Evolutionary Ecology Research, 2008, **10**: 311–331

Testing quantitative genetic hypotheses about the evolutionary rate matrix for continuous characters

Liam J. Revell[1]* and Luke J. Harmon[2]

[1]*Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, MA 02138 and* [2]*Department of Biological Sciences, University of Idaho, Moscow, ID 83844, USA*

ABSTRACT

Aim: Provide a new multivariate hypothesis testing approach for the evolution of continuous characters in a phylogenetic context.

Background: Brownian motion is the most commonly used model for the evolution of quantitative traits. Under multivariate Brownian motion, the evolution of multiple continuous traits can be described by an evolutionary rate matrix in which the diagonal elements determine the rate of evolution for individual characters, while the off-diagonal elements determine the extent to which different characters co-evolve.

Method: We present likelihood tests for two simple hypotheses about the evolutionary rate matrix: (1) equality or proportionality to a hypothetical matrix; and (2) concerted change in the rate matrix in a certain portion or portions of the phylogenetic tree. In case (1), the hypothetical matrix might be estimated from the within-species additive genetic variance–covariance matrix. In case (2), concerted change in the evolutionary rate matrix might result from shifts in ploidy level or the mutation rate. We illustrate these hypothesis tests using data from individual-based quantitative genetic simulations on stochastic phylogenetic trees.

Results: Our evolutionary rate matrix estimator exhibited minimal bias. Type I errors in our likelihood-based tests for matrix equality and proportionality were very close to appropriate levels. Power to detect selection was also sufficient, except in the case of the weakest selection simulated in this study. Type I error and the accuracy of parameter estimation in the test for rate matrix heterogeneity were also adequate.

Keywords: Brownian motion, comparative method, evolutionary constraint, genetic constraint, phylogenetic generalized least squares, maximum likelihood, phylogenetics.

INTRODUCTION

In recent years, the phylogenetic approach has become *de rigueur* in comparative biology (Felsenstein, 1985, 2004; Harvey and Pagel, 1991). This stems from the recognition that the character

* Author to whom all correspondence should be addressed. e-mail: lrevell@fas.harvard.edu

states for extant taxa related by a bifurcating history are non-independent due to common ancestry (Felsenstein, 1985; Harvey and Pagel, 1991). Methods developed for continuous characters, such as Felsenstein's (1985) independent contrasts method and the phylogenetic generalized least-squares approach (PGLS) (Grafen, 1989), take this phylogenetic non-independence into account – enabling analysis by standard statistical methods. Both independent contrasts and PGLS have been used primarily to address hypotheses concerning correlated evolution between pairs of characters. However, both have subsequently been extended to a broader set of applications, including the estimation of evolutionary rates (Garland, 1992), the comparison of rates among related groups (O'Meara *et al.*, 2006), the comparison of different models for the evolutionary process (Hansen, 1997; Martins and Hansen, 1997), and the estimation of ancestral states (Martins and Hansen, 1997; Rohlf, 2001).

Most comparative analyses focus on the evolution of a single character, or evaluate the relationship between two characters. However, since selection and drift affect populations of whole organisms each composed of numerous intercorrelated traits, evolution is inherently multivariate. Nonetheless, a thoroughly multivariate phylogenetic approach to the analysis of comparative data has not yet been proposed.

In contrast to the univariate or bivariate approaches adopted in most comparative analyses, evolutionary quantitative genetics has long been concerned with the analysis of multiple correlated characters (e.g. Lynch and Walsh, 1998). Theoretical quantitative genetics makes numerous predictions about the evolution of correlated characters under genetic drift and natural selection (e.g. Lande, 1979; Lande and Arnold, 1983; Schluter, 1996). Until quite recently, however, most quantitative geneticists essentially dismissed phylogeny from their analyses, with a few exceptions (Baker and Wilkinson, 2003; Bégin and Roff, 2004; Revell *et al.*, 2007). Among phylogeneticists, Felsenstein (1988, 2005) provides the connection between theoretical quantitative genetics and the assumptions of several comparative methods.

Quantitative genetics and the analysis of continuous characters in a phylogenetic context could be more closely integrated. For example, several new phylogenetic methods (e.g. McPeek, 1995; Garland and Ives, 2000; O'Meara *et al.*, 2006; Revell, in press) are concerned with the rate of evolution for continuous characters. Theoretical quantitative genetics has very specific predictions about the rate of evolution of continuous characters under various models of evolutionary change, including drift and selection. Nonetheless, few studies have attempted to unify the quantitative genetic and phylogenetic comparative approaches to the analysis of continuous characters (but see Felsenstein, 1988, for an early review).

In this paper, we provide a multivariate method for the analysis of evolutionary rates. Unlike previous studies of evolutionary rate, in which the properties of the rate estimator were evaluated using data simulated by univariate Brownian motion simulations (e.g. O'Meara *et al.*, 2006), we explore our multivariate approach using inherently more realistic individual-based, quantitative genetic simulations in a phylogenetic context. In so doing, we provide a first glimpse at the power of the phylogenetic comparative method in testing quantitative genetic hypotheses.

Modelling the evolutionary process

Statistical hypothesis testing of phenotypic data in a phylogenetic context requires that we first specify a model of the evolutionary process under which those data were produced. The most common such model for continuous data is Brownian motion (Edwards and Cavalli-Sforza, 1964; Felsenstein, 1973, 1981, 1985, 1988; Harvey and Pagel, 1991). Under Brownian motion, the mean

phenotype of the population or lineage being modelled can increase or decrease in any time interval. The amount of increase or decrease is determined by the variance of a Gaussian distribution from which the change is drawn. This variance is equal to the product of the length of the time interval and the Brownian motion rate parameter. The expected net evolutionary change is zero, but the expected variance among lineages increases linearly with time. In multivariate space, the Brownian motion process is governed by a rate matrix in which the diagonal elements determine the rate of evolution for individual characters, and are each equivalent to the rate parameter in the univariate case, while the off-diagonal elements determine the extent to which the different continuous characters co-evolve (Felsenstein, 1988).

Brownian motion is a standard evolutionary model used in a variety of phylogenetic approaches for the evolution of continuous phenotypic traits. These include the independent contrasts method (Felsenstein, 1985, 1988), the estimation of ancestral states (Schluter *et al.*, 1997), and the estimation of the evolutionary rate for a single character (Garland, 1992; O'Meara *et al.*, 2006; Revell, in press). Although Brownian motion is also a suitable model for some deterministic processes such as fluctuating directional selection and a shifting position of the fitness optimum (Hansen and Martins, 1996; O'Meara *et al.*, 2006), it is an inappropriate model in circumstances in which the population is under consistent stabilizing selection, consistent directional selection, or near the bounds of natural phenotype space (Butler and King, 2004; O'Meara *et al.*, 2006).

Brownian motion is a suitable model for genetic drift under many, but not all, circumstances. Brownian motion is not an appropriate model for drift when the phenotypic trait is determined by one or a few genes, particularly if the mutation rate is low relative to the population size, and if the alleles have finite possible states (Felsenstein, 1988). It is also not suitable for drift if genes behave highly non-additively. However, Brownian motion is a suitable drift model under the assumptions of most classical quantitative genetic studies [e.g. that the genes underlying a particular phenotypic trait are many, and that they have strictly or roughly additive effects on the phenotype (Lande, 1979; Felsenstein, 1988)].

The phylogenetic comparative approach

Several phylogenetic comparative methods for the analysis of data from continuous traits have been proposed. The most popular and widely implemented of these is the phylogenetic independent contrasts method of Felsenstein (1985). In this method the difference between phenotypic states at nodes (known or estimated) are subtracted and standardized by the branch length separating them. In so doing, the method provides a set of differences ('contrasts') unfettered by the common history that creates statistical dependence among observations at the tips of the tree.

Grafen (1989; see also Martins and Hansen, 1997; Rohlf, 2001) provided an alternative method based on the standard statistical approach of generalized least-squares (GLS) – phylogenetic generalized least-squares, or PGLS. Generalized least-squares is like ordinary least-squares, except that the error matrix is not necessarily assumed to be equal to $\sigma_e^2\mathbf{I}$, where $\mathbf{I}$ is the identity matrix (a matrix with 1.0s on the diagonal, and 0.0s elsewhere) and σ_e^2 is the common error variance. In PGLS, the error is instead assumed to be multivariate normal with a variance–covariance matrix proportional to $\mathbf{C}$, where, for n taxa, $\mathbf{C}$ is an $n \times n$ matrix constructed from the phylogenetic tree and the evolutionary model. Under Brownian motion, the expected covariance among tips is proportional to common ancestry, and as such the

off-diagonal elements of **C**, C_{ij}, are computed as the sum of the shared branch lengths between the root and the common ancestor of taxa i and j (Felsenstein, 1973). Diagonal elements of **C**, C_{ii}, are computed as the sum of the branch lengths from the root to each tip i. Alternative evolutionary models can be incorporated if **C** is calculated by different formulae (e.g. Butler *et al.*, 2000).

Using Brownian motion as the evolutionary model, $\sigma^2\mathbf{C}$ provides the expected variances and covariances among the values at the tips for a single trait, in which σ^2 is a scalar equivalent to the Brownian motion rate parameter of the trait. This is also the structure of the error term in the PGLS regression model, which can be used to transform the dependent and independent variables to a form amenable to standard statistical analyses such as multiple regression, ANOVA, MANOVA, ANCOVA, and so on (for details, see Martins and Hansen, 1997; Butler *et al.*, 2000; Rohlf, 2001).

Rohlf (2001, 2006) pointed out that the independent contrasts method of Felsenstein (1985), as commonly implemented, is a special case of the PGLS method for the situation in which Brownian motion is the assumed evolutionary model. The advantages of the PGLS approach over phylogenetic independent contrasts are several. For example: PGLS can explicitly incorporate different evolutionary models aside from Brownian motion (e.g. Butler *et al.*, 2000; see above); the intercept of regression and other statistical models can be estimated (Rohlf, 2001); and, relevant to this study, the likelihood equations for the evolutionary rate for a single character or rate matrix for many characters are readily derived (e.g. Felsenstein, 1973; O'Meara *et al.*, 2006).

In this study, we present a PGLS approach for the estimation of the evolutionary rate matrix. We also provide likelihood tests about the rate matrix for several simple quantitative genetic hypotheses. We use individual-based, genetically explicit numerical simulations to explore the properties of the rate matrix and hypothesis tests under more realistic simulation conditions than have been used previously.

METHODS AND RESULTS

Estimating the evolutionary rate matrix

The maximum-likelihood (ML) estimate for the Brownian motion rate parameter for a single continuous trait measured in n species can be calculated as:

$$\hat{\sigma}^2 = \frac{(\mathbf{x} - \hat{a}\mathbf{1})'\mathbf{C}^{-1}(\mathbf{x} - \hat{a}\mathbf{1})}{n}$$

(Felsenstein, 1973; Garland and Ives, 2000; Freckleton *et al.*, 2002; O'Meara *et al.*, 2006). Here, $\hat{\sigma}^2$ is our estimate of the true Brownian motion rate parameter, σ^2, $\mathbf{x}$ is an $n \times 1$ column vector of the trait values at the tips, $\hat{a}$ is the phylogenetic mean for the trait, $\mathbf{1}$ is an $n \times 1$ column vector of 1.0s, and **C** is the $n \times n$ matrix with elements proportional to co-ancestry, as described above. The phylogenetic mean is equivalent to the ML estimate for the state at the root of the tree. Its specific calculation is also provided below.

For a simple tree containing few taxa, **C** is relatively easy to calculate. For example, Fig. 1 provides a four-taxon tree for which **C** is as follows:

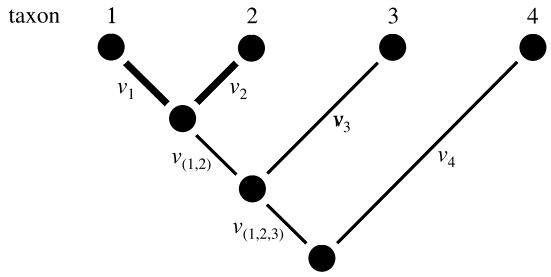

Fig. 1. A hypothetical four-taxon tree. $v_{i,j,\text{etc.}}$ indicates the branch length below the node containing descendants, i,j, etc.

$$\mathbf{C} = \begin{bmatrix} v_1 + v_{(1,2)} + v_{(1,2,3)} & v_{(1,2)} + v_{(1,2,3)} & v_{(1,2,3)} & 0.0 \\ v_{(1,2)} + v_{(1,2,3)} & v_2 + v_{(1,2)} + v_{(1,2,3)} & v_{(1,2,3)} & 0.0 \\ v_{(1,2,3)} & v_{(1,2,3)} & v_3 + v_{(1,2,3)} & 0.0 \\ 0.0 & 0.0 & 0.0 & v_4 \end{bmatrix}.$$

Here and in Fig. 1, v_i indicates the length of the branch below the tip i, and $v_{(i,j,\text{ etc.})}$ indicates the length of branch below the internal node with descendants i,j, and so on. For the case of an ultrametric tree, **C** can easily be obtained by subtracting each element of a patristic distance matrix divided by two from the total tree length. A general algorithm for the computation of **C** involves first determining the set of descendants of each node. Then, working through the branches of the tree starting from the root, each branch length, v, is added to C_{ij} if both i and j are among the descendants of the node after the branch.

Note that the scale of **C** is relevant in the interpretation of $\hat{\sigma}^2$. For example, if the branch lengths of the tree are measured in generations, then $\hat{\sigma}^2$ is an estimate of the evolutionary rate (specifically the expected rate of variance accumulation among lineages) per generation. By contrast, if the branch lengths are merely proportional to time, $\hat{\sigma}^2$ will have an expected value equal to $c \cdot \sigma^2$, where σ^2 is the true Brownian motion rate parameter and c is the proportionality constant scaling time in generations to branch length (Felsenstein, 1973).

The multivariate version of this equation is straightforward to derive. Each element of the evolutionary rate matrix, **R**, can be estimated as follows:

$$\hat{R}_{ij} = \frac{(\mathbf{x}_i - \hat{a}_i\mathbf{1})'\mathbf{C}^{-1}(\mathbf{x}_j - \hat{a}_j\mathbf{1})}{n}.$$

Here, $\mathbf{x}_i$ and $\hat{a}_i\mathbf{1}$ are $n \times 1$ vectors composed of the data for trait i for the species in the study and the phylogenetic mean for trait i, respectively. When $i = j$, the equations for $\hat{\sigma}^2$ and $\hat{R}_{ij}$ are equivalent expressions.

In matrix form, the evolutionary rate matrix, **R**, can be estimated as follows:

$$\hat{\mathbf{R}} = \frac{(\mathbf{X} - \mathbf{1}\hat{\mathbf{a}}')'\mathbf{C}^{-1}(\mathbf{X} - \mathbf{1}\hat{\mathbf{a}}')}{n}. \quad (1)$$

Here, **X** is an $n \times r$ array of trait values for n species and r traits and $\hat{\mathbf{a}}$ is an $r \times 1$ vector of the estimated phylogenetic means for each trait. **1** and **C** are as previously defined.

The equation for the calculation of $\hat{\mathbf{a}}$ (the $r \times 1$ vector of phylogenetic means) is $\hat{\mathbf{a}} = [(\mathbf{1}'\mathbf{C}^{-1}\mathbf{1})^{-1}(\mathbf{1}'\mathbf{C}^{-1}\mathbf{X})]'$ (Rohlf, 2001), in which $\mathbf{1}$, $\mathbf{C}$, and $\mathbf{X}$ are also as previously defined. In this case, unlike for $\hat{\sigma}^2$ and $\hat{\mathbf{R}}$, the same value for $\hat{\mathbf{a}}$ is obtained regardless of the scale of $\mathbf{C}$ [because $[\mathbf{1}'(k\mathbf{C})^{-1}\mathbf{1}]^{-1}[\mathbf{1}'(k\mathbf{C})^{-1}\mathbf{X}] = \frac{1}{k}(\mathbf{1}'\mathbf{C}^{-1}\mathbf{1})^{-1}k(\mathbf{1}'\mathbf{C}^{-1}\mathbf{X})$ for all $k \neq 0$].

Like many ML estimators, this estimate of $\mathbf{R}$ is expected to be biased by a factor of $(n-1)/n$. However, the bias will be slight when the number of taxa is large.

Likelihood equation for the estimator

The likelihood formula for the rate matrix estimator first requires the calculation of the expected variance–covariance matrix, $\mathbf{V}$, of the states for all traits at all tips. For a single trait and rate, $\mathbf{V}$ is equal to $\sigma^2\mathbf{C}$. In the multivariate case, $\mathbf{V}$ is a matrix of size $n \cdot r \times n \cdot r$ composed of r^2, $n \times n$ submatrices each of which is equal to $R_{ij}\mathbf{C}$. This is equivalent to the Kronecker tensor product of $\mathbf{R}$ and $\mathbf{C}$ in which every element of $\mathbf{R}$ is multiplied by every element of $\mathbf{C}$, i.e. $\mathbf{V} = \mathbf{R} \otimes \mathbf{C}$. $\mathbf{V}$ is the expected variance–covariance matrix among all observations at all tips for all traits. For example, for four taxa and two traits, $\mathbf{V}$ is computed as follows (dashed lines to demarcate submatrices are included for clarity only):

$$\mathbf{V} = \left[\begin{array}{cccc:cccc}
R_{11}C_{11} & R_{11}C_{12} & R_{11}C_{13} & R_{11}C_{14} & R_{12}C_{11} & R_{12}C_{12} & R_{12}C_{13} & R_{12}C_{14} \\
R_{11}C_{21} & R_{11}C_{22} & R_{11}C_{23} & R_{11}C_{24} & R_{12}C_{21} & R_{12}C_{22} & R_{12}C_{23} & R_{12}C_{24} \\
R_{11}C_{31} & R_{11}C_{32} & R_{11}C_{33} & R_{11}C_{34} & R_{12}C_{31} & R_{12}C_{32} & R_{12}C_{33} & R_{12}C_{34} \\
R_{11}C_{41} & R_{11}C_{42} & R_{11}C_{43} & R_{11}C_{44} & R_{12}C_{41} & R_{12}C_{42} & R_{12}C_{43} & R_{12}C_{44} \\
\hdashline
R_{21}C_{11} & R_{21}C_{12} & R_{21}C_{13} & R_{21}C_{14} & R_{22}C_{11} & R_{22}C_{12} & R_{22}C_{13} & R_{22}C_{14} \\
R_{21}C_{21} & R_{21}C_{22} & R_{21}C_{23} & R_{21}C_{24} & R_{22}C_{21} & R_{22}C_{22} & R_{22}C_{23} & R_{22}C_{24} \\
R_{21}C_{31} & R_{21}C_{32} & R_{21}C_{33} & R_{21}C_{34} & R_{22}C_{31} & R_{22}C_{32} & R_{22}C_{33} & R_{22}C_{34} \\
R_{21}C_{41} & R_{21}C_{42} & R_{21}C_{43} & R_{21}C_{44} & R_{22}C_{41} & R_{22}C_{42} & R_{22}C_{43} & R_{22}C_{44}
\end{array}\right].$$

As suggested by the preceding text, the value C_{ij} is the i,jth elements from the co-ancestry matrix, $\mathbf{C}$, while R_{ij} is the i,jth elements of the evolutionary rate matrix, $\mathbf{R}$. Obviously, for many traits and/or taxa, $\mathbf{V}$ will be a very large matrix.

As the values at the tips for all traits are expected to be distributed according to a multivariate normal distribution with variance–covariance matrix, $\mathbf{V}$, the likelihood equation for the rate matrix, given a particular value of $\mathbf{C}$ and the data at the tips, $\mathbf{X}$, is based on the multivariate normal equation and can be expressed as follows:

$$L(\mathbf{R}\,|\,\mathbf{X}, \mathbf{C}) = \frac{\exp[-(\mathbf{y} - \hat{\mathbf{b}})'\mathbf{V}^{-1}(\mathbf{y} - \hat{\mathbf{b}})/2]}{\sqrt{(2\pi)^{n \cdot r} \cdot |\mathbf{V}|}} \tag{2}$$

(Felsenstein, 1973, 1981). Here, $L(\mathbf{R}\,|\,\mathbf{X}, \mathbf{C})$ is the likelihood of the evolutionary rate matrix, $\mathbf{R}$, given a particular value for $\mathbf{X}$ and $\mathbf{C}$ [henceforward L or $L(\mathbf{R})$ will be used, for brevity]. $\mathbf{y}$ is a column vector of length $n \cdot r$ (for n taxa and r traits) composed of the trait values for all taxa for all traits. In $\mathbf{y}$, elements 1 through n are the species values for trait 1 in all taxa (column 1 in matrix $\mathbf{X}$, above); elements $n+1$ through $2n$ are the trait values for trait 2 in all taxa (column 2 in matrix $\mathbf{X}$, above); and so on. Similarly, $\hat{\mathbf{b}}$ is a column vector of length $n \cdot r$ composed of the phylogenetic means for each trait. In $\hat{\mathbf{b}}$, as for $\mathbf{y}$, above,

elements 1 through n are the phylogenetic mean for trait 1 [i.e. $(\mathbf{1}'\mathbf{C}^{-1}\mathbf{1})^{-1}(\mathbf{1}'\mathbf{C}^{-1}\mathbf{x}_1)$]; elements $n+1$ through $2n$ are the phylogenetic mean for trait 2; and so on. $\hat{\mathbf{b}}$ can be obtained from the vector of phylogenetic means, $\hat{\mathbf{a}}$, using the transformation $\hat{\mathbf{b}} = \mathbf{D}\hat{\mathbf{a}}$, in which $\mathbf{D}$ is an $n \cdot r \times r$ design matrix in which each entry D_{ij} is 1.0 if $(j-1) \cdot n < i \leq j \cdot n$ and 0.0 otherwise (Freckleton *et al.*, 2002). For example, in our four-taxon, two-trait example, $\mathbf{D}$ would be constructed as:

$$\mathbf{D} = \begin{bmatrix} 1.0 & 0.0 \\ 1.0 & 0.0 \\ 1.0 & 0.0 \\ 1.0 & 0.0 \\ 0.0 & 1.0 \\ 0.0 & 1.0 \\ 0.0 & 1.0 \\ 0.0 & 1.0 \end{bmatrix}.$$

A similar likelihood equation is used for the additive genetic variance–covariance matrix under the animal model in quantitative genetics (e.g. Shaw, 1991, equation 1). In that case, the co-ancestry matrix is based on the pedigree of the organisms in the study.

A more convenient expression for the likelihood, L, is obtained by computing the natural logarithms of the left- and right-hand sides of equation (2), in which case:

$$\log(L) = -(\mathbf{y} - \hat{\mathbf{b}})'\mathbf{V}^{-1}(\mathbf{y} - \hat{\mathbf{b}})/2 - \log|\mathbf{V}|/2 - n \cdot r \cdot \log(2\pi)/2. \quad (3)$$

In the univariate case, equation (2) (and the exponential of equation 3) reduce to the equation for the likelihood of the evolutionary rate parameter estimate, $\hat{\sigma}^2$, provided in O'Meara *et al.* (2006).

Testing hypotheses about the evolutionary rate matrix

Many hypotheses concerning multivariate trait evolution can be tested in a maximum-likelihood context. Here we focus on three very simple models, present them in some detail, discuss their quantitative genetic implications, and perform a limited set of individual-based, quantitative genetic numerical simulations to explore the power, bias, and in particular the error rates of our proposed tests.

Rate matrix equality to a hypothesized matrix

The simplest hypothesis that can be tested by this approach is a test of the hypothesis that the evolutionary rate matrix $\mathbf{R}$ is equal to some specified value, against the alternative hypothesis that $\mathbf{R}$ is equal to an arbitrary positive definite matrix (for example, its ML estimate for which the calculations are provided above). We will call the specific value of our null hypothesis $\mathbf{R}_0$: possible values for this matrix will be discussed in a later section. Suppose that some reasonable prior hypothesis for $\mathbf{R}_0$ has been proposed and we are interested in testing whether the ML estimator of $\mathbf{R}$, $\hat{\mathbf{R}}$, yields a significantly higher likelihood of our observed data than does the hypothesized rate matrix $\mathbf{R}_0$. One first calculates two likelihoods, one based on $\hat{\mathbf{R}}$ and the other on $\mathbf{R}_0$. These likelihoods can then be compared using a likelihood ratio test.

In particular, this calculation is performed by first calculating the Kronecker products $\hat{\mathbf{V}} = \hat{\mathbf{R}} \otimes \mathbf{C}$ and $\mathbf{V}_0 = \mathbf{R}_0 \otimes \mathbf{C}$, as described above. The following function for the log-likelihood ratio is then evaluated:

$$\log[L(\hat{\mathbf{R}})/L(\mathbf{R}_0)] = -(\mathbf{y}-\hat{\mathbf{b}})'\hat{\mathbf{V}}^{-1}(\mathbf{y}-\hat{\mathbf{b}})/2 - \log|\hat{\mathbf{V}}|/2 + (\mathbf{y}-\hat{\mathbf{b}})'\mathbf{V}_0^{-1}(\mathbf{y}-\hat{\mathbf{b}})/2 + \log|\mathbf{V}_0|/2\ . \quad (4)$$

Here, $\hat{\mathbf{V}}$ and $\mathbf{V}_0$ are expected variance–covariance matrices for the values for all traits at all tips given either the ML estimate of the evolutionary rate matrix (here, $\hat{\mathbf{R}}$) or the hypothesized evolutionary rate matrix, $\mathbf{R}_0$, respectively.

A couple of things are notable in equation (4). First, the choice of evolutionary rate matrix affects only the calculation of $\mathbf{V}$, not that of $\hat{\mathbf{b}}$. Second, the term $n \cdot r \cdot \log(2\pi)/2$, found in the expression for the log-likelihood (equation 3), cancels in the calculation of the log-likelihood ratio.

$2 \cdot \log[L(\hat{\mathbf{R}})/L(\mathbf{R}_0)]$ is expected to be approximately χ^2-distributed with degrees of freedom $r \cdot (r-1)/2 + r$ for r traits. This is because $r \cdot (r-1)/2 + r$ more parameters (the unique elements of $\hat{\mathbf{R}}$) are estimated in the alternative than in the fully specified null model.

Rate matrix proportionality to a hypothesized matrix

More commonly, for reasons discussed in a subsequent section, we have a matrix with which we hypothesize our evolutionary rate matrix is proportional. Under these circumstances, we must determine the value of the proportionality constant, k, that maximizes the likelihood of our data given the rate matrix $\mathbf{R}_0$. To do so, we need to evaluate the following expression for the log-likelihood:

$$\log[L(k\mathbf{R}_0)] = -(\mathbf{y}-\hat{\mathbf{b}})'(k\mathbf{V}_0)^{-1}(\mathbf{y}-\hat{\mathbf{b}})/2 - \log|k\mathbf{V}_0|/2 - n \cdot r \cdot \log(2\pi)/2$$

for possible values of k to find the value of k that maximizes the likelihood of our observed data. The following equivalent expression prevents us from having to evaluate the inverse of $k\mathbf{V}_0$ and $|k\mathbf{V}_0|$ whenever we want to evaluate the likelihood for a new value of k:

$$\log[L(k\mathbf{R}_0)] = -k^{-1}(\mathbf{y}-\hat{\mathbf{b}})'\mathbf{V}_0^{-1}(\mathbf{y}-\hat{\mathbf{b}})/2 - \log|\mathbf{V}_0|/2 - n \cdot r \cdot \log(k) - n \cdot r \cdot \log(2\pi)/2. \quad (5)$$

Once the value of k that maximizes equation (5) has been found, we can calculate the likelihood ratio $2 \cdot \log[L(\hat{\mathbf{R}})/L(k\mathbf{R}_0)]$ as in equation (4), above. Its value can be compared to a χ^2-distribution with degrees of freedom $r \cdot (r-1)/2 + r - 1$ (one fewer than in the previous model because one extra parameter, the value of k, is estimated in the denominator hypothesis).

Multiple rate matrices

Finally, we can test the hypothesis of a concerted change in all the elements of $\mathbf{R}$ on specific branches of the phylogeny or in all the lineages in a given clade or clades. This test differs from that of prior models because it must involve recalculating the co-ancestry matrix, $\mathbf{C}$. $\mathbf{C}$ is recomputed by rescaling some branches of the phylogeny by a constant, h, to be estimated using likelihood. Returning to our previous example, consider again the tree in Fig. 1 but imagine that the branches leading to taxa 1 and 2 (highlighted in **bold**) have been rescaled by a constant, h. This is similar to the test for unequal rates of O'Meara *et al.* (2006), but in that case the implementation was for a single evolving character.

$\mathbf{C}$ can be recalculated for a given value of h (or for a variety of different values for h, but this is beyond the scope of the present study) as follows:

$$\mathbf{C}_h = \begin{bmatrix} h \cdot v_1 + v_{(1,2)} + v_{(1,2,3)} & v_{(1,2)} + v_{(1,2,3)} & v_{(1,2,3)} & 0.0 \\ v_{(1,2)} + v_{(1,2,3)} & h \cdot v_2 + v_{(1,2)} + v_{(1,2,3)} & v_{(1,2,3)} & 0.0 \\ v_{(1,2,3)} & v_{(1,2,3)} & v_3 + v_{(1,2,3)} & 0.0 \\ 0.0 & 0.0 & 0.0 & v_4 \end{bmatrix}.$$

We call this matrix $\mathbf{C}_h$ to indicate that it refers to the case in which the evolutionary rate matrix is heterogeneous over time. Obviously, more complicated concerted rate heterogeneity than that portrayed in Fig. 1 and $\mathbf{C}_h$ above, perhaps affecting multiple clades or non-terminal branches, can also easily be modelled in this framework.

The likelihood equation for the rate matrix and scaling factor given the data and phylogeny is as follows:

$$\log[L(\mathbf{R}_h)] = -(\mathbf{y} - \hat{\mathbf{b}}_h)' \mathbf{V}_h^{-1} (\mathbf{y} - \hat{\mathbf{b}}_h)/2 - \log|\mathbf{V}_h|/2 - n \cdot r \cdot \log(2\pi)/2 \,. \quad (6)$$

Here, $\mathbf{V}_h = \mathbf{R}_h \otimes \mathbf{C}_h$ is the variance–covariance matrix for the values of all traits at all tips, based on the evolutionary rate matrix $\mathbf{R}_h$ estimated using equation (1). $\hat{\mathbf{b}}_h$ is a vector of the phylogenetic means, but is here calculated using $\mathbf{C}_h$ instead of $\mathbf{C}$. $\hat{\mathbf{b}}_h$ will usually be different than $\hat{\mathbf{b}}$. Unlike the situation for matrix proportionality, in this case there is no obvious way to avoid calculating the inverse of $\mathbf{V}_h$ and $|\mathbf{V}_h|$ for each value of h. One can slightly improve computation of $\mathbf{V}$ by separately calculating $\mathbf{C}$ matrices for the two rates by summing branch lengths separately into each matrix depending on their hypothesized rate category. For each value of h for which we want to obtain the likelihood, $\mathbf{C}_h$ is then calculated as $\mathbf{C}_h = \mathbf{C}_0 + h \cdot \mathbf{C}^*$, in which $\mathbf{C}_0$ is the co-ancestry matrix computed for the unscaled branches, and $\mathbf{C}^*$ is the co-ancestry matrix computed for the branches to be scaled by the unknown constant, h (e.g. Thomas *et al.*, 2006; Revell, in press).

To determine the scaling constant, a value of h is found that maximizes equation (6). The statistic $2 \cdot \log[L(\mathbf{R}_h)/L(\hat{\mathbf{R}})]$ can then be computed. Here, $\hat{\mathbf{R}}$ is the ML estimator of $\mathbf{R}$ given $\mathbf{C}$, with elements strictly proportional to branch lengths, and the data from the tips, while $\mathbf{R}_h$ is the ML estimator of $\mathbf{R}$ given $\mathbf{C}_h$, a matrix computed from a phylogenetic tree in which some branches have been rescaled by the factor h. The log-likelihood ratio statistic is expected to be distributed as a χ^2 with 1 degree of freedom, because only one further parameter, h, is estimated in the more complex model.

Error in the estimation of species means

In most studies, species means for traits are estimated with error. If an estimate of this error is available, it can be incorporated into tests about the evolutionary rate matrix. In particular, for the situation in which there is known (or estimated) error associated with species means for traits, equation (3) for the likelihood of the evolutionary rate matrix still applies. However, $\mathbf{V}$, the $n \cdot r \times n \cdot r$ expected variance–covariance matrix for the species means for all traits at all tips of the tree, is now computed as $\mathbf{V} = \mathbf{R} \otimes \mathbf{C} + \mathbf{E}$. $\mathbf{R}$ and $\mathbf{C}$ are as defined above. $\mathbf{E}$ is an $n \cdot r \times n \cdot r$ matrix containing error variances and covariances, to be described below. Given a particular value for $\mathbf{E}$, and the phylogenetic covariance matrix $\mathbf{C}$, the likelihood of any given value for the evolutionary rate matrix, $\mathbf{R}$, can be easily evaluated.

The error matrix, $\mathbf{E}$, is an $n \cdot r \times n \cdot r$ matrix. It is composed of r^2 (usually diagonal) submatrices, each $n \times n$ in size. For four taxa and two characters, as above, it will usually be of the form:

$$\mathbf{E}=\left[\begin{array}{cccc:cccc}
\sigma^2_{e,sp=1}(1) & 0.0 & 0.0 & 0.0 & \sigma_{e,sp=1}(1,2) & 0.0 & 0.0 & 0.0 \\
0.0 & \sigma^2_{e,sp=2}(1) & 0.0 & 0.0 & 0.0 & \sigma_{e,sp=2}(1,2) & 0.0 & 0.0 \\
0.0 & 0.0 & \sigma^2_{e,sp=3}(1) & 0.0 & 0.0 & 0.0 & \sigma_{e,sp=3}(1,2) & 0.0 \\
0.0 & 0.0 & 0.0 & \sigma^2_{e,sp=4}(1) & 0.0 & 0.0 & 0.0 & \sigma_{e,sp=4}(1,2) \\
\hdashline
\sigma_{e,sp=1}(2,1) & 0.0 & 0.0 & 0.0 & \sigma^2_{e,sp=1}(2) & 0.0 & 0.0 & 0.0 \\
0.0 & \sigma_{e,sp=2}(2,1) & 0.0 & 0.0 & 0.0 & \sigma^2_{e,sp=2}(2) & 0.0 & 0.0 \\
0.0 & 0.0 & \sigma_{e,sp=3}(2,1) & 0.0 & 0.0 & 0.0 & \sigma^2_{e,sp=3}(2) & 0.0 \\
0.0 & 0.0 & 0.0 & \sigma_{e,sp=4}(2,1) & 0.0 & 0.0 & 0.0 & \sigma^2_{e,sp=4}(2)
\end{array}\right]$$

in which $\sigma^2_{e,sp=i}(j)$ is the error variance for species i and character j, and $\sigma_{e,sp=i}(j,k)$ is the error covariance for species i between characters j and k. The off-diagonals of each submatrix can generally be assumed to be zero so long as error is not correlated among species. Submatrix off-diagonals might be non-zero if, for example, a subset of species were measured by one investigator and the remainder by a second (Ives *et al.*, 2007).

The likelihood of our null hypothesis for the evolutionary rate matrix, $\mathbf{R_0}$, will be very straightforward to calculate using equation (3). However, the ML estimate for the multivariate **R** will generally be difficult to obtain. Ives *et al.* (2007) discuss this problem in considerably greater detail and provide methods for optimization in an appendix. O'Meara *et al.* (2006) also discuss the univariate case, in which numerical methods for optimization are more straightforward. Suffice it say that for large r, the ML estimate for **R** will need to be found using numerical methods beyond the scope of this paper. Harmon and Losos (2005) discuss conditions under which estimation error can be ignored in comparative analyses – these might also apply to our hypothesis tests, although this should be the subject of further investigation.

Quantitative genetic implications

The approaches outlined above relate closely to key concepts in evolutionary quantitative genetics. The most obvious of these is that, under drift, the evolutionary rate matrix is expected to be proportional to the additive genetic variance–covariance matrix, or **G** matrix (Lande, 1979; Felsenstein, 1988). Specifically, if the branches of the tree are in units of organismal generations, under drift the evolutionary rate matrix has an expected value of $\mathbf{G}/\mathrm{N_e}$ (Lande, 1979).

Supposing that **G** and $\mathrm{N_e}$ are precisely known, as well as known to be constant over time, then a significant likelihood ratio ($2\cdot\log[L(\hat{\mathbf{R}})/L(\mathbf{G}/\mathrm{N_e})]$) indicates significantly non-genetic-drift-like evolution (e.g. selection); whereas a non-significant log-likelihood ratio indicates that a hypothesis of evolution by genetic drift cannot be rejected.

If **G** and $\mathrm{N_e}$ are estimated rather than known precisely, a significant test statistic could suggest either selection or that our estimates of **G** or $\mathrm{N_e}$ are imprecise in some way (perhaps because **G** fluctuates over time, or is estimated with error).

A more common situation for quantitative geneticists is that in which an estimate of **G** is available, but the scaling constant ($\mathrm{t}/\mathrm{N_e}$) is unknown. This corresponds to a variety of situations, including: (a) the phylogeny is obtained by molecular phylogenetic methods, but no relationship between sequence divergence and generation time is known; (b) the branch lengths are available in appropriate units, but the effective population size ($\mathrm{N_e}$) is unknown;

or (c), most commonly, neither the branch lengths in units of generations nor the effective size are known for the taxa in the study.

Under any of these circumstances, the hypothesis that **R** is equal to $\mathbf{G}/N_e$ becomes impossible to test, and the hypothesis of proportionality, $\mathbf{R} = k \cdot \mathbf{G}$, must be substituted. If neither the branch lengths in units of generations nor the effective size are known, then the scaling constant has an expected value equivalent to the ratio of branch length by number of generations, divided by the effective population size, N_e. Unfortunately, without precisely knowing the ratio of generations per unit branch length, it is impossible to estimate N_e from the scaling constant (and vice versa).

Considerable empirical evidence exists suggesting that the **G** matrix may not be stable over all time scales (Roff, 1997). Although a variety of manners of change in **G** are conceivable, we restrict our attention to the specific case in which **G** differs by a constant factor along a particular branch or set of branches in the phylogeny. Under drift, concerted change (be it an increase or a decrease) in the elements of **G**, all else being equal, should be reflected by a concordant change in the rate matrix, **R**.

In this case, before considering the quantitative genetic implications of a concerted change in the evolutionary rate matrix, **R**, we should first address some manner of changes in **G** that are unlikely to induce concordant change in **R**. One obvious circumstance in which the elements of **G** are expected to change by a common factor at the same time is when the effective population size changes. Since the expected value of **G** under genetic drift is $\hat{\mathbf{G}} = 2N_e\hat{\mathbf{M}}$ (Lynch and Hill, 1986; Felsenstein, 1988; Falconer and Mackay, 1996), any change in N_e is expected to result in a proportional change in **G**.

However, also under drift, the evolutionary rate matrix, **R**, has an expected value of $\mathbf{G}/N_e$ and so any change to **G** due to a change in the effective size is predicted to be balanced exactly by a change in the strength of drift, which is proportional to $1/N_e$. Not yet stated is the obvious observation that if $\hat{\mathbf{G}} = 2N_e\hat{\mathbf{M}}$ and $\mathbf{R} = \mathbf{G}/N_e$, then, by extension, $\mathbf{R} = 2 \cdot \hat{\mathbf{M}}$. Thus, the evolutionary rate matrix under genetic drift has an expected value equivalent to two times the mutation matrix, **M** (Felsenstein, 1988).

The consequence of this observation is that under pure genetic drift, changes to the evolutionary rate matrix should result only from changes in the mutation matrix, **M**. Various factors can influence the mutation matrix, in which $M_{ij} = 2m\mu\sqrt{\alpha_i^2\alpha_j^2}P_{\mu ij}$. Here, M_{ij} is the i,jth element of **M**; m is the number of genetic loci; μ is the allelic mutation rate; α_i^2 is the variance of effects of pleiotropic mutations on the ith phenotypic trait; and, finally, $P_{\mu ij}$ is the i,jth element of the correlation matrix of pleiotropic mutations $\mathbf{P}_\mu$. m might be increased by genome or gene duplication, and decreased by gene loss; μ might be increased or decreased by environmental factors, such as higher solar radiation in tropical climes (Rohde, 1978, 1992). Variances of the effects of pleiotropic mutations and $\mathbf{P}_\mu$ are not well known empirically (but see Camara *et al.*, 2000). Two recent studies (Jones *et al.*, 2007; Revell, 2007a) model the evolution of **G** when $\mathbf{P}_\mu$ changes over time.

Simulation analysis of likelihood tests

Keeping in mind the quantitative genetic implications of our hypothesis tests above, we used individual-based numerical simulations to explore the performance of likelihood tests about the evolutionary rate matrix, **R**. We used data generated by genetically explicit, individual-based, quantitative genetic simulations. More information on the specific details of the simulations can be found in Revell (2007b).

To test the type I error rate of the likelihood test of the hypothesis that the ML rate matrix $\hat{\mathbf{R}}$ is significantly more likely than some *a priori* specified rate matrix, $\mathbf{R_0}$, we used 1000 pairs of phylogenetic trees and simulated data sets generated following the approach of Revell (2007b).

We simulated all phylogenetic trees using pure-birth continuous time phylogenetic simulations. We simulated the trees to each have $n = 100$ taxa. We then rescaled all trees to have a total length of 10^4 generations, and rounded all branch lengths to have integer values (now corresponding to number of generations).

We simulated the data for four traits determined by $m = 20$ unlinked pleiotropic loci. We set the mutation rate to $\mu = 0.0025$ at all loci, and a mutation produced a new allele with new effects on all four traits. We drew mutational effects from a multivariate normal distribution with means [0, 0, 0, 0], variances $\boldsymbol{\alpha}^2 = [0.05, 0.1, 0.15, 0.20]$, and correlational mutation matrix:

$$\mathbf{P}_{\mu} = \begin{bmatrix} 1.0 & 0.75 & 0.50 & 0 \\ 0.75 & 1.0 & 0.75 & 0.50 \\ 0.50 & 0.75 & 1.0 & 0.75 \\ 0 & 0.50 & 0.75 & 1.0 \end{bmatrix},$$

which corresponds to a continuum of alleles mutation model (Crow and Kimura, 1964), with correlated effects of pleiotropic mutation. This results in an expected mutational variance–covariance matrix, $\hat{\mathbf{M}}$, with elements $M_{ij} = 2m\mu\sqrt{\alpha_i^2\alpha_j^2}P_{\mu ij}$ (Falconer and Mackay, 1996) and an additive genetic variance–covariance matrix, $\mathbf{G}$, with an expected value equal to $\hat{\mathbf{G}} = 2\mathrm{N_e}\hat{\mathbf{M}}$. $\mathrm{N_e}$ was set to 100 for all simulations. More details are provided in Revell (2007b). The quantitative genetic component of these simulations is highly similar to those performed by Jones *et al.* (2003, 2004, 2007) and Revell (2007a, 2007b).

We first determined the bias in $\mathbf{R}$ by comparing the mean value of $\mathbf{R}$ to the predicted generating matrix ($\hat{\mathbf{G}}/\mathrm{N_e}$). We found that, as expected, $\mathbf{R}$ was slightly downwardly biased, but that correction by the factor $n/(n-1)$ rendered $\mathbf{R}$ unbiased.

To determine the type I error rate of the test under the null hypothesis (drift), we first evaluated equation (4) for all 1000 trees and data sets, setting $\mathbf{R_0} = \hat{\mathbf{G}}/\mathrm{N_e}$, where the numerator of the latter was calculated from the predicted mutational variance–covariance matrix, $\hat{\mathbf{M}}$, as above. We compared the likelihood ratio $(2 \cdot \log[L(\hat{\mathbf{R}})/L(\mathbf{R_0})])$ to a χ^2-distribution with 10 degrees of freedom.

Figure 2A shows the distribution of the test-statistic compared to a χ^2-distribution with 10 degrees of freedom. The type I error rate was 0.065. Although this error rate is very close to 0.05, it is marginally significantly elevated [binomial P(true $\alpha \leq 0.05$) = 0.03].

We also evaluated equation (4) for all trees and data, but in this case set $\mathbf{R_0} = \bar{\mathbf{G}} / \mathrm{N_e}$, where $\bar{\mathbf{G}}$ is the mean value of the $\mathbf{G}$ matrix across all tips. Figure 2B shows the distribution of the test-statistic compared to a χ^2 with 10 degrees of freedom. Although quite near 0.05, the type I error rate of the test was highly significantly elevated and larger than when the theoretical value of $\mathbf{G}$ was used [error rate 0.075; P(true $\alpha \leq 0.05$) < 0.001], and the deviation in the distribution of the log-likelihood ratio statistic from χ^2 is evident (Fig. 2B).

We also conducted simulations in which the assumption of pure drift was violated to varying degrees. We conducted these simulations as above, but in this case added stabilizing selection in which the position of the optimum shifted by Brownian motion over time. We used the following multivariate Gaussian fitness function to assign individual fitness:

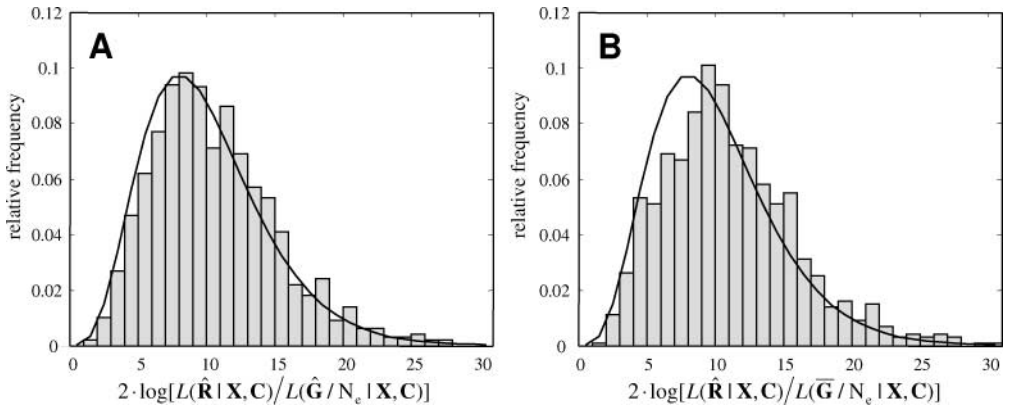

Fig. 2. Relative frequency distributions of two times the log-likelihood ratio for two hypothesis tests about the evolutionary rate matrix for continuous characters. In (A), the numerator of the log-likelihood ratios contains the likelihood of the maximum likelihood estimate of the evolutionary rate matrix, $\hat{\mathbf{R}}$, given the co-ancestry matrix, **C**, calculated from the phylogeny, and the values for phenotypic traits at the tips, **X**. The denominator of the likelihood ratios contains the likelihood of a hypothesized evolutionary rate matrix equal to the ratio of the predicted value of the additive genetic variance–covariance matrix, $\hat{\mathbf{G}}$, and the effective population size, N_e, given the data, **X**, and **C**. In (B), the numerator of the log-likelihood ratio is the same as in (A), but the denominator contains the likelihood of an evolutionary rate matrix calculated as the ratio of the mean **G** matrix calculated from the tips, $\overline{\mathbf{G}}$, and N_e, given **X** and **C**. In (A) and (B), the solid line indicates the expected distribution of $2 \cdot \log$(likelihood ratio) based on the χ^2 with degrees of freedom = 10. Results are based on 1000 individual-based, phylogenetic numerical simulations.

$$w = \exp[-\tfrac{1}{2}(\mathbf{z} - \boldsymbol{\theta})' \boldsymbol{\omega}^{-1} (\mathbf{z} - \boldsymbol{\theta})] \, .$$

Here, $\mathbf{z}$ is an $r \times 1$ column vector of individual trait values, $\boldsymbol{\theta}$ is an $r \times 1$ vector of the phenotypic optimum, and $\boldsymbol{\omega}$ is an $r \times r$ matrix that describes the shape (width and orientation) of the multivariate fitness surface.

In our simulations, the elements of the vector of the phenotypic optimum, $\boldsymbol{\theta}$, were allowed to change over time in each lineage according to independent Brownian motion processes with rate $\sigma^2_\theta = 0.01$ in each dimension. $\boldsymbol{\omega}$ was a matrix with diagonal elements, ω_{ii} set to 50, 10^2, 10^3, 10^4, and 10^5, and off-diagonal elements set to 0.0. These correspond to progressively weaker strengths of uncorrelated multivariate stabilizing selection around the optimum determined by $\boldsymbol{\theta}$. The univariate Gaussian fitness function corresponding to each value on the diagonal of the multivariate selection matrix, $\boldsymbol{\omega}$, ω_{ii}, is shown in Fig. 3 compared to the expected within-lineage phenotypic distribution for trait 1 under drift. It is clear that selection is very weak for large ω_{ii}. For each value of $\boldsymbol{\omega}$ we conducted 100 simulations.

Since selection renders the prediction of **G** impossible to obtain precisely, we evaluated equation (3) for all data setting $\mathbf{R}_0$ to $\overline{\mathbf{G}}/N_e$, with $\overline{\mathbf{G}}$ calculated from the values for **G** at the tips. Figure 4 shows the proportion of likelihood ratio tests that were significant when compared to a χ^2 with 10 degrees of freedom for each value of the diagonal elements of $\boldsymbol{\omega}$. Under all circumstances but the weakest strength of fluctuating stabilizing selection (left side of abscissa), the null hypothesis of genetic drift was almost invariably rejected by the

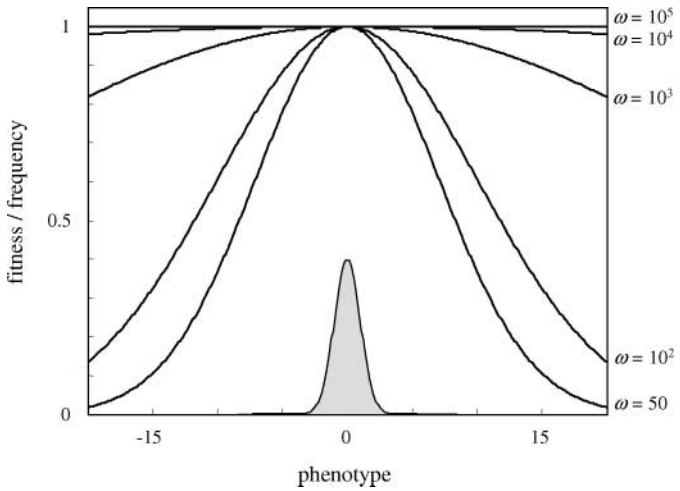

Fig. 3. Univariate representations of five multivariate Gaussian selection surfaces used in the selection simulations. The shape of multivariate stabilizing selection can be described by a matrix, **ω**. Shown is the multivariate surface along any axis for various values of the diagonal elements of **ω**. For comparison, the expected within-lineage phenotypic distribution for trait 1 is also shown as the shaded Gaussian curve.

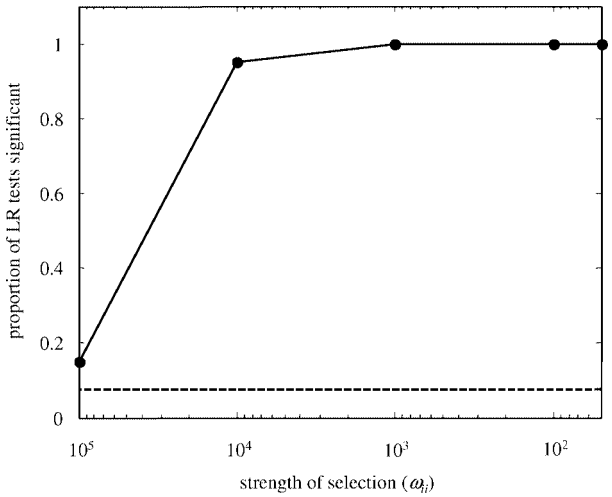

Fig. 4. The proportion of significant log-likelihood ratio (LR) tests from 100 simulations of each of several stabilizing selection simulations in which the position of the optimum moved by Brownian motion. A significant log-likelihood ratio indicates significantly non-drift-like evolution. Note that the strength of selection increases with decreasing stabilizing selection parameter, ω_{ii} (from left to right). The type I error rate under neutral drift (0.075 from simulation, in this case) is indicated by the horizontal dashed line.

likelihood ratio test (Fig. 4), indicating that the test is sensitive to detect even very weak stabilizing selection to a shifting optimum.

Thus, the ML estimate of the rate matrix is minimally biased for large numbers of taxa, n. This bias can be removed by scaling the rate matrix by $n/(n-1)$. The likelihood ratio test for

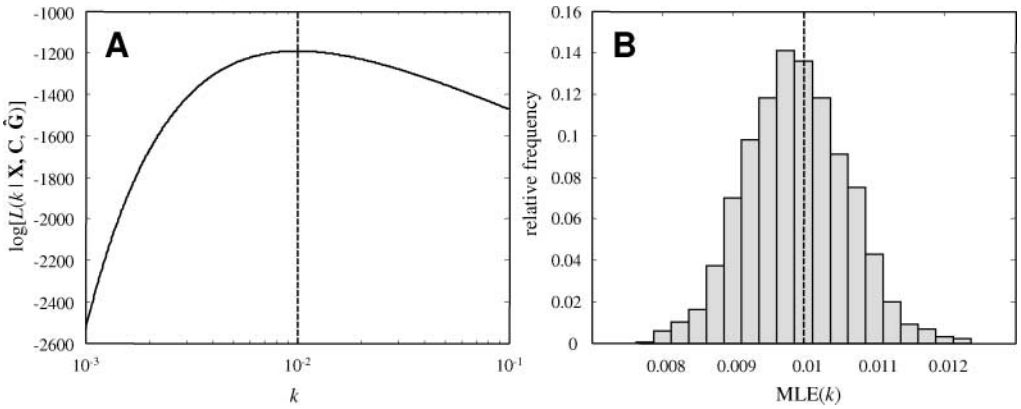

Fig. 5. (A) Example log-likelihood surface for the proportionality constant, k, in the test for matrix proportionality between the evolutionary rate matrix and the predicted value of the additive genetic variance covariance matrix, $\hat{\mathbf{G}}$. (B) Relative frequency distribution of the maximum likelihood estimate (MLE) for k from 1000 individual-based, phylogenetic numerical simulations. The expected value of k is $1/\mathrm{N_e} = 0.01$, and is indicated by the vertical dashed lines in both (A) and (B).

matrix equality between the ML estimates of $\mathbf{R}$, $\hat{\mathbf{R}}$, and $\mathbf{R_0} = \mathbf{G}/\mathrm{N_e}$ exhibited nearly appropriate type I error when the theoretical value of $\mathbf{G}$ ($\hat{\mathbf{G}} = 2\mathrm{N_e}\mathbf{M}$, for mutation matrix $\mathbf{M}$) was used (Fig. 2A). However, the error rate of the test was highly significantly elevated if the mean value of $\mathbf{G}$ for the tips ($\overline{\mathbf{G}}$) is substituted (Fig. 2B). This suggests that the theoretical value for $\mathbf{G}/\mathrm{N_e}$ provides a better estimate of the long-term average value of the evolutionary rate matrix than does $\overline{\mathbf{G}}/\mathrm{N_e}$, in which $\overline{\mathbf{G}}$ is estimated from the mean values at the tips. When stabilizing selection towards drifting phenotypic optima was simulated, the likelihood ratio test was almost invariably significant (indicating non-drift-like evolution) except for under the weakest conditions of stabilizing selection (Figs. 3, 4).

In addition to the tests of matrix equality, we also considered the more realistic situation in which an estimate of $\mathbf{G}$ (either theoretical or empirical) is available, but the factor scaling $\mathbf{G}$ to $\mathbf{R}$ is not known. This could be because the relationship between branch length and time is not known, because the effective population size is not known, or (most often) because neither is known.

Using the same data as for the previous analysis, above, and setting $\mathbf{R_0} = k\hat{\mathbf{G}}$, we evaluated equation (5) for a range of values for k. Since the expected value of k is known in this case (and is equal to $1/\mathrm{N_e}$), to find the ML estimate of the parameter k, we just tested various k, incremented and decremented by very small intervals, around the expected value.

Figure 5A shows an example log-likelihood surface for the scaling parameter, k. The mean value of k estimated by this procedure was slightly downwardly biased ($\bar{k} = 0.0099$; Fig. 5B). This is probably due to the fact that the ML estimate of $\mathbf{R}$ is also slightly biased. Rescaling k by $n/(n-1)$ rendered it unbiased.

Type I error of the likelihood ratio test was close to, but significantly higher than, 0.05, and comparable to that found for matrix equality [type I error rate: 0.065; P(true type I error ≤ 0.05) = 0.021].

In addition to comparing the evolutionary rate matrix to hypothetical matrices, we also considered the circumstance in which the evolutionary rate matrix is heterogeneous

throughout the tree. In particular, we considered the condition in which the rate matrix changes by a constant proportion along certain known branches in the phylogeny.

As previously alluded, a variety of factors might induce concerted change in all the elements of the evolutionary rate matrix. In particular, shifts in ploidy number and changes in the mutation rate are expected to induce proportionate changes in the **G** matrix and consequently in the evolutionary rate matrix. Since under an additive model a change in ploidy number will generally induce an instantaneous change in the phenotypic mean (by a factor of two), as well as in the rate of evolution, we restrict our simulations to conditions in which the mutation rate changes along certain branches of the phylogeny – because a change in the mutation rate should affect only the rate of character evolution by drift and not the population mean.

To evaluate the performance of our likelihood test for matrix heterogeneity, we used two types of simulated data. The first was 100 trees and data sets from the constant rate simulations used in the tests above. These data were used to evaluate parameter estimation (since the rate is homogeneous, the rate proportionality constant is expected to be 1.0), and to estimate the type I error of the test. The second set of simulated data was obtained using 100 individual-based phylogenetic quantitative genetic simulations in which the mutation rate was doubled along certain known branches in the phylogeny. Mutation rate increases were placed randomly in the tree at nodes with equivalent probability of 0.1, except for nodes in which a mutation rate change had preceded the node (in which case the probability of a rate change was 0.0) and for the root node (in which case the probability of a change in the mutation rate was also 0.0). To provide reasonable power for the rate heterogeneity test, only trees with a number of tips showing increased rate numbering between 30 and 70 were accepted. Otherwise, we performed these simulations using a procedure identical to that described above.

When the data were from previous, homogeneous rate simulations, the mean ML scaling factor for the rate matrix along randomly selected branches was unbiased ($\bar{h} = 1.03$). The type I error rate of the test (from 100 simulations) was not different from 0.05 [type I error rate: 0.05; P(true type I error ≤ 0.05) = 0.56].

In the heterogeneous rate simulations, the ML estimate of the branch length scaling factor, h, was also essentially unbiased ($\bar{h} = 2.02$). Figure 6A shows an example log-likelihood surface for the scaling factor, h, and Fig. 6B shows the distribution of the ML estimates of h from 100 simulations.

Thus, in the test for matrix proportionality, the proportionality constant was slightly downwardly biased by a factor of $(n - 1)/n$ (Fig. 5). The type I error of the test was low, but slightly significantly greater than 0.05. In the test for matrix heterogeneity, the scaling factors were unbiased – whether the data were generated under homogeneous or heterogeneous rate simulations (Fig. 6). The type I error of the test was appropriate in the homogeneous rate case.

DISCUSSION

We have proposed a new multivariate method for the comparative analysis of continuous trait data. The method is based on the phylogenetic generalized least-squares approach (Grafen, 1989; Martins and Hansen, 1997; Rohlf, 2001) and extends the univariate method for the analysis of evolutionary rate developed by O'Meara *et al.* (2006). We have illustrated this method

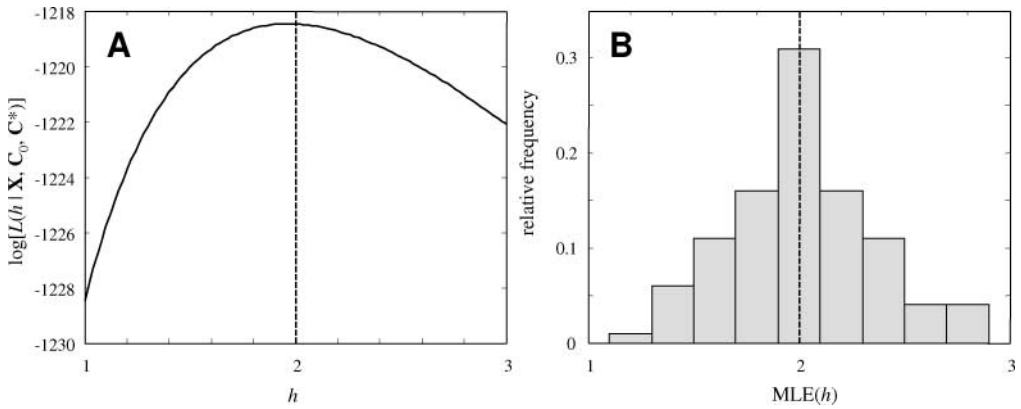

Fig. 6. (A) Example log-likelihood surface for the scaling factor, *h*, in the test for matrix heterogeneity. (B) Relative frequency distribution of the maximum likelihood estimate (MLE) for *h* from 100 individual-based, phylogenetic numerical simulations. Evolutionary rate matrix heterogeneity is due to a doubling in the mutation rate along certain branches in the phylogeny, and thus *h* has an expected value of 2.0, as shown by the vertical dashed lines in both (A) and (B).

using simulated data generated by individual-based, multivariate, quantitative genetic simulations on phylogenetic trees.

Properties of the rate matrix estimator

The evolutionary rate matrix provides, on its diagonal, the rate of evolution for each trait, and on its off-diagonal, the rates of co-evolution among characters. In this particular study we focus on the rate matrix as a quantitative genetic epiphenomenon of the within-species additive genetic variance–covariance matrix, **G**. However, this need not be the only source of the among-species rate matrix. If, for example, the position of the multivariate fitness optimum moves by Brownian motion, then the evolutionary rate matrix is appropriate to describe the movement of the optimum (O'Meara *et al.*, 2006). Nonetheless, the quantitative genetic predictions for the evolutionary rate matrix are simplest when evolution is by genetic drift, in which case the rate matrix has an expected value equivalent to the ratio of the additive genetic variance–covariance matrix and the effective population size.

The maximum likelihood (ML) estimate of the Brownian motion rate matrix is given by equation (1) in the text. We find that, like other maximum likelihood estimators, the evolutionary rate matrix of continuous characters is slightly downwardly biased – by a factor of $(n-1)/n$. This is the same bias as is found in the ML estimator of the variance and in O'Meara and colleagues' (2006) univariate rate estimator. Bias in the ML estimator for the variance comes from the fact that individual observations in a sample are expected to be closer to the sample mean than to the unknown population mean. The same is true of the Brownian motion rate matrix, although in this case the observations at tips are closer to the estimated phylogenetic mean for each trait than they are to the typically unknown true set of states for characters at the root node.

Likelihood tests about the rate matrix

We proposed two tests about the evolutionary rate matrix: a test for matrix equality and proportionality, and a test for matrix heterogeneity in different parts of the tree. In the former, the ML estimate of the evolutionary rate matrix is compared either to a hypothesized generating matrix or to a matrix hypothesized to be proportional to the generating matrix. In the latter, evolutionary rate matrices are estimated separately for different parts of the tree with the constraint that the second matrix must be a scalar multiple of the first.

Unlike previous similar studies, we chose to use individual-based quantitative genetic simulations to explore the properties of the estimator and those of our likelihood tests of the estimator. Previous studies, such as that of O'Meara *et al.* (2006), have relied on Brownian motion simulations. We chose to use individual-based genetic simulations because of the clear quantitative genetic implications of the rate matrix and tests (Felsenstein, 1988). This approach is also inherently more realistic, because real populations are composed of individuals with variable genotypes and phenotypes – not simply a wandering lineage mean. Nonetheless, increased reality of the simulations does not guarantee improved generality of the results. To some extent, results are more susceptible to be affected by the conditions of the simulation as they become more complicated. Furthermore, limits on computation require that some parameters of the simulation conditions be set with unrealistic values [although changing two or more parameters simultaneously can sometimes have a compensatory effect (e.g. Jones *et al.*, 2003; Revell, 2007a)].

Nonetheless, we conclude that the likelihood tests proposed in this study exhibited several desirable properties. First, when a theoretical value for the generating evolutionary rate matrix was available, tests of the null hypothesis that the rate matrix was equivalent to its theorized value exhibited nearly appropriate type I error (Fig. 2). Second, tests for matrix proportionality to a theorized generating matrix yielded appropriate values of the proportionality constant (Fig. 5) as well as appropriate type I error. Third, tests conducted on data simulated under selection conditions showed considerable power to detect non-drift evolution, except under the weakest level of fluctuating stabilizing selection (Figs. 3, 4). Finally, tests conducted to detect changes in the evolutionary rate matrix by a common factor both exhibited appropriate type I error under the null hypothesis of matrix constancy, and accurate parameter estimates under the alternative hypothesis (Fig. 6).

Type I error was marginally significantly elevated when a theoretical value of **G** was used. This is probably reflective in part of stochasticity in **G** around its expected value under genetic drift. To assess this possibility, we performed simple multivariate Brownian motion simulations with a constant generating matrix. We found that under these circumstances, type I error was elevated but not significantly so [type I error rate: 0.054; P(true type I error $\leq 0.05) = 0.30$].

Type I error was also slightly, but highly significantly, elevated when a theoretical value of **G** was not available for the null hypothesis, but in which the mean value of the **G** matrix from the tips of the phylogeny was substituted. This should provide a cautionary note to empiricists, to whom theoretical values of the generating evolutionary rate matrix are seldom available. In fact, since the empiricist will usually possess only estimates of **G** from a subset of the tips in the phylogeny, empirical studies using this approach will likely suffer from a considerably elevated probability of type I error – particularly if the **G** matrix is variable within the group of interest.

Finally, we also note that one could use model selection criteria [such as the Akaike information criterion (AIC)], in lieu of likelihood-ratio tests, to select among the models presented herein.

Future directions

Several future directions are suggested by this study. First, our approach assumes a Brownian motion model of evolution. Brownian motion is a suitable model for evolution by drift as well as for particular conditions of natural selection (O'Meara *et al.*, 2006). Nonetheless, Brownian motion has been criticized as representing an overly simplistic, frequently inappropriate model for the evolutionary process (Butler *et al.*, 2000; Butler and King, 2004; Estes and Arnold, 2007). Fortunately, Brownian motion is not a constraint of this method. Since the co-ancestry matrix, **C**, can be calculated from the phylogeny under a variety of evolutionary processes (Butler and King, 2004), so can comparative methods be extended to a multivariate context for non-Brownian motion models of evolution, at least in theory.

Second, our approach assumes that the species mean trait values, phylogeny, and phylogenetic branch lengths are known without error. This is true for numerical simulations, in which the generating phylogeny and data are known precisely, but it will not usually be true for empirical studies. We discuss error in species mean trait values in a prior section. Phylogenetic topology and branch length misestimation will probably affect the error rate of the tests. In fact, if the data and phylogenetic trees were assorted randomly, likelihood ratio tests would invariably suggest significantly non-drift-like evolution. To incorporate phylogenetic error, empiricists might consider conducting the likelihood tests for their data using a sample of trees from the posterior distribution in a Bayesian analysis (Revell *et al.*, 2007). Biased branch length estimation (e.g. Revell *et al.*, 2005) will also tend to affect the test for matrix heterogeneity. In particular, under circumstances in which early branches in the phylogeny are underestimated, rate heterogeneity calculations will tend to suggest that the rate of phenotypic evolution has decreased over time (Revell *et al.*, 2005).

Finally, the evolutionary rate matrix can change in more ways than simply by a scalar multiple. Not explored in this study are circumstances in which the evolutionary rate matrix retains all its eigenvectors, but changes its eigenvalues; changes some but not all eigenvectors; or changes entirely, retaining no common eigenstructure (Phillips and Arnold, 1999). There is no simple quantitative genetic model under which these changes to the evolutionary rate matrix are expected, although the empirical literature suggests that such changes do occur (Steppan *et al.*, 2002).

ACKNOWLEDGEMENTS

This work was supported in part by a grant from the National Science Foundation (DEB-0519777). P. Chesson provided very helpful comments on an earlier version of the article.

REFERENCES

Baker, R.H. and Wilkinson, G.S. 2003. Phylogenetic analysis of correlation structure in stalk-eyed flies (*Diasemopsis*, Diopsidae). *Evolution*, **57**: 87–103.

Bégin, M. and Roff, D.A. 2004. From micro- to macroevolution through quantitative genetic variation: positive evidence from field crickets. *Evolution*, **58**: 2287–2304.

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

Butler, M.A., Schoener, T.W. and Losos, J.B. 2000. The relationship between sexual size dimorphism and habitat use in Greater Antillean *Anolis* lizards. *Evolution*, **54**: 259–272.

Camara, M.D., Ancell, C.A. and Pigliucci, M. 2000. Induced mutations: a novel tool to study phenotypic integration and evolutionary constraints in *Arabidopsis thaliana. Evol. Ecol. Res.*, **2**: 1009–1029.

Crow, J.F. and Kimura, M. 1964. The theory of genetic loads. In *Proceedings of the XIth International Congress of Genetics* (S.J. Geerts, ed.), pp. 495–505. Oxford: Pergamon Press.

Edwards, A.W.F. and Cavalli-Sforza, L.L. 1964. Reconstruction of evolutionary trees. In *Phenetic and Phylogenetic Classification* (V.H. Heywood and J. McNeill, eds.), pp. 67–76. London: Systematics Association.

Estes, S. and Arnold, S.J. 2007. Resolving the paradox of stasis: models with stabilizing selection explain evolutionary divergence on all timescales. *Am. Nat.*, **169**: 227–244.

Falconer, D.S. and Mackay, T.F.C. 1996. *Introduction to Quantitative Genetics*, 4th edn. Harlow, UK: Prentice-Hall.

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

Felsenstein, J. 1981. Evolutionary trees from gene frequencies and quantitative characters: finding maximum likelihood estimates. *Evolution*, **35**: 1229–1242.

Felsenstein, J. 1985. Phylogenies and the comparative method. *Am. Nat.*, **125**: 1–15.

Felsenstein, J. 1988. Phylogenies and quantitative characters. *Annu. Rev. Ecol. Syst.*, **19**: 445–471.

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

Felsenstein, J. 2005. Using the quantitative genetic threshold model for inferences between and within species. *Phil. Trans. R. Soc. Lond. B*, **360**: 1427–1434.

Freckleton, R.P., Harvey, P.H. and Pagel, M. 2002. Phylogenetic analysis and comparative data: a test and review of evidence. *Am. Nat.*, **160**: 712–726.

Garland, T., Jr. 1992. Rate tests for phenotypic evolution using phylogenetically independent contrasts. *Am. Nat.*, **140**: 509–519.

Garland, T., Jr. and 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.

Grafen, A. 1989. The phylogenetic regression. *Phil. Trans. R. Soc. Lond. B*, **326**: 119–157.

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

Hansen, T.F. and Martins, E.P. 1996. Translating between microevolutionary process and macroevolutionary patterns: the correlation structure of interspecific data. *Evolution*, **50**: 1404–1417.

Harmon, L.J. and Losos, J.B. 2005. The effect of intraspecific sample size on Type I and Type II error rates in comparative studies. *Evolution*, **59**: 2705–2710.

Harvey, P.H. and Pagel, M.D. 1991. *The Comparative Method in Evolutionary Biology*. Oxford: Oxford University Press.

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

Jones, A.G., Arnold, S.J. and Bürger, R. 2003. Stability of the **G**-matrix in a population experiencing pleiotropic mutation, stabilizing selection, and genetic drift. *Evolution*, **57**: 1747–1760.

Jones, A.G., Arnold, S.J. and Bürger, R. 2004. Evolution and stability of the **G**-matrix on a landscape with a moving optimum. *Evolution*, **58**: 1639–1654.

Jones, A.G., Arnold, S.J. and Bürger, R. 2007. The mutation matrix and the evolution of evolvability. *Evolution*, **61**: 727–745.

Lande, R. 1979. Quantitative genetic analysis of multivariate evolution, applied to brain : body size allometry. *Evolution*, **33**: 402–416.

Lande, R. and Arnold, S.J. 1983. The measurement of selection on correlated characters. *Evolution*, **37**: 1210–1226.

Lynch, M. and Hill, W.G. 1986. Phenotypic evolution by neutral mutation. *Evolution*, **40**: 915–935.

Lynch, M. and Walsh, B. 1998. *Genetics and Analysis of Quantitative Traits*. Sunderland, MA: Sinauer Associates.

Martins, E.P. and 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.

McPeek, M.A. 1995. Testing hypotheses about evolutionary change on single branches of a phylogeny using evolutionary contrasts. *Am. Nat.*, **145**: 686–703.

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

Phillips, P.C. and Arnold, S.J. 1999. Hierarchical comparison of genetic variance–covariance matrices. I. Using the Flury hierarchy. *Evolution*, **53**: 1506–1515.

Revell, L.J. 2007a. The **G** matrix under fluctuating correlational mutation and selection. *Evolution*, **61**: 1857–1872.

Revell, L.J. 2007b. Testing the genetic constraint hypothesis in a phylogenetic context: a simulation study. *Evolution*, **61**: 2720–2727.

Revell, L.J. in press. On the analysis of evolutionary change along single branches in a phylogeny. *Am. Nat*.

Revell, L.J., Harmon, L.J. and Glor, R.E. 2005. Underparameterized model of sequence evolution leads to bias in the estimation of diversification rates from molecular phylogenies. *Syst. Biol.*, **54**: 973–983.

Revell, L.J., Harmon, L.J., Langerhans, R.B. and Kolbe, J.J. 2007. A phylogenetic approach to determining the importance of constraint on phenotypic evolution in the neotropical lizard *Anolis cristatellus*. *Evol. Ecol. Res.*, **9**: 261–282.

Roff, D.A. 1997. *Evolutionary Quantitative Genetics*. New York: Chapman & Hall.

Rohde, K. 1978. Latitudinal gradients in species diversity and their causes. I. A review of the hypotheses explaining the gradients. *Biologisches Zentralblatt*, **97**: 393–403.

Rohde, K. 1992. Latitudinal gradients in species diversity: the search for the primary cause. *Oikos*, **65**: 514–527.

Rohlf, F.J. 2001. Comparative methods for the analysis of continuous variables: geometric interpretations. *Evolution*, **55**: 2143–2160.

Rohlf, F.J. 2006. A comment on phylogenetic correction. *Evolution*, **60**: 1509–1515.

Schluter, D. 1996. Adaptive radiation along genetic lines of least resistance. *Evolution*, **50**: 1766–1774.

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

Shaw, R.G. 1991. The comparison of quantitative genetic parameters between populations. *Evolution*, **45**: 143–151.

Steppan, S.J., Phillips, P.C. and Houle, D. 2002. Comparative quantitative genetics: evolution of the **G** matrix. *Trends Ecol. Evol.*, **17**: 320–327.

Thomas, G.H., Freckleton, R.P. and Székely, T. 2006. Comparative analyses of the influence of developmental mode on phenotypic diversification rates in shorebirds. *Proc. R. Soc. Lond. B*, **273**: 1619–1624.