## APPLICATION

# mvMORPH: an R package for fitting multivariate evolutionary models to morphometric data

**Julien Clavel[1,2]\*, Gilles Escarguel[2] and Gildas Merceron[3]**

[1]*Ecole Normale Supérieure, IBENS UMR 8197, CNRS, 46 rue d'Ulm, 75005 Paris, France;* [2]*Laboratoire de Géologie de Lyon, UMR 5276, CNRS, UCB Lyon 1, ENS Lyon, Campus de la Doua, 2 rue Raphaël Dubois, 69622 Villeurbanne Cedex, France; and* [3]*IPHEP, UMR 7262, CNRS, Université de Poitiers, Bat. B35 – TSA-51106 – 6, rue M. Brunet, 86073 Poitiers Cedex 9, France*

### Summary

**1.** We present mvMORPH, a package of multivariate phylogenetic comparative methods for the R statistical environment. mvMORPH is freely available on the CRAN package repository (http://cran.r-project.org/web/packages/mvMORPH/).

**2.** mvMORPH allows fitting a range of multivariate evolutionary models under a maximum-likelihood criterion. Initially developed in the context of phylogenetic analysis of multiple morphometric traits, its use can be extended to any biological data set with one or multiple covarying continuous traits. All the fitting models include the possibility to use SIMMAP-like mapping, which may be useful for fitting changes along lineages at a given point in time. All models provide diagnostic metrics for convergence and reliability of estimates, as well as the possibility to include trait measurement errors in model estimates.

**3.** New features provided by the mvMORPH package include the possibility of fitting models with changes in the mode of evolution along the phylogeny, which will be particularly meaningful in comparative analyses that include extinct taxa, for example when testing changes in evolutionary mode associated with global biotic/abiotic events.

**4.** We briefly describe the models already included in mvMORPH and provide some demonstration of the use of the package with two simulated worked examples.

**Key-words:** Brownian motion, Early Burst, multivariate comparative methods, Ornstein–Uhlenbeck, stochastic mapping

## Introduction

Comparative methods for fitting evolutionary models to continuous data are becoming increasingly popular (e.g. Hansen 1997; Pagel 1999; Butler & King 2004; O'Meara *et al.* 2006; Thomas, Freckleton & Székely 2006; Hansen, Pienaar & Orzack 2008; Beaulieu *et al.* 2012; Thomas & Freckleton 2012; Adams 2013; Ingram & Mahler 2013). Over the last years, these methods used to study evolutionary patterns of continuous traits within phylogenies have been widely applied on a trait-by-trait basis for estimating shifts in evolutionary rates associated with environmental global changes, the timing of acquisition of key innovations, or the release of biotic interactions (Collar *et al.* 2009; Price *et al.* 2011, 2013; Slater 2013), as well as changes in modes of evolution and adaptation towards different optima (Scales, King & Butler 2009; Collar, Schulte & Losos 2011; Frédérich *et al.* 2013; Knope & Scales 2013; Mahler *et al.* 2013). However, phenomic evolution is multivariate by nature, as natural selection operates on functionally related organism's traits (Walsh 2007; Walsh & Blows 2009), and because traits may covary due to pleiotropic effects, genetic linkages or different sources of developmental constraints (Lande & Arnold 1983; Felsenstein 1988; Armbruster & Schwaegerle 1996; Armbruster *et al.* 2014). Thus, multivariate analyses should logically be favoured over univariate ones (Fig. 1) for describing the complex structure of multiple covarying traits and measurements as routinely studied in morphometrics. Moreover, by considering the evolutionary history of phylogenetically related species, we can further assess how and when phenotypic covariations evolved, either in relation or not with biotic/abiotic events that potentially drive evolution (Fig. 2). Here, we present a new R package, mvMORPH, dedicated to the phylogenetic comparative analysis of biological data sets with multiple covarying continuous traits.

Phenotypic multivariate evolution has been thoroughly studied over the last decades, particularly in quantitative genetics in relation to the evolution of genetic covariances (Lande 1979; Lande & Arnold 1983; Cheverud 1984; Felsenstein 1988;

*Correspondence author. E-mails: clavel@biologie.ens.fr, julien.clavel@hotmail.fr
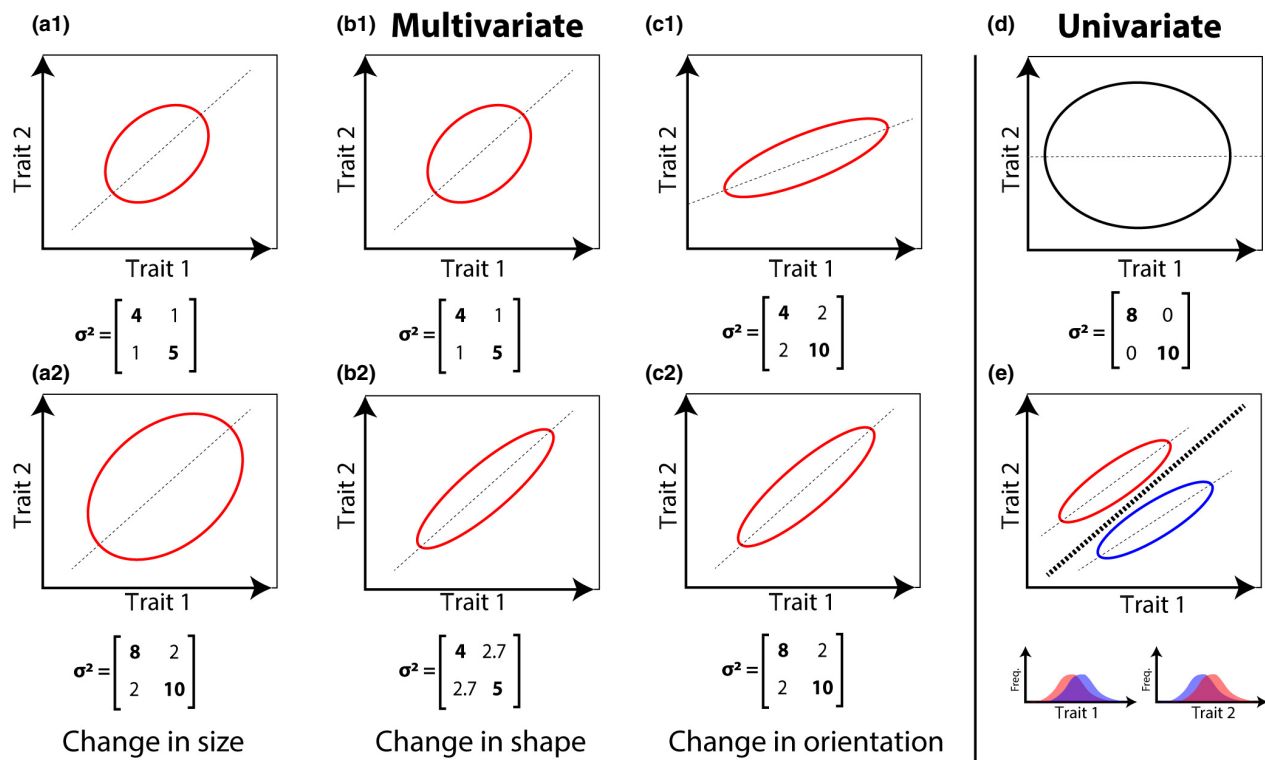
**Fig. 1.** Different aspects of trait variances and covariances (as recorded in the $\sigma^2$ matrices) described by multivariate analysis as the result of distinct selective, functional and/or genetic constraints. Bivariate cases are illustrated here for the sake of graphical simplicity, but same reasoning applies to any multivariate case. In (a1) and (a2), the correlations between traits 1 and 2 are similar (same covariance/variances ratio), but trait variances differ. Conversely, in (b1) and (b2), the variances are identical, but the correlations between traits 1 and 2 differ (higher in b2 than in b1). Finally, in (c1) and (c2) trait relationships differ in both trait variances and correlations. Contrasting with these multivariate cases, univariate analysis where only differences in trait variances can be assessed and interactions (i.e. covariances) between traits are simply ignored is illustrated in (d). Last, traits may also differ in their means rather than in their variance–covariance structures. When trait means differ along axes of low variance as in (e), trait-by-trait (univariate) analyses often fail at identifying differences between taxa (blue and red ellipses) due to lack of statistical power (see example 1 in Appendix S2).

Zeng 1988; Endler 1995; Arnold, Pfrender & Jones 2001; Steppan, Phillips & Houle 2002) and has led to particular expectations over macroevolutionary times (e.g. Schluter 1996; Steppan 1997; Arnold, Pfrender & Jones 2001; Baker & Wilkinson 2003; Polly 2004; Hohenlohe & Arnold 2008; Revell & Harmon 2008; Shoval *et al.* 2012; Goswami *et al.* 2014). Importantly, modelling multivariate phenotypic evolution through trait variances and pairwise covariances allows describing several aspects (e.g. correlated evolution, correlated selection) that separate univariate analyses cannot provide (Fig. 2). Indeed, univariate analyses are merely limited for describing complex structures and generally show less statistical power than multivariate counterparts (Zheng *et al.* 2009). Moreover, while complex multivariate data sets are often first reduced through ordination methods such as PCA and a subset of the first principal components is then treated as independent univariate traits, evolutionary model inferences from these transformed spaces can be severely biased (Revell 2009; Monteiro 2013; Polly *et al.* 2013; Uyeda, Caetano & Pennell 2015). Finally, it is a rather common approach in comparative analysis to regressing traits on each other by assuming, often arbitrarily, that one trait is the dependent variable and the other the independent variable. Nevertheless, such an approach implicitly assumes that one trait is evolutionarily responding to the other (Hansen, Pienaar & Orzack 2008; Revell 2010; Hansen & Bartoszek 2012), a critical assumption of biological causality which is actually relaxed in a multivariate setting focused on mere correlations among response variables.

The multivariate models provided by the mvMORPH package will be also of particular interest for evolutionary biologists integrating fossil specimens and taxa in their studies, as the package allows estimating models of changes in evolutionary modes and/or rates at a given point in time (e.g. Slater 2013). Indeed, integration of extinct clades in comparative studies is an important step forward in the study of morphological diversification (Slater, Harmon & Alfaro 2012; Slater 2013): while some studies have pointed towards the need to consider phylogenies in among-group comparisons of morphological rates of evolution (O'Meara *et al.* 2006; Ricklefs 2006; Adams *et al.* 2009; Revell & Collar 2009), it appears also necessary to consider underlying processes which may affect the realized variances (Hunt 2012). Actually, some processes such as those leading to shifts from the ancestral state to the optima can be estimated only when including fossil taxa in the analysis (Hansen 1997; Hunt, Bell & Travis 2008). Moreover, although
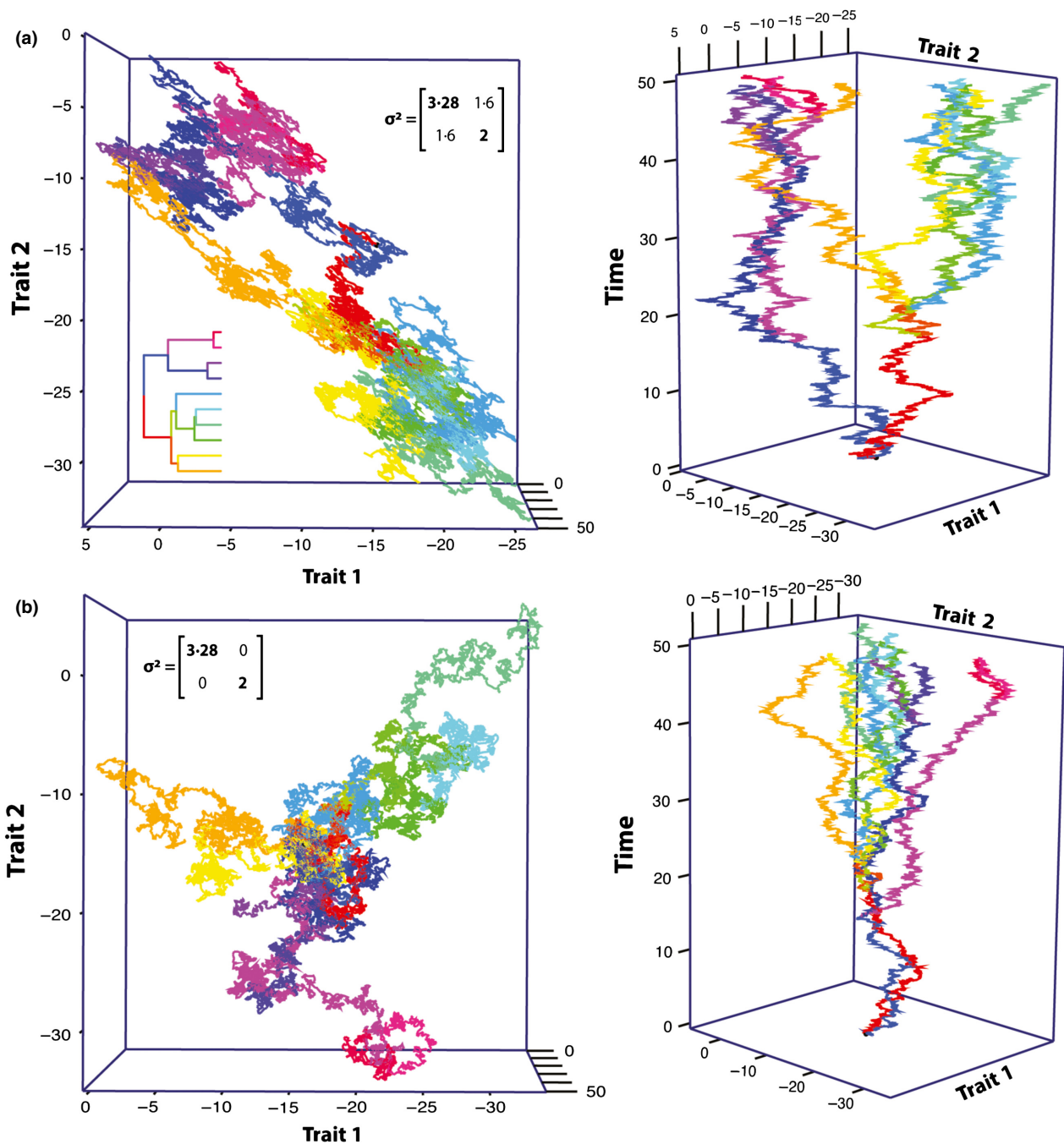
**Fig. 2.** Examples of two phenotypic traits evolving under a bivariate Brownian motion (BM) process within a 10 species tree [inset in (a)]. In (a), a positive ($s_{12}$ = 1·6) evolutionary covariance is accounted for by the BM process, making phenotypic changes in traits 1 and 2 correlated throughout the species evolutionary history (i.e. interspecific divergences in phenotype are more likely to occur along the correlation axis between traits 1 and 2). Conversely, in (b), traits evolve independently in the bivariate space (i.e. there is a null evolutionary covariance between traits 1 and 2). Importantly, the covariations between traits of phylogenetically related species may vary through time (e.g. when species are diversifying functionally during an adaptive radiation) or among different clades (e.g. release of competitive pressures) and may lead to particular expectations about changes in evolutionary correlations under various macroevolutionary scenarios that can be explicitly tested with *mvMORPH* by using various models and mapped histories (Table 1).

the use of non-ultrametric trees (whether including fossil taxa or fast evolving organisms such as viruses) enhances the accuracy and identifiability of parameter estimates (Slater, Harmon & Alfaro 2012; Ho & Ané 2013), it remains computationally limited to some of the current software implementations (Slater 2014).

## Description

### THE mvMORPH PACKAGE

The *mvMORPH* package provides in a unified environment useful tools to deal with ultrametric or non-ultrametric trees over

**Table 1.** Summary of evolutionary models and main functions currently available in the MVMORPH package (version 1.0.5)

| Function | Model | Number of parameters* | Short description of the associated evolutionary model |
|---|---|---|---|
| mvBM | BM1 | $(p(p + 1)/2) + p$ | BM on the whole tree (unique rate matrix) |
| | BMM | $(p(p + 1)/2)k + p$ | BM with multiple rates matrix (multiple selective regimes) |
| mvOU | OU1 | $2(p(p + 1)/2) + p$ | OU process with a unique adaptive optimum per trait |
| | OUM | $2(p(p + 1)/2) + kp$ | OU process with multiple adaptive optima per trait |
| mvEB | EB–ACDC | $(p(p + 1)/2) + p + 1$ | EB model or decelerating model of evolutionary rates. |
| mvSHIFT | BMEB/EBBM | $(p(p + 1)/2) + p + 1$ | Shift of a BM to/from EB process at a given point in time |
| | BMEBi/EBBMi | $2(p(p + 1)/2) + p + 1$ | Same as BMEB/EBBM with independent rates on each time slice |
| | BMOU/OUBM | $2(p(p + 1)/2) + p$ | Shift of a BM to/from OU process at a given point in time |
| | BMOUi/OUBMi | $3(p(p + 1)/2) + p$ | Same as BMOU/OUBM with independent rates on each time slice |
| | EBOU/OUEB | $2(p(p + 1)/2) + p + 1$ | Shift of a EB to/from OU process at a given point in time |
| | EBOUi/OUEBi | $3(p(p + 1)/2) + p + 1$ | Same as EBOU/OUEB with independent rates on each time slice |
| mvSIM | all | – | Simulate the evolution of traits under the models represented in MVMORPH |
| mvLL | – | – | Compute the log-likelihood of an user provided tree or variance–covariance matrix |
| half-life | OU | – | Compute the phylogenetic half-life for a multivariate OU process |
| stationary | OU | – | Compute the multivariate stationary distribution for a multivariate OU process |
| LRT | – | – | Compute the Log-ratio test for nested models in MVMORPH |

*p*, number of traits; k, number of distinct selective regimes. Model abbreviations: BM, Brownian motion; OU, Ornstein–Uhlenbeck; EB, Early Burst; ACDC, exponentially accelerating or decelerating.
*The number of parameters can be actually lower, depending on matrix constraints used.

a range of models of continuous trait evolution (Table 1). MVMORPH provides fast (see drawback below) multivariate implementations of comparative approaches allowing the incorporation of measurement errors (Ives, Midford & Garland 2007; Felsenstein 2008; Revell & Reynolds 2012; Silvestro *et al.* 2015), missing values and various parameterizations; it allows the user to handle trees in the SIMMAP-like format (Huelsenbeck, Nielsen & Bollback 2003; Bollback 2006) which provide a flexible way to map evolutionary hypothesis on a phylogeny thanks to existing functions from the phytools package (Revell 2012). Last, MVMORPH also allows simulating the multivariate evolution of traits (*mvSIM* function) and provides tools for users and developers with a general wrapper *mvLL* for computing log-likelihood of user-specified and customized models of trait evolution with various fast methods (see Appendix S1 for details, Supporting information).

## MULTIVARIATE MODELS IMPLEMENTED IN mvMORPH

The multivariate models for continuous trait evolution implemented in MVMORPH are summarized in Table 1. Most of these models have been already widely described in an univariate context (Hansen 1997; Blomberg, Garland & Ives 2003; Butler & King 2004; O'Meara *et al.* 2006; Thomas, Freckleton & Székely 2006; Harmon *et al.* 2010; Slater 2013). Below we give a short presentation of their multivariate extension as provided in the package (see Appendix S1 for detailed descriptions).

## LOG-LIKELIHOOD OF MULTIVARIATE MODELS

Most phylogenetic comparative analyses are done by fitting a generalized least square (GLS) model of the form:

$$Y = \beta X + \epsilon \qquad \text{eqn 1}$$

For a multivariate model with m traits and n species, $Y$ is a vector of length $mn$ and $\epsilon$ is a vector of $mn$ residuals normally distributed in $\mathcal{N}(0, V)$, with $V$ a $mn \times mn$ phylogenetic multivariate variance–covariance matrix. If we fit an intercept-only model (i.e. a correlational model where we are estimating the phylogenetic signal in $Y$), then $X$ is a $mn \times m$ design matrix, and $\beta$ – the (notional) ancestral state expectation – is a vector of parameters to estimate of length $m$ (assuming a unique ancestral state for each trait).

As for the univariate case, the fit of multivariate models of continuous trait evolution is assessed by maximization of the log-likelihood eqn 2 (e.g. Martins & Hansen 1997; Blomberg, Garland & Ives 2003; O'Meara *et al.* 2006).

$$\log(\mathcal{L}) = -\frac{1}{2}\left[mn\log(2\pi) + \log|V| + (Y - \beta X)^{\mathrm{T}} V^{-1} (Y - \beta X)\right]$$
$$\text{eqn 2}$$

Evolutionary parameters for the various models of phenotypic evolution [Brownian motion (BM), Ornstein–Uhlenbeck (OU), etc.] enter this equation through the phylogenetic variance–covariance matrix $V$, while $\beta$ is given by the GLS solution (eqn 3).

$$\beta = (X^T V^{-1} X)^{-1} X^T V^{-1} Y \qquad \text{eqn 3}$$

### Brownian motion

The BM model is the most widespread evolutionary model used in comparative methods (e.g. Felsenstein 1985). Its multivariate extension is straightforward and has been already thoroughly described (Felsenstein 2004; Revell & Harmon 2008; Revell & Collar 2009; Freckleton 2012). The variance–covariance matrix related to a multivariate BM process (eqn 4) is the Kronecker product of the phylogenetic variance–covariance matrix $C$ (for which each element $C_{ij}$ is the height above the root for the common ancestor of species $i$ and $j$; see Rohlf 2001; Revell & Collar 2009) with the maximum-likelihood estimate of the evolutionary rate matrix $R$ (Revell & Harmon 2008; Revell & Collar 2009; see also Appendix S1).

$$\text{Cov}\left[Y_{ik}, Y_{jl}\right] = C_{ij} R_{kl} \qquad \text{eqn 4}$$

### Ornstein–Uhlenbeck

The OU model is an appealing alternative to the BM model as it describes processes where the traits variance is constrained around one or several optima often referred as 'selective regime' or 'adaptive zone' optima at the macroevolutionary scale (Hansen 1997, 2012; Butler & King 2004). In comparative method studies, this process has often been misleadingly interpreted in terms of Lande's (1976) model of population undergoing stabilizing selection in a static landscape (see also discussion in Harmon *et al.* 2010 and Pennell & Harmon 2013).

The OU model for a multivariate trait matrix Y along a lineage is described by the differential equation:

$$dY(t) = A(\beta - Y(t))dt + R\,dW(t) \qquad \text{eqn 5}$$

where $A$ and $R$ are square matrices of the strength of selection and the random drift, respectively, β is a vector of mean (optimum) values, and $W$ is the diffusion parameter (Butler & King 2009; Bartoszek *et al.* 2012; Monteiro 2013). The eigenvalues and eigenvectors of the matrix $A$ describe the path towards the optimum, while the matrix $R$ describes the covariations of the random perturbations. The variance–covariance matrix computed between the *m* traits at tips *i* and *j* can be obtained in two different ways, depending on whether the value at the root is estimated or assumed at the stationary point (see Ho & Ané 2013 for the univariate counterpart). Particularly, when $A$ is symmetric positive definite (i.e. $A$ has only real positive eigenvalues), a stationary solution exists (Gardiner 2004) for which the variance–covariance matrix can be computed as follows:

$$\text{Cov}\left[Y_i, Y_j\right] = \int_0^{C_{ij}} e^{-A(C_{ii}-\upsilon)} R\, e^{-A^T(C_{jj}-\upsilon)} d\upsilon \qquad \text{eqn 6}$$

This is the implementation found in the R package OUCH (Butler & King 2009). Alternatively, if the covariance matrix is conditioned on the root, a more general form which also allows non-symmetric matrices $A$ may be used (Eq. B.3 in Bartoszek *et al.* 2012; see also Reitan, Schweder & Henderiks 2012 and Appendix S1 for details):

$$\text{Cov}\left[Y_i, Y_j\right] = e^{-A(C_{ii}-C_{ij})} \left( \int_0^{C_{ij}} e^{-A\upsilon} R\, e^{-A^T\upsilon} d\upsilon \right) e^{-A^T(C_{jj}-C_{ij})} \qquad \text{eqn 7}$$

This second approach is particularly useful for fitting multivariate models to non-ultrametric trees (e.g. including fossil species) for which the estimation of the root state is feasible. On ultrametric trees, both approaches should converge to the same estimates if the process is stationary.

As for the univariate case, the expectation of the OU process (eqn 8) can be computed as the weighted average of the optima and the common ancestral state at the root of the tree (Hansen 1997).

$$\mathbb{E}[Y_i(t)] = e^{-At}\mathbb{E}[Y_i(0)] + \int_0^t e^{A(\upsilon-t)} A\beta_i(\upsilon)d\upsilon \qquad \text{eqn 8}$$

This expectation is used to estimate the root and the optima by entering the coefficients of their relative contribution in the design matrix $X$ of the GLS equation (eqn 3). The various parameterizations implemented for the root state are described and discussed in the Appendix S1.

### Early Burst

The mvMORPH package allows computing a joint estimate for a decrease or increase in evolutionary rates in a multivariate data set, following Blomberg, Garland & Ives (2003) exponentially accelerating or decelerating (ACDC) model or Harmon *et al.*'s (2010) Early Burst (EB) model. These models have been used for studying adaptive radiations where evolutionary theory predicts a burst of evolutionary rates and an increase in morphological disparity, followed by a decrease in rates of morphological evolution as ecological niches are filled (Cooper & Purvis 2010; Harmon *et al.* 2010; Slater *et al.* 2010; Ingram, Harmon & Shurin 2012; Slater 2013). Here, the phylogenetic matrix $C$ is transformed by the exponential transformation (*g* parameter) of the ACDC model (Blomberg, Garland & Ives 2003), which is jointly estimated during likelihood maximization of the evolutionary rate matrix (eqn 9).

$$\text{Cov}\left[Y_{ik}, Y_{jl}\right] = R_{kl}\left( \frac{e^{gC_{ij}} - 1}{g} \right) \qquad \text{eqn 9}$$

### Shifts models

Slater (2013) recently introduced a heterogeneous model allowing for the shift from one evolutionary process to another in the same tree. Dealing with heterogeneous models was also

previously proposed by using tree-stretching approaches (O'Meara 2012; but see Slater 2014). Such approaches are particularly appealing when mixing fossil and extant data to assess changes in evolutionary modes in addition to temporal changes in evolutionary tempo (Hunt, Bell & Travis 2008; Slater & Harmon 2013; see Appendix S1). A combination of the various models implemented in MVMORPH can be fitted to two successive time slices or mapped era using the mvSHIFT function (Table 1). For instance, the variance–covariance for the 'ecological release' model proposed by Slater (2013), combining an OU process then followed by a BM process, is computed as follows:

$$\left[Y_i, Y_j\right] = \boldsymbol{Cbm}_{ij}\boldsymbol{R} + e^{-A\left(Cou_{ii}-Cou_{ij}\right)}\left(\int_0^{Cou_{ij}} e^{-A\upsilon}\boldsymbol{R}e^{-A^T\upsilon}d\upsilon\right)$$
$$e^{-A^T\left(Cou_{jj}-Cou_{ij}\right)}$$

eqn 10

where $\boldsymbol{Cbm}_{ij}$ represents the history between the taxa $i$ and $j$ shared during the time period where the BM process took place; $\boldsymbol{Cou}_{ij}$ represents the shared history between both taxa accrued during the OU process; and $\boldsymbol{Cou}_{jj}$ and $\boldsymbol{Cou}_{ii}$ represent the evolutionary histories (variances) accrued in each lineage during the OU process.

### *Constraining the parameter spaces*

Estimating parameters of multivariate models is a complex task because matrices of parameters must always fulfil specific conditions, for example positive definiteness for variance–covariance matrices. Generally, one of the efficient ways to solve this issue is to use an unconstrained optimization of the likelihood with a parameterization that enforces the matrices to a particular structure (Pinheiro & Bates 1996). Various parameterizations are used in MVMORPH (see Appendix S1) for ensuring positive definiteness of rate matrices (Pinheiro & Bates 1996), conditioning matrices on their eigenvalues (e.g. Sy, Taylor & Cumberland 1997; Jaffrézic, Thompson & Pletcher 2004; Bartoszek *et al.* 2012), or testing significance of interactions between traits by constraining the parameter space and taking advantage of joint likelihood optimization (e.g. Revell & Collar 2009; Adams 2013).

### Alternatives

Other R implementations of multivariate comparative approaches already exist that calculate parts of the same or alternative models to those provided in MVMORPH. Multivariate extensions of Pagel (1999) branch-length transformations are proposed in MOTMOT (Thomas & Freckleton 2012; no more on the CRAN repository but accessible from the archive); BM with multiple evolutionary rate matrices can be fitted with *evol.vcv* as well as a phylogenetic canonical correlation analysis with *phyl.cca* in phytools (Revell 2012), and geomorph allows computing BM rates for high-dimensional data sets (Adams & Otarola-Castillo 2013). Multivar-

iate OU is available from the OUCH package (Butler & King 2009) as well as MVSLOUCH that allows fitting multivariate OU models with a randomly evolving predictor variable (Bartoszek *et al.* 2012).

### Drawbacks

#### COMPUTATIONAL PERFORMANCES

One of the main drawback of working with multivariate dimensions is that computational time grows up exponentially with variance–covariance matrix construction, inversion and computation of the determinant (Felsenstein 1973; Henderson 1976; Hadfield & Nakagawa 2010; Freckleton 2012). Recent alternative algorithms have been published to cope with the problem of growing number of species and dimensions in comparative analyses (Hadfield & Nakagawa 2010; FitzJohn 2012; Thomas & Freckleton 2012; Ho & Ané 2014; Lartillot 2014). Nevertheless, no general solution emerged so far, and most of these fast algorithms remain actually limited with multivariate data. For instance, the algorithm proposed by Ho & Ané (2014) can be easily extended to multivariate case when phylogenetic variance–covariance submatrices are proportional and that the multivariate variance–covariance matrix maintains a three-point structure, which is unlikely for non-Brownian processes such as the multivariate OU. Similarly, the pruning algorithm used by Thomas & Freckleton (2012) in the MOTMOT package allows estimating a limited range of multivariate models such as a single optimum OU where selection acts independently on each traits.

Facing this lack of general solution, particularly for complex models such as multivariate OU processes, MVMORPH implements various approaches to speed up the computations (see Appendix S1 for details). The main implementations consist in different methods that avoid the explicit computation of the inverse and determinant of the variance–covariance matrix to solve the linear model in eqn 1 through Cholesky factorization and independent contrasts. Two efficient algorithms are used for computing the Cholesky factors. The first one (method = 'rpf') uses the rectangular full-packed format algorithm proposed by Gustavson *et al.* (2010). This algorithm uses half memory for the storage of symmetric matrices and reorganizes the upper triangular matrix in a rectangular format allowing the use of fast BLAS 3 matrix multiplication routines. The second method (method = 'sparse') takes advantage of the sparsity structure of phylogenetic variance–covariance matrices to use faster sparse methods and updating of the Cholesky factor (Furrer & Sain 2010; see Appendix S1). Finally, an extremely fast implementation of the likelihood computation using contrasts (Freckleton 2012; see Appendix S1) is also provided in MVMORPH for BM-based models (method = 'pic'). Overall, the computational solutions implemented in MVMORPH allow realistic computation times even for very large-sized data sets, with relative speeds several times faster than other available packages (see Appendix S1, Table S1).

## HIGH-DIMENSIONALITY

With increased number of traits and parameters (Table 1), over-parameterized multivariate models are often highly penalized by model selection criteria (Burnham & Anderson 2002; Revell & Collar 2009). Moreover, when the number of analysed traits reaches the number of studied species, the matrix computations involved in the likelihood calculation cannot be completed and parameter estimates become less accurate (Adams 2014a,b,c). Accordingly, Adams (2014a,b,c) has proposed methods based on matrix distances to allow the use of large *m*/small *n* data sets. Although these nonparametric approaches are appealing for comparing high-dimensional multivariate data sets, they do not fully address all the multivariate questions of trait evolution as they summarize the whole covariation information through a single metric, they assume a common phylogenetic structure for each trait, and they are limited to the BM model. Indeed, high-dimensional methods in phylogenetic comparative analysis should deserve more attention in the future as common approaches used to reduce the dimensionality of such data sets (e.g. principal component analysis) are prone to several biases (Monteiro 2013; Polly *et al.* 2013; Uyeda, Caetano & Pennell 2015).

## Worked examples

To illustrate how mvMORPH works, we describe and comment two simulated examples in the Appendix S2 (also included as vignettes in the mvMORPH package). In the first example, we show how multivariate models can out-compete separate univariate analyses for a phylogeny involving two distinct selective regimes corresponding to two separate species groups. In the second worked example, we illustrate how to assess significances of differences in evolutionary covariations by testing a hierarchy of nested models. In addition, the help pages embedded within the mvMORPH package also provide several simulated examples illustrating how the various functions work.

## Conclusion

Elaboration of the mvMORPH package was initially motivated by the need for a unified R package for fitting multivariate models of phenotypic evolution incorporating measurement errors and using flexible ways to map evolutionary scenarios for morphometric data. Obviously, such objectives can be extended to any biological data set made of covarying continuous traits, as well as univariate traits. mvMORPH also implements models that were not previously designed for the multivariate case, as well as new ones. Last, but not least, the fast likelihood implementations provided by the mvMORPH package may be particularly appealing when dealing with large data sets and assessing the statistical power of comparative methods through modelling approaches (Boettiger, Coop & Ralph 2012). The mvMORPH package is in continuous development and can be followed and contributed from the gitHub website (https://github.com/JClavel/

mvMORPH). mvMORPH (current version: 1.0.5) is an open-source R package available for download on the CRAN repository (http://cran.r-project.org/web/packages/mvMORPH/).

## Data accessibility

This manuscript does not include any data.

## References

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

Adams, D.C. (2014a) A generalized K statistic for estimating phylogenetic signal from shape and other high-dimensional multivariate data. *Systematic Biology*, **63**, 685–697.

Adams, D.C. (2014b) A method for assessing phylogenetic least squares models for shape and other high-dimensional multivariate data. *Evolution*, **68**, 2675–2688.

Adams, D.C. (2014c) Quantifying and comparing phylogenetic evolutionary rates for shape and other high-dimensional phenotypic data. *Systematic Biology*, **63**, 166–177.

Adams, D.C. & Otarola-Castillo, E. (2013) geomorph: an R package for the collection and analysis of geometric morphometric shape data. *Methods in Ecology and Evolution*, **4**, 393–399.

Adams, D.C., Berns, C.M., Kozak, K.H. & Wiens, J.J. (2009) Are rates of species diversification correlated with rates of morphological evolution? *Proceedings of the Royal Society B*, **276**, 2729–2738.

Armbruster, W.S. & Schwaegerle, K.E. (1996) Causes of covariation of phenotypic traits among populations. *Journal of Evolutionary Biology*, **9**, 261–276.

Armbruster, W.S., Pélabon, C., Bolstad, G.H. & Hansen, T.F. (2014) Integrated phenotypes: understanding trait covariation in plants and animals. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, **369**, 1–16.

Arnold, S.J., Pfrender, M.E. & Jones, A.G. (2001) The adaptive landscape as a conceptual bridge between micro and macroevolution. *Genetica*, **112–113**, 9–32.

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

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.

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

Boettiger, C., Coop, G. & Ralph, P. (2012) Is your phylogeny informative? Measuring the power of comparative methods. *Evolution*, **66**, 2240–2251.

Bollback, J.B. (2006) SIMMAP: Stochastic character mapping of discrete traits on phylogenies. *BMC Bioinformatics*, **7**, 1–7.

Burnham, K.P. & Anderson, D.R. (2002) *Model Selection and Multi-Model Inference: A Practical Information-Theoretic Approach*. Springer-Verlag, New York, USA.

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

Butler, M.A. & King, A.A. (2009) Multivariate comparative analysis using OUCH. *Integrative Comparative Biology*, **57.5**, e24.

Cheverud, J. (1984) Quantitative genetics and developmental constraints on evolution by selection. *Journal of Theoretical Biology*, **110**, 155–171.

Collar, D.C., Schulte, J.A. II & Losos, J.B. (2011) Evolution of extreme body size disparity in monitor lizards (varanus). *Evolution*, **65**, 2664–2680.

Collar, D.C., O'Meara, B.C., Wainwright, P.C. & Near, T.J. (2009) Piscivory limits diversification of feeding morphology in centrarchid fishes. *Evolution*, **63**, 1557–1573.

Cooper, N. & Purvis, A. (2010) Body size evolution in mammals: complexity in tempo and mode. *The American Naturalist*, **175**, 727–738.

Endler, J.A. (1995) Multiple-trait coevolution and environmental gradients in guppies. *Trends in Ecology & Evolution*, **10**, 22–29.

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

Felsenstein, J. (1985) Phylogenies and the comparative method. *The American Naturalist*, **125**, 1–15.

Felsenstein, J. (1988) Phylogenies and quantitative characters. *Annual Review of Ecology, Evolution and Systematics*, **19**, 445–471.

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

Felsenstein, J. (2008) Comparative methods with sampling error and within-species variation : contrasts revisited and revised. *The American Naturalist*, **171**, 712–725.

FitzJohn, R.G. (2012) Diversitree: comparative phylogenetic analyses of diversification in R. *Methods in Ecology and Evolution*, **3**, 1084–1092.

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

Frédérich, B., Sorenson, L., Santini, F., Slater, G.J. & Alfaro, M.E. (2013) Iterative ecological radiation and convergence during the evolutionary history of damselfishes (Pomacentridae). *The American Naturalist*, **181**, 94–113.

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*, 3rd edn. Springer, Berlin.

Goswami, A., Smaers, J.B., Soligo, C. & Polly, P.D. (2014) The macroevolutionary consequences of phenotypic integration: from development to deep time. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, **369**, 1–15.

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.

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

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

Hansen, T.F. (2012) Adaptive landscape and macroevolutionary dynamics. *The Adaptive Landscape in Evolutionary Biology* (eds E. Svensson & R. Calsbeek), pp. 205–226. Oxford University Press, Oxford.

Hansen, T.F. & Bartoszek, K. (2012) Interpreting the evolutionary regression : the interplay between observational and biological errors in phylogenetic comparative studies. *Systematic Biology*, **61**, 413–425.

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

Harmon, L.J., Losos, J.B., Davies, J.T., Gillespie, R.G., Gittleman, J.L., Jennings, B.W. *et al.* (2010) Early bursts of body size and shape evolution are rare in comparative data. *Evolution*, **64**, 2385–2396.

Henderson, C.R. (1976) A simple method for computing the inverse of a numerator relationship matrix used in prediction of breeding values. *Biometrics*, **32**, 69–88.

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) A linear-time algorithm for Gaussian and non-Gaussian trait evolution models. *Systematic Biology*, **63**, 397–408.

Hohenlohe, P.A. & Arnold, S.J. (2008) MIPoD: A hypothesis-testing framework for microevolutionary inference from patterns of divergence. *The American Naturalist*, **171**, 366–385.

Huelsenbeck, J.P., Nielsen, R. & Bollback, J.B. (2003) Stochastic mapping of morphological characters. *Systematic Biology*, **52**, 131–158.

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.

Ingram, T., Harmon, L.J. & Shurin, J.B. (2012) When should we expect early bursts of trait evolution in comparative data? Predictions from an evolutionary food web model. *Journal of Evolutionary Biology*, **25**, 1902–1910.

Ingram, T. & Mahler, D.L. (2013) SURFACE: detecting convergent evolution from comparative data by fitting Ornstein-Uhlenbeck models with stepwise Akaike Information Criterion. *Methods in Ecology and Evolution*, **4**, 416–425.

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.

Knope, M.L. & Scales, J.A. (2013) Adaptive morphological shifts to novel habitats in marine sculpin fishes. *Journal of Evolutionary Biology*, **26**, 472–482.

Lande, R. (1976) Natural selection and random genetic drift in phenotypic evolution. *Evolution*, **30**, 314–334.

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

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

Lartillot, N. (2014) A phylogenetic Kalman filter for ancestral trait reconstruction using molecular data. *Bioinformatics*, **30**, 488–496.

Mahler, D.L., Ingram, T., Revell, L.J. & Losos, J.B. (2013) Exceptional convergence on the macroevolutionary landscape in island lizard radiations. *Science*, **341**, 292–295.

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

Monteiro, L.R. (2013) Morphometrics and the comparative method: studying the evolution of biological shape. *Hystrix, the Italian Journal of Mammalogy*, **24**, 1–8.

O'Meara, B.C. (2012) Evolutionary inferences from phylogenies: a review of methods. *Annual Review of Ecology, Evolution and Systematics*, **43**, 267–285.

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.

Pagel, M.D. (1999) Inferring the historical patterns of biological evolution. *Nature*, **401**, 877–884.

Pennell, M.W. & Harmon, L.J. (2013) An integrative view of phylogenetic comparative methods: connections to population genetics, community ecology, and paleobiology. *Annals of the New York Academy of Sciences*, **1289**, 90–105.

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

Polly, P.D. (2004) On the simulation of the evolution of morphological shape: multivariate shape under selection and drift. *Palaeontologia Electronica*, **7.2.7A**, 1–28.

Polly, P.D., Lawing, M.A., Fabre, A.-C. & Goswami, A. (2013) Phylogenetic principal components analysis and geometric morphometrics. *Hystrix*, **24**, 1–9.

Price, S.A., Holzman, R., Near, T.J. & Wainwright, P.C. (2011) Coral reefs promote the evolution of morphological diversity and ecological novelty in labrid fishes. *Ecology Letters*, **14**, 462–469.

Price, S.A., Tavera, J.J., Near, T.J. & Wainwright, P.C. (2013) Elevated rates of morphological and functional diversification in reef-dwelling haemulid fishes. *Evolution*, **67**, 417–428.

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

Revell, L.J. (2009) Size-correction and principal components for interspecific comparative studies. *Evolution*, **63**, 3258–3268.

Revell, L.J. (2010) Phylogenetic signal and linear regression on species data. *Methods in Ecology and Evolution*, **1**, 319–329.

Revell, L.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.

Revell, L.J. & Reynolds, R.G. (2012) A new bayesian method for fitting evolutionary models to comparative data with intraspecific variation. *Evolution*, **66**, 2697–2707.

Ricklefs, R.E. (2006) Time, species, and the generation of trait variance in clades. *Systematic Biology*, **55**, 151–159.

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

Scales, J.A., King, A.A. & Butler, M.A. (2009) Running for your life or running for your dinner: what drives fiber-type evolution in lizard locomotor muscles? *The American Naturalist*, **173**, 543–553.

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

Shoval, O., Sheftel, H., Shinar, G., Hart, Y., Ramote, O., Mayo, A., Dekel, E., Kavanagh, K. & Alon, U. (2012) Evolutionary trade-offs, pareto optimality, and the geometry of phenotype space. *Science*, **336**, 1157–1160.

Silvestro, D., Kostikova, A., Litsios, G., Pearman, P.B. & Salamin, N. (2015) Measurement errors should always be incorporated in phylogenetic comparative analysis. *Methods in Ecology and Evolution*, **6**, 340–346.

Slater, G.J. (2013) Phylogenetic evidence for a shift in the mode of mammalian body size evolution at the Cretaceous-Palaeogene boundary. *Methods in Ecology and Evolution*, **4**, 734–744.

Slater, G.J. (2014) Correction to 'Phylogenetic evidence for a shift in the mode of mammalian body size evolution at the Cretaceous-Palaeogene boundary', and a note on fitting macroevolutionary models to comparative paleontological data sets. *Methods in Ecology and Evolution*, **5**, 714–718.

Slater, G.J. & Harmon, L.J. (2013) Unifying fossils and phylogenies for comparative analyses of diversification and trait evolution. *Methods in Ecology and Evolution*, **4**, 699–702.

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., Price, S.A., Santini, F. & Alfaro, M.E. (2010) Diversity versus disparity and the radiation of modern cetaceans. *Proceedings of the Royal Society B*, **277**, 3097–3104.

Steppan, S.J. (1997) Phylogenetic analysis of phenotypic covariance structure. II. Reconstructing matrix evolution. *Evolution*, **51**, 587–594.

Steppan, S.J., Phillips, P.C. & Houle, D. (2002) Comparative quantitative genetics: evolution of the G matrix. *Trends in Ecology and Evolution*, **17**, 320–327.

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.

Thomas, G.H. & Freckleton, R.P. (2012) MOTMOT: models of trait macroevolution on trees. *Methods in Ecology and Evolution*, **3**, 145–151.

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

Uyeda, J.C., Caetano, D.S. & Pennell, M.W. (2015) Comparative analysis of principal components can be misleading. *Systematic Biology*, **64**, 677–689.

Walsh, B. (2007) Escape from flatland. *Journal of Evolutionary Biology*, **20**, 36–38.

Walsh, B. & Blows, M.W. (2009) Abundant genetic variation + strong selection = multivariate genetic constraints: a geometric view of adaptation. *Annual Review of Ecology, Evolution and Systematics*, **40**, 41–59.

Zeng, Z.-B. (1988) Long-term correlated response, interpopulation covariation, and interspecific allometry. *Evolution*, **42**, 363–374.

Zheng, L., Ives, A.R., Garland, T.J., Larget, B.R., Yu, Y. & Cao, K. (2009) New multivariate tests for phylogenetic signal and trait correlations applied to eco-physiological phenotypes of nine Manglietia species. *Functional Ecology*, **23**, 1059–1069.

## Supporting Information

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

**Appendix S1.** Details on computational methods used by MVMORPH.

**Appendix S2.** Two worked examples of evolutionary model fitting with MVMORPH.