Theoretical Population Biology xxx (xxxx) xxx

Contents lists available at ScienceDirect

Theoretical Population Biology

journal homepage: www.elsevier.com/locate/tpb

Fast likelihood calculation for multivariate Gaussian phylogenetic models with shifts

Venelin Mitov [a,b,*], Krzysztof Bartoszek [c], Georgios Asimomitis [a], Tanja Stadler [a,b]

[a] Department of Biosystems Science and Engineering, Eidgenössische Technische Hochschule Zürich, Basel, Switzerland
[b] Swiss Institute of Bioinformatics, Lausanne, Switzerland
[c] Department of Computer and Information Science, Linköping University, Linköping, Sweden

ARTICLE INFO

Article history:
Received 10 April 2019
Available online xxxx

Keywords:
Pruning
Drift
Selection
Punctuated equilibrium
Lévy process
Jumps

ABSTRACT

Phylogenetic comparative methods (PCMs) have been used to study the evolution of quantitative traits in various groups of organisms, ranging from micro-organisms to animal and plant species. A common approach has been to assume a Gaussian phylogenetic model for the trait evolution along the tree, such as a branching Brownian motion (BM) or an Ornstein–Uhlenbeck (OU) process. Then, the parameters of the process have been inferred based on a given tree and trait data for the sampled species. At the heart of this inference lie multiple calculations of the model likelihood, that is, the probability density of the observed trait data, conditional on the model parameters and the tree. With the increasing availability of big phylogenetic trees, spanning hundreds to several thousand sampled species, this approach is facing a two-fold challenge. First, the assumption of a single Gaussian process governing the entire tree is not adequate in the presence of heterogeneous evolutionary forces acting in different parts of the tree. Second, big trees present a computational challenge, due to the time and memory complexity of the model likelihood calculation.

Here, we explore a sub-family, denoted $\mathcal{G}_{LInv}$, of the Gaussian phylogenetic models, with the transition density exhibiting the properties that the expectation depends **L**inearly on the ancestral trait value and the variance is **Inv**ariant with respect to the ancestral value. We show that $\mathcal{G}_{LInv}$ contains the vast majority of Gaussian models currently used in PCMs, while supporting an efficient (linear in the number of nodes) algorithm for the likelihood calculation. The algorithm supports scenarios with missing data, as well as different types of trees, including trees with polytomies and non-ultrametric trees. To account for the heterogeneity in the evolutionary forces, the algorithm supports models with "shifts" occurring at specific points in the tree. Such shifts can include changes in some or all parameters, as well as the type of the model, provided that the model remains within the $\mathcal{G}_{LInv}$ family. This contrasts with most of the current implementations where, due to slow likelihood calculation, the shifts are restricted to specific parameters in a single type of model, such as the long-term selection optima of an OU process, assuming that all of its other parameters, such as evolutionary rate and selection strength, are global for the entire tree.

We provide an implementation of this likelihood calculation algorithm in an accompanying R-package called **PCMBase**. The package has been designed as a generic library that can be integrated with existing or novel maximum likelihood or Bayesian inference tools.

1. Introduction

Since Felsenstein's (1985) work describing the independent contrasts algorithm, phylogenetic comparative methods (PCMs) have steadily been generalized with respect to available models and implementations of them. Following Felsenstein (1988)'s suggestion, Hansen (1997) described the Ornstein–Uhlenbeck (OU) process in the PCM setting. This led to the implementation of OU models in various packages such as **ouch** (Butler and King, 2004) or **geiger** (Harmon et al., 2008) to name a few. Nowadays, the OU process has become a standard model in the community, alongside the Brownian motion (BM) process popularized previously by Felsenstein (1985) (but see also Edwards (1970) and Lande (1976)). For species being characterized by multiple traits, the multivariate OU processes were introduced by R packages such as **ouch**, **slouch** (Hansen et al., 2008), **mvSLOUCH** (Bartoszek et al., 2012), **mvMORPH** (Clavel et al., 2015), **Rphylopars**

* Corresponding author at: Department of Biosystems Science and Engineering, Eidgenössische Technische Hochschule Zürich, Basel, Switzerland.
E-mail addresses: venelin.mitov@bsse.ethz.ch (V. Mitov), krzysztof.bartoszek@liu.se, krzbar@protonmail.ch (K. Bartoszek), asimomig@mskcc.org (G. Asimomitis), tanja.stadler@bsse.ethz.ch (T. Stadler).

https://doi.org/10.1016/j.tpb.2019.11.005

(Goolsby et al., 2016), again, to name a few. At the core of these methods, the likelihood of the model parameters and tree for given quantitative trait data at the tips is calculated, meaning the probability density of the tip trait values given the parameters and tree is calculated.

At present, the mathematical frameworks for PCMs are applied to situations that are very different from the original motivation of a between-species analysis within a small clade. For example, traits being gene expression levels (Bedford and Hartl, 2009; Rohlfs et al., 2013) or epidemiological measurements (the tree connects the epidemic's outbursts, Pybus et al., 2012) are analysed. With large and diverse clades, such as HIV transmission trees having thousands of tips, e.g. in Hodcroft et al. (2014), Bertels et al. (2017) and Mitov and Stadler (2018), there is a need to vary the parameters of the models across different clades or epochs in the tree. Already e.g. Bartoszek et al. (2012), Butler and King (2004) and Hansen (1997) showed the possibility of varying the deterministic optimum of OU processes. Beaulieu et al. (2012), Eastman et al. (2011), Clavel et al. (2015) and Manceau et al. (2016) went further to allow all parameters of the model to change at *known* "shift" points in the tree. The computationally harder task of inferring the branch and time of shift points has been addressed by Eastman et al. (2011), Ingram and Mahler (2013), Khabbazian et al. (2016), Gill et al. (2016), Caetano and Harmon (2017), and Bastide et al. (2018) with implementations in **AUTEUR**, **SURFACE**, **l1ou**, **BEAST**, **ratematrix**, and **PhylogeneticEM** software packages, respectively.[1] However, each of these tools has specific limitations with respect to the number of traits, the type and size of supported trees, the types of supported models and the parameters that can undergo shifts.

A main obstacle hindering the statistical inference of multivariate Gaussian phylogenetic models, in the general setting of correlated traits evolving over a (possibly big and non-ultrametric) phylogenetic tree, with inference of the branch and time of shifts in some or all model parameters, has been the lack of an efficient (i.e. linear in the number of nodes) likelihood calculation algorithm. Technically, the likelihood calculation requires integrating conditional probability density functions over unobserved values of the traits at the internal nodes. In this article, we propose a computationally efficient general solution to this integration problem, enabling in a longer term both, direct maximum likelihood as well as Bayesian inference. Specifically, we generalize Felsenstein's pruning algorithm to a multivariate Gaussian phylogenetic model with shifts and to any type of tree, including non-ultrametric trees (i.e. phylogenetic trees with tips in the past corresponding to extinct species) and polytomies (i.e. trees with any number of branches descending from a parent node). This algorithm enables the calculation of the log-likelihood of such models, in time proportional to the number of nodes in the tree. We prove that our approach applies to a large class of models, namely the family, hereby denoted $\mathcal{G}_{LInv}$, of all models where, conditional on the ancestral trait value, the descendant's trait value is normally distributed, the expectation of this normal distribution depends linearly on the ancestral trait, and the variance of this normal distribution does not depend on the ancestral value. From a mathematical point of view, our approach can be seen as an equivalent derivation of the Gaussian moment propagation procedure introduced recently by Bastide et al. (2018), see Discussion. More intuitively, we propose a generalization of the analytical integration technique described previously by Hadfield and Nakagawa (2010), FitzJohn (2012), Freckleton (2012), Pybus et al. (2012), Lartillot (2014), Cybis et al. (2015) and Mitov and Stadler (2019) to any type of tree and for multivariate Gaussian models with shifts, both, in the model parameters, as well as the type of the model, provided the new model type belongs to the $\mathcal{G}_{LInv}$ family. Pybus et al. (2012) pointed out that for such a method to work, it is needed "to keep track of partial" means and precisions. Here, we propose a very general, computationally efficient, and developer friendly way of doing this by recursively updating the polynomial representation of the multivariate normal density function. In order to use our approach for a new $\mathcal{G}_{LInv}$ model, one has to define functions for the variance of the transition along a branch, the shift in the mean along a branch (i.e. a vector), and the linear dependency (i.e. a matrix) on the ancestral state. Thus, in our framework, one needs to understand only the process dynamics of a single branch (lineage), something that is usually present at the model formulation stage.

It is important to stress here two points about the presented method and accompanying implementation. First, this work does not cover non-Gaussian phylogenetic models. To the best of our knowledge, for some non-Gaussian models, likelihood calculation methods have been proposed by Ho and Ané (2014a), using a 3-point structure algorithm, and by Hiscott et al. (2016), using numerical integration over the internal nodes. Second, it is beyond the scope of this work to provide a complete inference framework. Rather, this work is limited to providing a comprehensive theoretical description and a software tool for specifying complex multivariate Gaussian models with shifts and for efficiently calculating the likelihood of such models, given a comparative dataset and a phylogenetic tree. A complete use case implementing maximum likelihood inference in real and simulated datasets is provided in a separate paper (Mitov et al., 2019). There, building on top of the framework presented here, we infer a mixed Gaussian phylogenetic model of brain- and body-mass evolution in a phylogeny of 630 mammal species. Statistical properties of the $\mathcal{G}_{LInv}$ models, such as identifiability, model uncertainty, type I and II errors, and invariance to rigid rotations of the trait data (Adams and Collyer, 2018) are also evaluated by Mitov et al. (2019).

The $\mathcal{G}_{LInv}$ family contains the models encountered in a number of contemporary frameworks. In particular all OU type models (with implementations in **ouch**, **slouch**, **mvSLOUCH**, **mvMORPH**) belong to $\mathcal{G}_{LInv}$. The **RPANDA** SDE framework (excluding interaction between lineages) is also covered as are current punctuated equilibrium models (OU along a branch with a normal jump, denoted JOU Bartoszek, 2014; Bokma, 2002). To the best of our knowledge, our implementation provides efficient likelihood calculation for the widest class of models, including but not limited to BM, OU, BM with trend, drift, early burst/Accelerating–decelerating (EB/ACDC) or white noise (Harmon et al., 2008), on general types of phylogenetic trees.

The rest of the paper is organized as follows. In Section 2, we define the $\mathcal{G}_{LInv}$ family and describe the analytical integration algorithm for fast likelihood calculation. In Section 3, we describe how one can handle issues such as shifts, missing values, measurement error, punctuated components, trees with polytomies, as well as sequentially sampled data (such as fossil data) leading to non-ultrametric trees. Next, in Section 4, we discuss the standard OU setup and describe examples of $\mathcal{G}_{LInv}$ model types that have already been implemented in our framework. Two widely used models – the multivariate BM and OU processes and a novel model – a multivariate OU model with jumps are provided. Further, in Section 5, we discuss our method in the light of current approaches. In Appendix A we describe our software implementation—the **PCMBase** R-package. In Appendix B, we use code examples and output from the **PCMBase** package to provide

[1] With respect to the chronological order of the works, here, we omit the R package **PCMFit** (Mitov et al., 2019)—despite its earlier publication, the inference method introduced in Mitov et al. (2019) is based on the here presented likelihood calculation algorithm and software tool.

Please cite this article as: V. Mitov, K. Bartoszek, G. Asimomitis et al., Fast likelihood calculation for multivariate Gaussian phylogenetic models with shifts. Theoretical Population Biology (2019), https://doi.org/10.1016/j.tpb.2019.11.005.

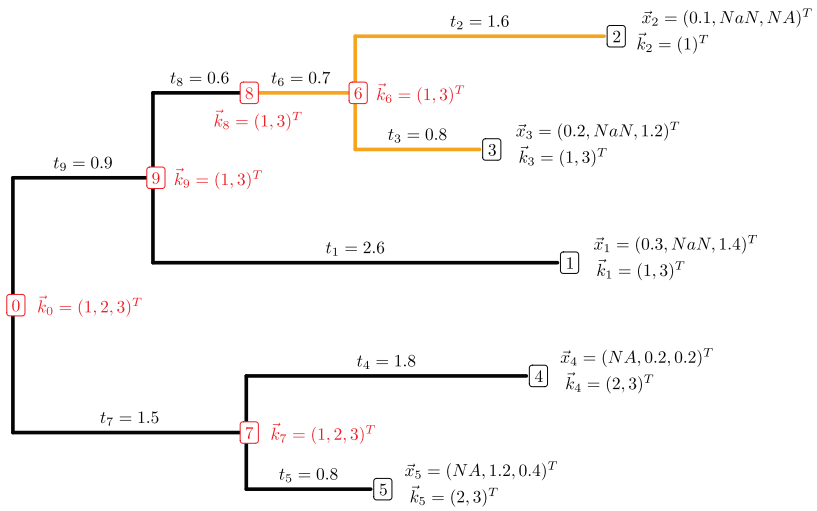

Fig. 1. A phylogenetic tree with observations at the tips. Numbered labels in black indicate the tips with observed trait vectors $\vec{x}_1, \ldots, \vec{x}_{N=5}$. Missing measurements are denoted as `NA` (Not Available), while non-existing traits are denoted as `NaN` (Not a Number). Numbered labels in red indicate the root, 0, and the internal nodes $6, \ldots, 9$, for which the trait vectors are unknown. The vectors, $\vec{k}_i$, denote the active coordinates for every node—for a tip-node these are all coordinates with a trait measurement (neither `NA` nor `NaN`); for an internal node, these are all the coordinates denoting traits that exist (are not `NaN`) for at least one of the tips descending from that node. The length of a branch leading to a tip or an internal node is known and denoted by t_i, $i = 1, \ldots, 9$. The change in branch colour from black to orange at the beginning of the branch leading to node 6 denotes the shift to a different evolutionary regime. This can be a change in the values of the model parameters, or a change to a new type of model within the $\mathcal{G}_{Llnv}$ family.

a step-by-step example of the log-likelihood calculation on a small tree. In Appendix C, we report a technical validation test of the implementation. In Appendix D, we report a performance benchmark evaluating the likelihood calculation time for different numbers of traits, tree sizes and $\mathcal{G}_{Llnv}$ model types.

2. Fast phylogenetic computational framework

2.1. Phylogenetic notation

We assume that we are given a rooted phylogenetic tree $\mathbb{T}$ representing the ancestral relationship between N species associated with the tips of the tree (Fig. 1). We denote the tips of the tree by the numbers $1, \ldots, N$, the internal nodes by the numbers $N+1, \ldots, M-1$ (where M is the total number of nodes in the tree) and the root-node by 0. For any internal node j, we denote by $Desc(j)$ the set of its direct descendants. We denote by t_j the known length of the branch in the tree leading to any tip or internal node j. By convention, we assume that time increases in the direction from the root to the tips of the tree, and t_j are positive scalars.

The object of all phylogenetic models discussed here will be a suite of k quantitative (real-valued) traits characterizing the N species. Associated with each tip, i, there is a real k-vector, $\vec{x}_i$, of measured values for the k traits. For some species, some trait measurements can be missing, reflecting two possible cases:

- the trait exists but was not measured for that species, denoted as `NA` (Not Available);
- the trait does not exist for that species denoted as `NaN` (Not a Number) (Fig. 1).

We introduce algebraic notation that will hold for the rest of the paper. Scalars are denoted by lower case letters, e.g. f, vectors are indicated by the arrow notation, e.g. $\vec{\theta}$, while matrices are denoted as upper case bold letters, e.g. $\mathbf{H}$. An exception to this is $\mathbf{X}_j$, meaning the set of measurements at the tips descending from an internal node j of the tree.

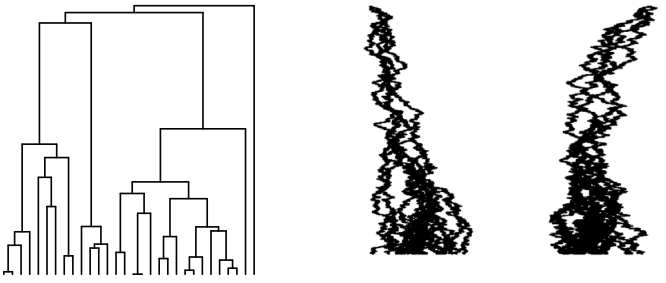

Fig. 2. Simulation of a bivariate OU process on top of a pure birth tree with 30 tips. The two traits are displayed on separate panels. The tree was simulated using the **TreeSim** package (Stadler, 2009, 2011), its height is 3.201. The bivariate OU process was simulated using **mvSLOUCH** (Bartoszek et al., 2012) with parameters (matrices are represented by their rows) $\mathbf{H} = \{\{1, 0.25\}, \{0, 2\}\}$, $\mathbf{\Sigma}_x = \{\{0.5, 0.25\}, \{0, 0.5\}\}$, $\vec{\theta} = (1, -1)^T$ and $\vec{x}_0 = (0, 0)^T$ (see Section 4.1 for the definition of these parameters).

2.2. Phylogenetic models of continuous trait evolution

We assume that the trait values measured at the tips of the tree are a realization of a continuous-time-continuous-state Markovian process evolving on top of the branching pattern in the tree. By this we mean that along any given branch we have a trajectory following the law of the process. Then, at speciation, the process "splits" into two processes. Both processes inherit the last value of their parent process. After the branching points, there is no interaction between the processes. This entails that all the dependencies between the values at the tips come from the time between the origin of the tree and the most recent common ancestor for each pair of species. Exactly how this shared time of evolution is translated into a dependency depends on the assumed process. A widely used example of such trait process is the Ornstein–Uhlenbeck process illustrated in Fig. 2.

Such stochastic processes are used as models of continuous trait evolution at the macro-evolutionary time scale, that is, when the time-units are in the order of hundreds to thousands of generations. Further in the text, we use the term "(trait evolutionary) model" to denote such kind of stochastic processes. We now turn

Please cite this article as: V. Mitov, K. Bartoszek, G. Asimomitis et al., Fast likelihood calculation for multivariate Gaussian phylogenetic models with shifts. Theoretical Population Biology (2019), https://doi.org/10.1016/j.tpb.2019.11.005.

to describing a family of models for which we will then provide an efficient way to calculate the likelihood of their parameters given the tree and the trait data observed at its tips.

2.3. The $\mathcal{G}_{LInv}$ family of models

The following definition specifies all requirements needed for a trait evolutionary model to be integrated within the fast computational framework:

Definition 1 (*The $\mathcal{G}_{LInv}$ Family*)**.** We say that a trait evolutionary model with parameters $\vec{\Theta}$ belongs to the $\mathcal{G}_{LInv}$ family if it satisfies the following

1. after branching, the traits evolve independently in all descending lineages;
2. the distribution of the trait vector at time t, $\vec{x}(t) \in \mathbb{R}^{k_t}$, conditional on a trait vector at time $s < t$, $\vec{x}(s) \in \mathbb{R}^{k_s}$ and, possibly, some other variables denoted by $\vec{\eta}$, is Gaussian with the mean and variance satisfying

   (2.a) $\mathrm{E}\left[\vec{x}(t)|\vec{x}(s)\right] = \vec{\omega}(s,t,\vec{\Theta},\vec{\eta}) + \mathbf{\Phi}(s,t,\vec{\Theta},\vec{\eta})\vec{x}(s)$ (the expectation is a linear function of the ancestral value),

   (2.b) $\mathrm{Var}\left[\vec{x}(t)|\vec{x}(s)\right] = \mathbf{V}(s,t,\vec{\Theta},\vec{\eta})$ (variance is invariant with respect to the ancestral value),

   for some positive but not necessarily equal dimensions k_t and k_s, a vector $\vec{\omega}(s,t,\vec{\Theta},\vec{\eta}) \in \mathbb{R}^{k_t}$, a matrix $\mathbf{\Phi}(s,t,\vec{\Theta},\vec{\eta}) \in \mathbb{R}^{k_t \times k_s}$, and a symmetric positive definite matrix $\mathbf{V}(s,t,\vec{\Theta},\vec{\eta}) \in \mathbb{R}^{k_t \times k_t}$, all of which may depend on s, t, $\vec{\Theta}$ and $\vec{\eta}$ but do not depend on the trait trajectory $\vec{x}(\cdot)$.

For simplicity, further in the text, we will write $\vec{\omega}$, $\mathbf{\Phi}$ and $\mathbf{V}$ remembering that they are functions of s, t, $\vec{\Theta}$, and, possibly, other variables $\vec{\eta}$ (e.g. a branch location in a phylogenetic tree).

Later, in Section 4, we show that the $\mathcal{G}_{LInv}$ family contains many well-known contemporary models such as BM, multivariate OU, BM or OU with normally distributed jumps. Now we derive an important property of the $\mathcal{G}_{LInv}$ family playing a key role for the fast likelihood calculation:

Theorem 1. *Let $\mathcal{M}$ be a trait model from the $\mathcal{G}_{LInv}$ family. Let i be a tip or internal node and j be its parent node in a tree $\mathbb{T}$, and let $\vec{x}_i \in \mathbb{R}^{k_i}$, $\vec{x}_j \in \mathbb{R}^{k_j}$ ($k_i, k_j \in \mathbb{Z}^+$) be the trait-vectors at the nodes i and j under a realization of $\mathcal{M}$ on $\mathbb{T}$. Let $\vec{\omega}_i$, $\mathbf{\Phi}_i$ and $\mathbf{V}_i$ denote the terms $\vec{\omega}$, $\mathbf{\Phi}$ and $\mathbf{V}$ from Definition 1 specific for node i. Then, the probability density function (pdf) of $\vec{x}_i$ conditioned on $\vec{x}_j$ can be expressed as the following exponential of a quadratic polynomial*

$$pdf(\vec{x}_i|\vec{x}_j) = \exp\left[\vec{x}_i^T\mathbf{A}_i\vec{x}_i + \vec{x}_i^T\vec{b}_i + \vec{x}_j^T\mathbf{C}_i\vec{x}_j + \vec{x}_j^T\vec{d}_i + \vec{x}_j^T\mathbf{E}_i\vec{x}_i + f_i\right], \quad (1)$$

where $\mathbf{A}_i$ is a symmetric negative-definite matrix, and all of the terms $\mathbf{A}_i$, $\vec{b}_i$, $\mathbf{C}_i$, $\vec{d}_i$, $\mathbf{E}_i$, f_i are constants with respect to $\vec{x}_j$, specified by the equations:

$$\begin{aligned}
\mathbf{A}_i &= -\tfrac{1}{2}\mathbf{V}_i^{-1} \in \mathbb{R}^{k_i\times k_i}\\
\vec{b}_i &= \mathbf{V}_i^{-1}\vec{\omega}_i \in \mathbb{R}^{k_i}\\
\mathbf{C}_i &= -\tfrac{1}{2}\mathbf{\Phi}_i^T\mathbf{V}_i^{-1}\mathbf{\Phi}_i \in \mathbb{R}^{k_j\times k_j}\\
\vec{d}_i &= -\mathbf{\Phi}_i^T\mathbf{V}_i^{-1}\vec{\omega}_i \in \mathbb{R}^{k_j}\\
\mathbf{E}_i &= \mathbf{\Phi}_i^T\mathbf{V}_i^{-1} \in \mathbb{R}^{k_j\times k_i}\\
f_i &= -\tfrac{1}{2}\vec{\omega}_i^T\mathbf{V}_i^{-1}\vec{\omega}_i - \tfrac{k_i}{2}\log(2\pi) - \tfrac{1}{2}\log|\mathbf{V}_i| \in \mathbb{R}.
\end{aligned} \quad (2)$$

Proof. Substituting $\vec{\omega}_i + \mathbf{\Phi}_i\vec{x}_j$ and $\mathbf{V}_i$ for the mean and variance in the formula for the pdf of a multivariate Gaussian distribution, we obtain:

$$pdf(\vec{x}_i|\vec{x}_j) = \exp\left[-\frac{1}{2}\left(\vec{x}_i - \left(\vec{\omega}_i + \mathbf{\Phi}_i\vec{x}_j\right)\right)^T\mathbf{V}_i^{-1}\left(\vec{x}_i - \left(\vec{\omega}_i + \mathbf{\Phi}_i\vec{x}_j\right)\right) - \frac{k_i}{2}\log(2\pi) - \frac{1}{2}\log|\mathbf{V}_i|\right] \quad (3)$$

By expanding and reordering the terms in parentheses, Eq. (3) can be rewritten as

$$\begin{aligned}
pdf(\vec{x}_i|\vec{x}_j) = \exp\Big[& \vec{x}_i^T\left(-\tfrac{1}{2}\mathbf{V}_i^{-1}\right)\vec{x}_i +\\
& \vec{x}_i^T\left(\mathbf{V}_i^{-1}\vec{\omega}_i\right) +\\
& \vec{x}_j^T\left(-\tfrac{1}{2}\mathbf{\Phi}_i^T\mathbf{V}_i^{-1}\mathbf{\Phi}_i\right)\vec{x}_j +\\
& \vec{x}_j^T\left(-\mathbf{\Phi}_i^T\mathbf{V}_i^{-1}\vec{\omega}_i\right) +\\
& \vec{x}_j^T\left(\mathbf{\Phi}_i^T\mathbf{V}_i^{-1}\right)\vec{x}_i +\\
& \left(-\tfrac{1}{2}\vec{\omega}_i^T\mathbf{V}_i^{-1}\vec{\omega}_i - \tfrac{k_i}{2}\log(2\pi)\right.\\
& \left.-\tfrac{1}{2}\log|\mathbf{V}_i|\right)\Big].
\end{aligned} \quad (4)$$

We can see the correspondence with the quadratic forms $\vec{x}_i^T(\ldots)\vec{x}_i$, $\vec{x}_j^T(\ldots)\vec{x}_j$ and the other terms in Eq. (1). Eq. (2) follows immediately. Furthermore, $\mathbf{A}_i$ is a symmetric negative-definite matrix, because $\mathbf{V}_i$ is a symmetric positive-definite matrix as it is a variance–covariance matrix. Finally, all of the terms $\mathbf{A}_i$, $\vec{b}_i$, $\mathbf{C}_i$, $\vec{d}_i$, $\mathbf{E}_i$, f_i are constant with respect to $\vec{x}_j$, because they are functions of $\vec{\omega}_i$, $\mathbf{\Phi}_i$ and $\mathbf{V}_i$ which are constants with respect to $\vec{x}_j$ by Definition 1. □

2.4. Calculating the likelihood of $\mathcal{G}_{LInv}$ models

Let $\mathcal{M}$ be a trait evolutionary model realized on a tree $\mathbb{T}$ and $\vec{\Theta}$ denotes the parameters of $\mathcal{M}$. The likelihood of $\mathcal{M}$ for given trait data $\mathbf{X}$ associated with the tips of $\mathbb{T}$ is defined as the function

$$\ell(\vec{\Theta}) = pdf(\mathbf{X}|\mathbb{T},\vec{\Theta}) = \int_{\Omega(\mathbf{Z})} pdf(\mathbf{Z},\mathbf{X}|\mathbb{T},\vec{\Theta})d\mathbf{Z}, \quad (5)$$

where $\mathbf{Z}$ denotes the unobserved trait values at the internal nodes of the tree (excluding the root[2]) and $\Omega(\mathbf{Z})$ denotes the space of possible values of $\mathbf{Z}$.

The representation of Eq. (1) allows for linear (in terms of the number of nodes, M) calculation of the likelihood of any trait model in the $\mathcal{G}_{LInv}$ family, given a phylogeny and measured data at its tips. This follows from the next theorem.

Theorem 2. *Let $\mathcal{M}$ be a trait evolutionary model from the $\mathcal{G}_{LInv}$ family and $\mathbb{T}$ be a phylogenetic tree. Let $\vec{\Theta}$ be the parameters of $\mathcal{M}$. For the root (0) or any internal node j in $\mathbb{T}$, there exists a $k_j \times k_j$ matrix $\mathbf{L}_j$, a k_j-vector $\vec{m}_j$ and a scalar r_j, such that the likelihood of $\mathcal{M}$ for the data $\mathbf{X}_j$, conditioned on $\vec{x}_j \in \mathbb{R}^{k_j}$ and $\mathbb{T}$ is expressed as:*

$$pdf(\mathbf{X}_j|\vec{x}_j,\mathbb{T},\vec{\Theta}) = \exp\left(\vec{x}_j^T\mathbf{L}_j\vec{x}_j + \vec{x}_j^T\vec{m}_j + r_j\right). \quad (6)$$

The parameters $\mathbf{L}_j$, $\vec{m}_j$, r_j are functions of $\vec{\Theta}$, the observed data $\mathbf{X}_j$, and the tree $\mathbb{T}$, namely, Eqs. (9), (10), and (11).

Proof. Following Condition 1 of Definition 1 we can factorize the conditional likelihood at any internal or root node j. Splitting $Desc(j)$, i.e. the set of nodes descending from node j, into tips and

[2] The treatment of the trait value at the root is discussed later in the text.

non-tips, denoted as $Desc(j) \cap \{1, \ldots, N\}$ and $Desc(j) \setminus \{1, \ldots, N\}$, we can write:

$$\begin{aligned} pdf(\mathbf{X}_j|\vec{x}_j, \mathbb{T}, \vec{\Theta}) &= \left(\prod_{i \in Desc(j) \cap \{1,\ldots,N\}} pdf(\vec{x}_i|\vec{x}_j, \mathbb{T}, \vec{\Theta}) \right) \times \\ &\left(\prod_{i \in Desc(j) \setminus \{1,\ldots,N\}} \int_{\mathbb{R}_i^k} pdf(\vec{x}_i|\vec{x}_j, \mathbb{T}, \vec{\Theta}) \right. \\ &\left. \times pdf(\mathbf{X}_i|\vec{x}_i, \mathbb{T}, \vec{\Theta}) \mathrm{d}\vec{x}_i \right). \end{aligned} \tag{7}$$

We first prove the theorem for nodes where all descendants are tips. If all descendants of j are tips (e.g. $j \in 6, 7$, Fig. 1), then, according to Eq. (1)

$$\begin{aligned} pdf(\mathbf{X}_j|\vec{x}_j, \mathbb{T}, \vec{\Theta}) &= \prod_{i \in Desc(j)} pdf(\vec{x}_i|\vec{x}_j, \mathbb{T}, \vec{\Theta}) \\ &= \exp\left(\sum_{i \in Desc(j)} \vec{x}_i^T \mathbf{A}_i \vec{x}_i + \vec{x}_i^T \vec{b}_i + \vec{x}_j^T \mathbf{C}_i \vec{x}_j \right. \\ &\left. + \vec{x}_j^T \vec{d}_i + \vec{x}_j^T \mathbf{E}_i \vec{x}_i + f_i \right), \end{aligned}$$

resulting in

$$\begin{aligned} pdf(\mathbf{X}_j|\vec{x}_j, \mathbb{T}, \vec{\Theta}) = \exp\left(\vec{x}_j^T \left(\sum_{i \in Desc(j)} \mathbf{C}_i \right) \vec{x}_j + \vec{x}_j^T \left(\sum_{i \in Desc(j)} \vec{d}_i + \mathbf{E}_i \vec{x}_i \right) \right. \\ \left. + \sum_{i \in Desc(j)} \vec{x}_i^T \mathbf{A}_i \vec{x}_i + \vec{x}_i^T \vec{b}_i + f_i \right) \end{aligned} \tag{8}$$

Then, to obtain the representation from Eq. (6), we denote:

$$\begin{aligned} \mathbf{L}_j &= \sum_{i \in Desc(j)} \mathbf{C}_i \\ \vec{m}_j &= \sum_{i \in Desc(j)} \vec{d}_i + \mathbf{E}_i \vec{x}_i \\ r_j &= \sum_{i \in Desc(j)} \vec{x}_i^T \mathbf{A}_i \vec{x}_i + \vec{x}_i^T \vec{b}_i + f_i \end{aligned} \tag{9}$$

If not all of $Desc(j)$ are tips, then, for the descendants which are tips, we define:

$$\begin{aligned} \mathbf{L}_j^{tips} &= \sum_{i \in Desc(j) \cap \{1,\ldots,N\}} \mathbf{C}_i \\ \vec{m}_j^{tips} &= \sum_{i \in Desc(j) \cap \{1,\ldots,N\}} \vec{d}_i + \mathbf{E}_i \vec{x}_i \\ r_j^{tips} &= \sum_{i \in Desc(j) \cap \{1,\ldots,N\}} \vec{x}_i^T \mathbf{A}_i \vec{x}_i + \vec{x}_i^T \vec{b}_i + f_i \end{aligned} \tag{10}$$

We perform mathematical induction to prove the theorem for all nodes. We need to show that Eq. (6) holds for each non-tip descendant of j, that is, for each $i \in Desc(j) \setminus \{1, \ldots, N\}$ there exist a $k_i \times k_i$ matrix $\mathbf{L}_i$, a k_i-vector $\vec{m}_i$ and a scalar r_i such that $pdf(\mathbf{X}_i|\vec{x}_i, \mathbb{T}, \vec{\Theta}) = \exp(\vec{x}_i^T \mathbf{L}_i \vec{x}_i + \vec{x}_i^T \vec{m}_i + r_i)$. We proved the induction base case, namely, we proved above that Eq. (6) holds for all nodes which have only tip-descendants. Then, the induction hypothesis is that for an internal node j, the statement of the theorem has been proven for all $i \in Desc(j)$. Now in the inductive step using Eq. (1) and the induction hypothesis, we can write the integral in Eq. (7) as

$$\begin{aligned} &\int_{\mathbb{R}^{k_i}} pdf(\vec{x}_i|\vec{x}_j, \mathbb{T}, \vec{\Theta}) \times pdf(\mathbf{X}_i|\vec{x}_i, \mathbb{T}, \vec{\Theta}) \mathrm{d}\vec{x}_i \\ &= \int_{\mathbb{R}^{k_i}} \exp\left(\vec{x}_i^T \mathbf{A}_i \vec{x}_i + \vec{x}_i^T \vec{b}_i + \vec{x}_j^T \mathbf{C}_i \vec{x}_j + \vec{x}_j^T \vec{d}_i + \vec{x}_j^T \mathbf{E}_i \vec{x}_i + f_i \right. \\ &\quad \left. + \vec{x}_i^T \mathbf{L}_i \vec{x}_i + \vec{x}_i^T \vec{m}_i + r_i \right) \mathrm{d}\vec{x}_i \\ &= \exp\left(\vec{x}_j^T \mathbf{C}_i \vec{x}_j + \vec{x}_j^T \vec{d}_i + f_i + r_i \right) \times \int_{\mathbb{R}^{k_i}} \exp\left(\vec{x}_i^T (\mathbf{A}_i + \mathbf{L}_i) \vec{x}_i \right. \\ &\quad \left. + \vec{x}_i^T (\vec{b}_i + \vec{m}_i + \mathbf{E}_i^T \vec{x}_j) \right) \mathrm{d}\vec{x}_i \\ &= \exp\left(\vec{x}_j^T \mathbf{C}_i \vec{x}_j + \vec{x}_j^T \vec{d}_i + f_i + r_i \right) \left(\sqrt{2\pi} \right)^{k_i} \\ &\quad \times \left(\sqrt{|(-2)(\mathbf{A}_i + \mathbf{L}_i)|} \right)^{-1} \\ &\quad \times \exp\left(-(1/4) \left(\vec{b}_i + \vec{m}_i + \mathbf{E}_i^T \vec{x}_j \right)^T (\mathbf{A}_i + \mathbf{L}_i)^{-1} \right. \\ &\quad \left. \times \left(\vec{b}_i + \vec{m}_i + \mathbf{E}_i^T \vec{x}_j \right) \right) \\ &= \exp\left(\vec{x}_j^T \mathbf{C}_i \vec{x}_j + \vec{x}_j^T \vec{d}_i + f_i + r_i \right) \left(\sqrt{2\pi} \right)^{k_i} \\ &\quad \times \left(\sqrt{|(-2)(\mathbf{A}_i + \mathbf{L}_i)|} \right)^{-1} \\ &\quad \times \exp\left(-(1/4) \left(\vec{b}_i + \vec{m}_i \right)^T (\mathbf{A}_i + \mathbf{L}_i)^{-1} \left(\vec{b}_i + \vec{m}_i \right) \right. \\ &\quad -(1/2) \vec{x}_j^T \mathbf{E}_i (\mathbf{A}_i + \mathbf{L}_i)^{-1} \left(\vec{b}_i + \vec{m}_i \right) \\ &\quad \left. -(1/4) \vec{x}_j^T \mathbf{E}_i (\mathbf{A}_i + \mathbf{L}_i)^{-1} \mathbf{E}_i^T \vec{x}_j \right) \\ &= \exp\left(\vec{x}_j^T \left(\mathbf{C}_i - (1/4) \mathbf{E}_i (\mathbf{A}_i + \mathbf{L}_i)^{-1} \mathbf{E}_i^T \right) \vec{x}_j \right. \\ &\quad + \vec{x}_j^T \left(\vec{d}_i - (1/2) \mathbf{E}_i (\mathbf{A}_i + \mathbf{L}_i)^{-1} \left(\vec{b}_i + \vec{m}_i \right) \right) \\ &\quad + f_i + r_i + (k_i/2) \log(2\pi) - (1/2) \log(|(-2)(\mathbf{A}_i + \mathbf{L}_i)|) \\ &\quad \left. -(1/4) \left(\vec{b}_i + \vec{m}_i \right)^T (\mathbf{A}_i + \mathbf{L}_i)^{-1} \left(\vec{b}_i + \vec{m}_i \right) \right). \end{aligned}$$

We can then see that for a non-tip node we can define

$$\begin{aligned} \mathbf{L}_j^{non-tips} &= \sum_{i \in Desc(j) \setminus \{1,\ldots,N\}} \left(\mathbf{C}_i - (1/4) \mathbf{E}_i (\mathbf{A}_i + \mathbf{L}_i)^{-1} \mathbf{E}_i^T \right) \\ \vec{m}_j^{non-tips} &= \sum_{i \in Desc(j) \setminus \{1,\ldots,N\}} \left(\vec{d}_i - (1/2) \mathbf{E}_i (\mathbf{A}_i + \mathbf{L}_i)^{-1} \left(\vec{b}_i + \vec{m}_i \right) \right) \\ r_j^{non-tips} &= \sum_{i \in Desc(j) \setminus \{1,\ldots,N\}} \left(f_i + r_i + (k_i/2) \log(2\pi) \right. \\ &\quad -(1/2) \log(|(-2)(\mathbf{A}_i + \mathbf{L}_i)|) \\ &\quad \left. -(1/4) \left(\vec{b}_i + \vec{m}_i \right)^T (\mathbf{A}_i + \mathbf{L}_i)^{-1} \right. \\ &\quad \left. \times \left(\vec{b}_i + \vec{m}_i \right) \right). \end{aligned} \tag{11}$$

The representation of $\mathbf{L}_j^{non-tips}$, $\vec{m}_j^{non-tips}$ and $r_j^{non-tips}$ in Eq. (11) immediately implies the existence of the $\mathbf{L}_j$, $\vec{m}_j$ and r_j elements in Eq. (6) for internal or root nodes j, hence we obtain the claimed polynomial form in the inductive step and in consequence the theorem. □

The inductive proof of Theorem 2 defines a pruning-wise procedure for calculating $\mathbf{L}_0$, $\vec{m}_0$ and r_0 (we remind that 0 stands for the root of the tree). In order to calculate the likelihood of the tree conditioned on $\vec{x}_0$, we use Theorem 2 with j being the root node. In order to be able to calculate the full likelihood, it now only remains to specify how to deal with the unknown trait value at the root of the tree, $\vec{x}_0$, i.e. the ancestral state. This is

an implementation detail up to the user. Similar to Ho and Ané (2014b), our implementation provided in the **PCMbase** package (Appendix A) allows for maximizing the polynomial with respect to $\vec{x}_0$ or for treating it as a free parameter (like the elements of the parameter set $\vec{\Theta}$) that the user provides. A detailed example of the pruning likelihood calculation is provided in Appendix B.

2.5. Scope of the framework

We now investigate if there are other trait evolutionary models, beyond the $\mathcal{G}_{LInv}$ family, for which the likelihood can be calculated using the same recursive formulae, Eqs. (9), (10), and (11). First, since we calculate the likelihood in a recursive pruning fashion, we assume that evolution is independent across branches, meaning Condition 1 is a necessary condition. In Theorem 3, we prove that Condition 2 in Definition 1 is also a necessary condition. In other words, we show that if the likelihood can be calculated via recursion based on Eqs. (2), (8), then the model is in the $\mathcal{G}_{LInv}$ family.

Theorem 3. *Let $\mathcal{M}$ be a trait model satisfying Condition 1 of Definition 1 and realized on a tree $\mathbb{T}$. If for every parent–child pair of nodes $\langle j, i\rangle$ in $\mathbb{T}$, the trait-vector $\vec{x}_i \in \mathbb{R}^{k_i}$ ($k_i \in \mathbb{Z}^+$) has non-zero support on the whole of $\mathbb{R}^{k_i}$ and there exist a symmetric negative-definite matrix $\mathbf{A}_i \in \mathbb{R}^{k_i \times k_i}$ and components $\vec{b}_i \in \mathbb{R}^{k_i}$, $\mathbf{C}_i \in \mathbb{R}^{k_j \times k_j}$, $\vec{d}_i \in \mathbb{R}^{k_j}$, $\mathbf{E}_i \in \mathbb{R}^{k_j \times k_i}$, $f_i \in \mathbb{R}$, such that, for any vector of values at the parent node, $\vec{x}_j \in \mathbb{R}^{k_j}$ ($k_j \in \mathbb{Z}^+$), the pdf of $\vec{x}_i$ conditional on $\vec{x}_j$ can be expressed by Eq. (1), then $\mathcal{M}$ belongs to the G_{LInv} family and the terms $\vec{\omega}_i$, $\mathbf{\Phi}_i$ and $\mathbf{V}_i$ denoting the terms $\vec{\omega}$, $\mathbf{\Phi}$ and $\mathbf{V}$ from Definition 1 specific for node i satisfy Eq. (2).*

Proof. We rearrange the terms on the right-hand side of Eq. (1) as follows

$$\begin{aligned}
pdf(\vec{x}_i|\vec{x}_j) &= \exp\left[\vec{x}_i^T\mathbf{A}_i\vec{x}_i - 2\vec{x}_i^T\mathbf{A}_i\left((-\tfrac{1}{2}\mathbf{A}_i^{-1})\left(\vec{b}_i + \mathbf{E}_i^T\vec{x}_j\right)\right)\right.\\
&\left.+\left(\vec{x}_j^T\mathbf{C}_i\vec{x}_j + \vec{x}_j^T\vec{d}_i + f_i\right)\right]\\
&= \exp\left[\left(\vec{x}_i + \tfrac{1}{2}\mathbf{A}_i^{-1}\left(\vec{b}_i + \mathbf{E}_i^T\vec{x}_j\right)\right)^T\mathbf{A}_i\left(\vec{x}_i + \tfrac{1}{2}\mathbf{A}_i^{-1}\left(\vec{b}_i + \mathbf{E}_i^T\vec{x}_j\right)\right)\right.\\
&-\tfrac{1}{4}\left(\vec{b}_i + \mathbf{E}_i^T\vec{x}_j\right)^T\mathbf{A}_i^{-1}\left(\vec{b}_i + \mathbf{E}_i^T\vec{x}_j\right)\\
&\left.+\left(\vec{x}_j^T\mathbf{C}_i\vec{x}_j + \vec{x}_j^T\vec{d}_i + f_i\right)\right].
\end{aligned} \tag{12}$$

As the above is by definition a density on $\mathbb{R}^{k_i}$, integrating over $\vec{x}_i$ equals 1. Hence, after taking all constants with respect to $\vec{x}_i$ out of the integral and multiplying/dividing the integral by the constant $(\sqrt{|2\pi(-2)\mathbf{A}_i|})^{-1}$, we obtain Eq. (13) given in Box I:

When calculating the integral in Eq. (13), we have used the fact that the matrix $(-2)\mathbf{A}_i$ is a symmetric positive-definite matrix as it is the negative of the symmetric negative-definite matrix $2\mathbf{A}_i$. Hence, the so constructed function below the integral in Eq. (13) is a k_i-variate Gaussian pdf with mean vector $-\frac{1}{2}\mathbf{A}_i^{-1}\left(\vec{b}_i + \mathbf{E}_i^T\vec{x}_j\right)$ and variance–covariance matrix $(-2)\mathbf{A}_i$.

By definition, $\mathbf{A}_i$, $\vec{b}_i$, $\mathbf{C}_i$, $\vec{d}_i$, $\mathbf{E}_i$, f_i are constant with respect to $\vec{x}_j$. Therefore, Eq. (13) has to hold for all $\vec{x}_j$. This implies the relationships:

$$\begin{aligned}
\mathbf{C}_i &= \mathbf{E}_i\mathbf{A}_i^{-T}\mathbf{E}_i^T,\\
\vec{d}_i &= 2\mathbf{E}_i\mathbf{A}_i^{-1}\vec{b}_i,\\
f_i &= \tfrac{1}{4}b_i^T\mathbf{A}_i^{-1}b_i - \tfrac{k_i}{2}\log(2\pi) - \tfrac{1}{2}\log\left(|(-2)\mathbf{A}_i|\right).
\end{aligned} \tag{14}$$

Next, we define $\mathbf{V}_i := (-\frac{1}{2})\mathbf{A}_i^{-1}$, $\vec{\omega}_i := (-\frac{1}{2})\mathbf{A}_i^{-1}\vec{b}_i$ and $\mathbf{\Phi}_i := (-\frac{1}{2})\mathbf{A}_i^{-1}\mathbf{E}_i^T$. Since $\mathbf{A}_i$ is symmetric negative-definite, $\mathbf{V}_i$ is symmetric positive-definite. Combining the above three definitions with Eq. (14) and expressing $\mathbf{A}_i$, $\vec{b}_i$, $\mathbf{C}_i$, $\vec{d}_i$, $\mathbf{E}_i$, f_i in terms of $\vec{\omega}_i$, $\mathbf{\Phi}_i$ and $\mathbf{V}_i$, we obtain again Eq. (2). Then, we can follow the equivalences in backward direction (Eqs. (2)→(4)→(3)) to prove that the pdf defined in Eq. (1) is equivalent to the Gaussian pdf defined in terms of $\vec{\omega}_i$, $\mathbf{\Phi}_i$ and $\mathbf{V}_i$, Eq. (3). We note also that $\vec{\omega}_i$, $\mathbf{\Phi}_i$ and $\mathbf{V}_i$ defined above are constant with respect to $\vec{x}_j$, because they are defined in terms of $\mathbf{A}_i$, $\vec{b}_i$ and $\mathbf{E}_i$, which are constant with respect to $\vec{x}_j$ by definition. With that we proved Condition 2 of Definition 1. Since $\mathcal{M}$ satisfies Condition 1 of Definition 1 by the first sentence in the theorem, it follows that $\mathcal{M}$ belongs to the $\mathcal{G}_{LInv}$ family. □

Remark 1. In Eq. (1), it suffices to consider symmetric negative-definite matrices $\mathbf{A}$ only. We remind that, by definition, a matrix $\mathbf{A}$ is negative-definite iff $\vec{x}^T\mathbf{A}\vec{x} < 0$ for every $\vec{x} \neq \vec{0}$. Considering non-symmetric negative-definite matrices $\mathbf{A}$ does not extend the family of pdfs represented by Eq. (1). In particular, for any square negative-definite matrix $\mathbf{Q}$, and (of appropriate size) vector $\vec{u}$, it holds that $\vec{u}^T\mathbf{Q}\vec{u} = \vec{u}^T\left[\frac{1}{2}(\mathbf{Q} + \mathbf{Q}^T)\right]\vec{u}$ and the matrix $\left[\frac{1}{2}(\mathbf{Q} + \mathbf{Q}^T)\right]$ is symmetric negative-definite. Hence if one took in Eq. (1) a non-symmetric $\mathbf{A}$, then the value of the pdf would be the same as if one had taken the symmetric negative-definite matrix $\left[\frac{1}{2}(\mathbf{A} + \mathbf{A}^T)\right]$.

Based on the above theorem and remark, we conclude that the $\mathcal{G}_{LInv}$ family is identical with the scope of the fast likelihood computation framework. This implies that to define any new model within the framework, it is sufficient to define the functions $\vec{\omega}$, $\mathbf{\Phi}$ and $\mathbf{V}$ for each branch in the tree.

3. Special features of the framework

3.1. Shifts in the values of model parameters and the type of model

Given that the framework is abstract with respect to the functions $\vec{\omega}$, $\mathbf{\Phi}$ and $\mathbf{V}$, as long as the conditions of Definition 1 are met, it is possible to change the rule for calculating these functions at any internal node in the tree. This allows a change of all model parameters involved in the calculation of $\vec{\omega}$, $\mathbf{\Phi}$ and $\mathbf{V}$, as well as a switch to a different type of model, without compromising the computational efficiency. Such level of flexibility has proven necessary in previous works, e.g. Slater (2013) and Clavel et al. (2015). Slater (2013) used the **geiger** R-package to measure the statistical support for a shift from an OU to a BM process in the evolution of mammal body size occurring at the end of the Mesozoic. Later, though, in Slater (2014), he realized that the results of this study were compromised by an erroneous transformation of the branch lengths in the tree that causes inaccurate likelihood values for non-ultrametric trees. Clavel et al. (2015) implemented a non-pruning algorithm for multivariate likelihood calculation for shifts between BM, OU and the early burst (EB) model of adaptive radiation in their R-package **mvMORPH**. Due to a slow (quadratic in the number of tips) likelihood calculation algorithm, though, this implementation does not scale with tree size, while only big trees have the amount of data necessary to identify such shifts in the mode of evolution. Our approach should allow inferring such models with shifts on big trees exceeding several hundred to thousand species, see e.g. Mitov et al. (2019).

Please cite this article as: V. Mitov, K. Bartoszek, G. Asimomitis et al., Fast likelihood calculation for multivariate Gaussian phylogenetic models with shifts. Theoretical Population Biology (2019), https://doi.org/10.1016/j.tpb.2019.11.005.

$$
\begin{aligned}
1 &= \underbrace{\int_{\mathbb{R}^{k_i}} \frac{1}{\sqrt{|2\pi(-2)\mathbf{A}_i|}} \exp\left[-\frac{1}{2}\left(\vec{x}_i + \frac{1}{2}\mathbf{A}_i^{-1}\left(\vec{b}_i + \mathbf{E}_i^T\vec{x}_j\right)\right)^T (-2\mathbf{A}_i)\left(\vec{x}_i + \frac{1}{2}\mathbf{A}_i^{-1}\left(\vec{b}_i + \mathbf{E}_i^T\vec{x}_j\right)\right)\right] \mathrm{d}\vec{x}_i}_{=1} \\
&\quad \times\sqrt{|2\pi(-2)\mathbf{A}_i|} \times \exp\left[-\tfrac{1}{4}\left(\vec{b}_i + \mathbf{E}_i^T\vec{x}_j\right)^T \mathbf{A}_i^{-1}\left(\vec{b}_i + \mathbf{E}_i^T\vec{x}_j\right) + \left(\vec{x}_j^T\mathbf{C}_i\vec{x}_j + \vec{x}_j^T\vec{d}_i + f_i\right)\right] \\
&= \exp\left[\tfrac{k_i}{2}\log(2\pi) + \tfrac{1}{2}\log|(-2)\mathbf{A}_i|\right] \times \exp\left[-\tfrac{1}{4}\left(\vec{b}_i + \mathbf{E}_i^T\vec{x}_j\right)^T \mathbf{A}_i^{-1}\left(\vec{b}_i + \mathbf{E}_i^T\vec{x}_j\right) + \left(\vec{x}_j^T\mathbf{C}_i\vec{x}_j + \vec{x}_j^T\vec{d}_i + f_i\right)\right] \\
&= \exp\left[\vec{x}_j^T\left(\mathbf{C}_i - \tfrac{1}{4}\mathbf{E}_i\mathbf{A}_i^{-1}\mathbf{E}_i^T\right)\vec{x}_j + \vec{x}_j^T\left(\vec{d}_i - \tfrac{1}{2}\mathbf{E}_i\mathbf{A}_i^{-1}\vec{b}_i\right) + f_i + \tfrac{k_i}{2}\log(2\pi) + \tfrac{1}{2}\log\left(|(-2)\mathbf{A}_i|\right) - \tfrac{1}{4}\vec{b}_i^T\mathbf{A}_i^{-1}\vec{b}_i\right].
\end{aligned}
\tag{13}
$$

Box I.

3.2. Missing measurements and non-existing traits

The trait measurement data are the observations at the tips. If a tip is described by a suite of traits it can easily happen that some of them are missing, either due to missing measurement or because the corresponding trait does not exist for the species. Removing such a tip from any further analysis would be wasting data, i.e. the observed data for the tip. When a trait is known to exist for all nodes but has not been measured for some tip, it is practical to assume that this is a random value with a distribution defined by the general rule of the model for this tip. In this case, the likelihood function is the marginal probability density function of the observed trait measurements. A different situation arises when a trait is known to have not existed for some of the ancestral nodes in the tree—such absence reduces the dimensionality of the space of unobserved trait values at the internal nodes over which the integral in the definition of the likelihood function is calculated, see Eq. (5). In the example likelihood calculation in Appendix B, we show that the assumption of trait existence or non-existence at the internal nodes leads to difference in the likelihood values. Unlike other implementations where existence is being assumed for all traits at all ancestral nodes (see e.g. Bastide et al. (2018)), our computational framework keeps track of the case of non-existing traits by carefully accounting for the dimensionality of the trait vectors at the tips and the internal nodes (see Fig. 1 and Appendix B for examples).

We now turn to describing the technicalities of the mechanism taking care of the missing measurements and/or non-existing traits. We use a vector of positive integers, $\vec{k}_j$, to denote the ordered set of active coordinates for a node j. If j is a tip, then $\vec{k}_j$ gives the indices of all non-missing entries in the trait vector for j; for an internal (unmeasured) node this gives the possibility to make some trait inactive. The cardinality of a vector is denoted with $|\vec{k}|$. For a vector, the notation $\vec{\theta}[\vec{k}]$ means the vector of elements of $\vec{\theta}$ on the coordinates contained in $\vec{k}$, while for a matrix $\mathbf{H}[\vec{k}_1, \vec{k}_2]$ means the matrix $\mathbf{H}$ with only the rows on the coordinates contained in $\vec{k}_1$ and columns contained in $\vec{k}_2$. For example take $\vec{\theta} = (10, 11, 12, 13)$ and $\vec{k} = (1, 3)$, then $\vec{\theta}[\vec{k}] = (10, 12)$, while if $\vec{k}_1 = (1, 3)$, $\vec{k}_2 = (2, 4)$ and

$$
\mathbf{H} = \begin{bmatrix} 10 & 11 & 12 & 13 \\ 14 & 15 & 16 & 17 \\ 18 & 19 & 20 & 21 \\ 22 & 23 & 24 & 25 \end{bmatrix},
$$

then

$$
\mathbf{H}[\vec{k}_1, \vec{k}_2] = \begin{bmatrix} 11 & 13 \\ 19 & 21 \end{bmatrix}.
$$

If a vector or matrix does not have any indication on which entries it is retained, then it means that we use the whole vector or matrix. All of the above notation is graphically represented in Fig. 1.

In our framework, we have the representation that $\vec{x}_i \in \mathbb{R}^{k_i}$ conditional on $\vec{x}_j \in \mathbb{R}^{k_j}$ is $\mathcal{N}(\vec{\omega}_i + \mathbf{\Phi}_i\vec{x}_j, \mathbf{V}_i)$ distributed. The input data is passed as a matrix (rows—trait measurements, columns—different species) the missing measurements have to be indicated as `NA`s, whereas the non-existing traits have to be indicated as `NaN`s (Fig. 1). By default, **PCMBase** constructs the coordinate vectors $\vec{k}_i$ and $\vec{k}_j$ in the following way: for a tip-node, i, $\vec{k}_i$ contains all observed (neither `NA` nor `NaN`) coordinates; for an internal node, i or j, the corresponding coordinate vector ($\vec{k}_i$ or $\vec{k}_j$) contains the coordinates denoting traits that exist (are not `NaN`) for at least one of the tips descending from that node (Fig. 1). Biologically, this treatment reflects a scenario where all of the traits with at least one non-`NaN` entry for at least one species (i.e. tip) in the tree must have existed for the root but some of the traits have subsequently disappeared on some lineages of the tree. In particular, if a trait exists for a given tip in the tree, it is assumed that it has existed for all of its ancestors up to the root of the tree. Conversely, if the trait does not exist for a tip, then it has not existed for any of its ancestors up to the first ancestor shared with a tip for which the trait does exist. Different biological scenarios are possible, e.g. assuming that some of the traits did not exist at the root-node but have appeared later for some on the lineages. These can be implemented by accordingly specifying the coordinate vectors (see Appendix B for an example).

During likelihood calculation for given trait data, a tree and a trait evolutionary model, the elements $\vec{\omega}_i$, $\mathbf{\Phi}_i$ and $\mathbf{V}_i$ of appropriate dimension are calculated for each non-root node i in the tree. This is done in two steps:

1. The general rule of the model is used to calculate the elements $\tilde{\vec{\omega}}_i$, $\tilde{\mathbf{\Phi}}_i$, $\tilde{\mathbf{V}}_i$ of full dimensionality (k), i.e. assuming that all traits exist;
2. Denoting by j the parent node of i, the elements $\vec{\omega}_i$, $\mathbf{\Phi}_i$ and $\mathbf{V}_i$ specific for the data in question are obtained as:

$$
\begin{aligned}
\vec{\omega}_i &= \tilde{\vec{\omega}}_i[\vec{k}_i], \\
\mathbf{\Phi}_i &= \tilde{\mathbf{\Phi}}_i[\vec{k}_i, \vec{k}_j], \\
\mathbf{V}_i &= \tilde{\mathbf{V}}_i[\vec{k}_i, \vec{k}_i],
\end{aligned}
\tag{15}
$$

where $\vec{k}_i$ and $\vec{k}_j$ denote the corresponding coordinate vectors at nodes i and j.

3.3. Measurement error

Commonly in PCMs the observed values at the tips are averages from a number of individuals of each species. Using just these average values does not take into account the intra-species variability. Ignoring this can have profound effects on any further estimation (see Hansen and Bartoszek, 2012). Following the PCM tradition, we call this intra-species variability a measurement error, but one should remember that it can be due to true biological variability. Technically, it is straightforward to take into account this variability in the calculation of the likelihood function. One recognizes, which component of the quadratic polynomial representation corresponds to the variance of the tip and augments it by the measurement error variance matrix, see the formulae in Section 4. In practice, though, dealing with the measurement error can be tricky, because it is often partially or completely unknown. **PCMBase** provides a flexible interface to accommodate such scenarios:

- **Known measurement error.** If for each tip i, $\vec{x}_i$ has been estimated as the mean k-vector of a known sample of independent and identically distributed (i.i.d.) k-variate measurements, the $k\times k$ variance covariance matrix of this mean vector is estimated by the formula

$$\mathbf{V}_{err,i} = \mathbf{V}_{emp,i}/n_i, \tag{16}$$

  where $\mathbf{V}_{emp,i}$ denotes the $k\times k$ empirical (sample) variance–covariance matrix from the sample, and n_i denotes the number of individuals in the sample (but see also Hansen and Bartoszek (2012), Garamszegi (2014) and Cooper et al. (2015), and Appendix H in Mitov et al. (2019) for longer discussions on measurement error in PCMs). The functions `PCMLik()` and `PCMSim()` accept a $k\times k\times N$ array parameter `SE`, where the block `SE[,,i]` is an upper triangular matrix factor of $\mathbf{V}_{err,i}$ (following by convention Eq. S5, Section Appendix A.3.1). Often, in particular, when the mean estimates for the individual traits originate from different samples within the same species, calculating $\mathbf{V}_{err,i}$ is not possible. In such cases, it is practical to assume that the measurement errors for each trait are uncorrelated, that is $\mathbf{V}_{err,i}$ is diagonal, and to calculate $diag(\mathbf{V}_{err,i}) = diag(\mathbf{V}_{emp,i})/\vec{n}_i$, where $\vec{n}_i$ is the k-vector of the sample sizes for each trait for the given species. To facilitate this use case, **PCMBase** allows the argument `SE` to be specified as a $k\times N$ matrix, where each column `SE[,i]` corresponds to $\sqrt{diag(\mathbf{V}_{err,i})}$.
- **Unknown measurement error.** If there is no way to estimate the measurement error empirically, e.g. when the individual measurements and sample sizes have been lost, **PCMBase** allows to include the measurement error as a free parameter called $\mathbf{\Sigma}_e$. $\mathbf{\Sigma}_e$ can be specified as a global parameter shared by all regimes in the model, or as a local (regime specific) parameter. At the level of a **PCMBase** model object, this parameter is stored in the form of a global or regime-specific upper triangular matrix parameter `Sigmae_x` denoting $\mathbf{\Sigma}_e$'s upper triangular factor (following by convention Eq. S4, Section Appendix A.3.1).

3.4. Non-ultrametric trees and multifurcations

If one has only measurements from contemporary species, then the phylogeny with branch lengths representing time is assumed to be an ultrametric one. However, very often the phylogeny is not ultrametric, e.g. when branch lengths represent genetic distances, or when extinct species are included. Given the fact that the quadratic polynomial framework treats each branch separately, it does not matter whether the tree is or is not ultrametric. Therefore, there is no need to search for an appropriate branch-length transformation as in other pruning implementations, such as the 3-point structure algorithm (Ho and Ané, 2014a) (see Section 5). This we believe should make the framework very straightforward to use. Furthermore, from the proof of Theorem 2 it is obvious that the tree does not need to be binary. Therefore, multifurcations are naturally supported by the framework.

3.5. Punctuated equilibrium

It is an ongoing debate in evolutionary biology at what time does evolutionary change take place. Change may take place either at times of speciation (punctuated equilibrium Eldredge and Gould, 1972; Gould and Eldredge, 1993) or gradually accumulate (phyletic gradualism, see references in Eldredge and Gould, 1972). There seems to be evidence for both types of evolution. For example, Bokma (2002) discusses that punctuated equilibrium is supported by fossil records (see Eldredge and Gould, 1972) but on the other hand Stebbins and Ayala (1981) also indicate experiments supporting phyletic gradualism.

At an internal node in the tree something happens that drives species apart and then "The further removed in time a species from the original speciation event that originated it, the more its genotype will have become stabilized and the more it is likely to resist change." (Mayr, 1982). Between branching events (and jumps) we can have stasis—"fluctuations of little or no accumulated consequence" taking place (Gould and Eldredge, 1993). Therefore, one would want processes that incorporate both types of evolution and allow for testing if either of them dominates.

Any $\mathcal{G}_{LInv}$ model of evolution can be extended to have a punctuated component by including jumps. Jump mechanisms, like jumps at the start of specific lineages or common jumps for daughter lineages, have to be developed on a per model basis, see Section 4.2 for an example. Within the **PCMBase** package, this feature is implemented through a binary vector `edge.jump` attached to the tree object (an augmented `phylo` object from the **ape** R-package). The length of this vector equals the number of edges in the tree. A 0 entry in this vector indicates that no jump took place at beginning of the corresponding branch, while a 1 entry that it did. While this reduces the locations of possible jumps to the branch-starts in the tree, **PCMBase** provides functions for inserting singleton nodes at user specified branches (function `PCMTreeInsertSingletons()`) or at all branches intersecting with a user specified time (function `PCMTreeInsertSingletonsAtEpoch()`).

4. Common $\mathcal{G}_{LInv}$ models

4.1. The multivariate Ornstein–Uhlenbeck process

The Ornstein–Uhlenbeck process (hereby abbreviated OU) is the workhorse in most contemporary PCMs. Since its introduction by Hansen (1997) it has been considered in detail with multiple software implementations (e.g. Butler and King, 2004; Hansen et al., 2008; FitzJohn, 2010; Bartoszek et al., 2012; Beaulieu et al., 2012; Ho and Ané, 2014a; Clavel et al., 2015; Goolsby et al., 2016; Mitov and Stadler, 2019).

In the most general form, the multivariate OU process describes the evolution of a k-dimensional suite of traits $\vec{x}\in\mathbb{R}^k$ over a period of time by the following stochastic differential equation

$$d\vec{x}(t) = -\mathbf{H}\left(\vec{x}(t) - \vec{\theta}(t)\right)dt + \mathbf{\Sigma}_x d\vec{W}(t), \tag{17}$$

$\mathbf{H}\in\mathbb{R}^{k\times k}$, $\vec{\theta}(t)\in\mathbb{R}^k$ and $\mathbf{\Sigma}_x\in\mathbb{R}^{k\times k}$. Notice that when $\mathbf{H}=\mathbf{0}$, we obtain a Brownian motion model.

Please cite this article as: V. Mitov, K. Bartoszek, G. Asimomitis et al., Fast likelihood calculation for multivariate Gaussian phylogenetic models with shifts. Theoretical Population Biology (2019), https://doi.org/10.1016/j.tpb.2019.11.005.

There is no current software package, in the case of phylogenetic OU models, that allows for an arbitrary form of the matrix **H**. Except for the Brownian motion case, nearly all assume that **H** has to be symmetric-positive-definite (note that this encompasses the single trait case). **mvMORPH** (Clavel et al., 2015), **SLOUCH** (Hansen et al., 2008) and **mvSLOUCH** (Bartoszek et al., 2012) seem to be the only exceptions. **mvMORPH** and **mvSLOUCH** allow for a general eigendecomposable **H** (with options to restrict it to diagonal, triangular, symmetric positive-definite, positive eigenvalues, real eigenvalues or general invertible). Furthermore, **mvSLOUCH** allows for a special singular structure of **H**. The matrix has to have in the upper-left-hand corner an invertible matrix (**SLOUCH**, the univariate predecessor of **mvSLOUCH** has a scalar here), arbitrary values to the right and **0** below. This type of model is called an Ornstein–Uhlenbeck–Brownian motion (OUBM) model. In contrast, when **H** is non-singular the model is called an Ornstein–Uhlenbeck–Ornstein–Uhlenbeck (OUOU) one, some variables are labelled as predictors while the rest as responses.

It is of course not satisfactory to have restrictions on the form of **H**, unless these are motivated by purely biological reasons. Different setups have different biological interpretations with regard to modelling causation (see Bartoszek et al., 2012; Reitan et al., 2012). In particular singular matrices will be interesting as they will correspond to certain linear combinations of traits under selection pressures while other linear combinations are free of this. The OUBM model is a special case where a pre-defined group of traits is assumed to evolve marginally as a Brownian motion. Of course a more general setup is desirable and actually, as we show in this work, possible.

With respect to computational feasibility, the only assumption we impose on **H** is that it possesses an eigendecomposition, $\mathbf{H} = \mathbf{P}\mathbf{\Lambda}\mathbf{P}^{-1}$ ($\mathbf{\Lambda}$ is a diagonal matrix, and the i-th element of the diagonal is denoted as λ_i). In particular $\mathbf{\Lambda}$ can be singular, i.e. some eigenvalues are 0 and furthermore the eigenvalues/eigenvectors are allowed to be complex.

In this work we assume that $\mathbf{\Sigma}_x$ is upper triangular. Despite how it looks at first sight, this is not any sort of restriction, as in the likelihood we have only $\mathbf{\Sigma} := \mathbf{\Sigma}_x\mathbf{\Sigma}_x^T$. We furthermore assume that $\mathbf{\Sigma}$ is non-singular, otherwise the whole model would be singular from a statistics point of view.

If we assume that the process starts at a value $\vec{x}(0) = \vec{x}_0$, then after evolution over time t (assuming all parameters are constant on this interval) it will be normally distributed with mean vector and variance–covariance matrix (Eqs. (A.1, B.2) Bartoszek et al., 2012)

$$\begin{aligned}
\mathrm{E}\left[\vec{x}\right](t) &= e^{-\mathbf{H}t}\vec{x}_0 + \left(\mathbf{I} - e^{-\mathbf{H}t}\right)\vec{\theta} \in \mathbb{R}^k \\
\mathrm{Var}\left[\vec{x}\right](t) &= \int_0^t e^{-\mathbf{H}v}\mathbf{\Sigma}e^{-\mathbf{H}^T v}\mathrm{d}v \\
&= \mathbf{P}\left(\left[\frac{1}{\lambda_i+\lambda_j}\left(1-e^{-(\lambda_i+\lambda_j)t}\right)\right]_{1\le i,j\le k} \odot \mathbf{P}^{-1}\mathbf{\Sigma}\mathbf{P}^{-T}\right)\mathbf{P}^T \\
&\equiv \mathbf{V}(t) \in \mathbb{R}^{k\times k},
\end{aligned} \tag{18}$$

where **I** is the identity matrix of appropriate size and $\odot$ represents the Hadamard, entrywise, matrix product. Notice that in the above, **H** only enters the moments through its exponential. Therefore the moments can be calculated (and hence the distribution is well defined) for all **H**, including defective ones. However, if **H** has (as we assumed) an eigendecomposition, then the exponential and in turn variance formula can be calculated effectively. If $\lambda_i = \lambda_j = 0$, then the term in the variance has to be treated in the limiting sense $\lambda^{-1}(1-e^{-\lambda t}) \to t$ with $\lambda \to 0$. Therefore, the variance matrix is always well defined and never singular for $t > 0$.

We assumed that **H** has to have an eigendecomposition while the process is well defined for any **H**, including defective ones. Calculation of the matrix exponential for a defective matrix can be done using Jordan block decomposition. However, we do not provide such functionality, as Jordan block decomposition is numerically unstable and in fact, we are not aware of any `R` implementation of it. Hence, our framework cannot calculate the model likelihood when **H** is a defective matrix. Therefore, in the current implementation, the matrix **H** is checked (by checking if the eigenvector matrix from `eigen()`'s output is non-singular, Corollary 7.1.8., p. 353 Golub and Van Loan, 2013) before evaluating the likelihood and, if defective, an `NA` likelihood value is returned. It is important to stress here that defectiveness is the exception and not the rule for matrices.

We now turn to showing how to construct the composite parameters found in the proof of Theorem 1 from the OU process representation of Eq. (17).

To simplify the notation, for each tip $i \in \{1,\ldots,N\}$, we denote by $\mathbf{\Sigma}_e^i$ the sum of the measurement error for i (i.e. $\mathbf{V}_{err,i}$, see Eq. (16)) and/or any intra-species variability variance that has not been accounted for by the phylogenetic OU process (i.e. the global or regime-specific parameter $\mathbf{\Sigma}_e$, see Section 3.3). Then, based on the definition of the function $\mathbf{V}(t)$, Eq. (18), for any tip or internal node $i \in \{1,\ldots,M-1\}$, we define

$$\tilde{\mathbf{V}}_i \equiv \begin{cases} \mathbf{V}(t_i) + \mathbf{\Sigma}_e^i & \text{if } i \text{ is a tip} \\ \mathbf{V}(t_i) & \text{otherwise.} \end{cases} \tag{19}$$

Theorem 4. *Let $\vec{k}_i$ be the vector of coordinates on which $\vec{x}_i$ is observed, $\vec{k}_j$ be the vector of coordinates for $\vec{x}_j$ and $\vec{k}$ the full vector of coordinates. Using the parametrization found in the proof of Theorem 1 a multivariate Ornstein–Uhlenbeck process of evolution can be represented as*

$$\begin{aligned}
\mathbf{V}_i &= \tilde{\mathbf{V}}_i[\vec{k}_i,\vec{k}_i] \in \mathbb{R}^{|\vec{k}_i|\times|\vec{k}_i|}, \\
\vec{\omega}_i &= \left(\mathbf{I}[\vec{k}_i,\vec{k}] - e^{-\mathbf{H}t_i}[\vec{k}_i,\vec{k}]\right)\vec{\theta}[\vec{k}] \in \mathbb{R}^{|\vec{k}_i|}, \\
\mathbf{\Phi}_i &= e^{-\mathbf{H}t_i}[\vec{k}_i,\vec{k}_j] \in \mathbb{R}^{|\vec{k}_i|\times|\vec{k}_j|}.
\end{aligned} \tag{20}$$

Proof. In the multivariate OU case, Eq. (1) will be

$$\begin{aligned}
pdf(\vec{x}_i|\vec{x}_j,t_i) = \mathcal{N}\Big(&e^{-\mathbf{H}t_i}[\vec{k}_i,\vec{k}_j]\vec{x}_j \\
&+ \left(\mathbf{I}[\vec{k}_i,\vec{k}] - e^{-\mathbf{H}t_i}[\vec{k}_i,\vec{k}]\right)\vec{\theta}[\vec{k}], \mathbf{V}_i[\vec{k}_i,\vec{k}_i]\Big). \quad \square
\end{aligned}$$

These formulae do not depend on whether the eigenvalues of **H** are positive, negative or 0. Only with $\mathbf{V}_i$ will we need to take an appropriate limit as an eigenvalue becomes 0, see comments after Eq. (18).

Corollary 1. *For a multivariate Brownian motion process of evolution, we have $\mathbf{H} = \mathbf{0}$ and $\tilde{\mathbf{V}}_i = t_i\mathbf{\Sigma} + \delta_{i\in\{1,\ldots,N\}}\mathbf{\Sigma}_e^i$. Hence, using the parametrization found in the proof of Theorem 1 one can represent it as*

$$\begin{aligned}
\mathbf{V}_i &= \tilde{\mathbf{V}}_i[\vec{k}_i,\vec{k}_i] \in \mathbb{R}^{|\vec{k}_i|\times|\vec{k}_i|}, \\
\vec{\omega}_i &= \vec{0}[\vec{k}_i] \in \mathbb{R}^{|\vec{k}_i|}, \\
\mathbf{\Phi}_i &= \mathbf{I}[\vec{k}_i,\vec{k}_j] \in \mathbb{R}^{|\vec{k}_i|\times|\vec{k}_j|}.
\end{aligned} \tag{21}$$

In Fig. 3, panel C, one can see an example collection of tip observations resulting from simulating of a bivariate trait following a BM process on top of a phylogeny and in panel D following an OU process.

4.2. Multivariate Ornstein–Uhlenbeck with jumps (JOU)

OU processes with jumps (hereby denoted JOU) capture the key idea behind the theory of punctuated equilibrium (see Section 3.5). If the speed of convergence of the process is large

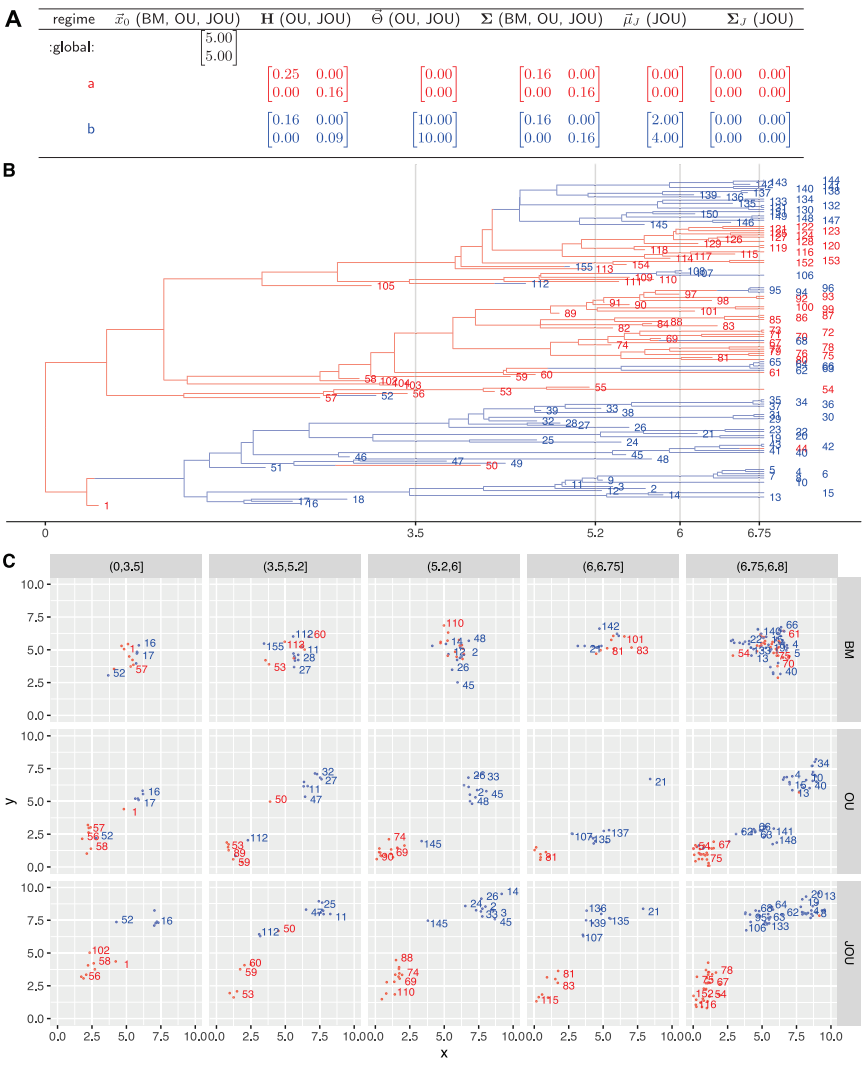

Fig. 3. Simulations of bivariate trait evolution under a BM, an OU and a JOU model with two regimes. The colour corresponds to the regime with red denoting regime "a" and blue denoting regime "b". A. Model parameters: each parameter has the same value for the model types written in parentheses, e.g. the parameters $\vec{x}_0$ and Σ are shared by all models, while the parameter $\vec{\mu}_J$ is used only in the JOU model. B. A birth–death tree generated using the function `pbtree()` and `sim.history()` from the package **phytools** (Revell, 2011). Parameters of the simulation: speciation-rate $\lambda = 1$, extinction-rate $\delta = 0.5$, migration rate from regime "a" to "b" $Q_{a\to b} = 0.1$; from regime "b" to "a" $Q_{b\to a} = 0.01$. C. Scatter plots of the trait values at the tips after random simulation using the function `PCMSim()` of the **PCMBase** package. The horizontal facets correspond to time intervals (also denoted by vertical grey lines in panel B). For BM, all observed values centre around the initial value, because the process does not have a directional component. The OU regimes have diagonal optima, reflected by the predominance of red dots in the bottom-left and blue dots in the top-right corners. There are several "clouds" of blue dots because the shift from "a" to "b" has not occurred at the same time for all lineages. For JOU, the shift from "a" to "b" is accompanied by a positive jump in the mean value for both traits. These jumps compensate for the delay in the shift from "a" to "b". Hence, most blue points form a cloud near the optimum for "b".

enough, then the stationary distribution is approached rapidly and the stationary oscillations around the (constant) mean can be interpreted as stasis between jumps.

The framework allows for calculating the likelihood of JOU models with normally distributed jumps in the trait's value occurring at user-specified points on the tree. The location of the jumps is specified by the user through the `edge.jump` of the tree object (see Section 3.5). Further, the jump distribution is specified by a k-vector parameter $\vec{\mu}_J$ denoting the mean of the jumps and a $k \times k$ matrix parameter Σ_J denoting the variance–covariance matrix of the jumps. These two parameters can be shared by all regimes or regime-specific. The following two corollaries describe how the definitions of $\mathbf{V}_i$, $\vec{\omega}_i$ and Φ_i change in the case of JOU and JBM processes (JBM standing for a Brownian motion with jumps process).

Corollary 2. *For a multivariate JOU process, jump distribution $\mathcal{N}(\vec{\mu}_J, \Sigma_J)$ and denoting by the indicator ξ_i the occurrence of a jump at the start of the branch leading to node i, we have*

$$\tilde{\mathbf{V}}_i = \int_0^{t_i} e^{-\mathbf{H}v}\Sigma e^{-\mathbf{H}^T v}\mathrm{d}v + \xi_i e^{-\mathbf{H}t_i}\Sigma_J e^{-\mathbf{H}^T t_i} + \delta_{i\in\{1,\dots,N\}}\Sigma_e^i. \quad (22)$$

Using the parametrization found in the proof of Theorem 1 one can represent it as

$$\begin{aligned}\mathbf{V}_i &= \tilde{\mathbf{V}}_i[\vec{k}_i,\vec{k}_i] \in \mathbb{R}^{|\vec{k}_i|\times|\vec{k}_i|},\\ \vec{\omega}_i &= \xi_i e^{-\mathbf{H}t_i}[\vec{k}_i,\vec{k}]\vec{\mu}_J[\vec{k}] + \left(\mathbf{I}[\vec{k}_i,\vec{k}] - e^{-\mathbf{H}t_i}[\vec{k}_i,\vec{k}]\right)\vec{\theta}[\vec{k}] \in \mathbb{R}^{|\vec{k}_i|},\\ \Phi_i &= e^{-\mathbf{H}t_i}[\vec{k}_i,\vec{k}_j] \in \mathbb{R}^{|\vec{k}_i|\times|\vec{k}_j|}.\end{aligned} \quad (23)$$

The multivariate JBM model follows as an immediate corollary ($\mathbf{H} \to \mathbf{0}$).

Corollary 3. *For a multivariate JBM model (jumps defined the same as in Corollary 2) the variance at a node i is $\tilde{\mathbf{V}}_i = t_i\Sigma[\vec{k}_i,\vec{k}_i] + \xi_i\Sigma_J[\vec{k}_i,\vec{k}_i] + \delta_{i\in\{1,\dots,N\}}\Sigma_e^i$. Using the parametrization found in the*

proof of Theorem 1 one can represent it as

$$\begin{aligned}\mathbf{V}_i &= \tilde{\mathbf{V}}_i[\vec{k}_i,\vec{k}_i] \in \mathbb{R}^{|\vec{k}_i|\times|\vec{k}_i|},\\ \vec{\omega}_i &= \xi_i\vec{\mu}_J[\vec{k}_i] \in \mathbb{R}^{|\vec{k}_i|},\\ \mathbf{\Phi}_i &= \mathbf{I}[\vec{k}_i,\vec{k}_j] \in \mathbb{R}^{|\vec{k}_i|\times|\vec{k}_j|}.\end{aligned} \tag{24}$$

In Fig. 3, panel E, one can see an example collection of tip observations resulting from simulating of a bivariate trait following a JOU process along a phylogeny.

4.3. Beyond the Ornstein–Uhlenbeck process

There are a number of popular PCM models that do not fall into the above described OU framework despite appearing very similar. In particular we mean the BM with trend, drift, early burst/Accelerating–decelerating (EB/ACDC) or white noise (implemented in the **geiger** R package Harmon et al., 2008). With the exception of white noise, they all can be represented by the SDE (cf. Eq. (1) of Manceau et al., 2016)

$$\begin{cases}\mathrm{d}\vec{x}(t) &= \left(\vec{h}(t)-\mathbf{H}\vec{x}(t)\right)\mathrm{d}t+\mathbf{\Gamma}(t)\mathrm{d}\vec{W}(t),\\ \vec{x}(0) &= \vec{x}_0.\end{cases} \tag{25}$$

Notice that setting $\vec{h}(t) = \vec{\theta}$ and $\mathbf{\Gamma}(t) = \mathbf{\Sigma}_x$ to constants we recover the "usual" OU process. Equations (4a) and (4b) in Manceau et al. (2016) provide the expectation and variance under the model. Localizing these two equations to a single branch, we can write

$$\begin{aligned}\mathrm{E}\left[\vec{x}_i|\vec{x}_j\right] &= e^{-t_i\mathbf{H}_i}\vec{x}_j+\int_{t_i^s}^{t_i^e}e^{(s-t_i^e)\mathbf{H}_i}\vec{h}_i(s)\mathrm{d}s,\\ \mathrm{Var}\left[\vec{x}_i|\vec{x}_j\right] &= \int_{t_i^s}^{t_i^e}e^{(s-t_i^e)\mathbf{H}_i}\mathbf{\Gamma}_i(s)\mathbf{\Gamma}_i^T(s)e^{(s-t_i^e)\mathbf{H}_i^T}\mathrm{d}s,\end{aligned} \tag{26}$$

where t_i^s is the time at the start of the branch and t_i^e at the end ($t_i = t_i^e - t_i^s$). This corresponds in our framework to

$$\begin{aligned}\vec{\omega}_i &= \int_{t_i^s}^{t_i^e}e^{(s-t_i^e)\mathbf{H}_i}\vec{h}_i(s)\mathrm{d}s,\\ \mathbf{\Phi}_i &= e^{-t_i\mathbf{H}_i},\\ \mathbf{V}_i &= \int_{t_i^s}^{t_i^e}e^{(s-t_i^e)\mathbf{H}_i}\mathbf{\Gamma}_i(s)\mathbf{\Gamma}_i^T(s)e^{(s-t_i^e)\mathbf{H}_i^T}\mathrm{d}s.\end{aligned} \tag{27}$$

Hence, *in the subcase of non-interacting lineages*, our framework covers Manceau et al. (2016)'s. As the initially mentioned models are subcases (cf. Tab. 1 of Manceau et al., 2016) they are available in our framework. To obtain the values of the $\vec{\omega}_i$, $\mathbf{\Phi}_i$ and $\mathbf{V}_i$ parameters in Eq. (27) one has to either analytically calculate the integrals for specific $\vec{h}_i(\cdot)$ and $\mathbf{\Gamma}(\cdot)$ functions or consider a general numerical integration scheme.

Apart from the previously considered OU model, for some other typical PCM models the integrals can be solved analytically.

1. ACDC model (after generalizing the one dimensional model presented by Blomberg et al., 2003; Harmon et al., 2010, to the multivariate case)

$$\begin{aligned}\vec{\omega}_i &= \vec{0},\\ \mathbf{\Phi}_i &= \mathbf{I},\\ \mathbf{V}_i &= \int_{t_i^s}^{t_i^e}e^{s\mathbf{R}_i}\mathbf{\Sigma}_i\mathbf{\Sigma}_i^Te^{s\mathbf{R}_i^T}\mathrm{d}s.\end{aligned} \tag{28}$$

   See Eq. (18) for how to calculate the integral, for $\mathbf{V}_i$, when the matrix $\mathbf{R}$ is eigendecomposable.
2. BM with drift

$$\begin{aligned}\vec{\omega}_i &= \vec{h}_it_i,\\ \mathbf{\Phi}_i &= \mathbf{I},\\ \mathbf{V}_i &= \mathbf{\Sigma}_i\mathbf{\Sigma}_i^Tt_i.\end{aligned} \tag{29}$$

3. BM with trend—in the most general setup of a linear form under the integral for $\mathbf{V}_i$ (based on the Supporting Information of Harmon et al., 2010, in the one dimensional case)

$$\begin{aligned}\vec{\omega}_i &= \vec{0},\\ \mathbf{\Phi}_i &= \mathbf{I},\\ \mathbf{V}_i &= \int_{t_i^s}^{t_i^s+t_i}(\mathbf{U}s+\mathbf{W})\,\mathrm{d}s = \mathbf{U}\tfrac{t_i^2}{2}+\mathbf{W}t_i.\end{aligned} \tag{30}$$

4. The white noise process corresponds to a situation, where the phylogeny does not contribute to the covariance structure between the species, so that all species are regarded as independent identically distributed observations of the same multivariate Gaussian distribution with global mean $\vec{x}_0 = \vec{\mu}$ and same variance–covariance matrix $\mathbf{\Sigma_e}$.

Naturally everything should be appropriately (as described in Section 3.2) adjusted if missing values are present.

5. Discussion

In this article, we have studied the sub-family of the Gaussian phylogenetic models $\mathcal{G}_{LInv}$. We have shown that, excluding the case of interacting lineages discussed previously in Manceau et al. (2016), $\mathcal{G}_{LInv}$ contains most, if not all, of the contemporary Gaussian models used in phylogenetic comparative methods. In mathematical terms, $\mathcal{G}_{LInv}$ includes the models of continuous trait evolution which are represented by a linear stochastic differential equation (SDE, see the representation by Eq. (1) of Manceau et al., 2016) whose drift matrix ("deterministic part" of the SDE) is piecewise constant with respect to the phylogeny, and diffusion matrix ("random part", sometimes referred to as "random drift part" in biological literature) does not depend on the trait. Furthermore, $\mathcal{G}_{LInv}$ includes models outside the above SDE representation, such as Lévy processes with jumps in the trait values occurring at random time instances (Landis et al., 2012; Duchen et al., 2017). Such models are of high biological relevance because they hold promise for attacking the longstanding question of whether "evolution is gradual or punctuated?".

We have formulated a computationally efficient framework for calculating the likelihood of $\mathcal{G}_{LInv}$ models. For that purpose, we have taken advantage of the absence of interactions between different lineages of the tree and derived an analytical integration of the multivariate conditional densities over the unobserved trait values at the internal nodes. Hence, the main contribution of this work is in formulating and implementing a general mathematical framework for linear in N likelihood calculation of $\mathcal{G}_{LInv}$ models (see also Eq. S4, Appendix D). However, for increasing numbers of traits, k, the likelihood calculation time, multiplied by the increasing number of needed likelihood calculations, will remain a key bottleneck preventing fast statistical inference. This is due to the fact that, complex $k \times k$ matrix operations are executed for each branch of the tree, even in the simplest case of a BM model. When the matrices $\mathbf{\Phi}$ and $\mathbf{V}$ (Definition 1) are dense, the asymptotically best known complexity of such operations is $O(k^{2.373})$ (Le Gall, 2014). We therefore do not believe that a probabilistic inference of such models would be practical beyond models of a few dozen traits. For bigger numbers of traits, the same theory presented here can be reused, but the models should be restricted to parametrizations where the matrices $\mathbf{\Phi}$ and $\mathbf{V}$ (Definition 1) are sparse and new implementations taking advantage of this sparsity need to be developed.

To further clarify the position of our work in the rapidly growing field of PCMs, we need to mention two previous publications that have contributed linear in N algorithms for multivariate PCM likelihood calculation. In the first work Bastide et al. introduce a phylogenetic EM method for maximum likelihood inference of shifts in a multivariate SCALAR OU model (Bastide et al., 2018).

Please cite this article as: V. Mitov, K. Bartoszek, G. Asimomitis et al., Fast likelihood calculation for multivariate Gaussian phylogenetic models with shifts. Theoretical Population Biology (2019), https://doi.org/10.1016/j.tpb.2019.11.005.

The upward recursion part of the E-step of the phylogenetic EM method includes an equivalent pruning algorithm for propagating the moments of the conditional Gaussian distributions from the tips to the root of the tree (Appendix 2, Section E step Bastide et al., 2018). The authors clarify that this procedure allows to calculate the log-likelihood in $O(N)$ time and generalizes to the same family of models ($\mathcal{G}_{Llnv}$). We acknowledge this fact, noting the different mathematical representation (i.e. Gaussian moment propagation versus quadratic polynomial integration). Another difference in the two works is the way missing measurements are treated. **PhylogeneticEM** uses a numerical approach by setting entries in the variance–covariance matrix corresponding to non-observed traits to infinity and then defining appropriate pseudo-inverse and lower-dimensional determinant operators. Conversely, **PCMBase** takes care of missing measurements by considering only the observed dimensions (i.e. appropriate rows and columns) of the vector- and matrix-terms involved in the likelihood calculation (see Section 3.2). Further, as conceptual novelties in our work, we emphasize the included support for non-existing traits (apart from missing measurements) and the formal proof that the $\mathcal{G}_{Llnv}$-family is identical with the scope of the pruning procedure (Theorems 1 and 3). Finally, the implementations resulting from the two works differ substantially in their intended usage: the **PhylogeneticEM** R-package targets the ML inference of shift points in the optimum ($\vec{\theta}$) of a SCALAR OU model, while the goal of the **PCMBase** package is to provide a convenient programming interface for specifying and calculating the likelihood of any $\mathcal{G}_{Llnv}$-model.

Another recent work is the 3-point structure likelihood calculation algorithm developed by Ho and Ané (2014a) and extended to multivariate OU-models by Goolsby et al. (2016). Like our framework, the 3-point structure algorithm also has $O(N)$ complexity. Unlike the 3-point structure algorithm, though, our framework does not require any model-specific transformation of the branch lengths. To help appreciating this simplicity, consider the case of a univariate OU process realized on a non-ultrametric tree. To apply the 3-point structure algorithm, the $N \times N$ variance covariance matrix, $\mathbf{V}$, resulting from the tree and the model parameters must satisfy a 3-point structure property, as defined on p. 400 in Ho and Ané (2014a).[3] In the case of a BM model the 3-point structure property holds for any type of tree. In the case of a univariate OU model, though, the 3-point structure property holds only if the tree is ultrametric. Each element V_{ij} of $\mathbf{V}$ is a function of the distance from the root to (i, j)'s most recent common ancestor, t_{ij}, and the phylogenetic distance between the two taxa, d_{ij} (see Equation on p. 402 in Ho and Ané (2014a)). *"The part involving t_{ij} satisfies the 3-point condition but the part with d_{ij} does not necessarily. It does if the tree is ultrametric because then d_{ij} only depends on the age of the most recent common ancestor of i and j, as we can write $d_{ij} = 2(T - t_{ij})$ where T is the tree height"* (p. 402 Ho and Ané, 2014a). Hence, to apply the 3-point structure algorithm in the case of a non-ultrametric tree, one needs to show that the process satisfies a *"generalized"* 3-point structure property, meaning that there exist diagonal $N \times N$ matrices $\mathbf{D_1}$ and $\mathbf{D_2}$, and a 3-point structure $N \times N$ variance covariance matrix $\tilde{\mathbf{V}}$, such that the original $N \times N$ variance covariance matrix can be expressed as $\mathbf{V} = \mathbf{D_1}\tilde{\mathbf{V}}\mathbf{D_2}$ (p. 401 Ho and Ané, 2014a). For the univariate OU process, the authors show that the *"generalized"* 3-point structure property is satisfied by a matrix $\mathbf{D_1} = \mathbf{D_2} = diag(e^{\alpha u_i}; i = 1, \ldots, N)$, where α is a current estimate for the (univariate) selection strength OU parameter and *"the external branch to taxon i is extended (or shortened) by u_i, in such a way that the modified tree is ultrametric"* (p. 403 Ho and Ané, 2014a).

[3] Until the end of this paragraph, we use the notation in Ho and Ané (2014a).

However, it remains unclear how the matrices $\mathbf{D_1}$ and $\mathbf{D_2}$ should be defined in the case of a multivariate OU process, where the selection strength parameter is a matrix, rather than a scalar (see Section 4.1). Also, it is not clear whether the same approach can be generalized in the case of an OU process with shifts or in the case of other Gaussian models with shifts. This, we think is the main reason why, in its current version (0.2.9) the R-package **Rphylopars**, using a multivariate generalization of the 3-point structure algorithm, does not support OU processes on non-ultrametric trees.

In contrast, once a $\mathcal{G}_{Llnv}$ model is formulated in terms of the transition functions $\vec{\omega}$, $\mathbf{\Phi}$ and $\mathbf{V}$ (Definition 1), the framework can be used to calculate the model likelihood on any kind of tree, including non-ultrametric trees and trees with polytomies. This is a major improvement with respect to current multivariate implementations with support for non-ultrametric trees, such as the R-packages **mvMORPH** (Clavel et al., 2015) and **RPANDA** (Manceau et al., 2016), which calculate the model likelihood in quadratic in N time using the standard formula for calculating multivariate Gaussian densities (see also Appendix D). **mvSLOUCH** (Bartoszek et al., 2012) is another R-package that allows for non-ultrametric trees but now uses **PCMBase** as its likelihood calculation engine.

Furthermore, our method can naturally handle measurement error (intra-species variability), missing data, non-existing traits for some of the species, and shifts at arbitrary points along the tree, both in the parameters of a model as well as the type of the model within the $\mathcal{G}_{Llnv}$ family. The support for non-ultrametric trees extends the application scope of our framework to macroevolutionary studies including sequentially sampled data from the fossil record, as well as epidemiological studies, in which the phylogenetic tree represents an approximation of the transmission tree of infected patients, as in the works of Hodcroft et al. (2014), Bertels et al. (2017) and Mitov and Stadler (2018).

We implemented our computational framework within the R-package **PCMBase** (Appendix A). This offers the possibility for fast likelihood calculation for all models in the $\mathcal{G}_{Llnv}$ family, including mixed-type models, where different types of models are realized on different parts of the phylogenetic tree. Further, it is extremely flexible allowing the user to easily use it as a computational engine for their particular modelling setup/parametrization. Currently, the package includes numerous parametrizations of the BM and the OU model and their jump-enabled variants. Other model types within $\mathcal{G}_{Llnv}$ can be implemented using the package extension interface (Appendix A). Finally, we note that the functional scope of **PCMBase** has been limited to specification, simulation and likelihood calculation for $\mathcal{G}_{Llnv}$ models. While the package does not provide functions for model inference, such functionality has been provided in depending packages, such as the new version of **mvSLOUCH** and the new R-packages **PCMFit** (Mitov et al., 2019) and **PCMBayes** (https://venelin.github.io/PCMBayes).

Acknowledgements

V.M. and T.S. were supported by ETH Zürich. K.B. was supported by the Knut and Alice Wallenbergs Foundation, the G S Magnuson Foundation of the Royal Swedish Academy of Sciences (Grant No. MG2016–0010) and is supported by the Swedish Research Council (Vetenskapsrådet) Grant No. 2017–04951. The authors wish to acknowledge three anonymous reviewers and Dr. Paul Bastide for their invaluable comments.

Supplementary Data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.tpb.2019.11.005.

References

Adams, D.C., Collyer, M.L., 2018. Multivariate phylogenetic comparative methods: Evaluations, comparisons, and recommendations. Syst. Biol. 67, 14–31.

Bartoszek, K., 2014. Quantifying the effects of anagenetic and cladogenetic evolution. Math. Biosci. 254, 42–57.

Bartoszek, K., Pienaar, J., Mostad, P., Andersson, S., Hansen, T.F., 2012. A phylogenetic comparative method for studying multivariate adaptation. J. Theoret. Biol. 314, 204–215.

Bastide, P., Ané, C., Robin, S., Mariadassou, M., 2018. Inference of adaptive shifts for multivariate correlated traits. Syst. Biol. 113, 2158–2680.

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, 2369–2383.

Bedford, T., Hartl, D.L., 2009. Optimization of gene expression by natural selection. In: Proceedings of the National Academy of Sciences. Department of Organismic and Evolutionary Biology, Harvard University, 16 Divinity Avenue, Cambridge, MA 02138 National Academy of Sciences, pp. 1133–1138.

Bertels, F., Marzel, A., Leventhal, G., Mitov, V., Fellay, J., Günthard, H.F., Böni, J., Yerly, S., Klimkait, T., Aubert, V., Battegay, M., Rauch, A., Cavassini, M., Calmy, A., Bernasconi, E., Schmid, P., Scherrer, A.U., Müller, S., Kouyos, R., Regoes, R.R., Swiss HIV Cohort Study, 2017. Dissecting HIV virulence: Heritability of setpoint viral load, CD4+ T cell decline and per-parasite pathogenicity. Mol. Biol. Evol. 35 (1), 27–37.

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.

Bokma, F., 2002. Detection of punctuated equilibrium from molecular phylogenies. J. Evol. Biol. 15, 1048–1056.

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

Caetano, D.S., Harmon, L.J., 2017. Ratematrix: An rpackage for studying evolutionary integration among several traits on phylogenetic trees. Methods Ecol. Evol. 8, 1920–1927.

Clavel, J., Escarguel, G., Merceron, G., 2015. Mvmorph: an R package for fitting multivariate evolutionary models to morphometric data. Methods Ecol. Evol. 6, 1311–1319.

Cooper, N., Thomas, G.H., Venditti, C., Meade, A., Freckleton, R.P., 2015. A cautionary note on the use of Ornstein Uhlenbeck models in macroevolutionary studies. Biol. J. Linnean Soc. 118, 64–77.

Cybis, G.B., Sinsheimer, J.S., Bedford, T., Mather, A.E., Lemey, P., Suchard, M.A., 2015. Assessing phenotypic correlation through the multivariate phylogenetic latent liability model. Ann. Appl. Stat. 9, 969–991.

Duchen, P., Leuenberger, C., Szilágyi, S.M., Harmon, L., Eastman, J., Schweizer, M., Wegmann, D., 2017. Inference of evolutionary jumps in large phylogenies using Lévy processes. Syst. Biol. 66, 950–963.

Eastman, J.M., Alfaro, M.E., Joyce, P., Hipp, A.L., Harmon, L.J., 2011. A novel comparative method for identifying shifts in the rate of character evolution on trees. Evolution 65, 3578–3589.

Edwards, A.W.F., 1970. Estimation of the branch points of a branching diffusion process (with discussion). J. Roy. Statist. Soc. Ser. B Methodol. 32, 155–174.

Eldredge, N., Gould, S.J., 1972. Punctuated equilibria: an alternative to phyletic gradualism. In: Schopf, T.J.M., Thomas, J.M. (Eds.), Models in Paleobiology. Freeman Cooper, San Francisco, pp. 82–115.

Felsenstein, J., 1985. Phylogenies and the comparative method. Amer. Nat. 125, 1–15.

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

FitzJohn, R.G., 2010. Quantitative traits and diversification. Syst. Biol. 59, 619–633.

FitzJohn, R.G., 2012. Diversitree: comparative phylogenetic analyses of diversification in R. Methods Ecol. Evol. 3, 1084–1092.

Freckleton, R.P., 2012. Fast likelihood calculations for comparative analyses. Methods Ecol. Evol. 3, 940–947.

Garamszegi, L.Z., 2014. Uncertainties due to within-species variation in comparative studies: Measurement errors and statistical weights. In: Modern Phylogenetic Comparative Methods and their Application in Evolutionary Biology. Springer, Berlin, Heidelberg, pp. 157–199.

Gill, M.S., Tung Ho, L.S., Baele, G., Lemey, P., Suchard, M.A., 2016. A relaxed directional random walk model for phylogenetic trait evolution. Syst. Biol. 66, 299–319, http://oup.prod.sis.lan/sysbio/article-pdf/66/3/299/25423282/syw093pdf.

Golub, G.H., Van Loan, C.F., 2013. Matrix Computations. The Johns Hopkins University Press, Baltimore.

Goolsby, E.W., Bruggeman, J., Ané, C., 2016. Rphylopars: fast multivariate phylogenetic comparative methods for missing data and within–species variation. Methods Ecol. Evol. 8, 22–27.

Gould, S.J., Eldredge, N., 1993. Punctuated equilibrium comes of age. Nature 366, 223–227.

Hadfield, J.D., Nakagawa, S., 2010. General quantitative genetic methods for comparative biology: phylogenies, taxonomies and multi-trait models for continuous and categorical characters. J. Evol. Biol. 23, 494–508.

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

Hansen, T.F., Bartoszek, K., 2012. Interpreting the evolutionary regression: the interplay between observational and biological errors in phylogenetic comparative studies. Syst. Biol. 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., Jonathan Davies, T., Gillespie, R.G., Gittleman, J.L., Bryan Jennings, W., Kozak, K.H., McPeek, M.A., Moreno-Roark, F., Near, T.J., Purvis, A., Ricklefs, R.E., Schluter, D., Schulte I.I., J.A., Seehausen, O., Sidlauskas, B.L., Torres-Carvajal, O., Weir, J.T., Mooers, A.Ø., 2010. Early bursts of body size and shape evolution are rare in comparative data. Evolution 64, 2385–2396.

Harmon, L.J., Weir, J.T., Brock, C.D., Glor, R.E., Challenger, W., 2008. GEIGER: investigating evolutionary radiations. Bioinformatics 24, 129–131.

Hiscott, G., Fox, C., Parry, M., Bryant, D., 2016. Efficient recycled algorithms for quantitative trait models on phylogenies. Genome Biol. Evol. 8, 1338–1350.

Ho, L.s.T., Ané, C., 2014a. A linear–time algorithm for Gaussian and non-Gaussian trait evolution models. Syst. Biol. 63, 397–408.

Ho, L.s.T., Ané, C., 2014b. Intrinsic inference difficulties for trait evolution with Ornstein–Uhlenbeck models. Methods Ecol. Evol. 5, 1133–1146.

Hodcroft, E., Hadfield, J.D., Fearnhill, E., Phillips, A., Dunn, D., O'Shea, S., Pillay, D., Brown, A.J.L., 2014. The contribution of viral genotype to plasma viral set–point in HIV infection. PLoS Pathogens 10, e1004112.

Ingram, T., Mahler, D.L., 2013. SURFACE: detecting convergent evolution from comparative data by fitting Ornstein–Uhlenbeck models with stepwise Akaike Information Criterion. Methods Ecol. Evol. 4, 416–425.

Khabbazian, M., Kriebel, R., Rohe, K., Ané, C., 2016. Fast and accurate detection of evolutionary shifts in Ornstein–Uhlenbeck models. Methods Ecol. Evol. 7, 811–824.

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

Landis, M.J., Schraiber, J.G., Liang, M., 2012. Phylogenetic analysis using Lévy processes: finding jumps in the evolution of continuous traits. Syst. Biol. 62, 193–204.

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

Le Gall, F., 2014. Powers of tensors and fast matrix multiplication. In: Proceedings of the 39th International Symposium on Symbolic and Algebraic Computation. ACM, New York, NY, USA, pp. 296–303.

Manceau, M., Lambert, A., Morlon, H., 2016. A unifying comparative phylogenetic framework including traits coevolving across interacting lineages. Syst. Biol. 66, syw115–568.

Mayr, E., 1982. Speciation and macroevolution. Evolution 36, 1119–1132.

Mitov, V., Bartoszek, K., Stadler, T., 2019. Automatic generation of evolutionary hypotheses using mixed Gaussian phylogenetic models. Proc. Natl. Acad. Sci. USA 34, 201813823.

Mitov, V., Stadler, T., 2018. A practical guide to estimating the heritability of pathogen traits. Mol. Biol. Evol. 6, e1001123, –msx328 VL – IS –.

Mitov, V., Stadler, T., 2019. Parallel likelihood calculation for phylogenetic comparative models: The splitt c++ library. Methods Ecol. Evol. 10, 493–506.

Pybus, O.G., Suchard, M.A., Lemey, P., Bernardin, F.J., Rambaut, A., Crawford, F.W., Gray, R.R., Arinaminpathy, N., Stramer, S.L., Busch, M.P., Delwart, E.L., 2012. Unifying the spatial epidemiology and molecular evolution of emerging epidemics. Proc. Natl. Acad. Sci. USA 109, 15066–15071.

Reitan, T., Schweder, T., Henderiks, J., 2012. Phenotypic evolution studied by layered stochastic differential equations. Ann. Appl. Stat. 6, 1531–1551.

Revell, L.J., 2011. phytools: an R package for phylogenetic comparative biology (and other things). Methods Ecol. Evol. 3, 217–223.

Rohlfs, R.V., Harrigan, P., Nielsen, R., 2013. Modeling gene expression evolution with an extended Ornstein–Uhlenbeck process accounting for within–species variation. Mol. Biol. Evol. 31, 201–211.

Slater, G.J., 2013. Phylogenetic evidence for a shift in the mode of mammalian body size evolution at the Cretaceous-Palaeogene boundary. Methods Ecol. Evol. 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 Ecol. Evol. 5, 714–718.

Stadler, T., 2009. On incomplete sampling under birth-death models and connections to the sampling–based coalescent. J. Theoret. Biol. 261, 58–66.

Stadler, T., 2011. Simulating trees with a fixed number of extant species. Syst. Biol. 60, 676–684.

Stebbins, G.L., Ayala, F.J., 1981. Is a new evolutionary synthesis necessary? Science 213, 967–971.

Please cite this article as: V. Mitov, K. Bartoszek, G. Asimomitis et al., Fast likelihood calculation for multivariate Gaussian phylogenetic models with shifts. Theoretical Population Biology (2019), https://doi.org/10.1016/j.tpb.2019.11.005.