et al. (2012) advocate the use of parametric bootstrap where data are repeatedly sampled from a given trait evolution model and analyzed like the original data, to perform reliable model selection. Bootstrap methods are powerful but computationally heavy, and have much to gain from a fast likelihood calculation algorithm. Taken together, the number of iterations required to estimate the parameters of the trait evolution model, to integrate over tree uncertainty and to conduct bootstrap procedures can quickly increase the overall computing load.

Model Adequacy on very Large Trees

At very large taxonomic scales, traditional homogeneous models of trait evolution are often likely to be simplistic. To account for changes in modes of trait evolution, complex models are increasingly needed to adequately analyze very large data sets. To do so, fast algorithms become even more important, to deal both with large trees and models with large numbers of parameters. The computational gains presented here could foster the development of complex models appropriate for large taxonomic scales, which may require special estimation and model selection procedures.

Funding

This work was funded in part by NSF award DMS-1106483 and DEB-0949121.

Acknowledgments

We thank Luke Harmon for pointing us to the alternative methods published recently by Robert Freckleton and Richard FitzJohn. We also thank Joe Felsenstein, Robert Freckleton and two other anonymous reviewers for helpful suggestions. C.A. is grateful for the NIMBioS working group on species delimitation organized by David Weisrock and Ruriko Yoshida, which motivated the fast approach to likelihood calculations.

Appendix

Algorithm for Multiple Traits with Different Phylogenetic Signals

We focus here on the model proposed by Freckleton (2012), where each trait is assumed to follow Pagel's λ model with phylogenetic signal λ_k for trait k, using $\mathrm{cov}(Y_{i,k}, Y_{j,\ell}) = C_{k\ell} V_{ij} (\sqrt{\lambda_k \lambda_\ell})$ for traits k, ℓ and taxa i, j (Freckleton 2012, Supporting Information). Of interest is $\mathbf{C}$, the evolutionary covariance between the m traits. We will use the fact that $\mathbf{V}(\lambda) = \lambda\mathbf{V} + (1-\lambda)\mathbf{T}$ where $\mathbf{V}$ is the BM covariance matrix with $V_{ij} = t_{ij}$ and $\mathbf{T}$ is the diagonal matrix with diagonal elements t_{ii}. For this model, the full covariance matrix of all the Y values stacked as a single vector is $\mathbf{W} = \mathbf{C}_1 \otimes \mathbf{V} + \mathbf{C}_2 \otimes \mathbf{T}$ where $\mathbf{C}_1 = \mathrm{diag}(\sqrt{\lambda_k})\mathbf{C}\mathrm{diag}(\sqrt{\lambda_k})$ and $\mathbf{C}_2 = \mathbf{C} - \mathbf{C}_1$. Since $\mathbf{V}$ is 3-point structured, we have from (1) that $\mathbf{V} = t\mathbf{1}\mathbf{1}' + \mathbf{A}$ with $\mathbf{A}$ block diagonal, each diagonal block $\mathbf{V}_s$ corresponding to a subtree from the root. We can then write $\mathbf{W} = \mathbf{B} + t\mathbf{C}_1 \otimes \mathbf{1}.\mathbf{I} \otimes \mathbf{1}'$ where $\mathbf{B} = \mathbf{C}_1 \otimes \mathbf{A} + \mathbf{C}_2 \otimes \mathbf{T}$. Up to a permutation of its columns and rows to bring together all traits from the taxa belonging to each subtree s, $\mathbf{B}$ is also block diagonal, with each diagonal block corresponding to a subtree: $\mathbf{W}_s = \mathbf{C}_1 \otimes \mathbf{V}_s + \mathbf{C}_2 \otimes \mathbf{T}_s$. The algorithm can then be adapted to recursively calculate the following quantities: $\mathbf{p} = \mathbf{I} \otimes \mathbf{1}'.\mathbf{W}^{-1}.\mathbf{C}_1 \otimes \mathbf{1}$, $\hat{\boldsymbol{\mu}}_Y = \mathbf{p}^{-1}.\mathbf{I} \otimes \mathbf{1}'.\mathbf{W}^{-1}\mathbf{Y}$, $\tilde{\boldsymbol{\mu}}'_X = \mathbf{X}'\mathbf{W}^{-1}.\mathbf{C}_1 \otimes \mathbf{1}.\mathbf{p}^{-1}$ and finally both $\mathbf{Q} = \mathbf{X}'\mathbf{W}^{-1}\mathbf{Y}$ and $\log|\mathbf{W}|$. The algorithm is adapted as follows.

1. Initialization: for a tree with a single tip i and parent edge of length t, set $\mathbf{p} = (t_{ii}\mathbf{C}\mathbf{C}_1^{-1} - (t_{ii} - t)\mathbf{I})^{-1}$, $\hat{\boldsymbol{\mu}}_Y = \mathbf{C}_1^{-1}\mathbf{Y}$, $\tilde{\boldsymbol{\mu}}'_X = \mathbf{X}'$, $\mathbf{Q} = \tilde{\boldsymbol{\mu}}'_X\mathbf{p}\mathbf{Y}$, and $\log|\mathbf{W}| = \log|t_{ii}\mathbf{C} - (t_{ii} - t)\mathbf{C}_1|$.
2. For a tree with two or more tips and root edge length t, let $\mathbf{p}_s$, $\hat{\boldsymbol{\mu}}_{Y,s}$, $\tilde{\boldsymbol{\mu}}'_{X,s}$ and $\mathbf{Q}_s$ the values obtained recursively from each subtree stemming from the root. Define $\mathbf{p}_A = \sum \mathbf{p}_s$ and weights $\mathbf{w}_s = \mathbf{p}_A^{-1}\mathbf{p}_s$ and $\tilde{\mathbf{w}}_s = \mathbf{p}_s\mathbf{p}_A^{-1}$. Then $\mathbf{p} = \mathbf{p}_A(\mathbf{I} + t\mathbf{p}_A)^{-1}$, $\hat{\boldsymbol{\mu}}_Y = \sum_s \mathbf{w}_s\hat{\boldsymbol{\mu}}_{Y,s}$, $\tilde{\boldsymbol{\mu}}'_X = \sum_s \tilde{\boldsymbol{\mu}}'_{X,s}\tilde{\mathbf{w}}_s$, $\mathbf{Q} = \sum_s \mathbf{Q}_s - \tilde{\boldsymbol{\mu}}'_X t\mathbf{p}_A^2(\mathbf{I} + t\mathbf{p}_A)^{-1}\hat{\boldsymbol{\mu}}_Y$ and $\log|\mathbf{W}| = \sum_s \log|\mathbf{W}_s| + \log|\mathbf{I} + t\mathbf{p}_A|$.
3. At the root of the full tree: return $\log|\mathbf{W}|$ and $\mathbf{Q}$ then stop.

Algorithm for Multiple Traits with Different within and between-species Covariances

If the phylogenetic covariance is shared between all traits, but if the within-species phenotypic covariance $\mathbf{C}_w$ is different from the between-species phylogenetic covariance $\mathbf{C}_b$ between traits, the full covariance matrix of all response values (stacked in a single vector) is $\mathbf{W} = \mathbf{C}_b \otimes \mathbf{V} + \mathbf{C}_w \otimes \mathbf{I}$, where $\mathbf{V}$ is the phylogenetic covariance Felsenstein (2008). The within-species variance $\mathbf{C}_w$ may also include measurement error, with the model above assuming that this error is independent across species. If $\mathbf{V}$ is 3-point structured, we can again use (1) to write $\mathbf{V} = t\mathbf{1}\mathbf{1}' + \mathbf{A}$ with $\mathbf{A}$ block diagonal, each diagonal block $\mathbf{V}_s$ corresponding to subtree s. Then we get $\mathbf{W} = \mathbf{B} + t\mathbf{C}_b \otimes \mathbf{1}.\mathbf{I} \otimes \mathbf{1}'$ where $\mathbf{B} = \mathbf{C}_b \otimes \mathbf{A} + \mathbf{C}_w \otimes \mathbf{I}$. Again, up to a permutation of its columns and rows to cluster all data from each subtree, $\mathbf{B}$ is also block diagonal with each diagonal block s corresponding to subtree s: $\mathbf{W}_s = \mathbf{C}_b \otimes \mathbf{V}_s + \mathbf{C}_w \otimes \mathbf{I}$. Therefore, we can recursively calculate the following quantities: $\mathbf{p} = \mathbf{I} \otimes \mathbf{1}'.\mathbf{W}^{-1}.\mathbf{C}_b \otimes \mathbf{1}$, $\hat{\boldsymbol{\mu}}_Y = \mathbf{p}^{-1}.\mathbf{I} \otimes \mathbf{1}'.\mathbf{W}^{-1}\mathbf{Y}$, $\tilde{\boldsymbol{\mu}}'_X = \mathbf{X}'\mathbf{W}^{-1}.\mathbf{C}_b \otimes \mathbf{1}.\mathbf{p}^{-1}$, and as before $\mathbf{Q} = \mathbf{X}'\mathbf{W}^{-1}\mathbf{Y}$ and $\log|\mathbf{W}|$. The same recursion (step 2) applies here as in the previous section, with a different initialization step:

1. For a tree with a single tip (here an individual within a species), set $\mathbf{p} = \mathbf{C}_w^{-1}\mathbf{C}_b$, $\hat{\boldsymbol{\mu}}_Y = \mathbf{C}_b^{-1}\mathbf{Y}$, $\tilde{\boldsymbol{\mu}}'_X = \mathbf{X}'$, $\mathbf{Q} = \mathbf{X}'\mathbf{C}_w^{-1}\mathbf{Y}$ and $\log|\mathbf{V}| = \log|\mathbf{C}_w|$.

References

Bartoszek K., Pienaar J., Mostad P., Andersson S., Hansen T.F. 2012. A phylogenetic comparative method for studying multivariate adaptation. J. Theor. Biolog. 314:204–215.

Beaulieu J.M., Jhwueng D.-C., Boettiger C., O'Meara B.C. 2012. Modeling stabilizing selection: Expanding the Ornstein-Ühlenbeck model of adaptive evolution. Evolution 66:2369–2383.

Bininda-Emonds O., Cardillo M., Jones K.E., MacPhee R.D.E., Beck R.M.D., Grenyer R., Price S.A., Vos R.A., Gittleman J.L., Purvis A. 2007. The delayed rise of present-day mammals. Nature 446:507–512.

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

Boettiger C., Coop G., Ralph P. 2012. Is your phylogeny informative? measuring the power of comparative methods. Evolution 66:2240–2251.

Brawand D., Soumillon M., Necsulea A., Julien P., Csardi G., Harrigan P., Weier M., Liechti A., Aximu-Petri A., Kircher M., Albert F.W., Zeller U., Khaitovich P., Grutzner F., Bergmann S., Nielsen R., Pääbo S., Kaessmann H. 2011. The evolution of gene expression levels in mammalian organs. Nature 478:343–348.

Butler M.A., King A.A. 2004. Phylogenetic comparative analysis: A modeling approach for adaptive evolution. Am. Nat. 164:683–695.

Cai X.K., Spooner D.M., Jansky S.H. 2011. A test of taxonomic and biogeographic predictivity: Resistance to potato virus Y in wild relatives of the cultivated potato. Phytopathology 101:1074–1080.

Cooper N., Purvis A. 2010. Body size evolution in mammals: Complexity in tempo and mode. Am. Nat. 175:727–738.

Cressie N.A.C. 1993. *Statistics for Spatial Data*. New rev. ed. New York: Wiley-Interscience.

Dempster A.P., Laird N.M., Rubin D.B. 1977. Maximum likelihood from incomplete data via the EM algorithm. J. R. Stat. Soc., Series B 39: 1–38.

Eastman J.M., Alfaro M.E., Joyce P., Hipp A.L., Harmon L.J. 2011. A novel comparative method for identifying shifts in the rate of character evolution on trees. Evolution 65:3578–3589.

Felsenstein J. 1973. Maximum-likelihood estimation of evolutionary trees from continuous characters. Am. J. Hum. Genet. 25:471–492.

Felsenstein J. 2008. Comparative methods with sampling error and withinspecies variation: Contrasts revisited and revised. Am. Nat. 171:713–725.

FitzJohn R. 2012. Diversitree: comparative phylogenetic analyses of diversification in R. Methods Ecol. Evol. 3:1084–1092.

Freckleton R. 2012. Fast likelihood calculations for comparative analyses. Methods Ecol. Evol. 3:940–947.

Freckleton R.P., Harvey P.H. 2006. Detecting non-brownian trait evolution in adaptive radiations. PLoS Biol. 4:e373.

Garland T. Jr., Dickerman A.W., Janis C.M., Jones J.A. 1993. Phylogenetic analysis of covariance by computer simulation. Syst. Biol. 42:265–292.

Garland T. Jr., Ives A.R. 2000. Using the past to predict the present: Confidence intervals for regression equations in phylogenetic comparative methods. Am. Nat. 155:346–364.

Goff S.A., Vaughn M., McKay S., Lyons E., Stapleton A.E., Gessler D., Matasci N., Wang L., Hanlon M., Lenards A., Muir A., Merchant N., Lowry S., Mock S., Helmke M., Kubach A., Narro M., Hopkins N., Micklos D., Hilgert U., Gonzales M., Jordan C., Skidmore E., Dooley R., Cazes J., McLay R., Lu Z., Pasternak S., Koesterke L., Piel W.H., Grene R., Noutsos C., Gendler K., Feng X., Tang C., Lent M., Kim S.-J., Kvilekval K., Manjunath B.S., Tannen V., Stamatakis A., Sanderson M., Welch S., Cranston K., Soltis P., Soltis D., O'Meara B., An C., Brutnell T., Kleibenstein D.J., White J.W., Leebens-Mack J., Donoghue M.J., Spalding E.P., Vision T.J., Myers C.R., Lowenthal D., Enquist B.J., Boyle B., Akoglu A., Andrews G., Ram S., Ware D., Stein L., Stanzione D. 2011. The iplant collaborative: Cyberinfrastructure for plant biology. Front. Plant Sci. 2.

Goldberger A.S. 1962. Best linear unbiased prediction in the generalized linear regression model. J. Am. Stat. Assoc. 57: 369–375.

Grafen A. 1989. The phylogenetic regression. Phil. Trans. R. Soc. London. Series B, Biol. Sci. 326:119–157.

Hadfield J. 2010. MCMC methods for multi-response generalized linear mixed models: the MCMCglmm R package. J. Stat. Software 33:1–22.

Hadfield J.D., Nakagawa S. 2010. General quantitative genetic methods for comparative biology: phylogenies, taxonomies and multi-trait models for continuous and categorical characters. J. Evol. Biol. 23:494–508.

Hager W. 1989. Updating the inverse of a matrix. SIAM Rev. 31:221–239.

Hansen T.F. 1997. Stabilizing selection and the comparative analysis of adaptation. Evolution 51:1341–1351.

Hansen T.F., Bartoszek K. 2012. Interpreting the evolutionary regression: The interplay between observational and biological errors in phylogenetic comparative studies. Syst. Biol. 61:413–425.

Harmon L.J., Losos J.B., Jonathan Davies T., Gillespie R.G., Gittleman J.L., Bryan Jennings W., Kozak K.H., McPeek M.A., Moreno-Roark F., Near T.J., Purvis A., Ricklefs R.E., Schluter D., Schulte J.A. II, Seehausen O., Sidlauskas B.L., Torres-Carvajal O., Weir J.T., Mooers A. 2010. Early bursts of body size and shape evolution are rare in comparative data. Evolution 64:2385–2396.

Ho L.S.T., Ané C. 2013. Asymptotic theory with hierarchical autocorrelation: Ornstein-uhlenbeck tree models. Annals Stat. 41:957–981.

Ives A., Helmus M. 2011. Generalized linear mixed models for phylogenetic analyses of community structure. Ecol. Monographs 81:511–525.

Ives A.R., Garland T. Jr. 2010. Phylogenetic logistic regression for binary dependent variables. Syst. Biol. 59:9–26.

Ives A.R., Midford P.E., Garland T. Jr. 2007. Within-species variation and measurement error in phylogenetic comparative methods. Syst. Biol. 56:252–270.

Jetz W., Thomas G., Joy J., Hartmann K., Mooers A. 2012. The global diversity of birds in space and time. Nature 491:444–448.

King A.A., Butler M.A. 2009. *ouch: Ornstein-Uhlenbeck models for phylogenetic comparative hypotheses*. (R package), http://ouch.r-forge.r-project.org.
Lavergne S., Hampe A., Arroyo J. 2013. In and out of africa: how did the strait of gibraltar affect plant species migration and local diversification? J. Biogeography 40:24–36.
Lawson C.L., Hanson R.J. 1995. *Solving Least Squares Problems*. Classics in Applied Mathematics Society for Industrial and Applied Mathematics.
Liang J. 2007. *Linear and Generalized Linear Mixed Models and their Applications*. Springer Series in Statistics. New York: Springer.
Lynch M. 1991. Methods for the analysis of comparative data in evolutionary biology. Evolution 45:1065–1080.
Martins E.P., Hansen T.F. 1997. Phylogenies and the comparative method: A general approach to incorporating phylogenetic information into the analysis of interspecific data. Am. Nat. 149:646–667.
Motani R., Schmitz L. 2011. Phylogenetic versus functional signals in the evolution of form-function relationships in terrestrial vision. Evolution 65:2245–2257.
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. 1997. Inferring evolutionary processes from phylogenies. Zoologica Scripta 26:331–348.
Pagel M. 1999. Inferring the historical patterns of biological evolution. Nature 401:877–884.
Paradis E., Claude J. 2002. Analysis of comparative data using generalized estimating equations. J. Theor. Biol. 218:175–185.
Revell L.J. 2009. Size-correction and principal components for interspecific comparative studies. Evolution 63:3258–3268.
Revell L.J., Harmon L.J. 2008. Testing quantitative genetic hypotheses about the evolutionary rate matrix for continuous characters. Evol. Ecol. Res. 10:311–331.
Revell L.J., Mahler D.L., Peres-Neto P.R., Redelings B.D. 2012. A new phylogenetic method for identifying exceptional phenotypic diversification. Evolution 66:135–146.
Revell L.J., Reynolds R.G. 2012. A new Bayesian method for fitting evolutionary models to comparative data with intraspecific variation. Evolution 66:2697–2707.
Semple C., Steel M. 2003. *Phylogenetics*. No. 24 in Oxford Lecture Series in Mathematics and its Applications New York: Oxford University Press
Smith S.A., Donoghue M.J. 2008. Rates of molecular evolution are linked to life history in flowering plants. Science 322:86–89.
Thomas G.H., Freckleton R.P., Székely T. 2006. Comparative analyses of the influence of developmental mode on phenotypic diversification rates in shorebirds. Proc. R. Soc. B: Biol. Sci. 273:1619–1624.
Venditti C., Meade A., Pagel M. 2011. Multiple routes to mammalian diversity. Nature 479:393–396.