Fast likelihood calculations for comparative analyses

Robert P. Freckleton*

School of Animal & Plant Sciences, University of Sheffield, Sheffield S10 2TN, UK

Summary

1. Modern comparative approaches use model-based methods to describe evolutionary processes. Generalised least squares calculations lie at the heart of many methods; however, they can be computationally intensive. This is because it is necessary to form a variance–covariance matrix, then to calculate the inverse and determinant of this.

2. Based on an algorithm provided by Felsenstein (American Journal of Human Genetics, 1973, 25, 471), I show how to perform comparative calculations that avoid these computational steps.

3. I apply the method to several problems in comparative analysis, including calculating likelihoods, estimating Pagel's λ for one or several traits and fitting linear models.

4. R code is provided, which implements the algorithm described. Examples are included to demonstrate the computational gains possible for several commonly used comparative methods.

Key-words: Brownian motion, evolutionary model, generalised least squares, phylogenetic contrasts

Introduction

The comparative method is the basis for a range of analyses in ecology and evolutionary biology. The rationale for the approach is that as species evolve, traits adapt in response to changes in the biotic and abiotic environment, with the consequence that current distributions of trait values reflect the processes that shaped them in the past. Given the information on species' traits, together with phylogenetic information, it should be possible to reconstruct the evolutionary history of a group.

The most common approach to comparative analysis revolves around a group of closely related statistical modelling approaches (e.g. Felsenstein 1985; Grafen 1989; Lynch 1991; Martins & Hansesn 1997; Pagel 1997; Garland, Midford & Ives 1999). In broad terms, these are equivalents of the linear modelling methods (GLM, regression, ANOVA, etc.) that are routinely used throughout biology, but accounting for the influence of phylogeny. The approaches rely on a common approach to modelling the interdependence between species resulting from common evolutionary history. In broad terms, the method relies on fitting multivariate normal distributions to describe the interdependence between species. In evolutionary terms, this can be justified by assuming that traits evolve according to a Brownian model of trait evolution (Felsenstein 1985; Harvey & Pagel 1991).

This approach to modelling trait variation is useful because it is so flexible: in addition to the variety of approaches that

have been developed directly based on this model for trait variation, the method has been extended in various ways. Pagel (1997, 1999) outlined transformations that could be applied to a phylogeny and estimated as part of the model. These include parameters allowing for increases or decreases in the rate of evolution (κ and δ) or for variable levels of phylogenetic dependence (λ; Freckleton, Harvey & Pagel 2002). Hansen (1997) showed that phylogenetic constraint could be measured by a further transformation (α) that generates an Ornstein–Uhlenbeck model (Hansen, Pinaar & Orzack 2008). Thomas, Freckleton & Szekely (2006) outlined a transformation (θ) to allow for trait-dependent rates of evolution (O'Meara *et al.* 2006). It is possible to combine different models for nonindependence, and Freckleton & Jetz (2009) suggested how spatial and phylogenetic models could be used simultaneously.

One of the difficulties in applying these approaches is that one step in the analysis can be computationally demanding. As outlined in detail below, to apply the approach, a cophenetic matrix, **V**, is computed, which comprises the shared path lengths of all *n* species in the phylogeny. For *n* species, **V** has dimensions $n \times n$, that is, its size grows as the square of the number of species in the analysis. This has two consequences for the computation of the PGLS model. First, the matrix has to be generated in the first place. This requires allocating enough memory to hold all of the entries of **V** and then initiating one traversal (i.e. successively visiting all nodes) of the phylogeny per pair of species sharing an ancestor to measure the shared path lengths. Second, **V** has to be inverted at one point in the analysis. This is a numerical step, and the computational overhead can be considerable in addition to the burden of computing **V**.

*Corresponding author. E-mail: r.freckleton@sheffield.ac.uk

Figure 1 illustrates that these computational burdens can be large and increase nonlinearly with the size of the phylogeny. The time taken for the formation of **V** scales approximately to the power 2·5, with increasing the size of the phylogeny, whilst inverting the variance–covariance matrix scales to the power 2·87. The lower bound for the exponent of the time taken to invert **V** is probably 2, as there are n^2 entries in a matrix for n species. However, the fastest current algorithm for matrix inversion has an exponent of 2·376 (Robinson 2005). Irrespective of processor power, this scaling sets an effective limit to the size of phylogeny that can be analysed using this method, probably of the order of 1×10^5. A further issue is that memory requirements also are demanding: a variance matrix for a phylogeny of 1×10^5 species will require 1×10^{10} entries to be stored (requiring *c.* 80 gigabytes of memory for double data types). Although cophenetic matrices for phylogenies are frequently sparse (many entries are zero) and efficient methods tailored to sparse matrices could be brought to bear (e.g. Hadfield & Nakagawa 2010), there are undoubtedly considerable computational costs to be borne. These problems are not unique to the generalised least squares approach. For example, phylogenetic eigenvector regression (PVR; Diniz-Filho, de Sant Ana & Bini 1998) requires that a distance matrix is computed, from which eigenvectors are extracted. These calculations require
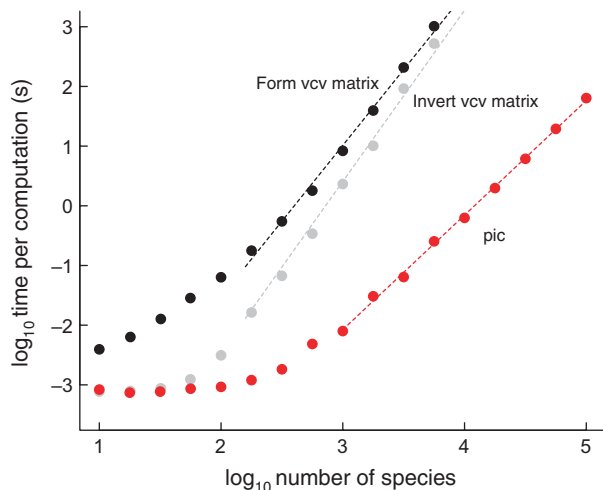
approximately the same time and memory as computing, storing and inverting **V**.

Modern comparative analyses can require considerable numbers of computations. For example, estimating parameters modelling different modes of evolution (e.g. the transformations described above) requires that for each value of the parameter examined, **V** is calculated from the matrix obtained from the phylogeny and solved for each parameter value. Because the parameters have to be estimated iteratively, a large number of values may have to be explored. In analyses in which phylogenetic uncertainty is analysed, **V** has to be computed individually for each candidate phylogeny. In Bayesian MCMC analyses, this number might be in the order of millions. Finally, simulations require large numbers of iterations across wide ranges of parameters, and slow computation can limit the range of parameters that can be explored.

The problems of computational constraints in analyses of this sort have long been recognised. Felsenstein (1973) struggled with the problem of calculating the likelihood of a set of data on a tree with a given set of branch lengths. This likelihood depends on a matrix, **V,** but given the computational constraints at that time, direct inversion of **V** was computationally impractical for even moderately sized problems. To get around this, Felsenstein (1973) presented a method of calculating likelihoods that did not require **V** or its inverse. Although the link has not been greatly stressed (Felsenstein 2004; Freckleton & Harvey 2006; Freckleton & Jetz 2009; Thomas & Freckelton 2011), this approach is essentially the same as the method of contrasts, which is the most widely used comparative method (Felsenstein 1985).

Here, I outline how the approach suggested in Felsenstein (1973), which has been largely overlooked in the comparative literature, can be used to greatly enhance the speed of computation in comparative analysis. Most codes and packages that are currently available use the slower matrix inversion method (Freckleton & Harvey 2006; Freckleton & Jetz 2009; Thomas & Freckelton 2011). We first outline the method for calculating the maximum likelihood estimates of parameters of a single trait. I then show how this can be generalised to calculate the likelihood for arbitrary parameters. I finally illustrate how this can be extended to problems of correlated evolution and PGLS. R code is supplied to demonstrate the computationally efficient methods.

Fig. 1. Computational times for calculations of log-likelihoods in generalised least squares. The calculation of a single likelihood via GLS requires two main computational steps: the formation of the variance–covariance matrix and the inversion of that matrix. Shown is the time taken to perform single calculations for randomly formed phylogenies of different sizes. The black points show the time taken to form the variance–covariance matrix (using the function VCV array in the R package CAPER: this is the faster available function for this operation). The grey points show the time taken to invert the matrix. The red points show the time taken to generate phylogenetically independent contrasts, which are required to calculate likelihoods using the method described in the text (using the pic function in the R package APE). The fitted dashed lines show that the timings scale with size of phylogeny as follows: matrix formation, time $\sim$ size$^{2\cdot54}$; matrix inversion, time $\sim$ size$^{2\cdot87}$; pics, time $\sim$ size$^{1\cdot92}$. Apart from being faster, as the size of the phylogeny increases, the pic method therefore has the slowest rate of increase in computational time.

Likelihood for single traits

BASIC MODEL

In this first section, I outline the method for calculating the maximum likelihood and parameters for a trait on a tree, following the description given in Felsenstein (1973). I then go on to outline how the approach can be generalised to calculate the likelihood for arbitrary parameters.

The model is a Brownian model of trait evolution. According to this model, traits accrue variance in direct proportion to the time they evolve. If the rate of evolution of trait y per unit time t is σ_y^2 and the state of x at the start of the process is μ_y then

for a an expected variance–covariance matrix **V**, the likelihood of the data given **V**, σ_y^2 and μ_y is (with **X** in this equation being a column of 1s):

$$\text{lik}\left(\mu_y, \sigma_y | \mathbf{V}\right) = (2\pi)^{-\frac{n}{2}} |\sigma^2 \mathbf{V}|^{-\frac{1}{2}}$$
$$\exp\left[-\frac{1}{2}\left(\mathbf{Y} - \mu_y \mathbf{X}\right)'\left(\sigma^2 \mathbf{V}\right)^{-1}\left(\mathbf{Y} - \mu_y \mathbf{X}\right)\right] \quad \text{eqn 1}$$

In eqn 1, **V** contains the shared path lengths for each pair of species. If two species do not share a common ancestor from the root of the phylogeny, then the corresponding entry of **V** is zero; otherwise, it is the shared path length from the root to the point at which they last shared an ancestor.

The corresponding log-likelihood for eqn 1 is:

$$L\left(\mu_y, \sigma_y | \mathbf{V}\right) = -\frac{1}{2}\left[n \log\left(2\pi\right) + \log\left|\sigma^2 \mathbf{V}\right|\right.$$
$$\left. + \left(\mathbf{Y} - \mu_y \mathbf{X}\right)'\left(\sigma^2 \mathbf{V}\right)^{-1}\left(\mathbf{Y} - \mu_y \mathbf{X}\right)\right] \quad \text{eqn 2}$$

The maximum likelihood estimates of the parameters of eqn 2 are:

$$\mu_y = \left(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}\right)^{-1}\left(\mathbf{X}'\mathbf{V}^{-1}\mathbf{Y}\right) \quad \text{eqn 3a}$$

$$\sigma_y^2 = \frac{1}{n}\left(\mathbf{Y} - \mu_y \mathbf{X}\right)'\mathbf{V}^{-1}\left(\mathbf{Y} - \mu_y \mathbf{X}\right) \quad \text{eqn 3b}$$

The denominator in eqn 3b is n for the maximum likelihood estimate or $n - 1$ for the restricted maximum likelihood (REML) estimator. More details of REML calculations are given below.

To generate the maximum likelihood parameter estimates, it is necessary to invert the variance–covariance matrix once. For a given fixed variance–covariance matrix **V**, it is not necessary to recalculate $|\mathbf{V}|$ to maximise the likelihood as other parameters are changed, because this is a constant. However, if **V** is altered in a manner other than multiplication by a constant, then $|\mathbf{V}|$ has to be recalculated each time the likelihood in (2) is calculated, further adding to the computational burden.

MAXIMUM LIKELIHOOD BY CONTRASTS

Figure 2 illustrates the principle underlying the method described here: Fig. 2a shows a simple bifurcating phylogeny of five species, with four internal nodes and given branch lengths. The variance–covariance matrix (**V**) is shown in Fig. 2b. This contains the shared path lengths from root to tip for each pair of species on the phylogeny. The shared path length represents shared evolutionary history: the longer the period of shared history, the more similar a pair of species is expected to be. The matrix **V** is the basis for the computations described above. Figure 2c shows the original tree (Fig. 2a) represented as four separate subtrees. The algorithm of Felsen-
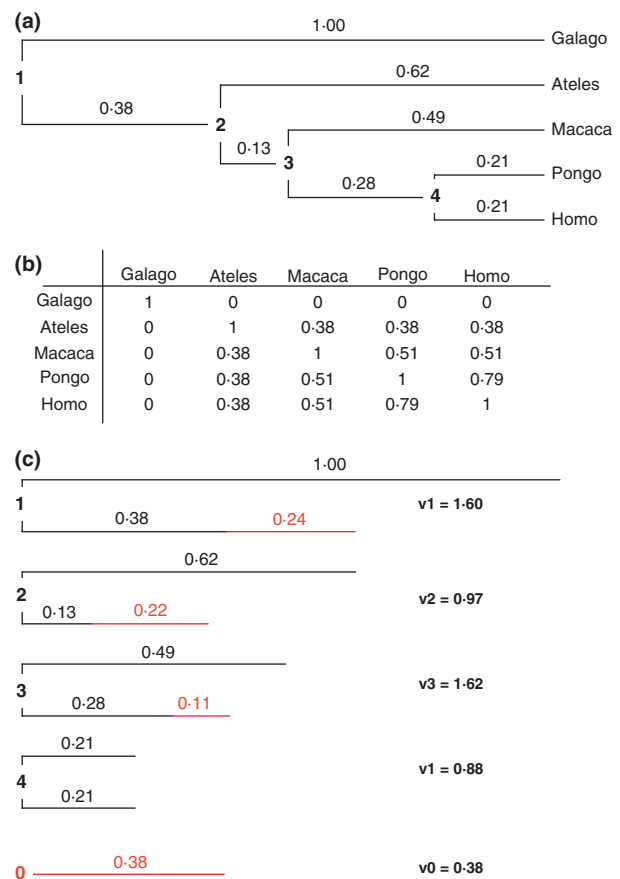

Fig. 2. Pruning the phylogeny and computing variances for calculating likelihoods. (a) The phylogeny is a simple example taken from Lynch (1991). The numbers above branches are path lengths. The numbers at nodes are node labels (referenced in c). (b) The expected variance–covariance matrix implied by the phylogeny. (c) The phylogeny pruned into independent components. These are essentially subtrees extracted from the original phylogeny. The black lines show the original path lengths. The red paths represent additional variance that is accumulated because the trait states at internal nodes are estimated with error that is proportional to the length of subtended branches (as described in the text). The final path (0) is the variance accumulated at the root of the phylogeny: this is the variance in the estimate of the state of the trait at the root, effectively the variance for the estimate of the mean. In fact the implementation here uses the pic() function in ape which is relatively inefficient, and in fact it is possible to increase the efficiency of the calculations so that computational time increases linearly with n (R. Fitzjohn, pers. comm.).

stein (1973) works by calculating likelihoods on these subtrees, rather than on the entire tree. This is computationally much more efficient.

Felsenstein (1973) noted that traits evolve in a Brownian model by accruing a series of changes from the root of the phylogeny to the set of extant species. Under a Brownian model, the changes that occur in a time period t are expected to have a mean of zero (i.e. no net change in the mean of the state) and variance $\sigma_y^2 t$ and be normally distributed. In overview, the algorithm estimates these changes on the phylogeny from ancestral reconstructions of traits and then calculates the likelihood of this set of changes for the phylogeny as a whole: in Fig. 2c the path lengths drawn in black correspond

exactly to those in the whole phylogeny in Fig. 2a. The black paths therefore represent the expected variance to accrue as a consequence of trait evolution. The red paths in Fig. 2c represent the statistical uncertainty in estimating ancestral trait values at the internal nodes of the phylogeny. When both sources of variance are combined for the subtrees in Fig. 2c, the likelihood for the set of changes is the same as the likelihood of observing the current states of the traits of the extant species (Felsenstein 1973, 2004).

More specifically, the method proceeds in the following steps (e.g. following Felsenstein 1973, 1985):

1 Beginning with a pair of adjacent tips (species i and j), which have trait values y_i and y_j, respectively, and with common ancestor k, the contrast $u_{ij} = y_i - y_j$ is computed. This value has expectation zero and variance $V_i = v_i + v_j$ where v_i and v_j are the lengths of the branches leading to nodes i and j, respectively.

2 Assign k the character state $y_k = \left(\frac{1}{v_i}y_i + \frac{1}{v_j}y_j\right) \Big/ \left(\frac{1}{v_i} + \frac{1}{v_j}\right)$ that is, the variance weighted mean of the two species observations.

3 To account for the statistical uncertainty involved in estimating y_k, the edge below k is increased from v_k to $v_k + v_i v_j/(v_i + v_j)$. In Fig. 2c, this uncertainty is represented by the paths drawn in red.

4 The two tips are removed from the tree, leaving k as a tip, and the process is repeated until all the tips on the tree have been removed.

5 The final node (i.e. the root) will have a zero contrast, by definition, but has a variance (v_0), which is the error in the ancestral state at the root, accumulated throughout the tree.

The contrasts, u_i, are expected to be normally distributed with mean zero and variance σ_y^2. Thus, at a single node, the log-likelihood is:

$$L = -\frac{1}{2}\left(\log(2\pi) + \log V_i + \frac{u_i^2}{\sigma_y^2 V_i}\right) \qquad \text{eqn 4}$$

The log-likelihood of trait y is then given by:

$$L\left(\mu_y, \sigma_y^2|\mathbf{V}, \mathbf{Y}\right) = -\frac{1}{2}\left(n\log(2\pi) + \sum_{i=0}^{n-1}\left[\log V_i + \frac{u_i^2}{\sigma_y^2 V_i}\right]\right) \qquad \text{eqn 5}$$

Equation 5 is exactly equal to eqn 1. The advantage of this approach is that it is computationally very much quicker as it does not require the inversion of $\mathbf{V}$. Assuming a nested data structure representing the phylogeny, the calculation can be achieved with two traversals of the phylogeny, the computational overheads of which are approximately linearly proportional to the square of the size of the phylogeny.

The mean, μ_y, is given by y_0, the estimated ancestral state of the trait at the root. The variance is estimated by:

$$\sigma_y^2 = \frac{1}{n}\sum_{i=0}^{n-1}\frac{u_i^2}{V_i} \qquad \text{eqn 6}$$

As described above, the REML estimate of variance would be given using $n - 1$ rather than n in the denominator. The

approach outlined here is exactly the same as used to generate phylogenetically independent contrasts (Felsenstein 1985) and emphasises that the two methods for calculating the likelihood (eqns 1 and 5) are identical in terms of the model they fit, the likelihood estimated and the parameters of that model (Garland & Ives 2000). Figure 3 gives a worked example to demonstrate this equivalence.

RESTRICTED MAXIMUM LIKELIHOOD

The restricted likelihood is the likelihood of the data free of the fixed effects. In the context of eqn 1, this is the likelihood of the data independent of the uncertainty associated with the estimation of the mean μ_y. Equation 5 can be used to calculate the REML, with two modifications: (i) the unbiased estimator of the variance is used rather than the ML (eqn 3b); (ii) the root variance, v_0, is the variance associated with the estimation of μ_y and is hence not included in the summation in eqn 5, so that the summation is from $i = 1$ to $n - 1$.

LIKELIHOOD FOR ARBITRARY PARAMETER VALUES

Equation 4 does not explicitly include the mean μ_y, as it is marginalised in the calculation. The likelihood of the model parameters for given values of the mean and variance of y is:

$$L\left(\mu_y, \sigma_y^2|\mathbf{V}, \mathbf{Y}\right) =$$
$$-\frac{1}{2}\left(n\log(2\pi) + \sum_{i=1}^{n}\left[\log v_i + \frac{u_i^2}{\sigma_y^2 V_i}\right] + \frac{(y_0 - \mu_y)^2}{\sigma_y^2 V_0}\right) \qquad \text{eqn 7}$$

The modification in eqn 7 is to the term estimating the likelihood at the root of the phylogeny: the difference between μ_y and y_0 is the difference between the mean implied by the traits and the phylogeny and μ_y, effectively the difference that would have accrued on the branch leading to the root of the phylogeny. Equation 7 would be useful, for example, if using Bayesian methods to sample from prior distributions of μ_y and σ_y^2.

ACCOUNTING FOR PHYLOGENY TRANSFORMATIONS

In analyses of trait evolution, transformations of the phylogeny are commonly used to model deviations from the basic Brownian model (Grafen 1989; Pagel 1997, 1999; Hansen 1997; Thomas, Freckleton & Szekely 2006; O'Meara *et al.* 2006; Hansen, Pinaar & Orzack 2008). Likelihoods were calculated, using eqns 4 or 5, by transforming the phylogeny. For example, Pagel's λ statistic (Pagel 1997, 1999) is a transformation of $\mathbf{V}$ in which the off diagonal elements of $\mathbf{V}$ are multiplied by λ, with λ usually lying between 0 and 1.

This model is effectively a random effect model for y, in which λ models a phylogenetically independent random component of the model (Freckleton, Harvey & Pagel 2002). To implement in eqns 4 or 6, the transformation is readily applied to a phylogeny, before calculation of the likelihood. In the R

(a) GLS calculations

$$\mathbf{Y} = \begin{bmatrix} -1{\cdot}46968 \\ 2{\cdot}02815 \\ 2{\cdot}37024 \\ 3{\cdot}61092 \\ 4{\cdot}09434 \end{bmatrix} \qquad \mathbf{X} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \qquad \mathbf{V}^{-1} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1{\cdot}25 & -0{\cdot}29 & -0{\cdot}18 & -0{\cdot}18 \\ 0 & -0{\cdot}29 & 1{\cdot}48 & -0{\cdot}36 & -0{\cdot}36 \\ 0 & -0{\cdot}18 & -0{\cdot}36 & 2{\cdot}80 & -1{\cdot}96 \\ 0 & -0{\cdot}18 & -0{\cdot}36 & -1{\cdot}96 & 2{\cdot}80 \end{bmatrix}$$

$$\mu_y = \left(\mathbf{X'V^{-1}X}\right)^{-1}\left(\mathbf{X'V^{-1}Y}\right) = (2{\cdot}66)^{-1} \times 3{\cdot}15 = 1{\cdot}18$$

$$\mathbf{Y} - \mu\mathbf{X} = \begin{bmatrix} -2{\cdot}65 \\ 0{\cdot}84 \\ 1{\cdot}18 \\ 2{\cdot}43 \\ 2{\cdot}91 \end{bmatrix}$$

$$\sigma_y^2 = \frac{1}{n}(\mathbf{Y}-\mu_Y\mathbf{X})'\mathbf{V}^{-1}(\mathbf{Y}-\mu_Y\mathbf{X}) = \frac{1}{5}\begin{bmatrix} -2{\cdot}65 & 0{\cdot}84 & 1{\cdot}18 & 2{\cdot}43 & 2{\cdot}91 \end{bmatrix}\begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1{\cdot}25 & -0{\cdot}29 & -0{\cdot}18 & -0{\cdot}18 \\ 0 & -0{\cdot}29 & 1{\cdot}48 & -0{\cdot}36 & -0{\cdot}36 \\ 0 & -0{\cdot}18 & -0{\cdot}36 & 2{\cdot}80 & -1{\cdot}96 \\ 0 & -0{\cdot}18 & -0{\cdot}36 & -1{\cdot}96 & 2{\cdot}80 \end{bmatrix}\begin{bmatrix} -2{\cdot}65 \\ 0{\cdot}84 \\ 1{\cdot}18 \\ 2{\cdot}43 \\ 2{\cdot}91 \end{bmatrix} = \frac{15{\cdot}77}{5} = 3{\cdot}15$$

$$L\left(\mu_y, \sigma_y \mid \mathbf{V}\right) = -\frac{1}{2}\left[n\log(2\pi) + \log|\sigma^2\mathbf{V}| + (\mathbf{Y}-\mu_y\mathbf{X})'(\sigma^2\mathbf{V})^{-1}(\mathbf{Y}-\mu_y\mathbf{X})\right] = -\frac{1}{2}\left[5\log(6{\cdot}28) + \log(66{\cdot}64) + 5\right] = -9{\cdot}19$$

(b) PIC calculations

$$u = [4{\cdot}25, 1{\cdot}17, 1{\cdot}48, 0{\cdot}48]$$
$$V = [1{\cdot}60, 0{\cdot}97, 0{\cdot}88, 0{\cdot}42], V_0 = 0{\cdot}38$$

$$\sigma^2 = \frac{1}{n}\sum_{i=0}^{n-1}\frac{u_i^2}{V_i^2} = \frac{1}{5}\left(\frac{18{\cdot}067}{1{\cdot}60} + \frac{1{\cdot}37}{0{\cdot}97} + \frac{2{\cdot}20}{0{\cdot}88} + \frac{0{\cdot}48}{0{\cdot}42} + \frac{0}{0{\cdot}38}\right) = \frac{15{\cdot}77}{5} = 3{\cdot}15$$

$$L\left(\mu_y, \sigma_y^2 \mid \mathbf{V}, \mathbf{Y}\right) = -\frac{1}{2}\left(n\log(2\pi) + \sum_{i=0}^{n}\left[\log V_i + \frac{u_i^2}{\sigma_y^2 V_i}\right]\right)$$

$$= -\frac{1}{2}\left[5\log(24{\cdot}77) + \log(1{\cdot}60) + \frac{18{\cdot}067}{3{\cdot}15 \times 1{\cdot}60} + \log(0{\cdot}97) + \frac{1{\cdot}37}{3{\cdot}94 \times 0{\cdot}97} + \log(0{\cdot}88) + \frac{2{\cdot}20}{3{\cdot}15 \times 0{\cdot}88} + \log(0{\cdot}44) + \frac{0{\cdot}48}{3{\cdot}15 \times 0{\cdot}42} + \log(0{\cdot}38) + \frac{0}{3{\cdot}15 \times 0{\cdot}38}\right] = -9{\cdot}19$$

Fig. 3. A worked example showing that the calculations performed with the two methods yield identical results, despite differing markedly in details. (a) The calculation of likelihoods using GLS methods. **Y** is the trait matrix, that is, the trait to be modelled. **X** is the design matrix, in this case a column of 1s, with the model thus being a mean and variance for this trait. **V**$^{-1}$ is the inverse of the variance–covariance matrix representing the phylogeny (Fig. 2b). The first step is the calculation of the mean, μ, using the formula given in the text. Using this, residuals (**Y** − μ**X**) are calculated, from which the variance is generated (σ²). Finally, these quantities are used to calculate the log-likelihood. Using the contrast method, contrasts (*u*) are generated using the pruning method (Fig. 2c) and the algorithm given in the text. The contrast variances (*V*), including the variance at the root (*V*$_0$), are calculated as shown in Fig. 2. The variance (σ²) is generated directly from the contrasts and their variances; then, finally these are all used to generate the (log) likelihood.

package ape (Paradis *et al.* 2004), this is achieved very quickly as internal and external branches are easily distinguished and referenced (see function lambda.trans() in the online supplement).

For other transformations, the calculations and phylogeny manipulations required may be slightly more involved. For example, Pagel's δ is a parameter that measures the degree to which the rate of evolution increases or decreases from the root of the phylogeny to the tips (Pagel 1997). This transformation raises node heights to the power δ, such that values of δ < 1 yield a relative increase in the length of branches near to the root (slowdown in evolution) and values >1 yield a relative increase in the length of branches near to the tips (increase in the rate of evolution). To use the algorithm described above would require the following steps: (i) generate a set of heights for all nodes and daughter nodes, (ii) transform these using a given value of δ, (iii) recalculate the branch lengths and transform the tree. However, although more involved, this approach relies on calculations that are much faster to perform than matrix inversion or computation of **V**. In general, the

approach described here can be applied to any model in which it is possible to represent the process as a transformation to the tree.

Likelihood for correlated traits

CORRELATIONAL MODEL

If **Y** is a $k \times n$ list of k traits observed on n species and **C** is the $k \times k$ variance–covariance matrix for the traits, then the combined likelihood for the traits on the phylogeny is:

$$\text{lik}(\mathbf{m}, \mathbf{C} \mid \mathbf{V}) = (2\pi)^{-\frac{kn}{2}}|\mathbf{C} \otimes \mathbf{V}|^{-\frac{1}{2}}$$
$$\exp\left[-\frac{1}{2}(\mathbf{Y}-\mathbf{mX})'(\mathbf{C} \otimes \mathbf{V})^{-1}(\mathbf{Y}-\mathbf{mX})\right]$$

eqn 8

In eqn 7, $\mathbf{C} \otimes \mathbf{V}$ is the Kronecker product of **C** and **V**, that is:

$$\mathbf{C} \otimes \mathbf{V} = \begin{pmatrix} c_{11}\mathbf{V} & \cdots & c_{1k}\mathbf{V} \\ \vdots & \ddots & \vdots \\ c_{k1}\mathbf{V} & \cdots & c_{kk}\mathbf{V} \end{pmatrix} \qquad \text{eqn 9}$$

This has dimensions $kn \times kn$, so that as the number of traits is increased, the computational burden using the direct maximisation of the likelihood in eqn 8 is expected to increase in proportion to both k^2 and n^2.

Using the logic described above, it is straightforward to derive an expression for this multivariate log-likelihood corresponding to eqn 4. This is done by calculating at each node on the phylogeny the contrasts, u, for each trait under consideration, such that at each node, a vector $\mathbf{u}$ of trait differences is estimated. The log-likelihood for a single node is then:

$$L = -\frac{1}{2}\left(k\log(2\pi) + \log|\mathbf{C}| + \log V_i + \frac{\mathbf{u}_i'\mathbf{C}^{-1}\mathbf{u}_i}{V_i} \right) \qquad \text{eqn 10}$$

So that for the whole dataset the log-likelihood is:

$$L\left(\mu_y, \sigma_y^2 | \mathbf{V}, \mathbf{Y}\right) = \\ -\frac{1}{2}\left(nk\log(2\pi) + n\log|\mathbf{C}| + \sum_{i=0}^{n-1}\left[\log V_i + \frac{\mathbf{u}_i'\mathbf{C}^{-1}\mathbf{u}_i}{V_i} \right] \right) \qquad \text{eqn 11}$$

Although eqn 10 requires the determinant and inverse of $\mathbf{C}$ to be calculated, the dimensions of this matrix are expected to be considerably smaller than those of $\mathbf{V}$.

RANDOM EFFECTS FOR CORRELATIONAL MODEL

If we assume that each trait has a separate associated random effect, then the net covariance matrix has the form:

$$\mathbf{V}_k = \begin{pmatrix} c_{11}\mathbf{V}(\lambda_1) & \cdots & c_{1k}\mathbf{V}(\lambda_1, \lambda_k) \\ \vdots & \ddots & \vdots \\ c_{k1}\mathbf{V}(\lambda_k, \lambda_1) & \cdots & c_{kk}\mathbf{V}(\lambda_k) \end{pmatrix} \qquad \text{eqn 12}$$

This is equivalent to assuming a separate variance–covariance matrix for each of the $k \times k$ variance–covariance estimates, in which each trait is assumed to have an individual random effect term, λ, which is equivalent to Pagel's λ above. This model allows for alternative variance structures in different traits, whereas the simpler model (Pagel's λ method; Freckleton, Harvey & Pagel 2002) assumes that all traits have the same variance structure. A likelihood ratio test could be used to compare these models.

Equation 11 is easily adapted to allow for each trait to have its own variance structure:

$$L\left(\mu_y, \sigma_y^2 | \mathbf{V}, \mathbf{Y}\right) = \\ -\frac{1}{2}\left(nk\log(2\pi) + n\log|\mathbf{C}| + \sum_{i=0}^{n}\sum_{j=1}^{k}\log V_{ik} + \sum_{i=1}^{n}(\mathbf{vu}_i)'\mathbf{C}^{-1}(\mathbf{vu}_i) \right) \qquad \text{eqn 13}$$

Likelihoods for Linear models by PGLS

One of the commonest applications of comparative analysis is to measure the effect of a set of predictors on some variable of interest. This is carried out by fitting a model of the form:

$$\mathbf{Y} = \mathbf{Xb} + \mathbf{e} \qquad \text{eqn 14}$$

The data $\mathbf{Y}$ are fitted as a function of predictors $\mathbf{X}$, parameters $\mathbf{b}$ and error term $\mathbf{e}$. $\mathbf{e}$ is assumed to be multivariate normally distributed with covariance matrix $\mathbf{V}$. The likelihood for this model is:

$$L(\mathbf{b}|\mathbf{X}, \mathbf{V}) = -\frac{1}{2}\Big[n\log(2\pi) + \log\left|\sigma_e^2\mathbf{V}\right| \\ + (\mathbf{Y} - \mathbf{Xb})'(\sigma_e^2\mathbf{V})^{-1}(\mathbf{Y} - \mathbf{Xb}) \Big] \qquad \text{eqn 15}$$

As is well known, the maximum likelihood parameters are given by:

$$\mathbf{b} = \left(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}\right)^{-1}\left(\mathbf{X}'\mathbf{V}^{-1}\mathbf{Y}\right) \qquad \text{eqn 16}$$

The single trait model (eqn 1) is a special case of this model in which $\mathbf{X}$ is a vector of 1s and the only parameter is the mean of y. Equation 14 can incorporate more complex designs, however, including covariates, predictors and interaction terms. These are included by specifying the appropriate structure for the design matrix $\mathbf{X}$.

In the current context, the main point I wish to make is that such models can be solved using the methodology described above. In this case, the statistical model is for the error term $\mathbf{e}$, not for $\mathbf{Y}$. At a single node on the phylogeny, $u_{y,i}$ is the contrast for y and $\mathbf{u}_{x,i}$ is the vector of contrasts for the predictors. V_i is the variance for this contrast, and σ_e^2 is the variance of the error term. The likelihood for $\mathbf{b}$ at this node is:

$$L = -\frac{1}{2}\left[\log(2\pi) + \log V_i + \frac{\left(u_{y,i} - \mathbf{u}_{x,i}\mathbf{b}\right)^2}{\sigma_e^2 V_i} \right] \qquad \text{eqn 17}$$

So that the likelihood for the whole tree is:

$$L(\mathbf{b}|\mathbf{X}, \mathbf{V}, \mathbf{Y}) =$$
$$-\frac{1}{2}\left(n\log(2\pi) + \sum_{i=0}^{n-1}\left[\log V_i + \frac{\left(u_{y,i} - \mathbf{u}_{x,i}\mathbf{b}\right)^2}{\sigma_e^2 V_i}\right]\right) \quad \text{eqn 18}$$

If $\mathbf{U}_x$ and $\mathbf{U}_y$ are matrices of contrasts of x and y, respectively, the maximum likelihood estimate of $\mathbf{b}$ is then given by:

$$\mathbf{b} = \left(\mathbf{U}'_x\mathbf{U}_x\right)^{-1}\left(\mathbf{U}'_x\mathbf{U}_y\right) \quad \text{eqn 19}$$

$\mathbf{U}_x$ does not include an intercept term. This is because an intercept is coded as a column of 1s in the design matrix $\mathbf{X}$, and hence, all contrasts for this will be zero. The intercept, in this formulation, is estimated from the grand mean of y, given by the mean of y at the root of the phylogeny.

Equation 15 is the log-likelihood that can be maximised to yield maximum likelihood (ML) parameter estimates. For the REML, the corresponding equation is:

$$L(\mathbf{b}|\mathbf{X}, \mathbf{V})$$
$$= -\frac{1}{2}[(n-k)\log(2\pi) + \log|\sigma_e^2\mathbf{V}| \quad \text{eqn 20}$$
$$+ (\mathbf{Y} - \mathbf{X}\mathbf{b})'(\sigma_e^2\mathbf{V})^{-1}(\mathbf{Y} - \mathbf{X}\mathbf{b}) + \log|\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}|]$$

Equation 18 then becomes:

$$L(\mathbf{b}|\mathbf{X}, \mathbf{V}, \mathbf{Y}) =$$
$$-\frac{1}{2}\left((n-k)\log(2\pi) + \log|\mathbf{u}_x\mathbf{u}_x| + \sum_{i=1}^{n-1}\left[\log V_i + \frac{\left(u_{y,i} - \mathbf{u}_{x,i}\mathbf{b}\right)^2}{\sigma_e^2 V_i}\right]\right) \quad \text{eqn 21}$$

The REML eqns 20 and 21 are more appropriate to use when comparing different models in which the random effects are varied, but the fixed effects are held constant (Pinheiro & Bates 2000).

Because the error is assumed to be contained in the residual term in eqn 14, no assumption is made about the distribution of $\mathbf{X}$. Hence, $\mathbf{X}$ can be continuous, ordinal or factorial. Factorial structures for predictors are particularly useful. These are achieved by appropriate specification of the design matrix $\mathbf{X}$ and dummy coding.

It is also possible to employ predictors that contain no phylogenetic structure in this model. It may seem incorrect to do this: for instance, if $\mathbf{X}$ is an environmentally driven variable with no phylogenetic structure, then the past values of this trait cannot be reconstructed, particularly if this is a variable that has not evolved as a trait. However, interpretation of nodal means as ancestral values is notional and not essential for the technique to work. With the phylogenetic structure in the residuals, the algorithm described will correctly model this variance, irrespective of the structure of $\mathbf{X}$.

R code

The accompanying R code provides functions to estimate parameters and calculate likelihoods via both the direct and rapid contrast methods, to show that the methods yield the same results and demonstrate the computational advantages of the contrast method. For a tree of 1000 tips, it is estimated that, using this code, the contrast method is 300–900 times faster. In simulations, I have generated trees of up to 1×10^6 species and performed analyses such as the maximum likelihood estimation of λ in reasonable times using currently available desktop computers.

Discussion

I have described a computationally efficient fast method for computing models accounting for phylogenetic structure that might normally be slow to solve owing to their numerical demands. The approach taken, deriving from Felsenstein (1973, 1985), is computationally and conceptually simple, and already widely used and understood by many practitioners of the comparative method in the form of phylogenetic contrasts. This approach is, in fact, a special case of the algorithm of peeling/pruning that is already widely used in phylogenetic reconstruction (Elston & Stewart 1971; Felsenstein 2004).

There are three main implications of the results I have presented. First, there is an increasing number of large (>1000 species) phylogenies becoming available, which are currently rather difficult to analyse. With large trees, it is likely that the assumptions of simple models will break down and that more complex models will need to be fitted. Second, there is an increase in the use of computationally demanding methods, such as Bayesian approaches, that require very large number of calculations: these could be greatly speeded up using the methods I have outlined. Finally, simulation models that require large numbers of replicates will be greatly speeded up using this approach.

The increase in computational speed is possible because the Brownian model of trait evolution for a group of species can be broken down into a sum of component changes that give rise to the final trait distribution. Alternative computational simplifications are possible; for example, approaches used in the analysis of pedigree data using the animal model can be applied to phylogenies, and significant computational gains can be made (e.g. Hadfield & Nakagawa 2010) using techniques for dealing with sparse matrices. The approach described here is also extremely efficient. Moreover, the memory requirements are extremely economical: in the online R code, I use the method to solve a problem for 1×10^6 species that could not easily be addressed using existing tools as this would require allocating enough memory for a matrix with 1×10^{12} entries, requiring somewhere around 8 TB of memory to store.

Given the equivalency of the contrast and GLS methods, an obvious question is, what is the use of expressing models in the more complex GLS form if they can be solved very easily using contrasts? The answer is expressed in the full form and it is clear what the model is and what assumptions are being made (Hadfield & Nakagawa 2010). This avoids misunderstandings in the presentation of the model. As an example, if we have two traits x and y, the relationship between them could be modelled by a correlational model (eqn 8) or a linear model (eqn 14). If we use eqn 8, we assume correlated Brownian motion and (unless using the more complex random effects model, eqn 12) both traits should have similar levels of phylogenetic dependence. On the other hand, if we are modelling y as a function of x, then the phylogenetic dependence of x is not important.

An obvious question is whether this approach can be applied to models based on models of trait change other than the Brownian process with normally distributed trait changes. For instance, hierarchical models with non-normal errors have been developed for phylogenetic analysis that are based on linear predictors with an essentially Brownian structure (Hadfield & Nakagawa 2010; Ives & Helmus 2011). The approach described by Hadfield & Nakagawa (2010) relies on an alternative computational simplification. Although probably not as efficient for the models described, their approach generalises to nontreelike variance structures. The approach described here relies on being able to model the changes in traits on a tree by calculating likelihoods at the internal nodes and will be a highly efficient approach for such problems.

In summary, the aim of this study has been to highlight the use of simple algorithms to speed up calculations in evolutionary models. These approaches will hopefully allow a step change in the size of datasets that can be modelled using comparative approaches compared with approaches currently widely used.

Acknowledgements

References

Diniz-Filho, J.A.F., de Sant Ana, C.E.R. & Bini, L.M. (1998) An eigenvector method for estimating phylogenetic inertia. *Evolution*, **52**, 1247–1262.

Elston, R.C. & Stewart, J. (1971) A general model for the genetic analysis of pedigree data. *Human Heredity*, **21**, 523–542.

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*, **126**, 1–25.

Felsenstein, J. (2004) *Inferring Phylogenies*. Sianuer, Sunderland, Massachusetts.

Freckleton, R.P. & Harvey, P.H. (2006) Detecting non-Brownian trait evolution in adaptive radiations. *PLoS Biology*, **4**, 2104–2111.

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.

Freckleton, R.P. & Jetz, W. (2009) Space versus phylogeny: disentangling phylogenetic and spatial signals in comparative data. *Proceedings of the Royal Society B-Biological Sciences*, **276**, 21–30.

Garland, T. Jr & Ives, A.R. (2000) Using the past to predict the present: confidence intervals for regression equations in phylogenetic comparative methods. *American Naturalist*, **155**, 346–364.

Garland, T.J., Midford, P.E. & Ives, A.R. (1999) An introduction to phylogenetically-based statistical methods, with a new method for confidence intervals based on ancestral values. *American Zoologist*, **39**, 374–388.

Grafen, A. (1989) The phylogenetic regression. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, **326**, 119–157.

Hadfield, J.D. & Nakagawa, S. (2010) General quantitative genetic methods for comparative biology: phylogenies, taxonomies and multi-trait models for continuous and categorical characters. *Journal of Evolutionary Biology*, **23**, 494–508.

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

Hansen, T.F., Pinaar, J. & Orzack, S.H. (2008) A comparative method for studying adaptation to a randomly evolving environment. *Evolution*, **62**, 1965–1977.

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

Ives, A.R. & Helmus, M.R. (2011) Generalized linear mixed models for phylogenetic analyses of community structure. *Ecological Monographs*, **81**, 511–525.

Lynch, M. (1991) Methods for the analysis of comparative data in evolutionary biology. *Evolution*, **45**, 1065–1080.

Martins, E.P. & Hansesn, T.F. (1997) Phylogenies and the comparative method: a general approach to incorporating phylogenetic information into the analysis of inter-specific data. *American Naturalist*, **149**, 646–667.

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

Pinheiro, J.C. & Bates, D.M. (2000) *Mixed-Effects Models in S and S-Plus*. Springer, New York.

Robinson, S. (2005) Towards an optimal algorithm for matrix multiplication. *SIAM News*, **38**, http://www.siam.org/news/news.php?issue = 0038.09.

Thomas, G. & Freckleton, R.P. (2011) MOTMOT: models of trait macroevolution on trees. *Methods in Ecology and Evolution*, **2**, 145–151.

Thomas, G.H., Freckleton, R.P. & Szekely, T. (2006) Comparative analyses of the influence of developmental mode on phenotypic diversification rates in shorebirds. *Proceedings of the Royal Society B-Biological Sciences*, **273**, 1619–1624.

Supporting Information

Additional Supporting Information may be found in the online version of this article.

Data S1. Code implementing fast likelihood calculations for comparative data.

Data S2. Example of application of fast likelihood calculations.

As a service to our authors and readers, this journal provides supporting information supplied by the authors. Such materials may be re-organized for online delivery, but are not copy-edited or typeset. Technical support issues arising from supporting information (other than missing files) should be addressed to the authors.