PhyloPars methodology

Introduction

PhyloPars reconstructs missing feature values of the nodes (species or strains) of a phylogeny, using a limited number of observations. The main idea behind the method is that the closer related a node is, the more it will tell us about a missing feature value. Specifically, feature values are assumed to change through genetic drift only, with the rate of change (mutation) being constant for the entire phylogeny. Evolutionary changes in the value of different features are allowed to be correlated, such as would be the case for the change in length and surface area, for instance. Additionally, the model allows for phenotypic variability of features: the fact that an observation or measurement is not necessarily the exact feature value of the representative node. This may be the case when there exists within-species or within-strain variation of the feature value, or when observed values are affected by measurement error.

The phylogenetic model used is the wide-spread "Brownian motion" model of evolution (1,2), formulated for multiple features (3-5). Phenotypic variability is added as an additional layer of variability between the species and sample level. The level of phenotypic variability is taken be feature-specific and constant across the phylogeny; its value is estimated from the observations (5). This contrasts with the approach by Ives et al (6), who allow for a variable (node-specific), pre-specified phenotypic variability. Conceptually the PhyloPars methodology is nearest to that employed by Felsenstein (5), but it differs in two aspects. First, PhyloPars assumes different observations on a single species to be independent, which formally implies that phenotypic correlations equal zero. Second, the model is extended with the facility of handling missing data: values in the feature matrix may be sampled 0, 1 or more times. This contrasts with the approach by Felsenstein, which requires for any sampled individual or population that all feature values are measured.

Broadly, there exist three approaches to mathematically develop and analyze the conceptual model introduced in the previous paragraph: Phylogenetically Independent Contrasts (PIC) (3,5), Generalized Least Squares (GLS) (7-9), and the Phylogenetic Mixed Model (PMM) (4,10). All build upon the same assumptions and tackle very similar conceptual problems. It is therefore not surprising that they can be shown to be closely related, and to deliver identical results for some models (5,10-12). Perhaps the main difference between the methods lies in the questions that they typically address: both PIC and GLS are often used directly to obtain either estimates of ancestral feature values (13,14) or to perform univariate phylogenetic regressions (12). Accordingly, mathematical theory and analytical formulae have been developed especially for these purposes. The PMM on the other hand aims to first (numerically) reconstruct the parameters of the full evolutionary model (specifically, the phylogenetic covariance matrix), which can then serve a variety of purposes including ancestral state reconstruction, univariate and multivariate phylogenetic regression analysis, and phylogenetic principal component analysis. This led Housworth et al. (10) to suggest that the PMM can be more informative than PIC and GLS-based counterparts. However, it should be stressed that the difference between the methods lies primarily in the typical application of

the methods, and not in their base assumptions and model formulation. For instance, Felsenstein (5) uses a PIC-based method to obtain all phylogenetic and phenotypic parameters of the evolutionary model.

In a sense, PhyloPars incorporates elements from both GLS- as well as PIC-based methods. First, the complete normal multivariate model is formulated for multiple nodes and multiple traits, describing the likelihood of observing all sampled variables. This would be the starting point for a GLS analysis. Second, contrasts (but not *independent* contrasts) are introduced to eliminate the phylogenetic mean of the features from the problem, not unlike the approach followed by Grafen (7). At this point, however, the inclusion of phenotypic variability has made analytical reconstruction of the model parameters impossible – commonly used analytical expressions for the value and confidence intervals of ancestral states and phylogenetic regression coefficients do not apply. As in comparable studies (5,6,10), numerical routines must be used to identify the parameters that maximize the likelihood.

Input

Let us start with a completely known phylogeny that contains a total of M nodes, and a limited set of observations on N features of a subset of nodes. Multiple observations on a single node are assumed to be independent (cf. 5). The whole set of observations is assumed to incomplete and may contain duplicates: any element of the $M \times N$ feature matrix may be sampled 0, 1 or more times. As the number of observations can differ between features, we define for each feature i its distinct set of observations as vector $\mathbf{y}_i = \left(y_{i1}, \ldots y_{iK_i}\right)^T$. The indices of the node to which each observation pertains will be denoted by corresponding vector $\mathbf{m}_i = \left(m_{i1}, \ldots m_{iK_i}\right)^T$. The ordering of observations is irrelevant: values in $\mathbf{m}_i$ do not need to increase monotonously. Also, as multiple observations on the value of a specific feature for a single node may be available, each $\mathbf{m}_i$ may contain duplicates.

Phylogenetic variability: genetic drift

If the value of features changes through genetic drift only, features can be considered to perform a random walk in evolutionary time. This corresponds to the Brownian motion model for evolution of continuous features (1). This is by no means the only model of feature evolution (15,16), but it is perhaps the simplest model possible, and it mathematical consequences can be motivated independently on first-principle statistical grounds (7,12,17). The Brownian motion model specifies that if feature values $(x_1, \ldots x_N)^T$ are known, the probability distribution of the feature values at time interval t later is described by a multivariate normal distribution with mean $(x_1, \ldots x_N)^T$ and covariances $a_{ij}\,t$, with a_{ij} defined as the covariance of evolutionary change between features i and j per unit time (i.e., branch length). Following Felsenstein (5), we will refer to these as phylogenetic covariances. These covariances are easily transformed into phylogenetic correlations and phylogenetic regression slopes. Since the intercept of phylogenetic regressions is identical to the feature value of the root node (12), which is reconstructed along with the values of all other nodes at the final step in the analysis, the present method also may be used to perform phylogenetic regression analyses.

Let us now consider the (rooted) phylogeny with the true value (cf., the observed value) of feature i for node k denoted by x_{ik}; $k = 0$ will be used to denote the root of the phylogeny. The feature values of any given node have been determined by the random walk that started from the root node. Lets us denote the set of branch segments that describes the evolutionary path from the root to a node k as P_k, and the sum of the lengths of the segments in such a set as $l(P_k)$. For this node k the feature values can now be described by a multivariate normal distribution with the mean equal to the feature values of the root node $(x_{10}, \dots x_{N0})^T$ and covariances equal to $a_{ij}\, l(P_k)$. It is not hard to see that the feature values of the different nodes must be correlated if the nodes' evolutionary paths from the root node (partially) overlap: these nodes have shared all changes in feature values from the root till their last common ancestor. This implies that the covariance between a feature value x_{ik} and any other feature value x_{jl} is given by $\mathrm{Cov}(x_{ik}, x_{jl}) = a_{ij}\, T_{kl}$, with $T_{kl} = l(P_k \cap P_l)$ simply equaling the length of the path from the root till the last common ancestor of both nodes.

Phenotypic variability: intraspecific variation and measurement error

In most cases, observations on the feature value of a given node will be subject to additional variability: different individuals of a node may have different feature values (intraspecific variability), and measurements of the feature values may be imperfect (measurement error). Following Felsenstein (5), we will refer to these as sources as phenotypic variability. We will assume that the uncertainty due to each phenotypic source of variation can be described by a normal distribution centered at the true feature value of the respective node. As a result, the combined effect of intraspecific variability and measurement error again be described by a normal distribution, which has a mean equal to the true mean of the feature value of the node and a variance that equals the sum of the variances of the two source of phenotypic variability. We will refer to these combined variances b_{ii} as phenotypic variances. It is worth noting that as we assume that the observations on a given node are independent, phenotypic correlations are all zero (cf. 5). The square root of the phenotypic variance may also be interpreted as the standard deviation expected for multiple samples taken from a single species, which thus is taken to be the same for all species (cf. 6).

Complete model

This completes the information needed for the model specification. We now combine all observations in a single vector $\mathbf{y} = (\mathbf{y}_1^T, \dots, \mathbf{y}_N^T)^T$, noting that observations on a single feature thus remain contiguous. The likelihood of this set of observations equals a multivariate normal distribution that combines phylogenetic and phenotypic components.

The phylogenetic model specifies that the expected feature value of any node equals the value of the corresponding feature for the root node, while the phenotypic model does not affect the observed mean. Thus the expectation of the distribution equals $\bar{\mathbf{y}} = (\bar{\mathbf{y}}_1^T, \dots, \bar{\mathbf{y}}_N^T)^T$, with each $\bar{\mathbf{y}}_i$ denoting a vector of length K_i and elements equal to $x_{i,0}$.

Since the effect of both phylogenetic and phenotypic processes can be described by a normal distribution and the phenotypic process has a zero mean, the chaining of these processes simply results in addition of the corresponding covariances (4-6,10). The base covariance

between two observations is specified by the phylogenetic model, and incorporates a phenotypic component only if the two observations are in fact the same (i.e., only variances incorporate a phenotypic component), as phenotypic correlations equal zero. The covariance between observation $p \in \{1, \dots, K_i\}$ on feature i and observation $q \in \{1, \dots, K_j\}$ on feature j thus is given by $\mathrm{Cov}(y_{ip}, y_{jq}) = a_{ij}\, T_{m_{ip}, m_{jq}} + \delta_{ij}\, \delta_{pq}\, b_{ij}$, with each δ representing the Kronecker delta that equals 1 if its subscript indices are equal and 0 otherwise. Thus the covariance matrix may be viewed as a $N \times N$ block matrix, with each feature-specific block $\mathbf{S}_{ij}$ representing a $K_i \times K_j$ matrix containing elements $\mathrm{Cov}(y_{ip}, y_{jq})$ with $p \in \{1, \dots, K_i\}, q \in \{1, \dots, K_j\}$. As **a**, **b** and **T** are all symmetric, it is not difficult to see that the resulting covariance matrix must be symmetric as well.

One could now maximize the likelihood in order to obtain the phylogenetic covariances a_{ij} and phenotypic variances b_{ii}, for $i, j = 1, \dots, N$. However, the likelihood also contains the N unknown feature values of the root node x_{i0}, which then would have to be estimated as well. This is possible and would not necessarily increase the number of free parameters, as one could "profile out" the root feature values by inserting their (analytically obtained) optimum values for any given estimate set of phylogenetic covariances and phenotypic variances. However, such joint estimation of the mean and covariance is well-known to induce a bias in the estimate of the covariances (18); in order to obtain an unbiased maximum likelihood estimator, we first rephrase the model in terms that do not include the root feature values.

Introducing contrasts

Following Felsenstein (1,3,5), we first rephrase the model in terms of contrasts: the difference between two observations on the same feature. Notably, however, we use contrasts with an arbitrarily chosen reference observation, rather than *independent* contrasts. We set aside one observation for each feature for use as reference: from each observation vector $\mathbf{y}_i$ we extract one element denoted as y_{i0}, reducing the vector length with 1; correspondingly we extract element m_{i0} from $\mathbf{m}_i$. Any element in the observation vector may be chosen as reference, which agrees with the fact the ordering of observations in $\mathbf{y}_i$ is irrelevant. We now define the contrast vector $\boldsymbol{\Delta}\mathbf{y}_i = \mathbf{y}_i - y_{i0}$ for each feature. The likelihood of observing contrasts $\boldsymbol{\Delta}\mathbf{y} = (\boldsymbol{\Delta}\mathbf{y}_1^T, \dots, \boldsymbol{\Delta}\mathbf{y}_N^T)^T$ again equals a multivariate normal distribution (19). The mean of this distribution is a null vector, as the expectation for each y_{ip} is the root feature value x_{i0}. The covariances of the contrasts can be derived from the covariances of $\mathbf{y}_i$ in a straightforward fashion:

$$\begin{aligned}\mathrm{Cov}(y_{ip} - y_{i0}, y_{jq} - y_{j0}) &= \mathbb{E}([y_{ip} - y_{i0}][y_{jq} - y_{j0}]) - \mathbb{E}(y_{ip} - y_{i0})\mathbb{E}(y_{jq} - y_{j0}) \\ &= \mathbb{E}(y_{ip} y_{jq}) - \mathbb{E}(y_{ip} y_{j0}) - \mathbb{E}(y_{i0} y_{jq}) + \mathbb{E}(y_{i0} y_{j0}) \\ &\quad - \mathbb{E}(y_{ip})\mathbb{E}(y_{jq}) + \mathbb{E}(y_{ip})\mathbb{E}(y_{j0}) + \mathbb{E}(y_{i0})\mathbb{E}(y_{jq}) - \mathbb{E}(y_{i0})\mathbb{E}(y_{j0}) \\ &= \mathrm{Cov}(y_{ip}, y_{jq}) - \mathrm{Cov}(y_{ip}, y_{j0}) - \mathrm{Cov}(y_{i0}, y_{jq}) + \mathrm{Cov}(y_{i0}, y_{j0})\end{aligned}$$

Inserting $\mathrm{Cov}(y_{ip}, y_{jq}) = a_{ij}\, T_{m_{ip}, m_{jq}} + \delta_{ij}\, \delta_{pq}\, b_{ij}$ that was obtained previously, we get

$$\mathrm{Cov}(y_{ip} - y_{i0}, y_{jq} - y_{j0}) = a_{ij}\left(T_{m_{ip} m_{jq}} - T_{m_{ip} m_{j0}} - T_{m_{i0} m_{jq}} + T_{m_{i0} m_{j0}}\right) + \delta_{ij}\left(1 + \delta_{pq}\right) b_{ij}$$

Thus the covariance matrix for $\mathbf{\Delta y}$ may be viewed as a $N \times N$ block matrix, with each feature-specific block $\mathbf{S}_{ij}$ representing a $K_i \times K_j$ matrix containing elements $\mathrm{Cov}(y_{ip} - yi0, yjq - yj0$ with $p \in 1,\ldots,Ki, q \in 1,\ldots,Kj$. It is worth remarking that the phenotypic variances contribute first to the block diagonal (due to the use of an observed feature value as reference in the contrasts), and again to the element diagonal (due to the individual observations).

The likelihood of observing contrasts $\mathbf{\Delta y}$ thus equals

$$L(\mathbf{\Delta y}|\mathbf{a},\mathbf{b}) = \frac{1}{(2\pi)^{n/2}|\mathbf{\Sigma}(\mathbf{a},\mathbf{b})|^{1/2}} e^{-\frac{1}{2}\mathbf{\Delta y}^T[\mathbf{\Sigma}(\mathbf{a},\mathbf{b})]^{-1}\mathbf{\Delta y}}$$

with $\mathbf{\Sigma}(\mathbf{a},\mathbf{b})$ denoting the covariance matrix and $|\mathbf{\Sigma}(\mathbf{a},\mathbf{b})|$ its determinant.

Estimating phylogenetic and phenotypic covariances

We now have an expression for the likelihood that depends only on the phylogenetic and phenotypic covariances. For optimization purposes it is often easier to work with the ln-likelihood, i.e.,

$$\Lambda(\mathbf{\Delta y}|\mathbf{a},\mathbf{b}) = -\tfrac{1}{2}n \ln 2\pi - \tfrac{1}{2}\ln|\mathbf{\Sigma}(\mathbf{a},\mathbf{b})| - \tfrac{1}{2}\mathbf{\Delta y}^T[\mathbf{\Sigma}(\mathbf{a},\mathbf{b})]^{-1}\mathbf{\Delta y}$$

The best estimates for $\mathbf{a}$ and $\mathbf{b}$ are given by those values that maximize the (ln-)likelihood. Note that the constant $-\frac{1}{2}n \ln 2\pi$ does not affect the position of the optimum of $\Lambda(\mathbf{\Delta y}|\mathbf{a},\mathbf{b})$ in parameter space, and can therefore be omitted.

To permit unconstrained maximization of the ln-likelihood, it is first rephrased in terms of the log Cholesky parameterization (20) of the phylogenetic covariances, and the logarithm of the phenotypic variances. This permits unconstrained optimization while ensuring that the phylogenetic covariance matrix remains positive definite, and phenotypic variances remain positive. The Broyden-Fletcher-Goldfarb-Shanno algorithm (21) is then used for unconstrained minimization of the negative ln-likelihood.

The inverse and determinant of the covariance matrix $\mathbf{\Sigma}(\mathbf{a},\mathbf{b})$ at each evaluation of the likelihood is calculated through Cholesky decomposition of the matrix. This is the most computationally expensive step in the procedure, since the size of the (square) matrix equals the total number of feature value observations, which may be very large (289 in the PhyloPars test case).

The likelihood maximization procedure must be provided with an initial estimate of phylogenetic covariances and phenotypic variances. To obtain such an estimate, we first calculate the optimal phylogenetic variances for each feature in the absence of phylogenetic correlations and phenotypic variability. In the absence of phylogenetic correlations, observations on different features are uncorrelated and the likelihood reduces to the product of feature-specific multivariate normal distributions. Each of these distributions has a zero mean and a covariance matrix $\tilde{\mathbf{S}}_{ii}$ that depends only on the feature's phylogenetic and phenotypic

variance; these components of the likelihood can therefore be maximized individually. If we neglect the phenotypic component, the feature-specific covariance matrix can be written as

$$\tilde{\mathbf{S}}_{ii} = \tilde{a}_{ii}\mathbf{T}_i \quad \text{with } T_{i,pq} = T_{m_{ip}m_{iq}} - T_{m_{ip}m_{i0}} - T_{m_{i0}m_{iq}} + T_{m_{i0}m_{i0}} \; \forall\, p,q \in \{1,\dots,K_i\}$$

The optimal $\tilde{a}_{ii}$ then is the one that maximizes

$$-\tfrac{1}{2}\ln|\tilde{a}_{ii}\mathbf{T}_i| - \tfrac{1}{2}\boldsymbol{\Delta}\mathbf{y}_i{}^T[\tilde{a}_{ii}\mathbf{T}_i]^{-1}\boldsymbol{\Delta}\mathbf{y}_i$$

which is found to equal

$$\tilde{a}_{ii} = \frac{\boldsymbol{\Delta}\mathbf{y}_i{}^T\mathbf{T}_i{}^{-1}\boldsymbol{\Delta}\mathbf{y}_i}{K_i}$$

recalling that K_i denotes the length of $\boldsymbol{\Delta}\mathbf{y}_i$. Further, we calculate the optimal phenotypic variances in the absence of phylogenetic variability. This is simply the variance of the observations for each feature, i.e., $\tilde{b}_{ii} = \mathrm{var}(\mathbf{y}_i)$, with $\mathbf{y}_i$ taken before reference element y_{i0} was extracted. As initial guess, we specify that half of the total variability is due to phylogenetic components, and half to phenotypic components. Initial estimates of phylogenetic and phenotypic variances are thus set to $\frac{1}{2}\tilde{a}_{ii}$ and $\frac{1}{2}\tilde{b}_{ii}$, respectively (recall that phylogenetic correlations are initially set to zero).

Reconstructing missing feature values

With phylogenetic covariances and phenotypic variances known, the mean values for all features of all nodes can be reconstructed. The values to estimate are denoted by vector $\mathbf{y}^*$, consisting of N feature-specific stacked vectors $\mathbf{y}_i^*$ of length M (one element per node). Again, contrasts are taken with the previously selected reference observations: $\boldsymbol{\Delta}\mathbf{y}_i^* = \mathbf{y}_i^* - y_{i0}$. These are combined in a single contrast vector $\boldsymbol{\Delta}\mathbf{y}^* = (\boldsymbol{\Delta}\mathbf{y}_1^{*T},\dots,\boldsymbol{\Delta}\mathbf{y}_N^{*T})^T$.

The desired feature values $\mathbf{y}^*$ can be estimated by finding the contrasts $\boldsymbol{\Delta}\mathbf{y}^*$ that maximize the likelihood, given the observed contrasts $\boldsymbol{\Delta}\mathbf{y}$, as well as the previously estimated phylogenetic covariances $\mathbf{a}$ and phenotypic variances $\mathbf{b}$. This likelihood is described by a multivariate normal distribution of the combined contrasts $\overline{\boldsymbol{\Delta}\mathbf{y}} = (\boldsymbol{\Delta}\mathbf{y}^{*T},\boldsymbol{\Delta}\mathbf{y}^T)^T$. As before, the mean of this distribution equals zero, since the expectation of the two terms in any contrast is identical (namely, the corresponding feature value of the root node). The covariance matrix of the distribution can be partitioned as

$$\overline{\boldsymbol{\Sigma}} = \begin{pmatrix} \boldsymbol{\Sigma}^* & \boldsymbol{\Sigma}^{\times} \\ \boldsymbol{\Sigma}^{\times T} & \boldsymbol{\Sigma} \end{pmatrix}$$

The lower right block describes the covariances between *observed* contrasts, and thus equals the covariance matrix $\boldsymbol{\Sigma}$ that was derived previously. The upper left block $\boldsymbol{\Sigma}^*$ describes the covariances between the *desired* contrasts:

$$\mathrm{Cov}\left(\mathrm{y}_{ip}^* - y_{i0}, \mathrm{y}_{jq}^* - y_{j0}\right) = a_{ij}\left(T_{pq} - T_{pm_{j0}} - T_{m_{i0}q} + T_{m_{i0}m_{j0}}\right) + \delta_{ij}\,b_{ij}$$

with $p, q \in \{1, \ldots, M\}$. It may be noted that phenotypic variability contributes at most once to the covariance: when both reference observations (y_{i0}, y_{j0}) are equal. It does not additionally contribute to the variance (diagonal elements of $\mathbf{\Sigma}^*$) if y^*_{ip} and y^*_{jq} are equal, as desired feature values (y^*_{ip}, y^*_{jq}) pertain to the nodes (the species mean), rather than to the individual samples; therefore they are not subject to phenotypic variability.

The off-diagonal block $\mathbf{\Sigma}^{\times}$ and its transpose $\mathbf{\Sigma}^{\times T}$ describe covariances between elements of $\mathbf{\Delta y}^*$ and elements of $\mathbf{\Delta y}$:

$$\mathrm{Cov}\left(y^*_{ip} - y_{i0}, y_{jq} - y_{j0}\right) = a_{ij}\left(T_{pm_{jq}} - T_{pm_{j0}} - T_{m_{i0}m_{jq}} + T_{m_{i0}m_{j0}}\right) + \delta_{ij}\, b_{ij}$$

with $p \in \{1, \ldots, M\}, q \in \{1, \ldots, K_j\}$. Since these covariances describe off-diagonal elements of the combined covariance matrix $\overline{\mathbf{\Sigma}}$ only, phenotypic variances contribute at most once to the covariance (when reference observations y_{i0} and y_{j0} are equal).

To obtain estimates for $\mathbf{\Delta y}^*$, the likelihood can be rephrased as the distribution of $\mathbf{\Delta y}^*$ conditional on $\mathbf{\Delta y}$. This distribution is again multivariate normal with mean $\mathbf{\Sigma}^{\times}\mathbf{\Sigma}^{-1}\mathbf{\Delta y}$ and covariance matrix $\mathbf{\Sigma}^* - \mathbf{\Sigma}^{\times}\mathbf{\Sigma}^{-1}\mathbf{\Sigma}^{\times T}$ (19). These directly specify estimates of the desired contrasts, and their covariances. From the estimates of the contrasts the estimates for the missing feature values are easily derived by taking for each feature the sum of the estimated contrast vector and the original reference value: $\mathbf{y}^*_i = \mathbf{\Delta y}^*_i + y_{i0}$. The variance of the estimates is directly equal to the diagonal of the contrast covariance matrix $\mathbf{\Sigma}^* - \mathbf{\Sigma}^{\times}\mathbf{\Sigma}^{-1}\mathbf{\Sigma}^{\times T}$, as the reference observations act as constants in this context.

References

1. Felsenstein, J. (1973) Maximum-likelihood estimation of evolutionary trees from continuous characters. *American Journal of Human Genetics*, **25**, 471-492.
2. Felsenstein, J. (1988) Phylogenies and quantitative characters. *Annual Review of Ecology and Systematics*, **19**, 445-471.
3. Felsenstein, J. (1985) Phylogenies and the comparative method. *American Naturalist*, **125**, 1-15.
4. Lynch, M. (1991) Methods for the analysis of comparative data in evolutionary biology. *Evolution*, **45**, 1065-1080.
5. Felsenstein, J. (2008) Comparative methods with sampling error and within-species variation: Contrasts revisited and revised. *American Naturalist*, **171**, 713-725.
6. Ives, A.R., Midford, P.E. and Garland, T. (2007) Within-species variation and measurement error in phylogenetic comparative methods. *Systematic Biology*, **56**, 252-270.
7. Grafen, A. (1989) The phylogenetic regression. *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences*, **326**, 119-157.
8. Pagel, M. (1997) Inferring evolutionary processes from phylogenies. *Zoologica Scripta*, **26**, 331-348.
9. 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. *American Naturalist*, **149**, 646-667.

10. Housworth, E.A., Martins, E.P. and Lynch, M. (2004) The phylogenetic mixed model. *American Naturalist*, **163**, 84-96.
11. Rohlf, F.J. (2001) Comparative methods for the analysis of continuous variables: Geometric interpretations. *Evolution*, **55**, 2143-2160.
12. Garland, T. and 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.
13. Garland, T., Midford, P.E. and Ives, A.R. (1999) An introduction to phylogenetically based statistical methods, with a new method for confidence intervals on ancestral values. *American Zoologist*, **39**, 374-388.
14. Schluter, D., Price, T., Mooers, A.O. and Ludwig, D. (1997) Likelihood of ancestor states in adaptive radiation. *Evolution*, **51**, 1699-1711.
15. Mooers, A.O., Vamosi, S.M. and Schluter, D. (1999) Using phylogenies to test macroevolutionary hypotheses of trait evolution in cranes (*Gruinae*). *American Naturalist*, **154**, 249-259.
16. Blomberg, S.P., Garland, T. and Ives, A.R. (2003) Testing for phylogenetic signal in comparative data: Behavioral traits are more labile. *Evolution*, **57**, 717-745.
17. Pagel, M. (1993) Seeking the evolutionary regression coefficient - an analysis of what comparative methods measure. *Journal of Theoretical Biology*, **164**, 191-205.
18. Revell, L.J. and Harmon, L.J. (2008) Testing quantitative genetic hypotheses about the evolutionary rate matrix for continuous characters. *Evolutionary Ecology Research*, **10**, 311-331.
19. Johnson, N.L. and Kotz, S. (1972) *Distributions in statistics: Continuous multivariate distributions*. Wiley, New York.
20. Pinheiro, J.C. and Bates, D.M. (1996) Unconstrained parametrizations for variance-covariance matrices. *Statistics and Computing*, **6**, 289-296.
21. Nocedal, J. and Wright, S.J. (2006) *Numerical optimization*. 2nd ed. Springer, New York.