MVMORPH: A PACKAGE FOR FITTING MULTIVARIATE EVOLUTIONARY MODELS TO MORPHOMETRIC DATA

JULIEN CLAVEL, GILLES ESCARGUEL & GILDAS MERCERON

DETAILS ON COMPUTATIONAL METHODS USED WITHIN MVMORPH

1 - Technical notes on the package

mvMORPH uses internally C codes and general purpose BLAS and LAPACK routines to speed up the computations; therefore it needs compilation. mvMORPH calls functions from the ape and phytools packages (Paradis *et al.* 2004 and Revell 2012, respectively); these functions are used for working with trees in R. mvMORPH also uses modified C codes from the OUCH (Butler & King 2004, 2009) and ape package, and uses SIMMAP-like mapped trees as provided by the phytools package. The spam package (Furrer & Sain 2010) is used to compute Cholesky factorization of sparse variance-covariance matrices; corpcor (Schäfer *et al.* 2013) is used to compute the Moore-Penrose pseudoinverse matrix related to the systems of linear equations used in the model fitting procedure (see section 8). Maximum log-likelihood search is performed using the non-linear optimization routines proposed by the *optim* function (Nelder-Mead, or gradient methods such as BFGS; Nelder & Mead 1965; Nocedal 1980; Byrd *et al.* 1995) from the stats package, as well as the subplex routine (a variant of the simplex Nelder-Mead that operates on a sequence of subspace) from the subplex package (King 2013). For each optimization procedure, diagnostics of convergence for the optimizing function are returned, as well as the hessian matrix computed for the maximum-likelihood estimate to get diagnostics on the reliability of the estimated parameters (e.g., Beaulieu *et al.* 2012). The hessian matrix can also be used to compute the variance-covariance matrix of ML parameter estimates for multivariate normal proposal functions in a Bayesian framework or to approximate the standard errors of the MLE.

2 - Computing the variance-covariance matrix for the multivariate Brownian motion process

The multivariate variance-covariance matrix for a Brownian motion process is computed using the kronecker product (⊗) of the rate (drift) matrix $\boldsymbol{R}$ with the phylogenetic variance-covariance matrix $\boldsymbol{C}$ (Felsenstein 2004; Revell & Harmon 2008; Revell & Collar 2009).

$$\boldsymbol{V} = \boldsymbol{R} \otimes \boldsymbol{C}$$

eqn S1

Otherwise the sum of kronecker products is used when k different selective regimes (clade, epochs…) have their own rate matrix $\boldsymbol{R}_k$. These regimes can be defined in the sub-matrices $\boldsymbol{C}_k$ (Revell & Collar 2009).

$$\boldsymbol{V} = \sum_{i=1}^{k} \boldsymbol{R_i} \otimes \boldsymbol{C_i}$$

eqn S2

The rate matrix $\boldsymbol{R}$ is parameterized by its Cholesky decomposition as the cross-product of upper (or lower) triangular matrices (see below, eqn S5) to ensure symmetric positive definiteness of the variance-covariance matrix during the log-likelihood optimization. Note that for the case where a unique BM process is assumed for each trait (i.e., same rate matrix for all selective regimes), there is an analytical solution for the maximum-likelihood estimate of $\boldsymbol{R}$ (eqn S3; see, e.g., Garland & Ives 2000; Felsenstein 2004; O'Meara *et al.* 2006; Freckleton 2012).

$$\boldsymbol{R} = \frac{(\boldsymbol{Y} - \boldsymbol{X}\beta)^T \boldsymbol{C}^{-1} (\boldsymbol{Y} - \boldsymbol{X}\beta)}{N}$$

eqn S3

However, in the multiple-rate case, or when measurement errors are added, a numerical optimization is necessary (Ives *et al.* 2007; Revell & Collar 2009).

3 - Computing the variance-covariance matrix for the multivariate Ornstein-Uhlenbeck process

The multivariate Ornstein-Uhlenbeck process (eqn S4) is described by $m \times m$ matrices of parameters $\boldsymbol{A}$ (also called the transition or pull matrix) and $\boldsymbol{\Sigma}$ (the drift or dispersion matrix, which is the scatter generator) (Gardiner 2004; Butler & King 2009; Meucci 2010; Bartoszek *et al.* 2012; Reitan *et al.* 2012). W is a vector of independent Brownian motions (independent and identically distributed normal variables) and β is a vector of optimum traits values (see eqn S15). $\boldsymbol{\Sigma}$ is a square matrix which can be lower triangular to generate correlated stochastic perturbations (eqn S5); $\boldsymbol{A}$ is a square matrix for which we only restrict its eigenvalues to be real numbers (see section 7). The main computational difficulties in the multivariate OU come from the matrix exponentiation of $\boldsymbol{A}$ and the integral decomposition with matrix exponential (Bartoszek *et al.* 2012; Reitan *et al.* 2012). The eigendecomposition of $\boldsymbol{A}$ (eqn S6) is used to

compute the matrix exponential (eqn S7) involved in the variance-covariance computation (eqns S8-S13) and in the design matrix (eqn S16).

$$dY(t) = \boldsymbol{A}\big(\beta(t) - Y(t)\big)dt + \boldsymbol{\Sigma}dW(t)$$

eqn S4

Because the covariance matrix of the stochastic infinitesimal changes in eqn S4 is symmetric positive definite (eqns S8-S13), we can express more generally the lower triangular matrix $\boldsymbol{\Sigma}$ which determines the scatter of the stochastic perturbations as the Cholesky decomposition of the variance-covariance matrix $\boldsymbol{R}$ (Pinheiro & Bates 1996).

$$\boldsymbol{R} = \boldsymbol{\Sigma}\boldsymbol{\Sigma}^T$$

eqn S5

(Note however that we can also use singular value decomposition (e.g. eqn S6) with $\boldsymbol{R} = \boldsymbol{P}\boldsymbol{D}\boldsymbol{P}^{-1}$ *and obtain* $\boldsymbol{\Sigma} = \boldsymbol{P}\boldsymbol{U}$ *with* $\boldsymbol{U} = diag\big(\sqrt{\lambda_1}, \dots, \sqrt{\lambda_m}\big)$*. In such a case,* $\boldsymbol{\Sigma}$ *is not triangular.)*

Assuming the transition matrix $\boldsymbol{A}$ has only real-part eigenvalues, we can express it by its eigenvalue decomposition (eqn S6; see also section 7) where $\boldsymbol{P}$ is the (orthogonal) matrix of eigenvectors of $\boldsymbol{A}$ and $\boldsymbol{D}$ is the diagonal matrix of eigenvalues (λ).

$$\boldsymbol{A} = \boldsymbol{P}\boldsymbol{D}\boldsymbol{P}^{-1}$$

eqn S6

The matrix exponential can be solved using the eigendecomposition of $\boldsymbol{A}$ as in eqn S7 (Moler & Van Loan 2003).

$$e^{-\boldsymbol{A}t} = \boldsymbol{P}diag\big(e^{-\lambda_1 t}, \dots, e^{-\lambda_m t}\big)\boldsymbol{P}^{-1}$$

eqn S7

The integral in the variance-covariance computation of the multivariate Ornstein-Uhlenbeck process conditioned on the root value (eqns S8, S9; Bartoszek *et al.* 2012) is also computed using the eigendecomposition of $\boldsymbol{A}$ (A.1 and B.3 in Bartoszek *et al.* 2012; see also Meucci 2010; Reitan *et al.* 2012).

$$Cov\big[Y_i, Y_j\big] = e^{-\boldsymbol{A}(C_{ii} - C_{ij})}\left(\int_0^{C_{ij}} e^{-\boldsymbol{A}v}\boldsymbol{R}e^{-\boldsymbol{A}^T v}\, dv\right)e^{-\boldsymbol{A}^T(C_{jj} - C_{ij})}$$

eqn S8

$$Cov\big[Y_i, Y_j\big] = e^{-\boldsymbol{A}(C_{ii} - C_{ij})}\left[\mathbf{P}\left(\left[\frac{1}{\lambda_k + \lambda_l}\left(1 - e^{-(\lambda_k + \lambda_l)C_{ij}}\right)\right]_{1 \le kl \le m} \odot \mathbf{P}^{-1}\boldsymbol{\Sigma}\boldsymbol{\Sigma}^T(\mathbf{P}^{-1})^T\right)\mathbf{P}^T\right]e^{-\boldsymbol{A}^T(C_{jj} - C_{ij})}$$

eqn S9

Alternatively, we can avoid computing the distance between the tips and the common ancestor explicitly and use eqn S10 instead.

$$e^{-\boldsymbol{A}C_{ii}}\left(\int_0^{C_{ij}} e^{\boldsymbol{A}v}\boldsymbol{R}e^{\boldsymbol{A}^T v}\,dv\right)e^{-\boldsymbol{A}^T C_{jj}} =$$
$$e^{-\boldsymbol{A}C_{ii}}\left[\mathbf{P}\left(\left[\frac{1}{\lambda_k+\lambda_l}\left(e^{(\lambda_k+\lambda_l)C_{ij}}-1\right)\right]_{1\leq kl\leq m}\odot \mathbf{P}^{-1}\boldsymbol{\Sigma}\boldsymbol{\Sigma}^T(\mathbf{P}^{-1})^T\right)\mathbf{P}^T\right]e^{-\boldsymbol{A}^T C_{jj}}$$

eqn S10

For the univariate case, Ho & Ané (2013) have shown that the computation of the covariances between tip species depends on the way the root value is treated. Assuming that the root is situated at the stationary point, the multivariate phylogenetic variance-covariance matrix can be expressed as in eqn S11. This is the formulation actually used, for instance, in the OUCH package (Butler & King 2009).

$$e^{-\boldsymbol{A}C_{ii}}\boldsymbol{\Xi}e^{-\boldsymbol{A}^T C_{jj}} + \int_0^{C_{ij}} e^{-\boldsymbol{A}(C_{ii}-v)}\boldsymbol{R}e^{-\boldsymbol{A}^T(C_{jj}-v)}\,dv$$
$$= \mathbf{P}\left(\left[\frac{1}{\lambda_k+\lambda_l}e^{-\lambda_k(C_{ii}-C_{ij})}e^{-\lambda_l(C_{jj}-C_{ij})}\right]_{1\leq kl\leq m}\odot \mathbf{P}^{-1}\boldsymbol{\Sigma}\boldsymbol{\Sigma}^T(\mathbf{P}^{-1})^T\right)\mathbf{P}^T$$

eqn S11

The first term in the left hand side of eqn S11 is the variance-covariance of the traits at the root where $\boldsymbol{\Xi}$ is the stationary distribution (see solution in eqn S18) that satisfy the following relationship (Gardiner 2004):

$$\boldsymbol{A}\boldsymbol{\Xi} + \boldsymbol{\Xi}\boldsymbol{A}^T = \boldsymbol{R}$$

eqn S12

Taking $\boldsymbol{A}$ symmetric positive definite, we can simplify the computation of the variance-covariance matrix using the transpose of the orthogonal matrix $\boldsymbol{P}$.

$$\mathbf{P}\left(\left[\frac{1}{\lambda_k+\lambda_l}e^{\left(-\lambda_k(C_{ii}-C_{ij})-\lambda_l(C_{jj}-C_{ij})\right)}\right]_{1\leq kl\leq m}\odot \mathbf{P}^T\boldsymbol{\Sigma}\boldsymbol{\Sigma}^T\mathbf{P}\right)\mathbf{P}^T$$

eqn S13

(A stacked vector form for this Lyapunov equation is: $\left[(\boldsymbol{A}\oplus\boldsymbol{A})^{-1}e^{-(\boldsymbol{A}\oplus\boldsymbol{A})t}\right]vec(\boldsymbol{\Sigma}\boldsymbol{\Sigma}^T)$*, where* $(\boldsymbol{A}\oplus\boldsymbol{A})$ *is the kronecker sum:* $(\boldsymbol{I}\otimes\boldsymbol{A})+(\boldsymbol{A}\otimes\boldsymbol{I})$*)*

This formulation is more stable (particularly when the eigenvalues of $\boldsymbol{A}$ are small) than the more general one (eqn S8) and is faster to compute as it avoids the need to compute the two matrices exponential outside of the integral in eqn S8. However, it assumes the optimum is at the stationary distribution and cannot be used for inferring a shift from the ancestral state (e.g., with fossil data). Although its computation is faster, the resulting variance-covariance matrix is a full matrix, making impossible to use algorithms based on sparse matrices to speed up the computation of the log-likelihood (see Section 8). Nevertheless, with ultrametric trees, both approaches should converge to the same estimates if the process is stationary (i.e., the alpha parameters or eigenvalues of $\boldsymbol{A}$ are large enough). The two approaches are allowed in mvMORPH by specifying the type of variance-covariance matrix to compute in the "param" list (`param=list(vcv="mvmorph")` or `param=list(vcv="ouch")`).

For the general univariate case, mvMORPH switch to an algorithm using the formulation of Hansen (1997) to simplify the computations and allowing the use of non-ultrametric trees with eqn S14.

$$Cov[Y_i, Y_j] = \frac{\sigma^2}{2\alpha}(1 - e^{-2\alpha c_{ij}})e^{-\alpha\left((c_{ii}-c_{ij})+(c_{jj}-c_{ij})\right)}$$

eqn S14

Note that the variance term $\left(\frac{\sigma^2}{2\alpha}\right)$ needs to be computed once.

4 - Expectation of the multivariate Ornstein-Uhlenbeck process

Assuming the selective regimes are piecewise-constant, the expectation of the Ornstein-Uhlenbeck process is the weighted sum of the selective regimes optimums and the root value (Hansen 1997; Butler & King 2004; see also eqn 8 in the main text):

$$\mathbb{E}[Y] = \boldsymbol{W}\beta$$

eqn S15

The design matrix (or weights matrix) $\boldsymbol{W}$ is computed following Butler & King (2004) but with matrix exponential of $\boldsymbol{A}$ (for the multivariate case, see also eqn S7):

$$W_{ik} = \psi_{ik}^{\gamma} e^{-\boldsymbol{A}t_i} + \sum_{\gamma=1}^{k(i)} \psi_{ik}^{\gamma} \left(e^{-\boldsymbol{A}\left(t_i - t_i^{\gamma-1}\right)} - e^{-\boldsymbol{A}(t_i - t_i^{\gamma})}\right)$$

eqn S16

Here, matrix $\boldsymbol{W}$ is $(m \times k)$ by $(m \times n)$ where k is the number of selective regimes, m the number of traits, and n the number of tip species. When the selective regime (eqn S16) operating in the time interval γ for the i^{th} lineage is the k^{th} regime, then the step function ψ has the value of 1, else it is 0. t_i is the time of the i^{th} tip. By explicitly taking into account the root state $e^{-\boldsymbol{A}t_i}$ with the step function ψ for the oldest selective regime, we are assuming that the root state cannot be estimated independently of the optimum state.

This parameterization allows ensuring convergence during the ML optimization with ultrametric trees (Butler & King 2009; Ho & Ané 2014). Alternatively, with non-ultrametric trees, the root state may be identifiable and it could be estimated by adding a column in matrix $\boldsymbol{W}$ with the expected coefficients for the root state $e^{-\boldsymbol{A}t_i}$ and removing the first term in eqn S16. Both options are allowed in the *mvOU* function with the "param" list argument by specifying `param=list(root=FALSE)` or `param=list(root=TRUE)`, respectively. A third alternative is to assume that the optima are at their stationary distributions by avoiding the computation of the root state ($e^{-\boldsymbol{A}t_i}$) and row-standardizing the matrix $\boldsymbol{W}$ to 1 (Beaulieu *et al.* 2012). This is the approach used in the OUwie package (Beaulieu *et al.* 2012); it can be achieved in mvMORPH by specifying `param=list(root="stationary")`. The optima (β) are then computed using the design matrix (eqn S16) in the GLS equation (eqn 3 in the main text).

5 - Computation of trait correlations from covariance matrices

It is possible to compute the evolutionary correlations or partial correlations between traits from the inverse of the ML estimate of the rate matrix (i.e., the precision matrix; see Lartillot & Delsuc 2012).

$$r_{ij} = \frac{-\Omega_{ij}}{\sqrt{\Omega_{ii}\Omega_{jj}}}; \quad \boldsymbol{\Omega} = \boldsymbol{R}^{-1}$$

eqn S17

This corresponds to a standardization of the covariance matrix that is easily done in R with the base *cov2cor* function or *cor2pcor* in "corpcor".

For the Ornstein-Uhlenbeck model the variance accrued along the branches of the phylogeny is related to both the rate and alpha matrices (matrix $\boldsymbol{A}$ in eqn S4; see also eqn S12). Therefore, for the multivariate OU process, the correlation matrix between traits must be computed from the multivariate stationary distribution (eqn. S18; Bartoszek *et al.* 2012). The stationary covariance matrix can be computed using the *stationary* function in mvMORPH.

$$Cov_{t\to\infty}[Y] = \mathbf{P}\left(\left[\frac{1}{\lambda_k + \lambda_l}\right]_{1\leq kl\leq m} \odot \mathbf{P}^{-1}\boldsymbol{\Sigma}\boldsymbol{\Sigma}^T(\mathbf{P}^{-1})^T\right)\mathbf{P}^T$$

eqn S18

(Alternatively $Cov_{t\to\infty}[Y] = (\boldsymbol{A}\oplus\boldsymbol{A})^{-1}vec(\boldsymbol{\Sigma}\boldsymbol{\Sigma}^T)$ *as the solution of the Lyapunov equation S12)*

6 - Phylogenetic half-life

Hansen (1997) related the α parameter of the Ornstein-Uhlenbeck process (e.g., eqn S14) to the time it takes to move from one optimum to the other through the phylogenetic half-life equation:

$$t_{1/2} = \frac{\ln(2)}{\alpha}$$

eqn S19

This characteristic time was proposed as a measure of the phylogenetic signal (for a unique optimum) or as an estimation of the rate of adaptation from an optimum to another on the adaptive landscape (Hansen 2012). The multivariate counterpart of the phylogenetic half-life for each trait is computed from the eigenvalues of the matrix $\boldsymbol{A}$ while the eigenvectors describe the path toward the optimum (Bartoszek *et al.* 2012). The phylogenetic half-life can be computed using the *halflife* function in mvMORPH.

Similarly, the phylogenetic half-life concept can be extended to other exponential models. For instance, in order to improve optimization convergence during parameter searches for the Early-Burst model, Slater & Pennell (2014) proposed to fix a meaningful maximum bound by considering the half-life BM rate (r) which, for an exponentially decaying process, is $t_{1/2} = \frac{ln\,(2)}{r}$. They proposed to use a range of r values comprised between 1 and 10 half-lives elapsing over the age of the tree.

7 - Matrix parameterization

We introduced above the use of the Cholesky parameterization of covariance matrices such as the drift (rate) matrix $\boldsymbol{R}$ (eqn S5). Cholesky parameterization is an efficient strategy to ensure the symmetry and positive definiteness of the covariance matrix during the ML optimization (Pinheiro & Bates 1996). Parameterization of the rate matrix can also be useful for testing nested models such as in the Flury tests hierarchy (Flury 1988; Phillips & Arnold 1999). For instance, multiplying the $\boldsymbol{R}_i$ matrices of a multiple rate matrix model (eqn S2) by a scalar can be used to test for proportionality. This is done through the "param" list in mvBM (`param=list(constraint="proportional")`). For testing for shared eigenvectors (i.e., same variances but different covariances), or common correlations but different variances, the mvMORPH package uses a spherical parameterization with $m(m\text{-}1)/2$ coordinates θ_i (Pinheiro & Bates 1996) to obtain the Cholesky factors $\boldsymbol{U}$ of a correlation matrix $\boldsymbol{C}$ (by fixing the first element of $\boldsymbol{U}_{11}$ to 1; see also Lu & Ades 2009). The $\boldsymbol{C}$ matrix is then scaled with the square root of the variances contained in the diagonal matrix $\boldsymbol{D}$.

$$\boldsymbol{R} = \boldsymbol{DCD} = \boldsymbol{DU}^T\boldsymbol{UD}$$

eqn S20

$$\boldsymbol{R} = \begin{bmatrix} \sqrt{v_1} & 0 \\ 0 & \sqrt{v_2} \end{bmatrix} \left(\begin{bmatrix} 1 & \cos\theta_1 \\ 0 & \sin\theta_1 \end{bmatrix}^T \begin{bmatrix} 1 & \cos\theta_1 \\ 0 & \sin\theta_1 \end{bmatrix} \right) \begin{bmatrix} \sqrt{v_1} & 0 \\ 0 & \sqrt{v_2} \end{bmatrix}$$

eqn S20b

This spherical parameterization of Cholesky factors and scaling (see example for a bivariate case in eqn S20b) is obtained through the "param" list (`param=list(constraint="shared")`, or `param=list(constraint="correlation")`). Adams (2013) presented a constrained Cholesky parameterization allowing comparison of rates of evolution between covarying traits. This option is accessible in mvMORPH through the "param" list in mvBM (`param=list(constraint=TRUE)`). Note that matrix parameterization could be useful in model such as Early-Burst to allow each trait to have its own phylogenetic covariance structure (see also the random effect model in Freckleton 2012); however there is no obvious way to parameterize such matrix, a possibility which is therefore not yet implemented in mvMORPH.

The multivariate Ornstein-Uhlenbeck process is described by the eigenvalues of matrix $\boldsymbol{A}$ (Hansen & Martins 1996; Gardiner 2004; Bartoszek *et al.* 2012; Reitan *et al.* 2012). General matrix parameterization conditioned on eigenvalues have been already proposed for the bivariate case (e.g., Sy *et al.* 1997; Jaffrézic *et al.* 2004). mvMORPH uses the Schur decomposition for parameterizing the matrix $\boldsymbol{A}$ for the multivariate Ornstein-Uhlenbeck process. This parameterization allows restraining the optimization to non-symmetric matrices with real part eigenvalues and avoid the use of complex numbers machinery more difficult to implement and to interpret biologically. The Schur parameterization is done by computing first an orthogonal matrix $\boldsymbol{P}$, which is done in mvMORPH through Givens rotations (Golub & Van Loan 2013). This orthogonal matrix is then multiplied to an upper triangular matrix $\boldsymbol{U}$ for which the diagonal values represent the eigenvalues (eqn S21; see eqn S21b for a bivariate illustration).

$$\boldsymbol{A} = \boldsymbol{PUP}^T$$

eqn S21

$$\boldsymbol{A} = \begin{bmatrix} \cos\theta_1 & -\sin\theta_1 \\ \sin\theta_1 & \cos\theta_1 \end{bmatrix} \begin{bmatrix} \lambda_1 & \theta_2 \\ 0 & \lambda_2 \end{bmatrix} \begin{bmatrix} \cos\theta_1 & -\sin\theta_1 \\ \sin\theta_1 & \cos\theta_1 \end{bmatrix}^T$$

eqn S21b

For the symmetric case we can use the SVD (Singular Value Decomposition) parameterization with the same orthogonal matrix $\boldsymbol{P}$ and a diagonal matrix of eigenvalues $\boldsymbol{D}$ instead of **U** (note that compared to the Cholesky, the SVD parameterization allows negative eigenvalues – see Bartoszek et al. 2012 on negative values interpretations). The solution of the Schur decomposition is not unique (note that it is also the case for the Cholesky parameterization; Pinheiro & Bates 1996), and there is the possibility to use instead an alternative algorithm based on QR decomposition (K. Bartoszek, personal communication 2014). Advantage of this parameterization is that it avoids the need to explicitly compute the eigendecomposition of the matrix $\boldsymbol{A}$ in the covariance calculation as the eigenvectors are provided by the matrix $\boldsymbol{P}$ and eigenvalues by the diagonal elements of $\boldsymbol{U}$ or $\boldsymbol{D}$ (in counterpart, user time to convergence and number of iteration is often greater for SVD than when using spherical or Cholesky parameterization; Pinheiro & Bates 1996). The various parameterization of $\boldsymbol{A}$ are done through the "param" list with the *decomp* argument (e.g., `param=list(decomp="symmetricPositive")` for the Cholesky parameterization, `param=list(decomp="symmetric")` for the SVD parameterization with Givens rotations, and `param=list(decomp="nsymmetric")` for the Schur decomposition). Finally, matrix parameterization should be worth of attention for extending multivariate models to Bayesian approaches, as the priors have to be set on transformed spaces.

8 - Likelihood computation using factorized matrices

The linear model used for fitting phenotypic traits evolution is of the form:

$$b = \boldsymbol{A}x + w$$

eqn S22

Where $w \sim \mathcal{N}(0, \boldsymbol{V})$ is a vector of normally-distributed error terms with mean 0 and variance-covariance matrix $\boldsymbol{V}$. The GLS solution to this equation is:

$$x = (\boldsymbol{A}\boldsymbol{V}^{-1}\boldsymbol{A}^T)^{-1}\boldsymbol{A}^T\boldsymbol{V}^{-1}b$$

eqn S23

and the log-likelihood function to maximize is:

$$\log(\mathcal{L}) = -\frac{1}{2}[n\,log(2\pi) + log|\boldsymbol{V}| + (\boldsymbol{b} - x\boldsymbol{A})^T\boldsymbol{V}^{-1}(\boldsymbol{b} - x\boldsymbol{A})]$$

eqn S24

This GLS solution and the log-likelihood function involve both the computation of the inverse of the variance-covariance matrix $\boldsymbol{V}$ as well as its determinant. Such linear problem are computationally intensive (Freckleton 2012), particularly for multivariate models. We can solve this issue by avoiding the computation of the matrix inverse by factorizing $\boldsymbol{V}$ with its Cholesky decomposition (eqn S25; Seber & Lee 2003; Golub & Van Loan 2013) for solving the OLS model (eqn S26).

$$\boldsymbol{V} = \boldsymbol{L}\boldsymbol{L}^T$$

eqn S25

$$x = (\widetilde{\boldsymbol{A}}\widetilde{\boldsymbol{A}}^T)^{-1}\widetilde{\boldsymbol{A}}^T\tilde{b}$$

eqn S26

Using forward (or backward for upper triangular factor) substitution we can easily solve the following systems without inverting explicitly the $\boldsymbol{L}$ matrix:

$$\tilde{b} = \boldsymbol{L}^{-1}b \; ; \widetilde{\boldsymbol{A}} = \boldsymbol{L}^{-1}\boldsymbol{A}$$

eqn S27

We can use the pseudoinverse $\boldsymbol{A}^+$ for solving a part of the linear system in eqn S26:

$$\boldsymbol{A}^+ = (\widetilde{\boldsymbol{A}}^T\widetilde{\boldsymbol{A}})^{-1}\widetilde{\boldsymbol{A}}^T$$

eqn S28

This can be done by using Singular Value Decomposition (SVD) of the "thin" matrix $\widetilde{\boldsymbol{A}}$.

$$\boldsymbol{A}^+ = \boldsymbol{V}\boldsymbol{\Sigma}^+\boldsymbol{U}^T$$

eqn S29

with

$$\boldsymbol{\Sigma}^+ = diag(\frac{1}{\lambda_1}, \dots)$$

eqn S30

mvMORPH uses the fast SVD routine from the "corpcor" package (Schäfer *et al.* 2013) for this computation. The GLS solution is finally obtained from:

$$x = \boldsymbol{A}^+\tilde{b}$$

eqn S31

Last, the log-determinant of the variance-covariance matrix $\boldsymbol{V}$ involved in the log-likelihood calculation is easy to obtain from the factorized matrix:

$$|\boldsymbol{V}| = \sum log\big(diag(\boldsymbol{L})\big)$$

eqn S32

While the quadratic product $w^T\boldsymbol{V}^{-1}w$ is simply computed by solving through forward substitution (i.e., without inverting the matrix) the following linear system:

$$\boldsymbol{L}y = w$$

eqn S33

Then the estimated values *y* are squared and summed. In this way, several expensive explicit computations such as the log-determinant and the inverse of $\boldsymbol{V}$ are thus avoided. Although the

building of the variance-covariance matrix itself is still time consuming, the use of the direct solver is ~10 times faster than the use of the traditional GLS solution through matrix inversion (Fig. S1.1).

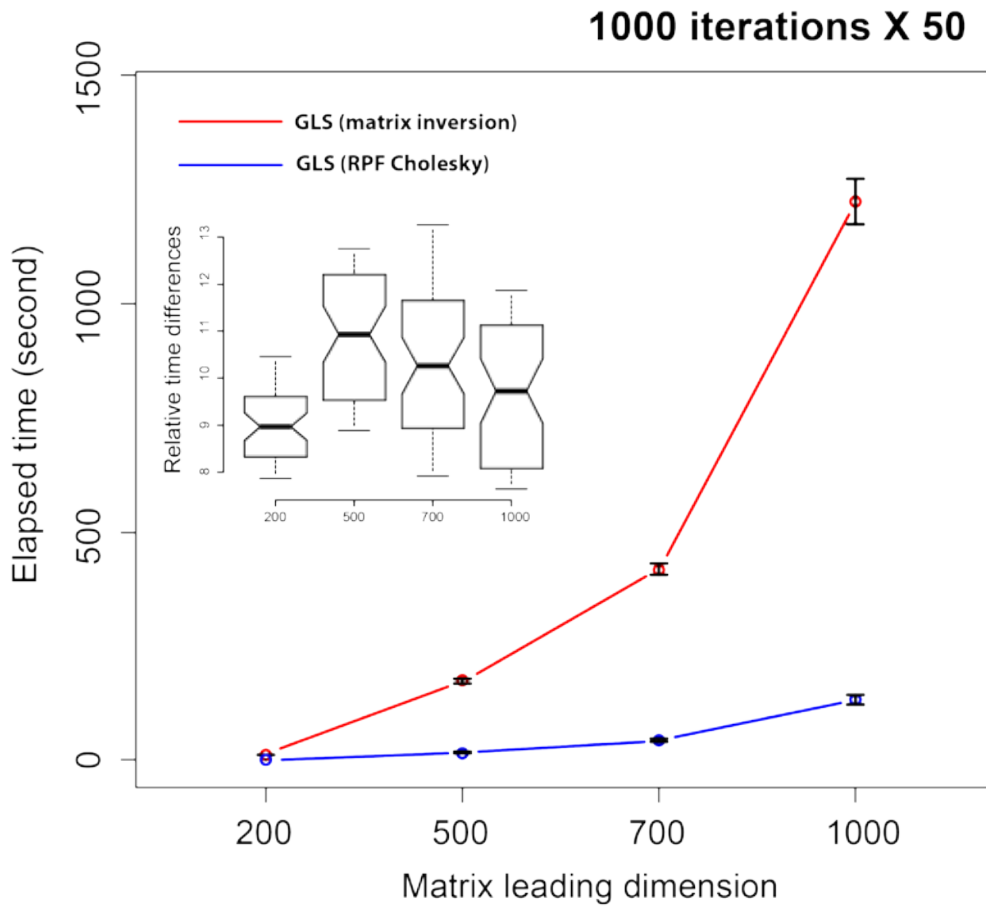

Figure S1.1. Comparison of computational time for the log-likelihood estimate of the variance-covariance matrix of 50 simulated Yule trees of 200, 500, 700, and 1000 species with the standard approach (red) and the RPF approach (blue). Simulations done on a CPU Intel® 3.16 Ghz and 8Go RAM.

The remaining computational bottleneck is the Cholesky decomposition of the variance-covariance matrix that requires $n^3/3$ operations. To improve the computation of the Cholesky factorization we used a half storage algorithm (for the lower triangular part only) that rearranges the triangular matrix in a rectangular full packed format (Fig. S1.2), then allowing the use of fast BLAS 3 routines for matrix to matrix multiplications (Andersen *et al.* 2001; Gustavson *et al.* 2010). This algorithm is more efficient than the R base Cholesky algorithm and can be used with full variance-covariance matrices (e.g., as obtained with the "ouch" algorithm for computing the Ornstein-Uhlenbeck variance-covariance matrix). This approach is available in all the functions of the mvMORPH package through the "method" argument (`method="rpf"`).

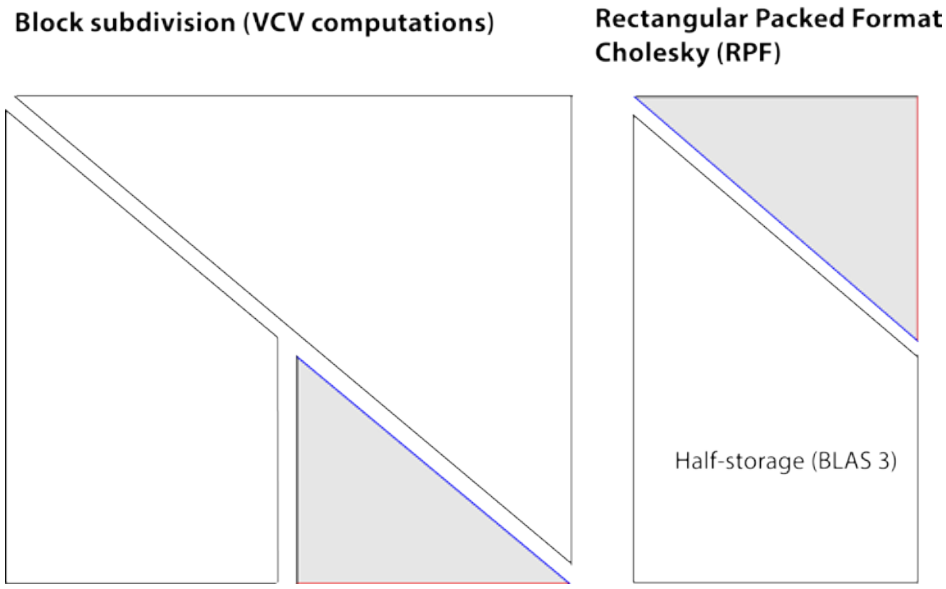

Figure S1.2. Rectangular (Full) Packed Format configuration of the upper triangular variance-covariance matrix used to compute the Cholesky decomposition using fast BLAS 3 subroutines and half-storage.

It remains possible to improve the computation of the Cholesky factor during the ML optimization iterations by updating the existing factor rather than computing it from scratch. This is done by taking advantage of the sparsity structure of a phylogenetic variance-covariance matrix (Fig. S1.3) using the R package "spam" (Furrer & Sain 2010). Note that the efficiency (Fig. S1.4) of this sparse method actually relies on the sparsity structure of the phylogenetic variance-covariance matrix (i.e., on the imbalance of the tree). This method can be used through the argument `method="sparse"`.

Finally, mvMORPH also implements a fast pruning algorithm based on Independents contrasts (Felsenstein 1973; Freckleton 2012) available through the argument `method="pic"`. This method cannot be used with the *mvOU* function (multiple optimum and multivariate OU are not allowed). Nevertheless this fast algorithm can be particularly useful for parametric bootstrap simulations on BM-based processes, Bayesian mcmc, and user-specified tree-branch length optimizations (see Table S1 for a benchmark with current fast implementations for computing the log-likelihood).

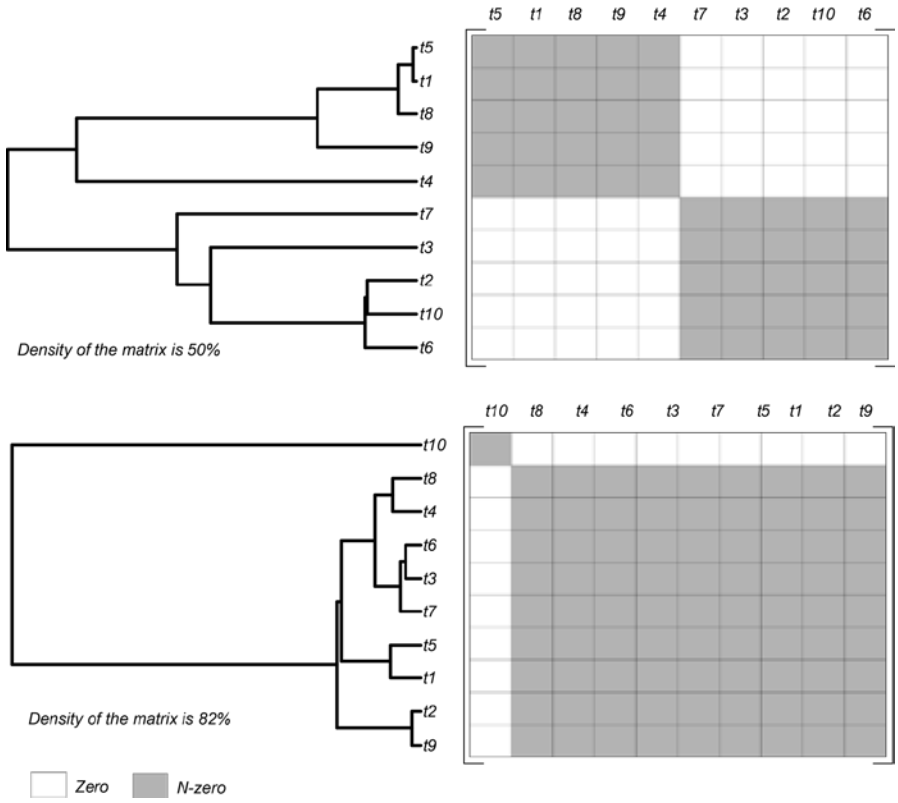

Figure S1.3. Sparsity structure of a phylogenetic variance-covariance matrix. In the "best" case, half of the entries are null values, otherwise on highly unbalanced trees there is only n-1 null entries on each triangle.

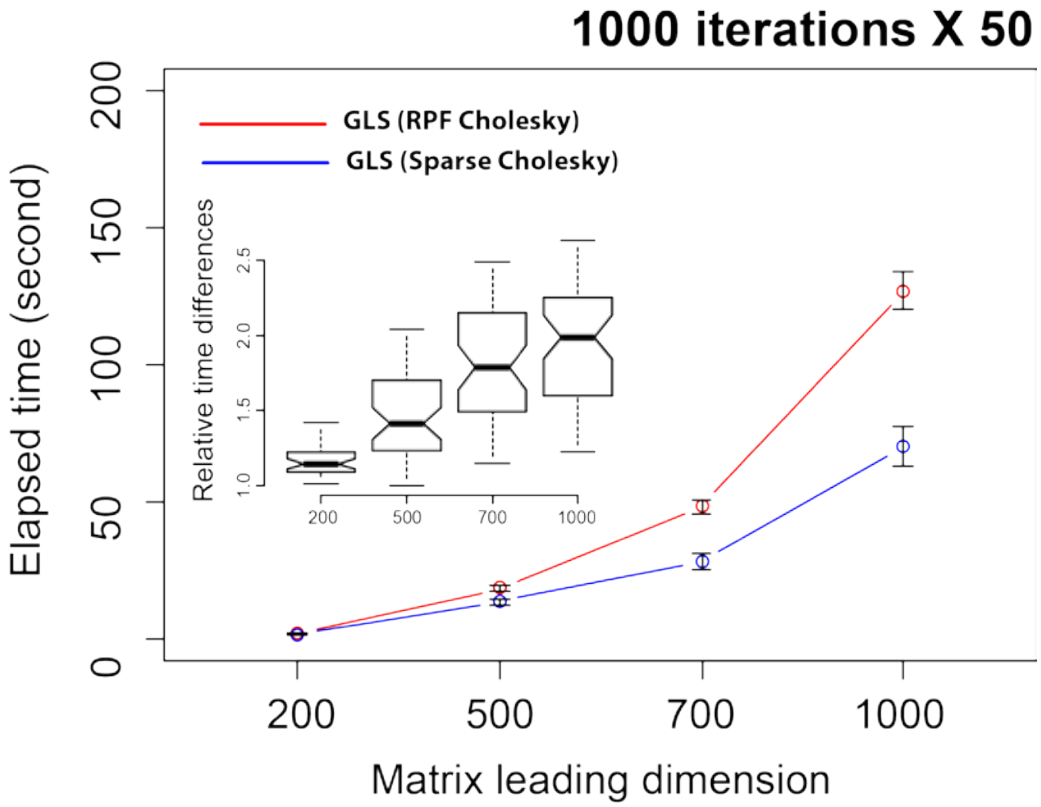

Figure S1.4. Comparison of the two Cholesky approaches on the variance-covariance matrices of 50 random Yule trees of 200, 500, 700, and 1000 species. Depending on the sparsity structure of the phylogenetic variance-covariance matrix, the gain using the sparse or rpf methods can be low for low-dimension matrices. Simulations done on a CPU Intel® 3.16 Ghz and 8 Go RAM.

Table S1. Benchmark of the « pic » method in mvMORPH (function *mvLL*) for computing the log-likelihood of a univariate BM process relative to the packages MOTMOT (1.0.1) function *transformPhylo.ll*, geiger (2.0.3) *fitContinuous*, and phylolm (2.2) *phylolm* for differently-sized random trees (n = 500, 5000, 10 000, and 1 000 000 species). Simulations done on a 2.5 Ghz Intel ® Core i7 – 16Go 1600 Mhz DDR3 and R 3.1.2. Note that *fitContinuous* does not use the analytical MLE for the BM log-likelihood but optimization. Elapsed time is expressed in seconds.

N = 500

	Functions	replications	elapsed	relative
1	mvLL(phy, data, method="pic")	100	**0.022**	1
2	transformPhylo.ll(data, phy = tree, model="bm")	100	0.563	25.591
3	phylolm(trait ~ 1, data, tree, model="BM")	100	2.006	91.182
4	fitContinuous(tree, data, model = "BM")	100	53.783	2444.682

N = 5000 (~ mammal phylogeny)

	Functions	replications	elapsed	relative
1	mvLL(phy, data, method="pic")	100	**0.095**	1
2	transformPhylo.ll(data, phy = tree, model="bm")	100	7.221	76.011
3	phylolm(trait ~ 1, data, tree, model="BM")	100	18.864	198.568
4	fitContinuous(tree, data, model = "BM")	100	172.317	1813.863

N = 10 000 (without *fitContinuous*; ~bird phylogeny)

	Functions	replications	elapsed	relative
1	mvLL(phy, data, method="pic")	100	**0.240**	1
2	transformPhylo.ll(data, phy = tree, model="bm")	100	19.862	82.758
3	phylolm(trait ~ 1, data, tree, model="BM")	100	40.443	168.512

N = 1 000 000 (without *fitContinuous*; ~just for fun – up to now!…)

	Functions	replications	elapsed	relative
1	mvLL(phy, data, method="pic")	1	**0.148**	1
2	transformPhylo.ll(data, phy = tree, model="bm")	1	889.109	6007.493
3	phylolm(trait ~ 1, data, tree, model="BM")	1	1094.32	7394.101

9 - mvSHIFT: Identifiability of parameters with ultrametric and non-ultrametric trees

Estimation of parameters with phylogenetic comparative methods can be enhanced with the inclusion of fossil data (e.g., Ané 2008; Slater *et al.* 2012; Pyron & Burbrink 2012; Ho & Ané 2013). However, an aspect remains too often neglected here: in addition to parameter identifiability, fossil data also allow more accurate detection of changes in evolutionary modes (e.g., Hunt *et al.* 2008; Hunt 2012). Indeed, most of the models are phenomenological and try to describe the pattern of phenotypic variance based on a stochastic process according to a given inter-species (covariances) evolutionary history. Several processes such as the Ornstein-Uhlenbeck one tend to erase the phylogenetic signal (closely related species are as different as distant ones), and the accurate estimation of parameters becomes difficult in a polyphased model because the underlying process is no longer stationary in that case. This is particularly true for the « Ecological release » or « Release and radiate » models (model="OUBM" and "OUBMi" in mvSHIFT), where a flat log-likelihood surface is usually attained with ultrametric trees (Fig. S1.5). In those cases, fossil data added to extant ones in non-ultrametric trees frequently greatly improve the reliability and accuracy of parameter estimates and are mandatory for detecting changes in evolutionary modes.

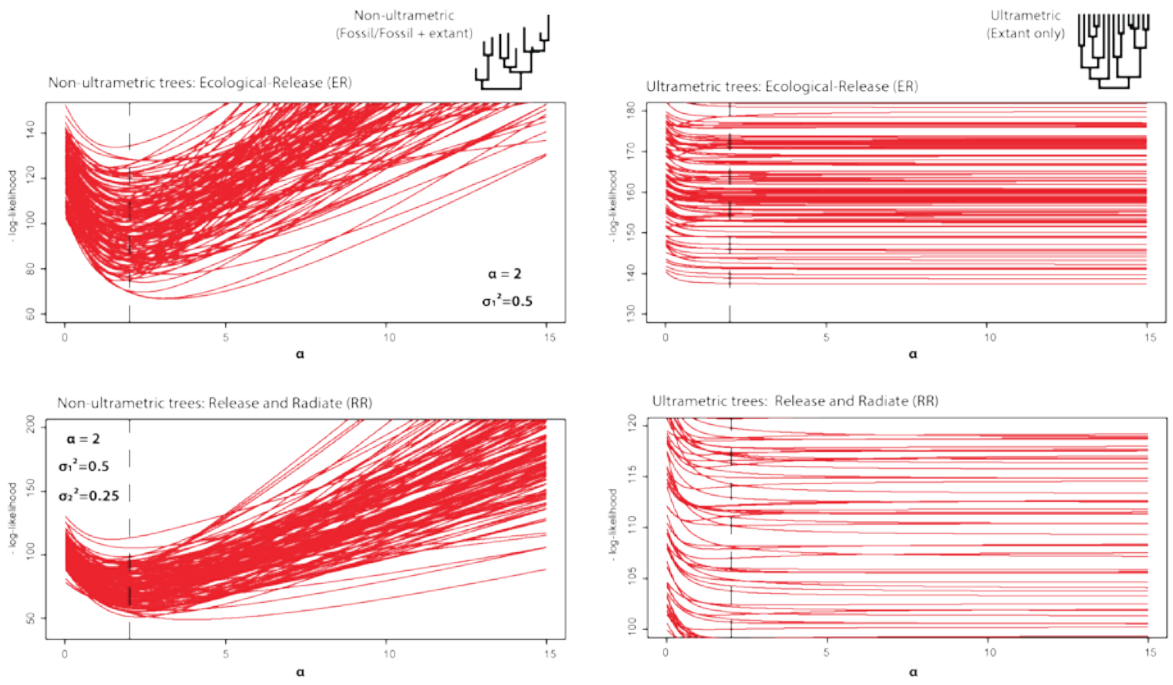

Figure S1.5. Log-likelihood profile for the alpha parameter of "Ecological release" and "Release and radiate" models in mvSHIFT. Root age fixed at 8 Ma and shift fixed at 4 Ma for 100 simulated trees with 32 tip species. Generating values are given in bold characters within the left graphs. The (negative) log-likelihood clearly shows an optimum around the generating value of the alpha parameter (dashed vertical line) when the analysis includes fossil data (graphs on the left), while analysis on extant species alone shows flat likelihood surfaces with no identifiable optimum around the expected (generating) value.

References

Adams, D.C. (2013). Comparing evolutionary rates for different phenotypic traits on a phylogeny using likelihood. *Systematic Biology*, **62**, 181–192.

Andersen, B.S., Waśniewski, J. & Gustavson, F.G. (2001). A recursive formulation of Cholesky factorization of a matrix in packed storage. *ACM Transactions on Mathematical Software*, **27**, 214–244.

Ané, C. (2008). Analysis of comparative data with hierarchical autocorrelation. *The Annals of Applied Statistics*, **2**, 1078–1102.

Bartoszek, K., Pienaar, J., Mostad, P., Andersson, S. & Hansen, T.F. (2012). A phylogenetic comparative method for studying multivariate adaptation. *Journal of Theoretical Biology*, **314**, 204–215.

Beaulieu, J.M., Jhwueng, D.-C., Boettiger, C. & O'Meara, B.C. (2012). Modeling stabilizing selection: expanding the Ornstein-Uhlenbeck model of adaptive evolution. *Evolution*, **66**, 2369–2389.

Butler, M.A. & King, A.A. (2009). Multivariate comparative analysis using OUCH. *Integrative Comparative Biology*, **e24**.

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

Byrd, R.H., Lu, P., Nocedal, J. & Zhu, C. (1995). A limited memory algorithm for bound constrained optimization. *SIAM Journal of Scientific Computing*, **16**, 1190–1208.

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

Massachusetts, USA.

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

Flury, B.D. (1988). *Common principal components and related multivariate models*. John Wiley & Sons, New York.

Freckleton, R.P. (2012). Fast likelihood calculations for comparative analyses. *Methods in Ecology and Evolution*, **3**, 940–947.

Furrer, R. & Sain, S.R. (2010). Spam: A sparse matrix R package with emphasis on MCMC methods for Gaussian Markov random fields. *Journal of Statistical Software*, **36**, 1–25.

Gardiner, C.W. (2004). *Handbook of Stochastic Methods for Physics, Chemistry and the Natural Sciences*, Third Editionn. Springer, Verlag Berlin Heidelberg.

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

Golub, G.H. & Van Loan, C.F. (2013). *Matrix computations*, 4th edn. The John Hopkins University Press, Baltimore.

Gustavson, F.G., Waśniewski, J., Dongarra, J.J. & Langou, J. (2010). Rectangular full packed format for Cholesky's algorithm: factorization, solution and inversion. *ACM Transactions on Mathematical Software*, **37**, 1–33.

Hansen, T.F. (2012). Adaptive landscape and macroevolutionary dynamics. *The adaptive landscape in evolutionary biology*, pp. 205–226. Oxford University Press, Oxford.

Hansen, T.F. (1997). Stabilizing selection and the comparative analysis of adaptation.

Evolution, **51**, 1341–1351.

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

Ho, L.S.T. & Ané, C. (2013). Asymptotic theory with hierarchical autocorrelation: Ornstein-Uhlenbeck tree models. *The Annals of Statistics*, **41**, 957–981.

Ho, L.S.T. & Ané, C. (2014). Intrinsic inference difficulties for trait evolution with Ornstein-Uhlenbeck models. *Methods in Ecology and Evolution*, **5**, 1133–1146.

Hunt, G. (2012). Measuring rates of phenotypic evolution and the inseparability of tempo and mode. *Paleobiology*, **38**, 351–373.

Hunt, G., Bell, M.A. & Travis, M.P. (2008). Evolution toward a new adaptive optimum: phenotypic evolution in a fossil stickleback lineage. *Evolution*, **62**, 700–710.

Ives, A.R., Midford, P.E. & Garland, T.J. (2007). Within-species variation and measurement error in phylogenetic comparative methods. *Systematic Biology*, **56**, 252–270.

Jaffrézic, F., Thompson, R. & Pletcher, S.D. (2004). Multivariate character process models for the analysis of two or more correlated function-valued traits. *Genetics*, **168**, 477–487.

King, A.A. (2013). *Subplex optimization algorithm*.

Lartillot, N. & Delsuc, F. (2012). Joint reconstruction of divergence times and life-history evolution in placental mammals using a phylogenetic covariance model. *Evolution*, **66**, 1773–1787.

Lu, G. & Ades, A.E. (2009). Modeling between-trial variance structure in mixed treatment comparisons. *Biostatistics*, **10**, 792–805.

Meucci, A. (2010). Review of statistical arbitrage, cointegration, and multivariate Ornstein-Uhlenbeck. *SYMMYS*, Working paper, 1–20.

Moler, C. & Van Loan, C.F. (2003). Nineteen dubious ways to compute the exponential of a matrix, twenty-five years later. *SIAM Review*, **45**, 1–46.

Nelder, J.A. & Mead, R. (1965). A simplex method for function minimization. *The Computer Journal*, **7**, 308–313.

Nocedal, J. (1980). Updating quasi-Newton matrices with limited storage. *Mathematics of Computation*, **35**, 773–782.

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

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

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

Pinheiro, J.C. & Bates, D.M. (1996). Unconstrained parameterizations for variance-covariance matrices. *Statistics and Computing*, **6**, 289–296.

Pyron, R.A. & Burbrink, F.T. (2012). Trait-dependent diversification and the impact of palaeontological data on evolutionary hypothesis testing in New World ratsnakes (tribe Lampropeltini). *Journal of Evolutionary Biology*, **25**, 497–508.

Reitan, T., Schweder, T. & Henderiks, J. (2012). Phenotypic evolution studied by layered stochastic differential equations. *The Annals of Applied Statistics*, **6**, 1531–1551.

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

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

Revell, L.J. & Harmon, L.J. (2008). Testing quantitative genetic hypotheses about the evolutionary rate matrix for continuous characters. *Evolutionary Ecology Research*, **10**, 311–331.

Schäfer, J., Opgen-Rhein, R., Zuber, V., Ahdesmäki, M., Silva, P.D. & Strimmer, K. (2013). *Corpcor: Efficient estimation of covariance and (partial) correlation*.

Seber, G.A.F. & Lee, A.J. (2003). *Linear Regression Analysis*, second editionn. John Wiley & Sons, Inc., Hoboken, Ney Jersey.

Slater, G.J., Harmon, L.J. & Alfaro, M.E. (2012). Integrating fossils with molecular phylogenies improves inference of trait evolution. *Evolution*, **66**, 3931–3944.

Slater, G.J. & Pennell, M. (2014). Robust regression and posterior predictive simulation increase power to detect early bursts of trait evolution. *Systematic Biology*, **63**:293-308.

Sy, J.P., Taylor, J.M.G. & Cumberland, W.G. (1997). A stochastic model for the analysis of bivariate longitudinal AIDS data. *Biometrics*, **53**, 542–555.