

A Unifying Comparative Phylogenetic Framework Including Traits Coevolving Across Interacting Lineages

MARC MANCEAU^{1,3,4,*}, AMAURY LAMBERT^{2,3}, AND HÉLÈNE MORLON⁴

¹Muséum National d'Histoire Naturelle, 75005 Paris, France; ²Laboratoire Probabilités et Modèles Aléatoires, UPMC—University of Paris 06, 75005 Paris, France; ³Center for Interdisciplinary Research in Biology, Collège de France, CNRS UMR 7241, 75005 Paris, France; and ⁴Institut de Biologie, École Normale Supérieure, CNRS UMR 8197, 75005 Paris, France

*Correspondence to be sent to: Marc Manceau, Institut de Biologie de l'École Normale Supérieure, 46 rue d'Ulm, 75005 Paris, France; E-mail: marc.manceau@ens.fr

Received 20 May 2016; reviews returned 23 November 2016; accepted 15 December 2016
Associate Editor: Mark Holder

Abstract.—Models of phenotypic evolution fit to phylogenetic comparative data are widely used to make inferences regarding the tempo and mode of trait evolution. A wide range of models is already available for this type of analysis, and the field is still under active development. One of the most needed development concerns models that better account for the effect of within- and between-clade interspecific interactions on trait evolution, which can result from processes as diverse as competition, predation, parasitism, or mutualism. Here, we begin by developing a very general comparative phylogenetic framework for (multi)-trait evolution that can be applied to both ultrametric and nonultrametric trees. This framework not only encapsulates many previous models of continuous univariate and multivariate phenotypic evolution, but also paves the way for the consideration of a much broader series of models in which lineages coevolve, meaning that trait changes in one lineage are influenced by the value of traits in other, interacting lineages. Next, we provide a standard way for deriving the probabilistic distribution of traits at tip branches under our framework. We show that a multivariate normal distribution remains the expected distribution for a broad class of models accounting for interspecific interactions. Our derivations allow us to fit various models efficiently, and in particular greatly reduce the computation time needed to fit the recently proposed phenotype matching model. Finally, we illustrate the utility of our framework by developing a toy model for mutualistic coevolution. Our framework should foster a new era in the study of coevolution from comparative data. [Character displacement; coevolution; comparative phylogenetics; interspecific interactions; linear stochastic differential equations; trait evolution.]

Evolutionary biologists have long been interested in the long-term evolution of phenotypic traits (Simpson 1944). Felsenstein (1973) introduced one of the first models of phenotypic evolution, with the initial goal to account for shared ancestry when testing for statistical correlation between pairs of traits in extant species. In this founding paper, Felsenstein proposed that a one-dimensional quantitative trait evolving on a tree could be modeled as a Brownian process that splits into two independent Brownian processes at branching times. This model mimics a trait that would evolve as a mere effect of stochastic drift; it is now often used as a null model, but also to estimate the relative lability (or rate of evolution) of various traits in a given group of organisms or of a given trait across different groups of organisms (Thomas et al. 2006; Harmon et al. 2010).

Since these early developments, evolutionary biologists have designed a series of models to better understand the evolutionary processes that shape phenotypic evolution [see Pennell and Harmon (2013) for a review]. The Ornstein–Uhlenbeck (OU) process has been proposed to model evolution under stabilizing selection, that is, with a selective pressure pushing trait values toward a given optimum (Hansen and Martins 1996; Hansen 1997; Butler and King 2004). The ACDC model has been proposed to account for accelerating (AC) or decelerating (DC) rates of phenotypic evolution through time (Blomberg et al. 2003). The latter scenario, where the evolutionary rate is high early in the history of a clade and subsequently declines toward the present,

well known as the early burst (EB) model, has often been used to test support for adaptive radiation theory (Harmon et al. 2010; Moen and Morlon 2014). These univariate models representing the evolution of a single trait have been extended to multivariate models representing the simultaneous evolution of multiple traits, thus allowing to directly test hypotheses about the coevolution between several phenotypic traits (Hansen et al. 2008; Bartoszek et al. 2012; Jhwueng and Maroulas 2014). Other extensions have been developed to account for variations in model parameters across clades (Butler and King 2004; O'Meara et al. 2006; Revell and Collar 2009; Eastman et al. 2011; Beaulieu et al. 2012). Finally, some of these models have been developed in the context of phylogenies including fossil data (i.e., nonultrametric trees, see Ruta et al. 2006; Slater 2015) in addition to phylogenies with only extant taxa (i.e., ultrametric trees). Most of these models have been implemented in open-access packages (Butler and King 2004; Martins 2004; Harmon et al. 2008; Thomas and Freckleton 2012; Clavel et al. 2015; Morlon et al. 2015), allowing their application to a broad variety of questions and data sets (see e.g., Labra et al. 2009; Mahler et al. 2010; Dale et al. 2015; Quintero et al. 2015; Slater 2015).

Despite these developments, most currently available models ignore the effect of interspecific interactions on trait evolution. Given the importance of species interactions in classical evolutionary theories, such as Simpson's adaptive radiation (Simpson 1944), Ehrlich and Raven's escape and radiate (Ehrlich and Raven 1964)

and Van Valen's Red Queen (Van Valen 1973) theories, building models that better account for such interactions is fundamental. In a first attempt to take into account the role of competition for niche space on character evolution, a diversity-dependent (DD) model has been introduced, where the rate of phenotypic evolution declines as the number of lineages in the clade increases (Mahler et al. 2010; Weir and Mursleen 2013). While this model represents an important first step, it still assumes that trait changes in one lineage are independent from the value of traits in other, interacting lineages, therefore ignoring the widespread idea of trait- (or ecologically-) driven interspecific interactions. More recently, the phenotype matching (PM) model relaxed these hypotheses and more explicitly accounted for interspecific interactions by modeling either the attraction or the repulsion of traits from a clade-wise average trait value. In the first case, referred to as matching mutualism, species traits tend to converge to similar values, whereas in the second case, referred to as matching competition, species traits tend to diverge (Nuismer and Harmon 2014; Drury et al. 2016).

The comparative phylogenetic approach developed by Drury et al. (2016) allows fitting a version of the model introduced by Nuismer and Harmon (2014) where the evolution of trait values in one lineage is influenced by the trait values of other lineages. This approach is focused on the evolution of traits within one clade. While within-clade interactions can be particularly relevant for some types of interactions (e.g., in the case of competitively driven character displacement, Brown and Wilson 1956), the effect of other types of antagonistic or mutualistic interactions on trait evolution is often most relevant between distantly related species. For example, host–parasite interactions are thought to drive a coevolutionary race between traits involved in host defence and parasite ability to infect (e.g., the production rate of a toxic compound v.s. the inhibitory concentration of parasite growth). Similarly, prey–predator interactions may lead to the coevolution of prey traits involved in camouflage, repulsion, or escape strategies, together with predator traits involved in the ability to detect and capture its prey (e.g., escape speed vs. hunting speed) (Ehrlich and Raven 1964; Dawkins and Krebs 1979). Mutualistic plant–pollinator interactions also are thought to drive the coevolution between plant traits involved in pollen accessibility or flower attractiveness to their pollinator (secondary metabolites, floral traits), and pollinator traits involved in the ability to detect suitable plants and to exploit plant rewards (Fenster et al. 2004; Weiblen 2004; Sletvold et al. 2016). While these types of biotic interactions likely play a key role in trait evolution and have been crucial in the development of coevolutionary theories (Ehrlich and Raven 1964; Van Valen 1973), there currently exists no framework for fitting models of phenotypic evolution incorporating the effect of clade–clade interactions.

The current article expands the work of Bartoszek et al. (2012) who presented a unified framework for studying

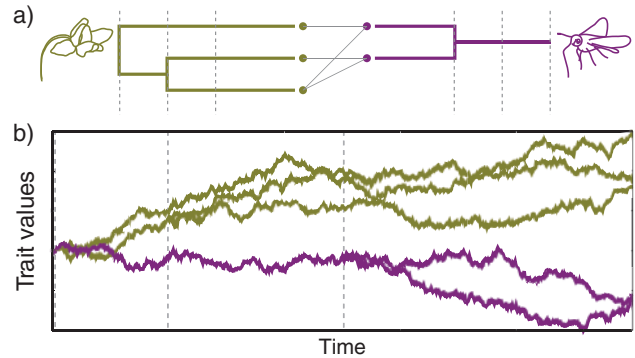


FIGURE 1. Phenotypic evolution in coevolving clades. We aim at developing a framework for analyzing how traits evolve as a response of potentially complex ecological interactions. a) The phylogenies of plants (left) and pollinators (right), together with the present-day interaction network (middle). b) The mean proboscis length of pollinators and the mean floral tube length of pollinated flowers coevolve as a result of mutualistic interactions.

coevolving traits in independently evolving lineages by providing a unified framework for coevolving traits in coevolving lineages. Throughout the article, the term coevolution refers to the evolution of traits on fixed phylogenies, that is, without any effect on the speciation–extinction dynamics. In short, we aim to provide the phylogenetic comparative tools allowing to study coevolutionary scenarios such as the one depicted on Figure 1, where a plant trait and a pollinator trait coevolve as a result of mutualistic interactions. We focus on quantitative rather than discrete traits (Pagel 1994) and on gradual rather than punctuated evolution (Bokma 2002; 2008; Landis et al. 2013; Bartoszek 2014). Our framework, based on linear stochastic differential equations (SDE), encompasses all models of continuous (multi-)trait evolution mentioned above, and allows the treatment of a broad set of coevolutionary models. We show that the tip trait distribution under all these models is Gaussian, and we highlight a general procedure that allows one to compute its expectation and covariance structure. We also provide analytical developments that speed up the computation of the likelihood in comparison with the general procedure. This leads, for example, to a faster algorithm for the likelihood computation of the PM model. The goal of the article is two-fold: first, by providing general solutions to the distribution of traits at tip branches under our unified framework, we hope to help users find their way in a dense and potentially overwhelming literature; second, by showing how the framework can be used to treat a broad class of within-clade and clade–clade coevolutionary scenarios, we hope to foster the development of models to test long-standing hypotheses on the role of competition, predation, parasitism, and mutualism in evolution.

We begin by presenting our framework and showing how previous models as well as novel clade–clade coevolutionary models fit within this framework; next, we provide general solutions for the distribution of tip trait values under this framework; then, we illustrate

how the framework can be used to formalize and study a toy model of clade–clade coevolution.

A GENERAL FRAMEWORK FOR PHENOTYPIC EVOLUTION

We introduce a general formalism to study (multi-)trait (co)evolution when the interaction between distinct phenotypic traits, distinct lineages within a clade, and/or distinct lineages among several clades potentially affects how phenotypic traits evolve.

Trait Evolution Through Time

We begin by considering the evolution of n traits at a given time t ; these can, for example, represent n distinct traits evolving on a single lineage or a single trait evolving in n lineages. We denote by X_t the column vector of the n trait values at time t . Throughout the article we assume that the evolution of the traits is driven by a linear stochastic differential equation of the form:

$$\begin{cases} dX_t = (a(t) - AX_t)dt + \Gamma(t)dW_t \\ X_0 = x_0 \end{cases} \quad (1)$$

where a is a vector of \mathbb{R}^n whose coefficients can vary with time, A is a constant square matrix of size $\mathbb{R}^n \times \mathbb{R}^n$, Γ is a square matrix of the same size whose coefficients can vary with time, and W_t is a n -Brownian motion (BM) (i.e., a vector composed of n independent standard BMs). Schematic examples are presented in Figure 2.

The formulation above implies that we consider interaction effects that operate gradually in time rather than as punctuated events. Intuitively, the deterministic part $(a(t) - AX_t)dt$ reflects the direct effects of trait values on the evolution of these traits, including the effect of a trait value in one lineage on both

its own evolution (as in the OU process) and the evolution of traits in the other lineages (as in the PM process). The stochastic part $\Gamma(t)dW_t$ reflects drift and the environmental noise influencing trait evolution. It has been proposed that correlations within the covariance matrix Γ represent noncausal correlations, for example, linked to joint evolutionary responses to shared environmental conditions, while correlations within the interaction matrix A represent causal effects (Reitan et al. 2012; Liow et al. 2015). For simplicity, we avoid the “causal/non-causal” dichotomy here; we stick to the term “correlations” and consider only models making the simplifying assumption that Γ is diagonal. The framework is, however, equally adapted to deal with correlations incorporated through Γ .

Notation for Trees and Traits

We now consider a single or several clades, each of them represented by a single, fixed, binary, time-calibrated phylogenetic tree. Our framework could easily be modified to treat nonbinary trees including polytomies. Trees are not necessarily ultrametric, meaning that they may include noncontemporary tips (fossils). When considering multiple clades, all associated trees share the same absolute time calibration. Time t runs from the root of the oldest tree ($t = \tau_0 = 0$) to the most recent tip of all trees ($t = T$ is the present if at least one of the phylogenies includes extant species). The K successive branching and extinction times when considering the various trees altogether are denoted by $(\tau_1, \tau_2, \dots, \tau_K)$ and the time intervals between two such events are called epochs, following Butler and King (2004). We denote by n_t the total number of lineages that arose before (and at) time t .

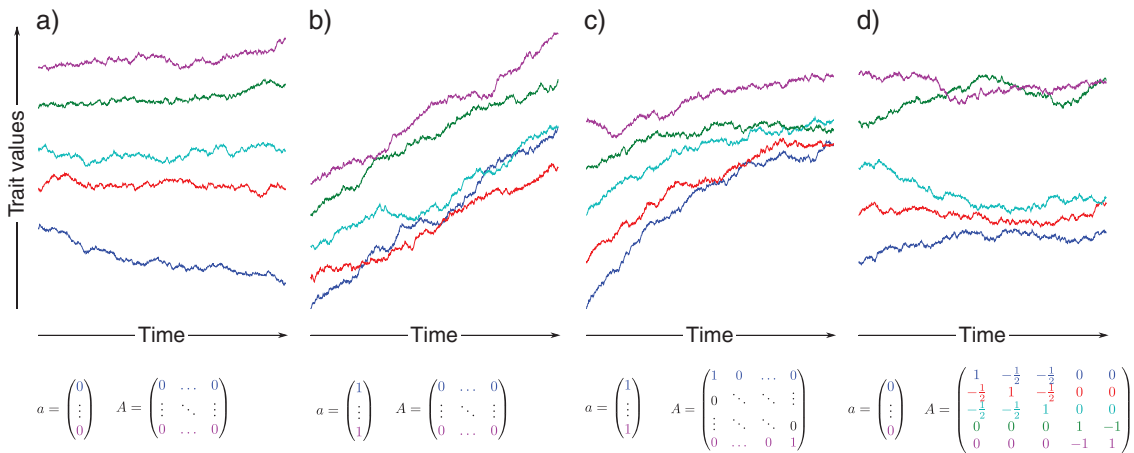


FIGURE 2. Schematic examples of trait evolution under various models covered by our framework. a) BM model b) BM model with drift c) OU model. In (a), (b), and (c), traits evolve independently from one another. d) illustrates a new class of model that can be handled in our framework, where a given trait value can influence the evolution of other traits: the two top traits evolve independently from the three bottom traits, but within each of these independently evolving groups of traits, trait values are attracted to the mean of the interacting traits, as in the PM model. The a vector and the A matrix corresponding to each model are represented, with trivial parameter values (e.g., the strength of attraction and optimal value of the OU model are set to 1). The parameters in each row dictate the evolution of the trait to which the specific row is associated. Each column represents the effect of the corresponding trait on each of the evolving traits. Under all these models, Γ is the identity matrix.

In the case of trait evolution within a single clade, we assign numbers (from 1 to n_t) to lineages by order of origination. At each branching event τ , one daughter lineage inherits the number assigned to the ancestral lineage while the other one is assigned n_τ .

We model the evolution of d one-dimensional quantitative traits. We denote by $X_t^{(i,j)}$ the value of trait j ($1 \leq j \leq d$) on branch i at time t and X_t the column vector containing the values of all traits on all lineages at time t , ordered as follows: $X_t = {}^{tr}(X_t^{(1,1)}, X_t^{(1,2)}, \dots, X_t^{(1,d)}, X_t^{(2,1)}, \dots, X_t^{(n_t,d)})$, where tr stands for the transposition.

In the case of trait evolution in c distinct (co)evolving clades, we begin by arbitrarily ordering the clades from 1 to c ; then, we assign numbers to lineages following the formalism introduced above, first numbering lineages from clade 1, then clade 2, and so on. As above, we denote by X_t the column vector containing the values of all traits on all lineages at time t , which now is a concatenation of the c column vectors corresponding to each clade.

Trait Evolution on Trees

Given one (or several) phylogenetic tree(s), a model of phenotypic evolution is entirely defined by initial conditions X_0 on the trait values at the root(s) and a set of rules dictating how the vector of traits X_t is updated (i) at branching times, (ii) through each epoch (i.e., between two branching or extinction times), and (iii) after a death time.

In line with most models of phenotypic evolution, we consider anagenetic character evolution, meaning that traits do not change at cladogenesis. Hence, at a given branching time τ , each of the daughter lineages inherits the trait value of their mother lineage. In practice, in the case of evolution within a single clade, the new vector X_τ is obtained by concatenating the d trait values of the branching lineage at time τ at the end of $X_{\tau-}$ (where $\tau-$ is the time just preceding the branching event). In the case of evolution in several clades, the new vector X_τ is obtained by inserting the d trait values

of the branching lineage at time τ at the appropriate location in $X_{\tau-}$ (i.e., at the end of the part of $X_{\tau-}$ corresponding to the clade in which the branching event is occurring).

On each given epoch (τ_i, τ_{i+1}) ($i \in \{0, 1, \dots, K-1\}$), we assume that the evolution of the d traits on the n lineages is driven by a linear stochastic differential equation of the form introduced earlier in Equation (1):

$$\begin{cases} dX_t = (a_i(t) - A_i X_t)dt + \Gamma_i(t) dW_t \\ X(\tau_i) = X_{\tau_i} \end{cases} \quad (2)$$

where a , A , and Γ are now indexed by i , the label of the focal period. The content, as well as the size of a , A , and Γ can hence vary with the period. Here, a_i is a vector of \mathbb{R}^{nd} whose coefficients can vary with time, A_i is a constant square matrix of size $\mathbb{R}^{nd} \times \mathbb{R}^{nd}$, Γ_i is a square matrix of the same size whose coefficients can vary with time, and W_t is a nd -BM.

Finally, when a lineage goes extinct at a given time τ , its d trait values no longer evolve (i.e., they are frozen at the extinction time), and they no longer have any influence on the evolution of the traits of other lineages until reaching the end of the process at time $t = T$. In practice, this means that the vector X_τ is simply equal to $X_{\tau-}$, and that the d lines and columns in a_i , A_i , and Γ_i corresponding to the now extinct lineage are all set to zero.

We will show later that this general formulation encapsulates many classical models of phenotypic evolution, ensures analytical tractability, and further allows the incorporation of a broad set of interspecific coevolutionary scenarios.

Given the above, initial conditions on X_0 , and the collection of (a_i) , (A_i) , and (Γ_i) associated to each epoch fully define a process of trait evolution on one or several trees. This formalism is illustrated in Figure 3 for a single trait evolving on a single small tree.

All models written under the formalism that we propose can easily be simulated numerically. First, the whole trajectory of the process can be simulated using a numerical scheme for SDE such

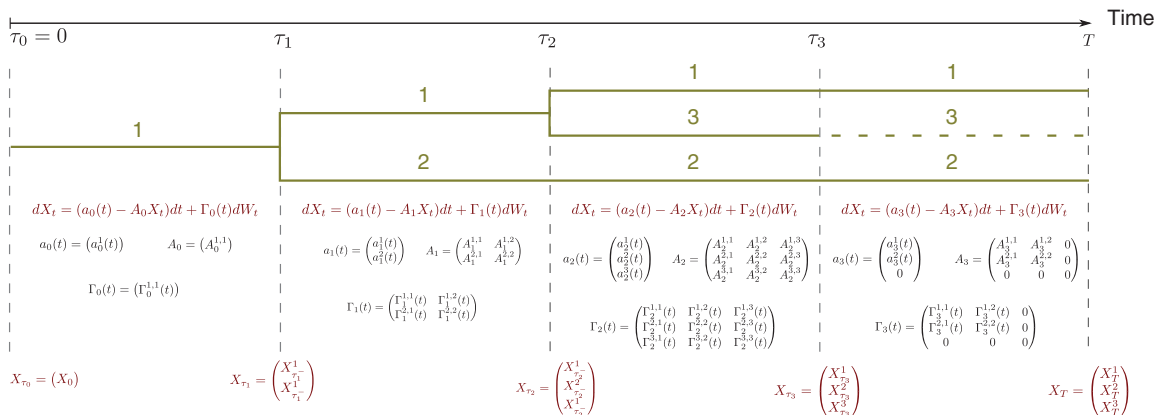


FIGURE 3. Formalism used throughout the article to model the evolution of one trait on a nonultrametric tree. Epochs are separated with vertical dashed lines.

as the Euler–Maruyama scheme (Gardiner et al. 1985) through each epoch, and augmenting the vector of traits at branching times with traits corresponding to the branching lineage (see Online Appendix D.1 available as Supplementary Material on Dryad at <http://dx.doi.org/10.5061/dryad.52636>). Second, we show in the next section how to compute numerically the tip distribution. Tip values can then directly be drawn in a fast way from the tip distribution.

Application: Existing and Novel Models of Trait Evolution

We first show that the general formulation above encapsulates many classical models of phenotypic evolution, before showing how it further allows considering a much broader set of models, including models of within- and between-clades coevolution.

Models of phenotypic evolution have traditionally been characterized by a stochastic differential equation specifying how a given trait evolves along a single lineage. Applying Equation (2) to trait k on epoch i yields:

$$dX_t^{(k)} = \left(a_i^{(k)}(t) - \sum_{l=1}^{n_t d} A_i^{(k,l)} X_t^{(l)} \right) dt + \sum_{l=1}^{n_t d} \Gamma_i^{(k,l)}(t) dW_t^{(l)} \quad (3)$$

where the two sums are taken over all traits and all lineages. The term $\sum_{l=1}^{n_t d} A_i^{(k,l)} X_t^{(l)}$ is the term that specifies how the value of trait k and all other traits in all other lineages influence the evolution of trait k . Given a well-known differential equation specifying how a given trait evolves along a single lineage for a previously proposed model of phenotypic evolution (second column in Table 1), deriving the corresponding expressions for a , A , and Γ using Equation (3) is straightforward. Table 1 summarizes these expressions

for existing univariate models running on ultrametric trees.

The first three models (BM, ACDC, and DD) are models in which trait evolution along a lineage is influenced neither by the trait value of this lineage nor the trait value of any other lineage. The corresponding A matrices are null matrices, as would be the case for any model with the latter property. The fourth model (OU) is a model in which trait evolution along a lineage is influenced by its own trait value, but not the trait values of other lineages. The corresponding A matrix is diagonal, as would be the case for any model with this property. Finally, the last model (PM) is a model in which trait evolution along a lineage is influenced by its own trait value and the trait values of other lineages, such that A has nonnegative off-diagonal values. A remarkable property of A under this model is that all its off-diagonal values are identical. This is explained by the fact that the PM model is a neutral model, in the sense that the effect $A^{(k,l)}$ of lineage l on lineage k is the same for all lineages $k \neq l$. All other models in which the off-diagonal elements of A are identical would have this same property, known in probability theory as exchangeability.

Several variations around these models can still be embedded in our general framework: (i) Models in which the rate of phenotypic evolution depends on a variable $Y(t)$ that itself varies through time (see e.g., global temperature $T(t)$ in Clavel J., Morlon H., unpublished data) can be formalized similarly to ACDC, with time t replaced by $Y(t)$. (ii) Models accounting for the biogeographic background in which species coevolved (e.g., all the “+GEO” models in Drury et al. 2016) can be incorporated in our framework through the design of the A matrix when ancestral geographic ranges are known or reconstructed (see details in Online Appendix C.3, available on Dryad). (iii) Considering subclades in

TABLE 1. Examples of classical models of trait evolution fitting our framework

Key	Model name Evolution along lineage k	a	A	Γ
BM	Brownian motion, random genetic drift $dX_t^{(k)} = \sigma dW_t^{(k)}$	0	0	σI
ACDC	Accelerating or decelerating rate, early burst $dX_t^{(k)} = \sigma_0 e^{rt} dW_t^{(k)}$	0	0	$\sigma_0 e^{rt} I$
DD	Diversity-dependent $dX_t^{(k)} = \sigma_0 e^{r n_t} dW_t^{(k)}$	0	0	$\sigma_0 e^{r n_t} I$
OU	Ornstein–Uhlenbeck, stabilizing selection $dX_t^{(k)} = \psi(\theta - X_t^{(k)})dt + \sigma dW_t^{(k)}$	$\psi\theta V$	ψI	σI
PM	Phenotype matching $dX_t^{(k)} = \psi(\theta - X_t^{(k)})dt + S\left(\frac{1}{n_t} \sum_{l=1}^{n_t} X_t^{(l)} - X_t^{(k)}\right)dt + \sigma dW_t^{(k)}$	$\psi\theta V$	$(\psi + S)I - \frac{S}{n_t} U$	σI

Notes: The unity vector (vector full of 1) is denoted by V , I refers to the identity matrix (diagonal matrix with diagonal values equal to 1), and U refers to the unity matrix (matrix full of 1). Their size is the same as the size of the vector of traits X_t considered. Parameters are σ : rate of neutral phenotypic evolution; ψ : strength of stabilizing selection; θ : optimal phenotype; S : strength of between-lineage competition driving individual phenotypes away from clade-wise average phenotype; σ_0 : rate of phenotypic evolution at the root of the tree; r : parameter controlling the exponential rise or decay of the rate of phenotypic evolution with time (ACDC) or with the number of lineages (DD). Considering nonultrametric trees including fossils amounts to replacing vector V and matrices I and U by their homologs V_{alive} , I_{alive} and U_{alive} , where the subscript specifies that the vector and matrices have 0 on lines and columns corresponding to lineages that are extinct in the given epoch.

which trait evolution follows distinct models or similar models with distinct parameter values (as in [Butler and King 2004](#)) is also straightforward. One just needs to specify distinct parameters in a , A , and Γ on the lines and columns corresponding to lineages in the distinctive subclade. (iv) Multivariate trait evolution models, in which several distinct traits evolve in a correlated manner ([Hansen et al. 2008](#); [Bartoszek et al. 2012](#)) are easily written in our framework, as shown with some examples in Online Appendix B.2, available on Dryad. In multivariate models with lineages evolving independently from one another (e.g., multivariate combinations of BM, ACDC, DD, and OU models), A , and Γ are block diagonal matrices, with blocks of size the number of traits, each of them describing correlated multivariate evolution along a particular lineage. In this case, trait–trait correlations introduced through the A matrix correspond, as in [Bartoszek et al. \(2012\)](#), to the case when a given trait on a lineage is attracted to (or repulsed from) a linear combination of other traits in this lineage. (v) Finally, accounting for observation errors when available only requires to adjust the variance–covariance matrix as described in [Hansen and Bartoszek \(2012\)](#).

By considering previous models under this light, it becomes very clear that the set of models that have been considered so far represents a very small fraction of all the models that could potentially be considered. In particular, the A matrix, which dictates how the value of a given trait influences the evolution of other traits—either different traits in the same lineage, or the same trait in other lineages, or yet different traits in other lineages—has so far been very constrained. It has been considered to be zero (BM, ACDC, DD), diagonal (OU), block diagonal (multivariate), and only recently with nonzero off-diagonal values (PM). Relaxing these constraints means that a much broader array of models incorporating the effect of interspecific interactions on phenotypic evolution can be considered. In particular, lineages do not need to be interchangeable. Evolution in complex networks of interactions can be considered by designing *a priori* the A matrix according to the known network. The effect of clade–clade interactions can be modeled by filling the A matrix with nonzero entries $A^{(k,l)}$ with k and l corresponding to lineages from different clades. For example, under a scenario of two clades coevolving with no effect of within-clade interactions, this leads to a A matrix with two off-diagonal blocks.

We can thus imagine a variety of coevolutionary scenarios, the only major constraint being that the effect of a trait value on the evolution of other traits is assumed to be linear [Equations (2) and (3)]. Given a scenario, we can write the corresponding evolution of each trait on a given lineage through each epoch [Equation (3)], and deduce the collection of (a_i) , (A_i) , and (Γ_i) defining the evolutionary process [Equation (2)]. Below, we first show how to derive the probabilistic distribution of traits at tip branches for any model that

can be written under this framework before illustrating the approach with a particular model of clade–clade interaction.

DISTRIBUTION OF TIP TRAIT VALUES

The Distribution of Traits is Gaussian

Deriving the probabilistic distribution of traits at tip branches is key to our ability to fit phenotypic models to comparative data using maximum likelihood or Bayesian approaches. It also provides a very efficient way to simulate tip values for specific models, by drawing from the expected tip distribution.

When X_0 has a Gaussian distribution (including the particular case when X_0 is constant) the linear equations considered in the framework ensure that X_t remains a Gaussian vector at each time t (see details in Appendix section “The Distribution is Gaussian”). The trait vector X_t , of size $n_t d$, is thus uniquely defined by its expectation vector m_t and covariance matrix Σ_t , and has the following density:

$$\forall x \in \mathbb{R}^{n_t d}, f(x) = \frac{1}{\sqrt{(2\pi)^{n_t d} \det(\Sigma_t)}} e^{-\frac{1}{2} tr(x - m_t) \Sigma_t^{-1} (x - m_t)}$$

In particular, the distribution of tip trait values at present time T is Gaussian with expectation vector m_T and covariance matrix Σ_T . We can compute m_T and Σ_T iteratively: starting with initial conditions m_0 and Σ_0 for $X_{\tau_0} = X_0$, we compute, until reaching the present:

1. $m_{\tau_{i+1}}^-$ and $\Sigma_{\tau_{i+1}}^-$ at the end of each epoch i
2. $m_{\tau_{i+1}}$ and $\Sigma_{\tau_{i+1}}$ at the branching time τ_{i+1}

Evolution of the Distribution Through Each Epoch

Knowing the expectation vector and covariance matrix $(m_{\tau_i}, \Sigma_{\tau_i})$ at the beginning of epoch i , we show (see Appendix section “Evolution through time with close formula”) that $m_{\tau_{i+1}}^-$ and $\Sigma_{\tau_{i+1}}^-$ at the end of epoch i are given by the following analytical expressions:

$$m_{\tau_{i+1}}^- = e^{(\tau_i - \tau_{i+1}) A_i} m_{\tau_i} + \int_{\tau_i}^{\tau_{i+1}} e^{(s - \tau_{i+1}) A_i} a_i(s) ds \quad (4a)$$

$$\begin{aligned} \Sigma_{\tau_{i+1}}^- &= \left(e^{(\tau_i - \tau_{i+1}) A_i} \right) \Sigma_{\tau_i}^{tr} \left(e^{(\tau_i - \tau_{i+1}) A_i} \right) \\ &+ \int_{\tau_i}^{\tau_{i+1}} \left(e^{(s - \tau_{i+1}) A_i} \Gamma_i(s) \right)^{tr} \left(e^{(s - \tau_{i+1}) A_i} \Gamma_i(s) \right) ds \end{aligned} \quad (4b)$$

Alternatively, we can write the evolution of m and Σ on epoch i as a set of ordinary differential equations (ODEs), and integrate these ODEs numerically, with initial conditions given by $(m_{\tau_i}$ and $\Sigma_{\tau_i})$. On each epoch i , each component k of the expectation vector evolves

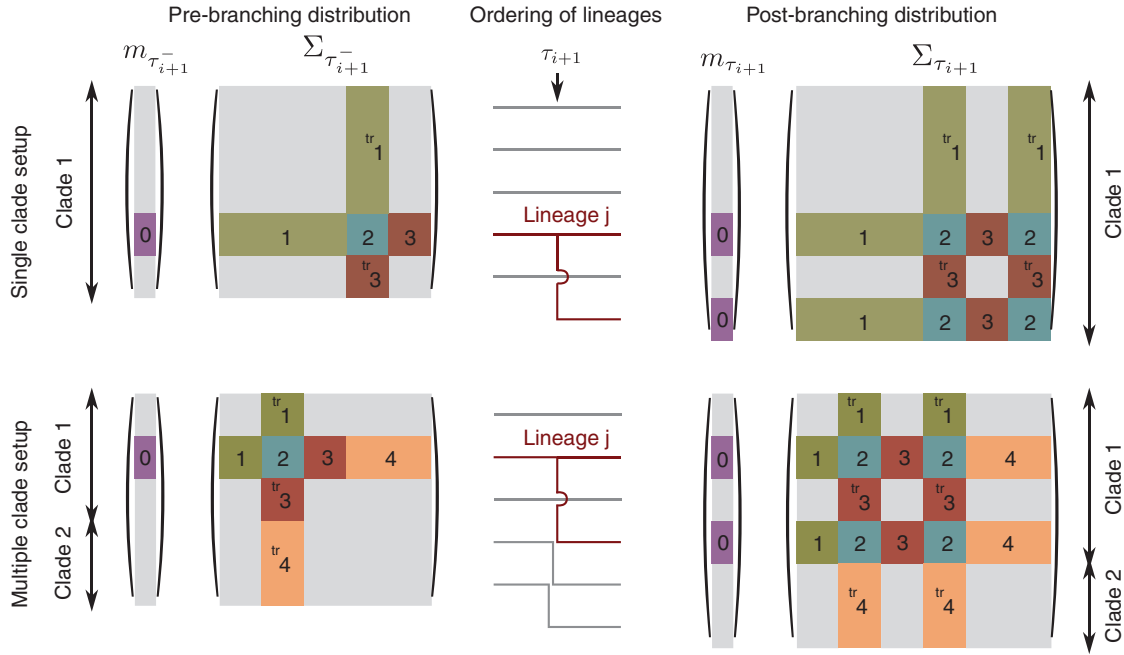


FIGURE 4. Update step for the expectation vector and covariance matrix at branching times when there is one (top row) or two (bottom row) clades. The middle panel highlights the branching lineage j , as well as the ordering of lineages before and after the branching event. The vector $m_{\tau_{i+1}}$ and matrix $\Sigma_{\tau_{i+1}}$ (displayed on the right) are constructed by augmenting $m_{\tau_{i+1}^-}$ and $\Sigma_{\tau_{i+1}^-}$ (displayed on the left) with copies of blocks corresponding to lineage j (materialized by numbers).

according to Equation (5a) and each component (k, l) of the covariance matrix evolves according to Equation (5b) (see derivation in Appendix section “Evolution through time with ODE system”):

$$\frac{d}{dt} m_t^{(k)} = a_i^{(k)}(t) - \sum_{m=1}^{n_t d} A_i^{(k, m)} m_t^{(m)} \quad (5a)$$

$$\begin{aligned} \frac{d}{dt} \Sigma_t^{(k, l)} = & - \sum_{m=1}^{n_t d} A_i^{(k, m)} \Sigma_t^{(m, l)} + A_i^{(l, m)} \Sigma_t^{(m, k)} \\ & - \Gamma_i^{(l, m)}(t) \Gamma_i^{(k, m)}(t) \end{aligned} \quad (5b)$$

Equations (4a, 4b) and the ODE system described by Equations (5a, 5b) are mathematically equivalent. The first formulation is more computationally efficient when the integrals can be simplified analytically. For example, when A is symmetric Equations (4a, 4b) can be simplified (Appendix section “For models considering a constant, A symmetric, and $\Gamma = \sigma I$ ”) and computed very efficiently. This first formulation also reveals that if Σ_{τ_i} is positive definite, then $\Sigma_{\tau_{i+1}}$ remains positive definite (and thus invertible) even when Γ_i is not. Inverting Σ is important for computation of the Gaussian distribution. The second formulation provides a more intuitive interpretation of the components that influence the evolution of trait distribution, and is easily implementable for any model.

Evolution of the Distribution at Branching Times

Knowing the expectation vector and covariance matrix $(m_{\tau_{i+1}^-}, \Sigma_{\tau_{i+1}^-})$ at the end of epoch i , which precedes the branching of a given lineage j , we build $m_{\tau_{i+1}}$ and $\Sigma_{\tau_{i+1}}$ at the branching event, as illustrated in Figure 4.

Recall that in the case of evolution within a single clade, $X_{\tau_{i+1}}$ is obtained by concatenating the d trait values of lineage j at time τ_{i+1}^- at the end of $X_{\tau_{i+1}^-}$. The d new components in $X_{\tau_{i+1}}$ are thus the exact copies of the trait values of lineage j , and have the same expectation and covariance matrix. Hence, the expectation vector $m_{\tau_{i+1}}$ is simply obtained by concatenating $m_{\tau_{i+1}^-}$ with the d components of $m_{\tau_{i+1}^-}$ corresponding to lineage j .

The covariance matrix $\Sigma_{\tau_{i+1}}$ is obtained as follows: the covariance matrix corresponding to the previously existing lineages is unchanged, given by $\Sigma_{\tau_{i+1}^-}$; to this main block, we add below a copy of the d lines corresponding to the covariances between the d traits in lineage j and all the other traits, and we add to the right the same components arranged in d columns; finally, we fill the last missing block in the bottom right corner of $\Sigma_{\tau_{i+1}}$ with the block corresponding to the covariance matrix among the d traits in lineage j (i.e., the $d \times d$ diagonal block of $\Sigma_{\tau_{i+1}^-}$ starting from line $(j-1)d+1$).

In the case of evolution in multiple clades, $m_{\tau_{i+1}}$ and $\Sigma_{\tau_{i+1}}$ are constructed following a similar procedure, by augmenting $m_{\tau_{i+1}^-}$ and $\Sigma_{\tau_{i+1}^-}$ with copies of blocks corresponding to lineage j , inserted at the appropriate location. We illustrate this update step in Figure 4.

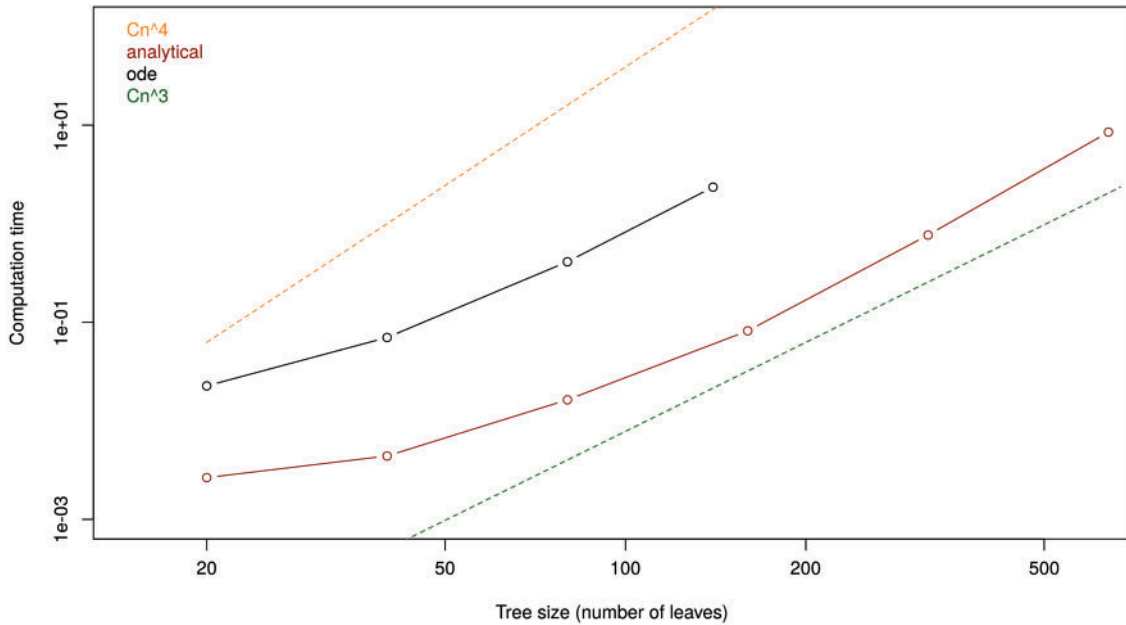


FIGURE 5. Time (in seconds) needed to compute the distribution of tip data following the PM model with parameters $(m_0, v_0, \theta, \psi, S, \sigma) = (0, 0, 1, 0.1, 1, 2)$. Trees are simulated under a pure-birth model conditioned on having a given number of leaves. Bottom dots: ‘analytical’ implementation; Top dots: ‘ode’ implementation. The top dashed curve represents the slope of time increase as a power 4 of the number of leaves while the bottom dashed curve represents the slope of time increase as a power 3 of the number of leaves.

Tip Trait Distribution for Particular Models

Applying this general iterative procedure along a phylogenetic tree provides closed analytic tip distribution formulae for a wide variety of models. In Appendix section “Deriving the analytical tip distribution for known models”, we re-derive known tip distributions for models without lineage–lineage interaction, thus providing a unified review of mathematical results associated to these models. Tip distributions for classical univariate models (BM, ACDC, DD, OU) on ultrametric and non-ultrametric trees are summarized in Appendix Table A1. We confirm, as has been shown before (Uyeda et al. 2015), that the OU and AC models have identical tip distributions on ultrametric trees. We also re-derive results that can be found in Bartoszek et al. (2012) providing tip distributions for multivariate models.

Analytical formulae of tip distributions for models with lineage–lineage interactions have not yet been proposed. Drury et al. (2016) developed the inference tools that allow fitting the PM model, using the ODE system given in Equations (5a, 5b) (thereafter referred to as “ode” method). Here, we develop the inference tools based on analytical reduction of Equations (4a, 4b), (thereafter referred to as “analytical” method, see Appendix section “For the PM model”), and compare the efficiencies of the two methods. Specifically, we simulated 50 pure-birth Yule trees with a per-lineage speciation rate of 1 per time unit, conditioned to having a given number of tips at present, using the “phytools” R package (Revell 2012). We then computed the tip distribution corresponding to the PM model with

parameters fixed at $(m_0, v_0, \theta, \psi, S, \sigma) = (0, 0, 1, 0.1, 1, 2)$ using both the analytical and the ode methods. The new analytical method is much more efficient than the previous ode method (Fig. 5). While we were previously limited to fitting the PM model to trees of less than 150 tips due to memory issues, the analytical method allows fitting trees with up to 600 tips on a desktop computer.

Drury et al. (2016) also proposed an extension of the PM model accounting for the biogeographic history of lineages. In the case when each lineage is present in at most one location, the “analytical” method can be extended, providing fast likelihood computation (see Online Appendix C.3 available on Dryad). When there are lineages occurring in more than one location at the same time, we need to resolve numerically the ODE system in order to compute the likelihood of tip traits. While this is more time consuming than finding a good “analytical” reduction, the new implementation is more efficient than the one we previously proposed (Drury et al. 2016).

MODELING TRAIT EVOLUTION ON COEVOLVING CLADES

We illustrate how our framework can be used to study trait coevolution in scenarios of clade–clade interactions. We consider a simple model with two interacting clades (numbered 1 and 2), in which a given trait in clade 1 coevolves with another given trait in clade 2. Following the approach introduced above, we define X_t the vector of trait values containing first the trait values for clade 1, and then the trait values for clade 2, and we write a

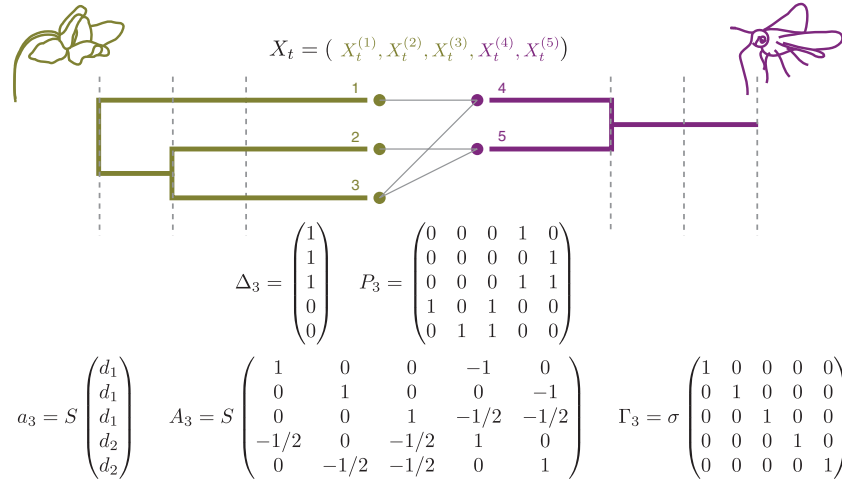


FIGURE 6. Hypothetical clade-clade coevolutionary scenario. Vertical dashed lines delimitate the successive epochs. The vector X_t contains the trait values on the third (last) epoch, P_3 is the matrix of network interactions, and a_3 , A_3 and Γ_3 together define trait evolution according to the clade-clade matching model defined in Equations (6) and (7).

stochastic differential equation specifying how each trait value evolves along a single lineage k . In the spirit of the PM model (Nuismer and Harmon 2014), we propose here a formulation in which the trait of lineage k is attracted to (or repelled from, depending on the sign of S) the average trait value of the lineages it interacts with, plus (or minus) a shift:

$$dX_t^{(k)} = S \left(\delta_k d_1 + (1 - \delta_k) d_2 + \frac{1}{n_k} \sum_{l=1}^n p_{k,l} X_t^{(l)} - X_t^{(k)} \right) dt + \sigma dW_t^{(k)} \quad (6)$$

where S represents the attracting or repelling strength of species interactions on trait evolution, d_1 represents the shift for lineages from clade 1, d_2 the shift for lineages from clade 2, σ is the drift parameter, δ_k equals one if lineage k belongs to clade 1 and zero if it belongs to clade 2, $p_{k,l}$ equals one if lineages k and l interact and zero otherwise, $n_k = \sum_l p_{k,l}$ is the number of lineages interacting with lineage k , and n is the total number of lineages.

When S is positive, the trait value of lineage k is attracted to an optimal trait value given by the average trait value of the interacting species (plus a shift d_1 or d_2).

An example of such a scenario of clade-clade matching mutualism is the coevolution between the length of floral tubes and the length of butterfly proboscis in a plant-pollinator mutualistic network (illustrated in Fig. 6). In this example, we assume that the optimal length of a butterfly proboscis is the average length of the plant floral tubes it pollinates plus a shift d_2 , while the optimal length of a plant floral tube is the average proboscis length of its butterfly pollinators plus a shift d_1 . With $d_1 + d_2 = 0$, both traits can reach their optimal state, leading to a stable situation with butterfly proboscis a bit longer (if $d_1 > 0$) or shorter (if $d_1 < 0$) than plant floral tubes. With $d_1 + d_2 \neq 0$, traits

cannot reach their optimal state, resulting in a runaway process where both traits tend to evolve toward an ever-moving optimum. For example, with positive d_1 and d_2 , the butterflies proboscis tends to get longer to better access the nectar, while the floral tube also tends to get longer to force the butterfly's body to touch the stamen. The parameters S and σ control respectively the strength of the interaction effect and the rate of stochastic phenotypic change. The bigger S , the closer the traits will track the optimum; the bigger σ , the bigger the fluctuations around this optimum.

When S is negative, the traits are repelled from the average trait value of the interacting species (plus a shift d_1 or d_2). This may capture natural situations of clade-clade competition driving trait displacement. Finally, some antagonistic interactions between traits could require to introduce two parameters $S_1 > 0$ and $S_2 < 0$ to capture match-versus-escape scenarios. For example, parasites might tend to develop cues matching those of their hosts whereas hosts develop cues to escape their parasites in a coevolutionary arms race.

From Equation (6) we deduce the corresponding a , A and Γ through each epoch:

$$\begin{aligned} a &= S(\Delta d_1 + (V - \Delta) d_2) \\ A_{k,l} &= S \left(\mathbb{1}_{k=l} - \frac{p_{k,l}}{n_k} \right) \\ \Gamma &= \sigma I \end{aligned} \quad (7)$$

where Δ is the vector of elements δ_k (see Fig. 6 for an illustration). Matrix A is in general not symmetric anymore, as all species k do not have the same number n_k of species that they interact with.

As shown by Equation (7), entirely defining a model of clade-clade coevolution requires introducing a constant network of interaction during each epoch (the P matrix with elements $p_{k,l}$). We can potentially re-define epochs to account for events of change in the interaction

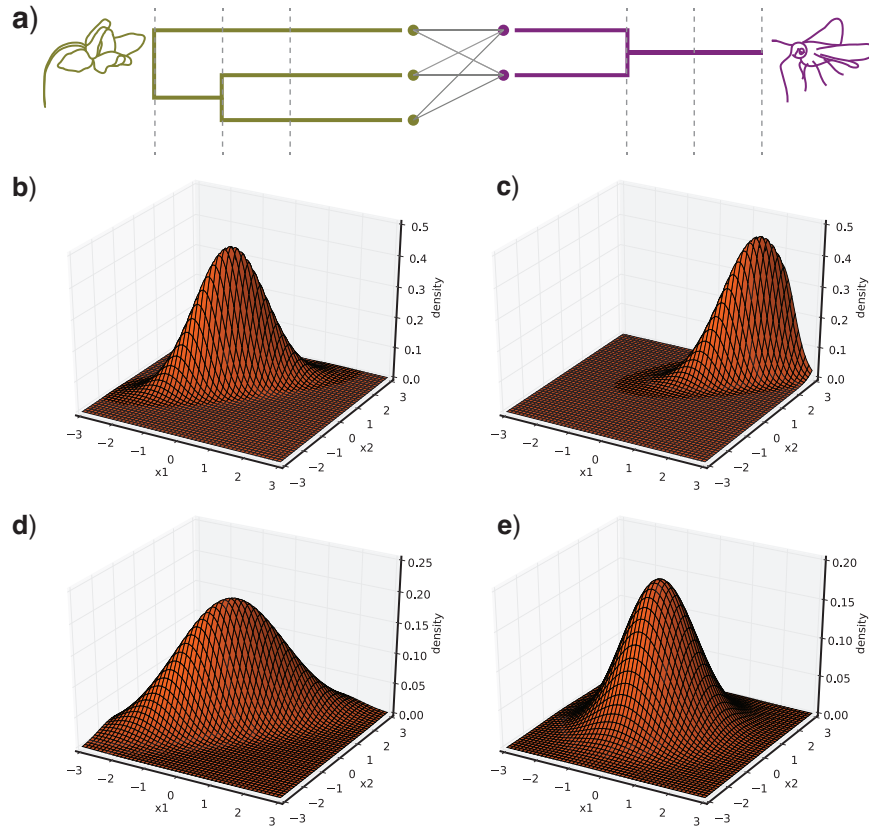


FIGURE 7. Trait evolution under the Generalist Matching Mutualism (GMM) model. a) an illustrative generalist network of interactions between two clades. Vertical dashed lines delimitate the successive epochs. b) c) d) e) Expected tip distribution for the average trait value in each clade, with parameter values $(S, d_1, d_2, \sigma) =$ b) $(2, -1, 1, 1)$, c) $(2, 0, 2, 1)$, d) $(2, -1, 1, 1.5)$, e) $(0.2, -1, 1, 1)$.

network in addition to speciation and extinction events, thus allowing interaction networks to evolve along branches. In practice, we typically (at best) have access to the current interaction network (Fig. 6), but not the ancestral networks. A solution to this would consist in treating the ancestral P matrices as parameters of the model, and searching the ancestral network(s) that maximize the fit to the data. Another approach would consist in reconstructing ancestral networks over each period according to rules regarding the inheritance of interactions at speciation times. Developing these approaches is outside the scope of the current study, but we have shown how to compute tip trait distributions once they are developed.

We illustrate the computation of tip trait distributions for a model in which the ancestral networks are known: a generalist model where all species from clade 1 interact with all species from clade 2. We consider a ‘Generalist Matching Mutualism’ model of trait evolution (thereafter referred to as GMM, and illustrated in Fig. 7a), which is captured by Equation (6) with S positive and $p_{k,l}=1$ for any two lineages k and l from different clades and $p_{k,l}=0$ for any two lineages k and l from the same clade. Given that the model fits within our framework, we know that the trait distribution at the tips is Gaussian, and we can compute the expectation vector and covariance matrix corresponding to the model using

Equations (4a, 4b), which we can reduce for this specific model in order to speed up the computation (Appendix “For the GMM model”).

The tip distribution is relatively fast to compute (e.g., in the order of 0.8 seconds with two 100-tip trees on a desktop computer), such that fitting the model by maximum likelihood or in a Bayesian framework should not be problematic for trees with a few hundred tips. However, we do not aim here to carry an in-depth study of this particular model, nor to fit it to empirical data. Rather, we use our ability to rapidly compute tip trait distribution to get a first glimpse of the model behavior under distinct sets of parameter values.

In Figure 7 (b,c,d,e), we plotted the distribution of the average \bar{X}^1 of trait values in clade 1 and the average \bar{X}^2 of trait values in clade 2 for traits evolving under the GMM model with four parameter sets chosen to lead to four distinct qualitative behaviours. From Equation (6), we can easily show that under GMM $\bar{X}^1 + \bar{X}^2$ is a drifted BM with drift term $S(d_1 + d_2)$ and $\bar{X}^1 - \bar{X}^2$ is an OU process with optimum $(d_1 - d_2)/2$ and selection strength S . The shift parameters d_1 and d_2 thus directly determine the position of the optimum of the distribution. An “equilibrium” scenario corresponds to $d_1 = -d_2$: the more likely values for the two average traits \bar{X}^1 and \bar{X}^2 are such that $\bar{X}^1 = \bar{X}^2 + d_1$ (see Fig. 7b). In contrast,

when $d_1 \neq -d_2$, the two communities have optimal trait values that are noncompatible, and the traits will tend to increase if $d_1 + d_2 > 0$ and decrease if $d_1 + d_2 < 0$ (see Fig. 7c) in a “runaway” process. In this case, the position of the peak in the tip distribution will also depend on the depth of the root, the trait values at the root, and the value of the parameter S . The parameter S plays an important role in the hump thickness: the bigger S , the more constrained $\bar{X}^1 - \bar{X}^2$ around $(d_1 - d_2)/2$ (see Fig. 7e). The parameter σ also plays a role in the thickness of the hump, but in the orthogonal direction: increasing σ flattens the distribution by allowing different \bar{X}^1 and \bar{X}^2 values while retaining the constraint on $\bar{X}^1 - \bar{X}^2$ (see Fig. 7d). In future work, it would be interesting to assess whether these results can be used to build a statistical test for distinguishing between the runaway and equilibrium scenarios.

Implementation

Our framework is implemented in the R package “RPANDA” (Morlon et al. 2015), including functions to compute tip distributions, to simulate trait evolutionary trajectories using the Euler–Maruyama scheme (see Online Appendix D.1 available on Dryad), and to simulate tip data by drawing from the expected tip distribution. We also implemented optimization functions to infer model parameters by maximum likelihood, and to compare the fit of distinct models using information criteria. In the most general user-defined use of our framework, the input is one or several potentially nonultrametric phylogenetic tree(s) and the collection of (a_i, A_i, Γ_i) matrices during each epoch that define a specific model. In this case, the tip distribution is computed using the most general “ode” method that solves numerically the ODEs. In addition, we implemented all models mentioned in Table 1 as well as GMM with the fastest described algorithm to compute their tip distribution. In Online Appendix E available on Dryad, we provide a tutorial explaining the structure of our code and illustrating how to use it. We however recommend potential users to thoroughly test the statistical properties of the models they design (parameter estimation, type I and II error rates) using simulations before applying them to empirical data. These properties are model dependent and thus should be assessed case by case. We are in the process of performing these tests for the GMM model.

DISCUSSION

We developed a modeling framework for traits coevolving in coevolving lineages and clades. We highlighted that under a wide variety of models where the evolution of a given trait on a given lineage is linearly related to its own value and the value of other traits on the same lineage, of the same trait on other lineages, and/or of other traits on other lineages, the

expected tip trait distribution is Gaussian. We showed how to compute this tip distribution in general, as well as for specific models, including classical models of phenotypic evolution and new models of clade–clade coevolution.

Many classical models of phenotypic evolution, such as univariate and multivariate BM, OU, ACDC, and DD fit within our framework. They correspond to the situation where the evolution of traits on a given lineage is independent of trait values on other lineages. For these models, we already know that the tip trait distribution is Gaussian. However, finding the relevant computation of the expectation vectors and covariance matrices associated with each model in the dense literature of comparative phylogenetics can be overwhelming for neophytes. Our Appendix section “Deriving the analytical tip distribution for known models” unifies these computations under a common formalism, providing both the expressions for the various existing models and their mathematical underpinning. This is done in the context of trees that are not necessarily ultrametric, meaning that all models can be applied to phylogenies including fossils. We hope that this Appendix can serve as a useful review for navigating phylogenetic approaches for understanding trait evolution.

The fact that the distribution of traits remains Gaussian when traits from different lineages coevolve is a convenient result, because it means that computing the tip distribution only requires computing the expectation vector and covariance matrix associated with the different models. For example, we used this result in Drury et al. (2016) to compute tip trait distributions for the PM model (Nuismer and Harmon 2014) and fit it to comparative data by maximum likelihood. Here we vastly extend the set of potential coevolutionary models for which tip trait distributions can be computed and provide two general approaches for computing the expectation vector and covariance matrix. One of these two approaches (the “ode” approach) consists in numerically integrating a set of ODEs. This is the approach that was used in Drury et al. (2016). The other approach (the “analytical” approach) involves computing integrals and is more efficient when these integrals can be analytically reduced, which depends on the form of the model. Applying the “analytical” approach to the PM model, we greatly improved its computational efficiency.

We provide a framework for computing tip trait distributions for a wide class of models accounting for within-clade and clade–clade interactions. We hope that this flexibility will foster the development and study of various models adapted to the specificities of particular scientific questions and biological systems. We did not study at length a particular coevolutionary model in this article, but the PM model was thoroughly studied elsewhere (Drury et al. 2016). The GMM model that we introduce here can be seen as a clade–clade analogue to the PM model (Nuismer and Harmon 2014). Both models are “generalist” in the sense that all lineages are assumed

to interact (within-clade in the case of PM and between clades in the case of GMM). This assumption can be relaxed by incorporating additional information. In our biogeographic models, for example, lineages can only interact if they are sympatric (Drury et al. 2016). More generally, any information or hypothesis concerning the network of interactions between lineages can be accounted for into the A matrices.

There are two main limitations to the modeling framework presented here. The first one is that trait evolution is always assumed to respond linearly to trait values in other lineages. Thus, nonlinear effects such as a stronger selection for divergence when phenotypes are similar cannot be accounted for. Nuismer and Harmon (2014) originally developed an individual-based model accounting for such nonlinear effects, and then linearized the effects to derive the model of phenotypic evolution emerging at the lineage level (their equation S38). This linear expression inspired the development of the present framework, and assures that the tip trait distribution is Gaussian. It would however be particularly interesting to model nonlinear effects. The second limitation is the issue of model and parameter identifiability, in particular in the absence of fossils. A Gaussian distribution in \mathbb{R}^{nd} can potentially allow identifying several models and parameters, but there are distinct combinations for which a similar (or even identical) distribution is expected. For example, we already know that parameters of the OU model are non identifiable on phylogenies with only extant species (Ho and Ané 2014) and that OU and AC have identical tip distributions on ultrametric trees (Uyeda et al. 2015). Thus, while we wrote our framework in all generality, with a , A , and Γ encompassing as many parameters as desired, and parameters that potentially vary between epochs, it is clear that simplifying assumptions need to be made in order to reduce this parameter space. Identifiability cannot always be checked analytically, as in the case of the OU and AC models. In addition, there can be differences between theoretical and *de facto* identifiability, with models that are identifiable in theory but are difficult to identify in practice. For example, we can show analytically that ψ and S from the PM model are theoretically identifiable, but in practice in most cases only $\psi + S$ can be estimated with precision. Also, *de facto* identifiability depends on the data available, such as the size and shape of a particular phylogeny, and whether it includes fossils or not (Slater et al. 2012). Furthermore, models taking into account interactions among lineages will have to assess the influence of extinct lineages in the past. This has been studied in Drury et al. (2016) for the PM model, by simulating trait evolution on trees including dead branches, before fitting the model on the reconstructed tree only. Our recommendation is to check identifiability on a case-by-case basis, by fitting the set of models under consideration to trait data sets simulated directly on the specific empirical phylogenies in hands. We provide the tools for rapidly simulating tip values under various models by sampling expected distributions.

One of the most challenging and exciting developments that we see ahead is to move from generalist models to models that account for specific interaction networks. We show in this article how to compute tip trait distributions for such models, assuming that the ancestral networks are known. While some fossil species interaction networks have been compiled (Dunne et al. 2008), such data is typically not available. Thus, if we are to really understand if and how species interactions affected long-term phenotypic evolution, we need to start developing models for reconstructing ancestral networks, analogous to the use of ancestral biogeographic models [see Ronquist and Sanmartín (2011) for a review] to incorporate biogeography into models of phenotypic evolution (see e.g., DD+GEO or MC+GEO, in Drury et al. 2016). Interestingly, our modeling framework could provide an approach to do so, informed by species phylogenies, the interaction network of present day species, and current species phenotypes. Indeed, rather than assuming that the ancestral networks are known, we could treat them as additional parameters to optimize upon, and find the ancestral networks that maximize the likelihood of the current data. These approaches have not been experimented yet and are not part of the available code in RPANDA. Whether there will be enough information in the data to distinguish the probability of alternative ancestral networks remains to be tested, but the observed phylogenetic signal in empirical networks of interactions is encouraging (Ives and Godfray 2006; Rafferty and Ives 2013; Hadfield et al. 2014; Hayward and Horton 2014; Martín González et al. 2015). Our ability to distinguish the probability of alternative ancestral networks will be increased by proposing various scenarios regarding the inheritance of interactions at speciation times, such as scenarios in which daughter species interact with many or few of the species that interacted with their mother lineage. These upcoming developments can draw upon the existing literature on the cophylogeny problem (Conow et al. 2010), and will certainly have an important role to play in the ongoing effort of understanding the evolution of species interaction networks (Loeuille and Loreau 2005; Martinez 2006; Nuismer et al. 2013).

Our framework for modeling continuous trait evolution on phylogenetic trees includes most previously proposed models and can be used to develop a series of new models of within-clade and clade-clade coevolution. We hope that this will motivate new theoretical and empirical applications aimed at unravelling how species interactions evolve and influence phenotypic evolution over macro-evolutionary timescales.

FUNDING

M.M. acknowledges his Ph.D. funding from the École Normale Supérieure; A.L. acknowledges funding from the Center for Interdisciplinary Research in Biology

(Collège de France); H.M. acknowledges funding from ERC, [grant ERC-CoG PANDA].

ACKNOWLEDGMENTS

We are very grateful to Jonathan Drury, Florence Débarre and Luke Harmon, as well as members of our research teams for helpful discussions and comments on previous versions of the manuscript. We also acknowledge three reviewers for their helpful comments that helped us improve this paper.

SUPPLEMENTARY MATERIAL

Data available from the Dryad Digital Repository: <http://dx.doi.org/10.5061/dryad.52636>.

APPENDIX

Evolution of the Distribution Through Each Epoch

The distribution is Gaussian.—Recall that a vector is Gaussian if all linear combinations of its components follow a normal distribution. We will thus show by induction that all linear combinations of the traits follow a normal distribution.

The process of trait evolution starts either at the stem root with a vector of size d defined by the initial conditions $X_{\tau_0} = {}^{tr}(X_0^1, \dots, X_0^d)$, or at the crown root with a vector of size $2d$ defined by the initial conditions : $X_{\tau_0} = {}^{tr}(X_0^1, \dots, X_0^d, X_0^1, \dots, X_0^d)$, or at any other step, provided the initial conditions are Gaussian by assumption.

Now, assume that X_{τ_i} is a Gaussian vector.

Then, $\forall t \in (\tau_i, \tau_{i+1})$, after integration we have the following closed expression for the value of the process X_t .

$$X_t = e^{-tA_i} \left(e^{\tau_i A_i} X_{\tau_i} + \int_{\tau_i}^t e^{sA_i} a_i(s) ds + \int_{\tau_i}^t e^{sA_i} \Gamma_i(s) dW_s \right) \quad (A1)$$

Moreover, we have, for any deterministic function Φ (Gardiner et al. 1985),

$$\int_{t_n}^t \Phi_s dW_s \sim \mathcal{N} \left(0, \int_{t_n}^t \Phi_s {}^{tr} \Phi_s ds \right)$$

Hence, X_t is a linear combination of Gaussian vectors, which makes it a Gaussian vector.

Last, suppose that at time τ_{i+1} , the j th branch splits, in which case the vector grows. All linear combinations of the components of X_t at time τ_{i+1}^- have a normal distribution. And the d additional components added at time τ_{i+1} belong to the components at time τ_{i+1}^- . It

follows that all linear combinations of the new vector still have a normal distribution.

Evolution through time with close formula.—Still assuming that we know the (Gaussian) distribution of X_{τ_i} at the beginning of an epoch (τ_i, τ_{i+1}) , a few more lines allow us to provide a closed formula for the distribution of X_t at all time $t \in (\tau_i, \tau_{i+1})$. Indeed, using Equation (A1), and the fact that, if X and Y are two independent Gaussian vectors with expectation vectors respectively m_X and m_Y and covariance matrices respectively Σ_X and Σ_Y , then:

$$DX + d \sim \mathcal{N}(Dm_X + d, D\Sigma_X {}^{tr} D)$$

$$X + Y \sim \mathcal{N}(m_X + m_Y, \Sigma_X + \Sigma_Y)$$

It thus follows that, $\forall t \in [\tau_i, \tau_{i+1}]$,

$$m_t = e^{(\tau_i - t)A_i} m_{\tau_i} + \int_{\tau_i}^t e^{(s-t)A_i} a_i(s) ds \quad (4a)$$

$$\Sigma_t = \left(e^{(\tau_i - t)A_i} \right) \Sigma_{\tau_i} {}^{tr} \left(e^{(\tau_i - t)A_i} \right) + \int_{\tau_i}^t \left(e^{(s-t)A_i} \Gamma_i(s) \right) {}^{tr} \left(e^{(s-t)A_i} \Gamma_i(s) \right) ds \quad (4b)$$

Applying these equations for $t = \tau_{i+1}$ thus gives the distribution of the trait vector at time τ_{i+1} , which is the result stated in Equations (4a, 4b) in the main text.

Remark that, unless one of the very first branches immediately dies at the beginning of the process at a fixed initial condition, the density of the tip distribution has support in \mathbb{R}^{nd} . One can check that Σ_t stays positive definite (implying that $\det \Sigma_t \neq 0$), even when some Γ_i are not positive definite (except the first one).

Evolution through time with ODE system.—The expectation and covariance formulae provided in Equations (4a, 4b) require to deal with an integral which is not always straightforward to compute. Alternatively, one can prefer to take the derivative of this expression, get a set of ODEs verified by the expectation and covariance elements through each epoch, and subsequently integrate the ODE system. We show now another way to derive this set of ODEs.

First, we write the stochastic differential equation on any epoch (τ_i, τ_{i+1}) and for each trait k , which is given in the most general setting by :

$$dX_t^{(k)} = \left(a_i^{(k)}(t) - \sum_{m=1}^{n_i d} A_i^{(k,m)} X_t^{(m)} \right) dt + \sum_{m=1}^{n_i d} \Gamma_i^{(k,m)}(t) dW_t^{(m)}$$

Itô's formula (Gardiner et al. 1985) then gives us:

$$\begin{aligned}
 d(X_t^{(k)} X_t^{(l)}) &= X_t^{(k)} dX_t^{(l)} + X_t^{(l)} dX_t^{(k)} + d\langle X_t^{(k)}, X_t^{(l)} \rangle \\
 &= \left(a_i^{(l)}(t) X_t^{(k)} - \sum_{m=1}^{n_t d} A_i^{(l,m)} X_t^{(m)} X_t^{(k)} \right) dt \\
 &\quad + \sum_{m=1}^{n_t d} \Gamma_i^{(l,m)}(t) X_t^{(k)} dW_t^{(m)} \\
 &\quad + \left(a_i^{(k)}(t) X_t^{(l)} - \sum_{m=1}^{n_t d} A_i^{(k,m)} X_t^{(m)} X_t^{(l)} \right) dt \\
 &\quad + \sum_{m=1}^{n_t d} \Gamma_i^{(k,m)}(t) X_t^{(l)} dW_t^{(m)} \\
 &\quad + \sum_{m=1}^{n_t d} \Gamma_i^{(l,m)}(t) \Gamma_i^{(k,m)}(t) dt
 \end{aligned}$$

Taking the expectation, it follows that

$$\begin{aligned}
 \frac{d}{dt} \mathbb{E}(X_t^{(k)} X_t^{(l)}) &= a_i^{(l)}(t) \mathbb{E}(X_t^{(k)}) + a_i^{(k)}(t) \mathbb{E}(X_t^{(l)}) \\
 &\quad - \sum_{m=1}^{n_t d} A_i^{(l,m)} \mathbb{E}(X_t^{(m)} X_t^{(k)}) \\
 &\quad - \sum_{m=1}^{n_t d} A_i^{(k,m)} \mathbb{E}(X_t^{(m)} X_t^{(l)}) \\
 &\quad + \sum_{m=1}^{n_t d} \Gamma_i^{(l,m)}(t) \Gamma_i^{(k,m)}(t)
 \end{aligned}$$

In the same fashion, we get

$$\frac{d}{dt} \mathbb{E}(X_t^{(k)}) = a_i^{(k)}(t) - \sum_{m=1}^{n_t d} A_i^{(k,m)} \mathbb{E}(X_t^{(m)}) \quad (5a)$$

This leads to

$$\begin{aligned}
 \frac{d}{dt} (\mathbb{E}(X_t^{(k)}) \mathbb{E}(X_t^{(l)})) &= \mathbb{E}(X_t^{(l)}) \frac{d}{dt} \mathbb{E}(X_t^{(k)}) + \mathbb{E}(X_t^{(k)}) \frac{d}{dt} \mathbb{E}(X_t^{(l)}) \\
 &= a_i^{(k)}(t) \mathbb{E}(X_t^{(l)}) - \sum_{m=1}^{n_t d} A_i^{(k,m)} \mathbb{E}(X_t^{(m)}) \mathbb{E}(X_t^{(l)}) \\
 &\quad + a_i^{(l)}(t) \mathbb{E}(X_t^{(k)}) - \sum_{m=1}^{n_t d} A_i^{(l,m)} \mathbb{E}(X_t^{(m)}) \mathbb{E}(X_t^{(k)})
 \end{aligned}$$

Putting together these different parts gives us the ODE satisfied by all covariances :

$$\begin{aligned}
 \frac{d}{dt} \text{Cov}(X_t^{(k)}, X_t^{(l)}) &= \frac{d}{dt} (\mathbb{E}(X_t^{(k)} X_t^{(l)}) - \mathbb{E}(X_t^{(k)}) \mathbb{E}(X_t^{(l)})) \\
 &= - \sum_{m=1}^{n_t d} [A_i^{(k,m)} \text{Cov}(X_t^{(m)}, X_t^{(l)}) \\
 &\quad + A_i^{(l,m)} \text{Cov}(X_t^{(m)}, X_t^{(k)}) \\
 &\quad - \Gamma_i^{(l,m)}(t) \Gamma_i^{(k,m)}(t)] \quad (5b)
 \end{aligned}$$

Note that in a vectorial formalism with the expectation vector m and covariance matrix Σ , these sets of ODEs can be written equivalently as follows

$$\frac{dm_t}{dt} = a_i(t) - A_i m_t \quad (A2)$$

$$\frac{d\Sigma_t}{dt} = -A_i \Sigma_t - {}^{tr}\Sigma_t {}^{tr}A_i + \Gamma_i {}^{tr}\Gamma_i \quad (A3)$$

Deriving the Analytical Tip Distribution for Known Models

All previously known results of analytic tip distribution of univariate models fit in our framework. They can be rediscovered by following the same scheme:

1. Reduce Equations (4a, 4b) or (5a, 5b) according to the model.
2. Look for an analytical solution at any time τ_i , by calculating manually the expectations and covariances at $\tau_1, \tau_2, \tau_3, \dots$
3. Prove by induction that the analytical solution holds at any time τ_i .

These steps are explicitly written in the Online Appendix (available on Dryad) for a number of models. We summarize in Table A1 the results for the most famous ones. We call hereafter $t_{k,l}$ the time of the most recent common ancestor to lineages k and l , and $t_{k,k}$ the death time of lineage k , equal to T if it survives until present (see notations in Fig. A1).

Speeding up the Computation of Tip Distribution

For models considering a constant, A symmetric, and $\Gamma = \sigma I$.—Equations (4a, 4b) become:

$$\mathbb{E}(X_t) = e^{(\tau_i - t)A_i} \mathbb{E}(X_{\tau_i}) + \int_{\tau_i}^t e^{(s-t)A_i} a_i(s) ds$$

$$\text{Var}(X_t) = (e^{(\tau_i - t)A_i}) \text{Var}(X_{\tau_i}) {}^{tr}(e^{(\tau_i - t)A_i}) + \sigma^2 \int_{\tau_i}^t e^{2(s-t)A_i} ds$$

If A_i is symmetric with coefficients in \mathbb{R} , it can be diagonalized by orthogonal passage matrices : we can

TABLE A1. Analytic tip distribution for models without interactions between traits or lineages

Code	m_0	Σ_0	$(m_T)^{(k)}$	$(\Sigma_T)^{(k,l)}$
BM	m_0	v_0	$m_0 + bt_{k,k}$	$v_0 + \sigma^2 t_{k,l}$
OU	θ	0	θ	$\frac{\sigma^2}{2\psi} e^{-\psi(t_{k,k}+t_{l,l}-2t_{k,l})} (1 - e^{-2\psi t_{k,l}})$
OU	θ	$\frac{\sigma^2}{2\psi}$	θ	$\frac{\sigma^2}{2\psi} e^{-\psi(t_{k,k}+t_{l,l}-2t_{k,l})}$
ACDC	m_0	v_0	m_0	$v_0 + \frac{\sigma_0^2}{2r} (e^{2rt_{k,l}} - 1)$
DD	m_0	v_0	m_0	$v_0 + \sigma_0^2 \sum_{j=0}^{N-1} e^{2rt_j} (\tau_{j+1} - \tau_j) \mathbb{1}_{t_{k,l} > \tau_j}$

Notes: We recall that $t_{k,l}$ is the absolute time of the most recent common ancestor to lineages k and l , and $t_{k,k}$ is the death time of lineage k , equal to T if it survives until present.

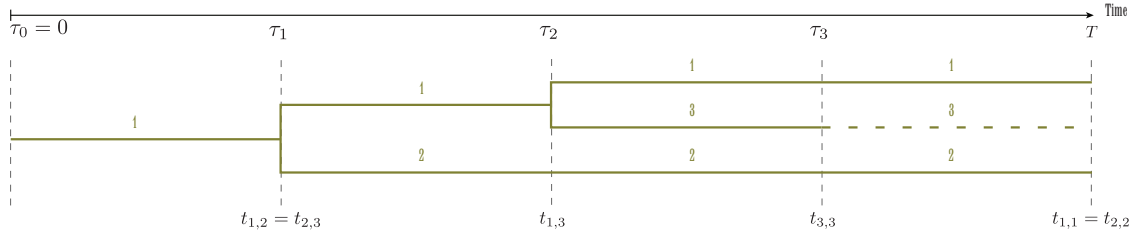


FIGURE A1. Formalism used in analytic formulae presented in Table A1.

exhibit a matrix Q verifying ${}^{tr}QA_iQ = \Lambda_i$ is diagonal and $Q^{-1} = {}^{tr}Q$.

$$\begin{aligned}\mathbb{E}(X_t) &= Qe^{(\tau_i-t)\Lambda_i} {}^{tr}Q \mathbb{E}(X_{\tau_i}) + Q \left(\int_{\tau_i}^t e^{(s-t)\Lambda_i} ds \right) {}^{tr}Q a_i \\ \text{Var}(X_t) &= Qe^{\Lambda_i(\tau_i-t)} {}^{tr}Q \text{Var}(X_{\tau_i}) Qe^{(\tau_i-t)\Lambda_i} {}^{tr}Q \\ &\quad + \sigma^2 Q \left(\int_{\tau_i}^t e^{2(s-t)\Lambda_i} ds \right) {}^{tr}Q\end{aligned}$$

This expression can be used for the numerical integration, in particular, of the PM model.

Note that with A diagonalizable but not symmetric, Equations (4a, 4b) can also be reduced, but the transposition of A is no longer A , and it does not lead exactly to the same expression.

For the PM model.—We consider again the PM model introduced in Nuismer and Harmon (2014). We introduce the line vector u , with value u_j that equals 1 if lineage j is alive, and 0 otherwise. In order to use our framework, we further want to express the model in the form given by Equation (2). This is achieved by taking:

$$\begin{aligned}a_i &= \psi \theta {}^{tr}u \\ A_i &= (\psi + S) \text{diag}(u) - \frac{S}{u {}^{tr}u} {}^{tr}uu \\ \Gamma_i &= \sigma \text{diag}(u)\end{aligned}$$

where $\text{diag}(u)$ is the diagonal matrix with diagonal elements the elements of the vector u .

First, the tip distribution can be computed using the general algorithm that numerically resolves the set of ODEs given in Equations (5a, 5b). Second, the PM model falls within the class of models studied in the previous section, that is, with a symmetric A matrix. The tip distribution can thus be numerically computed faster using this reduction.

We describe here a third (and faster) way to derive the tip distribution. It is based on an analytical reduction of Equations (4a, 4b) that is specific to the PM model.

Remark that $\text{diag}(u)$ and ${}^{tr}uu$ commute, leading to the following calculus,

$$\begin{aligned}e^{(\tau_i-\tau_{i+1})A_i} &= e^{(\tau_i-\tau_{i+1})((\psi+S)\text{diag}(u) - \frac{S}{u {}^{tr}u} {}^{tr}uu)} \\ &= e^{(\tau_i-\tau_{i+1})(\psi+S)\text{diag}(u)} e^{-(\tau_i-\tau_{i+1})\frac{S}{u {}^{tr}u} {}^{tr}uu} \\ &= \text{diag}\left(e^{(\tau_i-\tau_{i+1})(\psi+S)u}\right) \\ &\quad \times \left(\sum_{k \geq 0} \frac{\left(\frac{-(\tau_i-\tau_{i+1})S}{u {}^{tr}u}\right)^k}{k!} ({}^{tr}uu)^k \right)\end{aligned}$$

where e^w is the line vector with elements e^{w_j} . Further, remark that for any $k \geq 1$,

$$\begin{aligned}({}^{tr}uu)^k &= ({}^{tr}uu)({}^{tr}uu)({}^{tr}uu) \dots ({}^{tr}uu) \\ &= {}^{tr}u(u {}^{tr}u)(u {}^{tr}u) \dots (u {}^{tr}u)u \\ &= (u {}^{tr}u)^{k-1} ({}^{tr}uu)\end{aligned}$$

For simplicity, we will write in the following $\Delta = \tau_i - \tau_{i+1}$, leading us to

$$\begin{aligned}
 e^{\Delta A_i} &= \text{diag}\left(e^{(\psi+S)\Delta u}\right) \left(I + \sum_{k \geq 1} \frac{\left(\frac{-S\Delta}{u^{tr u}}\right)^k (u^{tr u})^{k-1} (tr uu)}{k!}\right) \\
 &= \text{diag}\left(e^{(\psi+S)\Delta u}\right) \left(I + \frac{1}{u^{tr u}} \left(\sum_{k \geq 1} \frac{(-\Delta S)^k}{k!}\right) tr uu\right) \\
 &= \text{diag}\left(e^{(\psi+S)\Delta u}\right) \left(I + \frac{1}{u^{tr u}} (e^{-S\Delta} - 1) tr uu\right) \\
 &= \text{diag}\left(e^{(\psi+S)\Delta u}\right) + \frac{1}{u^{tr u}} \text{diag}\left(e^{-S\Delta} e^{(\psi+S)\Delta u}\right) tr uu \\
 &\quad - \frac{1}{u^{tr u}} \text{diag}\left(e^{(\psi+S)\Delta u}\right) tr uu \\
 &= \text{diag}\left(e^{(\psi+S)\Delta u}\right) + \frac{1}{u^{tr u}} (e^{\psi\Delta} - e^{(\psi+S)\Delta}) tr uu \quad (\text{A4})
 \end{aligned}$$

where the last equality is due to the product by $tr uu$, allowing to forget the cases where $u_j = 0$ in the exponential.

We further need to compute

$$\begin{aligned}
 \int_{\tau_i}^{\tau_{i+1}} e^{(s-\tau_{i+1})A_i} a_i ds &= \psi \theta \int_{\tau_i}^{\tau_{i+1}} e^{\psi(s-\tau_{i+1})} ds tr u \\
 &= \theta (1 - e^{\psi\Delta}) tr u \quad (\text{A5})
 \end{aligned}$$

We thus get $m_{\tau_{i+1}}^-$ with the help of Equations (A4) and (A5).

Now, in order to simplify Equation (4b), remark that A_i and Γ_i are symmetric, and so are $e^{\Delta A_i}$ and $e^{\Delta A_i} \Gamma_i$. Moreover, Γ_i is diagonal, and commutes with any other matrix, leading to,

$$\Sigma_{\tau_{i+1}}^- = e^{\Delta A_i} \Sigma_{\tau_i} e^{\Delta A_i} + \int_{\tau_i}^{\tau_{i+1}} e^{2(s-\tau_{i+1})A_i} \Gamma_i \Gamma_i ds$$

The first term can be computed thanks to Equation (A4). For the second one, remark that $tr uu \text{diag}(u) = tr uu$, thus leading to

$$\begin{aligned}
 &\int_{\tau_i}^{\tau_{i+1}} e^{2(s-\tau_{i+1})A_i} \Gamma_i \Gamma_i ds \\
 &= \sigma^2 \int_{\tau_i}^{\tau_{i+1}} e^{2(\psi+S)(s-\tau_{i+1})} ds \text{diag}(u) \\
 &\quad + \frac{\sigma^2}{u^{tr u}} \int_{\tau_i}^{\tau_{i+1}} \left(e^{2\psi(s-\tau_{i+1})} - e^{2(\psi+S)(s-\tau_{i+1})}\right) ds \\
 &\quad \times tr uu \text{diag}(u) \\
 &= \sigma^2 \frac{(1 - e^{2(\psi+S)\Delta})}{2(\psi+S)} \text{diag}(u) \\
 &\quad + \frac{\sigma^2}{u^{tr u}} \left(\frac{1 - e^{2\psi\Delta}}{2\psi} - \frac{1 - e^{2(\psi+S)\Delta}}{2(\psi+S)}\right) tr uu \quad (\text{A6})
 \end{aligned}$$

We thus get $\Sigma_{\tau_{i+1}}^-$ with the help of Equations (A4) and (A6).

For the GMM model.—Assume that we rank first the n_1 plant traits, before the n_2 butterfly traits in the X vector. Traits evolve following the equation:

$$\begin{aligned}
 \forall k \in \{1, \dots, n_1\}, dX_t^{(k)} &= S \left(d_1 + \frac{1}{n_2} \sum_{l=n_1+1}^{n_1+n_2} X_t^{(l)} - X_t^{(k)} \right) dt + \sigma dW_t^{(k)} \\
 \forall l \in \{n_1+1, \dots, n_1+n_2\}, dX_t^{(l)} &= S \left(d_2 + \frac{1}{n_1} \sum_{k=1}^{n_1} X_t^{(k)} - X_t^{(l)} \right) dt + \sigma dW_t^{(l)}
 \end{aligned}$$

In the general framework formulation, this leads to :

$$\begin{aligned}
 a(t) &= tr(Sd_1, \dots, Sd_1, Sd_2, \dots, Sd_2) \\
 A &= \begin{pmatrix} S & 0 & \dots & 0 & \frac{-S}{n_2} & \dots & \dots & \frac{-S}{n_2} \\ 0 & \ddots & \ddots & \vdots & \vdots & & & \vdots \\ \vdots & \ddots & \ddots & 0 & \vdots & & & \vdots \\ 0 & \dots & 0 & \ddots & \frac{-S}{n_2} & \dots & \dots & \frac{-S}{n_2} \\ \frac{-S}{n_1} & \dots & \dots & \frac{-S}{n_1} & \ddots & 0 & \dots & 0 \\ \vdots & & & \vdots & 0 & \ddots & \ddots & \vdots \\ \vdots & & & \vdots & \vdots & \ddots & \ddots & 0 \\ \frac{-S}{n_1} & \dots & \dots & \frac{-S}{n_1} & 0 & \dots & 0 & S \end{pmatrix} \\
 \Gamma &= \sigma I
 \end{aligned}$$

We would like to be able to compute the expectation and variance easily during each epoch. We thus want to reduce Equations (4a, 4b). For simplicity, we will write in the following $\Delta = \tau_i - \tau_{i+1}$. With some work, we can find the generic element of the matrix $e^{\Delta A}$.

The main idea is to decompose $A = S(I + Z)$, where I is the identity matrix, and Z is made of two blocks with elements $\frac{-1}{n_2}$ and $\frac{-1}{n_1}$. I and Z commute, meaning that :

$$e^{\Delta A} = e^{\Delta S(I+Z)} = e^{\Delta S I} e^{\Delta S Z} = e^{\Delta S} e^{\Delta S Z}$$

Moreover, we show in the Online Appendix (available on Dryad) how to find by induction the generic element of the matrix Z^k , and further use this to find the generic element of the matrix $e^{\Delta S Z} = \sum_{k \geq 0} \frac{S^k \Delta^k Z^k}{k!} = I + \sum_{k \geq 1} \frac{S^k \Delta^k Z^k}{k!}$. From there, we get the main elements from which to derive the expectation vector and covariance matrix. Full derivations are shown in Online Appendix (available on Dryad).

REFERENCES

- 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. Theor. Biol.* 314:204–215.
- Beaulieu J.M., Jhwueng D.-C., Boettiger C., O'Meara B.C. 2012. Modeling stabilizing selection: expanding the Ornstein–Uhlenbeck model of adaptive evolution. *Evolution* 66:2369–2383.
- Blomberg S.P., Garland T., 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.
- Bokma F. 2008. Detection of “punctuated equilibrium” by Bayesian estimation of speciation and extinction rates, ancestral character states, and rates of anagenetic and cladogenetic evolution on a molecular phylogeny. *Evolution* 62:2718–2726.
- Brown W.L., Wilson E.O. 1956. Character displacement. *Syst. Zool.* 5:49–64.
- Butler, M. A. and A. A. King. 2004. Phylogenetic comparative analysis: a modeling approach for adaptive evolution. *Am. Nat.* 164:683–695.
- Clavel J., Escarguel G., Merceron G. 2015. mvMORPH: an R package for fitting multivariate evolutionary models to morphometric data. *Method. Ecol. Evol.* 6:1311–1319.
- Conow C., Fielder D., Ovadia Y., Libeskind-Hadas R. 2010. Jane: a new tool for the cophylogeny reconstruction problem. *Algorithms Mol. Biol.* 5:1–10.
- Dale J., Dey C.J., Delhey K., Kempenaers B., Valcu M. 2015. The effects of life history and sexual selection on male and female plumage colouration. *Nature* 527:367–370.
- Dawkins R., Krebs J.R. 1979. Arms races between and within species. *Proc. Roy. Soc. B* 205:489–511.
- Drury J., Clavel J., Manceau M., Morlon H. 2016. Estimating the effect of competition on trait evolution using maximum likelihood inference. *Syst. Biol.* 65:700–710.
- Dunne J.A., Williams R.J., Martinez N.D., Wood R.A., Erwin D.H. 2008. Compilation and network analyses of Cambrian food webs. *PLoS Biol.* 6:1–16.
- 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.
- Ehrlich P.R., Raven P.H. 1964. Butterflies and plants: a study in coevolution. *Evolution* 18:586–608.
- Felsenstein J. 1973. Maximum-likelihood estimation of evolutionary trees from continuous characters. *Am. J. Hum. Genet.* 25:471–492.
- Fenster C.B., Armbruster W.S., Wilson P., Dudash M.R., Thomson J.D. 2004. Pollination syndromes and floral specialization. *Annu. Rev. Ecol. Evol. Syst.* 35:375–403.
- Gardiner C.W. et al. 1985. *Handbook of stochastic methods* vol. 4. Berlin: Springer.
- Hadfield J.D., Krasnov B.R., Poulin R., Nakagawa S. 2014. A tale of two phylogenies: comparative analyses of ecological interactions. *Am. Nat.* 183:174–187.
- 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., Martins E.P. 1996. Translating between microevolutionary process and macroevolutionary patterns: the correlation structure of interspecific data. *Evolution* 50:1404–1417.
- Hansen T.F., Pienaar J., Orzack S.H. 2008. A comparative method for studying adaptation to a randomly evolving environment. *Evolution* 62:1965–1977.
- Hansen T.F. 1997. Stabilizing selection and the comparative analysis of adaptation. *Evolution* 51:1341–1351.
- Harmon L.J., Losos J.B., Jonathan Davies T., Gillespie R.G., Gittleman J.L., Bryan Jennings W., Kozak K.H., McPeck M.A., Moreno-Roark F., Near T.J., Purvis A., Ricklefs R.E., Schluter D., Schulte II 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.
- Hayward J., Horton T.R. 2014. Phylogenetic trait conservation in the partner choice of a group of ectomycorrhizal trees. *Mol. Ecol.* 23:4886–4898.
- Ho L.S.T., Ané C. 2014. Intrinsic inference difficulties for trait evolution with Ornstein–Uhlenbeck models. *Methods Ecol. Evol.* 5:1133–1146.
- Ives A.R., Godfray H.C.J. 2006. Phylogenetic analysis of trophic associations. *Am. Nat.* 168:1–14.
- Jhwueng D.-C., Maroulas V. 2014. Phylogenetic Ornstein–Uhlenbeck regression curves. *Stat. Probab. Lett.* 89:110–117.
- Labra A., Pienaar J., Hansen T.F. 2009. Evolution of thermal physiology in *Liolaemus* lizards: adaptation, phylogenetic inertia, and niche tracking. *Am. Nat.* 174:204–220.
- Landis M.J., Schraiber J.G., Liang M. 2013. Phylogenetic analysis using Lévy processes: finding jumps in the evolution of continuous traits. *Syst. Biol.* 62:193–204.
- Liow L.H., Reitan T., Harnik P.G. 2015. Ecological interactions on macroevolutionary time scales: clams and brachiopods are more than ships that pass in the night. *Ecol. Lett.* 18:1030–1039.
- Loeuille N., Loreau M. 2005. Evolutionary emergence of size-structured food webs. *Proc. Natl. Acad. Sci. USA* 102:5761–5766.
- Mahler D.L., Revell L.J., Glor R.E., Losos J.B. 2010. Ecological opportunity and the rate of morphological evolution in the diversification of Greater Antillean anoles. *Evolution* 64:2731–2745.
- Martinez N.D. 2006. Network evolution: exploring the change and adaptation of complex ecological systems over deep time. In: *Ecological networks: linking structure to dynamics in food webs*. Pascual M., Dunne J., editor. Oxford: Oxford University Press. p. 287–302.
- Martins E.P. 2004. Compare, version 4.6 b. computer programs for the statistical analysis of comparative data.
- Martín González A.M., Dalsgaard B., Nogués-Bravo D., Graham C.H., Schleuning M., Maruyama P.K., Abrahamczyk S., Alarcón R., Araujo A.C., Araújo F.P., Mendes de Azevedo Jr. S., Baquero A.C., Cotton P.A., Ingversen T.T., Kohler G., Lara C., Las-Casas F.M.G., Machado A.O., Machado C.G., Maglianesi M.A., McGuire J.A., Moura A.C., Oliveira G.M., Oliveira P.E., Ornelas J.F., da Cruz Rodrigues L., Rosero-Lasprilla L., Rui A.M., Sazima M., Timmermann A., Galarda Varassin I., Vizentin-Bugoni J., Wang Z., Watts S., Rahbek C., Martinez N.D. 2015. The macroecology of phylogenetically structured hummingbird–plant networks. *Glob. Ecol. Biogeogr.* 24:1212–1224.
- Moen D., Morlon H. 2014. From dinosaurs to modern bird diversity: extending the time scale of adaptive radiation. *PLoS Biol.* 12: 1–4.
- Morlon H., Lewitus E., Condamine F.L., Manceau M., Clavel J., Drury J. 2015. RPANDA: an R package for macroevolutionary analyses on phylogenetic trees. *Method. Ecol. Evol.* 7:589–597.
- Nuismer S. L., Harmon L.J. 2014. Predicting rates of interspecific interaction from phylogenetic trees. *Ecol. Lett.* 18:17–27.
- Nuismer S.L., Jordano P., Bascompte J. 2013. Coevolution and the architecture of mutualistic networks. *Evolution* 67:338–354.
- O'Meara B.C., Ané C., Sanderson M.J., Wainwright P.C. 2006. Testing for different rates of continuous trait evolution using likelihood. *Evolution* 60:922–933.
- Pagel M. 1994. Detecting correlated evolution on phylogenies: a general method for the comparative analysis of discrete characters. *Proc. Roy. Soc. B* 255:37–45.
- Pennell M.W., Harmon L.J. 2013. An integrative view of phylogenetic comparative methods: Connections to population genetics, community ecology, and paleobiology. *Ann. NY Acad. Sci.* 1289: 90–105.
- Quintero I., Keil P., Jetz W., Crawford F.W. 2015. Historical biogeography using species geographical ranges. *Syst. Biol.* 64: 1059–1073.
- Rafferty N.E., Ives A.R. 2013. Phylogenetic trait-based analyses of ecological networks. *Ecology* 94:2321–2333.
- Reitan T., Schweder T., Henderiks J. 2012. Phenotypic evolution studied by layered stochastic differential equations. *Ann. Appl. Stat.* 6:1531–1551.

- Revell L.J., Collar D.C. 2009. Phylogenetic analysis of the evolutionary correlation using likelihood. *Evolution* 63:1090–1100.
- Revell L.J. 2012. phytools: an R package for phylogenetic comparative biology (and other things). *Method. Ecol. Evol.* 3:217–223.
- Ronquist F., Sanmartín I. 2011. Phylogenetic methods in biogeography. *Annu. Rev. Ecol. Evol. Syst.* 42:441–464.
- Ruta M., Wagner P.J., Coates M.I. 2006. Evolutionary patterns in early tetrapods. I. rapid initial diversification followed by decrease in rates of character change. *Proc. Roy. Soc. B* 273:2107–2111.
- Simpson G.G. 1944. *Tempo and mode in evolution*. New York: Columbia University Press.
- Slater G.J., Harmon L.J., Alfaro M.E. 2012. Integrating fossils with molecular phylogenies improves inference of trait evolution. *Evolution* 66:3931–3944.
- Slater G.J. 2015. Iterative adaptive radiations of fossil canids show no evidence for diversity-dependent trait evolution. *Proc. Natl. Acad. Sci. USA* 112:4897–4902.
- Sletvold N., Trunschke J., Smit M., Verbeek J., Ågren J. 2016. Strong pollinator-mediated selection for increased flower brightness and contrast in a deceptive orchid. *Evolution* 70:716–724.
- Thomas G.H., Freckleton R.P. 2012. MOTMOT: models of trait macroevolution on trees. *Method. Ecol. Evol.* 3:145–151.
- Thomas G.H., Freckleton R.P., Székely T. 2006. Comparative analyses of the influence of developmental mode on phenotypic diversification rates in shorebirds. *Proc. Roy. Soc. B* 273:1619–1624.
- Uyeda J.C., Caetano D.S., Pennell M.W. 2015. Comparative analysis of principal components can be misleading. *Syst. Biol.* 64: 677–689.
- Van Valen L. 1973. A new evolutionary law. *Evol. Theor.* 1:1–30.
- Weiblen G.D. 2004. Correlated evolution in fig pollination. *Syst. Biol.* 53:128–139.
- Weir J.T., Mursleen S. 2013. Diversity-dependent cladogenesis and trait evolution in the adaptive radiation of the auks (Aves: Alcidae). *Evolution* 67:403–416.