Computational Statistics and Data Analysis 145 (2020) 106924

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

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

Modeling rate of adaptive trait evolution using Cox–Ingersoll–Ross process: An Approximate Bayesian Computation approach

Dwueng-Chwuan Jhwueng

Department of Statistics, Feng-Chia University, No. 100, Wenhwa Rd., Seatwen, Taichung, Taiwan 40724, R.O.C.

ARTICLE INFO

Article history:
Received 20 August 2018
Received in revised form 19 December 2019
Accepted 10 January 2020
Available online 22 January 2020

Keywords:
Phylogenetic comparative analysis
Gaussian process
Brownian motion
Ornstein–Uhlenbeck process
CIR process
Trait evolution
Approximate Bayesian Computation

ABSTRACT

Over the past decades, Gaussian processes have been widely used to study trait evolution. In particular, two members of Gaussian processes, Brownian motion and the Ornstein–Uhlenbeck process, have been frequently applied for describing continuous trait evolution. Several models (OUBM, OUOU, OUBMBM, OUOUBM) have been proposed to study the impact on the optimum of a trait by other traits. Applying the Cox–Ingersoll–Ross (CIR) process on rate of evolution, which prevents rates from becoming negative, is a potentially useful extension developed here as the OUBMCIR and OUOUCIR models. Since the likelihood functions of the OUBMCIR and the OUOUCIR models are intractable, a heuristic algorithm for parameter estimation and inference under Approximate Bayesian Computation (ABC) is proposed. Simulation studies show that new models perform well. Empirical analysis using several data sets from literature also provides evidence of the validity and utility of the new models. The relevant data sets and R scripts developed for this project can be accessed through the link.[1]

1. Introduction

Statistical methods for phylogenetics are key for the understanding of how species evolve: these methods both help understand linkage between traits and correct for the non-independence that comes from shared evolutionary history. An example of this was the study of Watanabe et al. (2015), who showed that having warmer muscles in fish (such as tuna) led to faster swimming rates, even after controlling for the few repeated origins of warm muscles in fish. These methods rely on a phylogenetic tree. A tree $\mathbb{T}$ is a branching diagram that displays evolutionary relationships among a group of species. The change of a trait value during evolutionary history of life can be viewed using a stochastic variable defined on the time domain. In the case of continuous trait evolution, the trait of a species observed at time t is defined as a stochastic variable y_t. The dynamic behavior of y_t is constrained by the following stochastic differential equation (SDE)

$$dy_t = \mu(y_t, \Theta, t)dt + \tau(y_t, \Theta, t)dW_t^y, \ t > 0. \tag{1}$$

The dy_t on the left hand side of Eq. (1) represents the quantity of change during an infinitesimal time dt. The deterministic term $\mu(y_t, \Theta, t)$ on the right hand side of Eq. (1) refers to a coefficient that measures the quantity of change in an infinitesimal time dt while $\tau(y_t, \Theta, t)$ is the diffusion coefficient that amplifies/reduces the trait change according

E-mail address: dcjhwueng@fcu.edu.tw.

[1] https://tonyjhwueng.info/ououcir.

https://doi.org/10.1016/j.csda.2020.106924
0167-9473/

to the random changing environment measured by dW_t^y where W_t^y is a Wiener process having continuous paths and independent Gaussian increments (i.e. $dW_t^y \sim \mathcal{N}(0, dt)$) and Θ is a vector containing model parameters.

Many statistical methods have been developed by applying continuous stochastic processes ranging from a Gaussian process (Felsenstein, 1985; Hansen and Martins, 1996; Butler and King, 2004; OMeara et al., 2006; Beaulieu et al., 2012; Pagel, 1999; Freckleton, 2012) or non-Gaussian processes (Blomberg, 0000; Jhwueng et al., 2020) to study trait evolution. The Ornstein–Uhlenbeck (OU) process is a continuous stochastic process for modeling trait evolution (Butler and King, 2004; Beaulieu et al., 2012; Hansen, 1997). An OU stochastic variable y_t solves the SDE in Eq. (1) with $\mu(y_t, \Theta, t) = \alpha(\theta - y_t)$ and $\tau(y_t, \Theta, t) = \tau$ where θ is interpreted as the evolutionary optimum of y_t, α is a constraining force that pulls trait back to the optimum θ and τ is a rate of evolution that measures the speed of random change, though this describes the phenomenon of trait evolution not necessarily the underlying biological mechanism. When α is zero, there is no constraining force and the model is equivalent to Brownian motion.

Many statistical methods have been proposed by expanding the OU model. Those methods use a generalized OU process established by assuming pertinent processes on the model parameters α_t^y, θ_t^y and τ_t^y. Since the continuous trait variable y_t solves the following SDE:

$$dy_t = \alpha_t^y(\theta_t^y - y_t)dt + \tau_t^y dW_t^y, \ t > 0, \tag{2}$$

several models have been developed by assuming $\alpha_t^y = \alpha_y$ a constant, θ_t^y a constant or a stochastic variable, and τ_t^y a constant or a stochastic variable (see Butler and King, 2004; OMeara et al., 2006, and Beaulieu et al., 2012). By assuming that θ_t^y is a continuous stochastic variable, θ_t^y solves the following SDE:

$$d\theta_t^y = \mu(\theta_t^y, \Theta, t)dt + \tau(\theta_t^y, \Theta, t)dW_t^\theta, \ t > 0, \tag{3}$$

where dW_t^θ is a Wiener process.

In a case where $\mu(\theta_t^y, \Theta, t) = 0$ and $\tau(\theta_t^y, \Theta, t) = \sigma_\theta$, Hansen et al. (2008) created an OUBM model for optimal linear regression. Jhwueng and Maroulas (2014) expanded the OUBM model to the OUOU model by allowing an Ornstein–Uhlenbeck process on θ_t^y (i.e. $\mu(\theta_t^y, \Theta, t) = -\alpha_\theta(\theta_t^y - \tilde{\theta})$ and $\tau(\theta_t^y, \Theta, t) = \sigma_\theta$) where α_θ is the force that pulls θ_t^y back to its optimum $\tilde{\theta}$ and σ_θ is the rate parameter for θ_t^y. Those models are applied to study the linear relationship between traits building upon its optimum with $\theta_t^y = \beta_0 + \sum_{i=1}^p \beta_i x_{i,t}$, $i = 1, 2, \ldots, p$ where $x_{i,t}$ is the ith covariate trait and β_i is the ith regression parameter (see application section in Hansen et al., 2008; Jhwueng and Maroulas, 2014).

By treating the rate of evolution τ_t^y in Eq. (2) as a stochastic variable, τ_t^y solves the following SDE in Eq. (4):

$$d\tau_t^y = \mu(\tau_t^y, \Theta, t)dt + \tau(\tau_t^y, \Theta, t)dW_t^\tau. \tag{4}$$

In particular when $\mu(\tau_t^y, \theta, t) = 0$ and $\tau(\tau_t^y, \theta, t) = \tau$ a constant, τ_t^y is Brownian motion. The model joint with Eqs. (2) and (3) and a Brownian motion variable τ_t^y is called OUBMBM model when θ_t^y is a Brownian motion variable and the model is called OUOUBM model when θ_t^y is an OU process variable (Jhwueng and Maroulas, 2016).

Kostikova et al. (2016) bridged inter- and intraspecific trait evolution with a hierarchical Bayesian approach, with the assumption that species-specific trait means evolve under a simple Brownian motion process, whereas species-specific trait variances are modeled with a Brownian motion or an OU process. An alternative approach, mentioned by Kostikova et al. (2016) could be based on geometric Brownian motion (GBM) and Cox–Ingersoll–Ross (CIR) processes, which both have a boundary condition at 0, as that bound could be appropriate for the evolution of variance.

Here I use the Cox–Ingersoll–Ross(CIR) process (Cox et al., 1985) to model the dynamic of τ_t^y where τ_t^y solves the SDE in Eq. (5)

$$d\tau_t^y = \alpha_\tau(\theta_\tau - \tau_t^y)dt + \sigma_\tau\sqrt{\tau_t^y}dW_t^\tau, \tag{5}$$

where $\theta_\tau > 0$ is the optimum of τ_t^y, $\alpha_\tau > 0$ is a constant that pulls τ_t^y back to θ_τ, $\sigma_\tau > 0$ is the rate of change for τ_t^y, and W_t^τ is a Wiener process. Note that the zero of τ_t^y is precluded if the condition $2\alpha_\tau\theta_\tau \geq \sigma_\tau^2$ is held. When τ_t^y is close to zero, the random effect is weakened by small values of $\sigma_\tau\sqrt{\tau_t^y}$. Then the process of τ_t^y is dominated by the pull factor (i.e. $\alpha_\tau(\theta_\tau - \tau_t^y) > 0$) which moves τ_t^y upward.

The distribution of future values of a CIR random variable τ_t^y conditioned on current value τ_s^y has a distribution: $c\chi^2(k, \lambda)$ where $c = \sigma_\tau^2(1 - e^{-\alpha_\tau t})/(4\alpha_\tau)$, $k = 4\theta_\tau\alpha_\tau/\sigma_\tau^2$ and $\lambda = 4\tau_s^y\alpha_\tau e^{-\alpha_\tau t}/(\sigma_\tau^2(1 - e^{-\alpha_\tau t}))$ and $\chi^2(k, \lambda)$ is a non-central chi-squared distribution with degree of freedom k and non-central parameter λ.

Furthermore, the optimal linear regression is allowed with interaction in this work where the relation between optimum θ_t^y and its covariates $x_{i,t}$, $i = 1, 2, \ldots, p$ is represented in Eq. (6)

$$\theta_t^y = \beta_0 + \sum_{i=1}^p \beta_i x_{i,t} + \sum_{i=1}^p \sum_{j \neq i}^p \beta_{ij} x_i x_j, \quad i, j = 1, 2, \ldots, p, \tag{6}$$

where β_is are the corresponding regression parameters to the ith covariate and β_{ij}s are the corresponding regression parameters to the interaction of the ith and the jth covariates.

When jointly modeling adaptive trait evolution using Eqs. (2), (3), (5) and (6), one of important steps is to specify the distribution of the trait variable y_t. It is known that the distribution of y_t in the OUBM model (Hansen et al., 2008) or

Table 1
Property of models used in adaptive trait evolution. The check symbol ✓indicates that the variable has the property, the letter n represents that the variable does not have the property, and the symbol — means that the variable is not available in the model. For instance, in the OUBMBM model the triple parameters (y_t, θ_t, τ_t) with (✓,✓,✓) in the linearity property (column Linear) indicate that each of the parameters is a solution to a linear SDE. On the other hand, in the OUOUCIR model the triple parameters (y_t, θ_t, τ_t) with (n,✓,n) in the normality property (column Normal) indicates that y_t and τ_t are not normally distributed stochastic variable while θ_t is a normal distributed stochastic variable.

Parameters Model	(y_t, θ_t, τ_t) Linear	(y_t, θ_t, τ_t) Autonomous	(y_t, θ_t, τ_t) Additive	(y_t, θ_t, τ_t) Normal	References
OUBM	(✓, ✓, –)	(n, ✓, –)	(✓, ✓, –)	(✓, ✓, –)	Hansen et al. (2008)
OUOU	(✓, ✓, –)	(n, ✓, –)	(✓, ✓, –)	(✓, ✓, –)	Jhwueng and Maroulas (2014)
OUBMBM	(✓, ✓, ✓)	(n, ✓, ✓)	(✓, ✓, ✓)	(n, ✓, ✓)	Jhwueng and Maroulas (2016)
OUOUBM	(✓, ✓, ✓)	(n, ✓, ✓)	(✓, ✓, ✓)	(n, ✓, ✓)	Jhwueng and Maroulas (2016)
OUBMCIR	(✓, ✓, n)	(n, ✓, n)	(✓, ✓, n)	(n, ✓, n)	This work
OUOUCIR	(✓, ✓, n)	(n, ✓, n)	(✓, ✓, n)	(n, ✓, n)	This work

in the OUOU model (Jhwueng and Maroulas, 2014) follows a multivariate normal distribution. However, the distribution of y_t is intractable when τ_t^y follows a CIR process. Approximation Bayesian Computation (ABC) (Blum, 2010; Blum and François, 2010; Csilléry et al., 2010; Jabot et al., 2013) for parameter estimation and inference is an appropriate approach to adopt in such cases.

The construction of a new model for adaptive trait evolution is shown in Section 2. The new model is called OUBMCIR model if y_t follows a generalized OU process, θ_t^y follows a BM and τ_t^y follows a CIR process. The model is called the OUOUCIR model if y_t follows a generalized OU process, θ_t^y follows an OU process and τ_t^y follows a CIR process. The explicit closed-form of y_t is derived for each model in Section 2.3. Those expressions are used to simulate trait values along a given phylogenetic tree under the tree traversal algorithm (Morris, 1979). The statistical inference to assess models of interest is provided in Section 3. Results from simulation to evaluate model performance as well as from empirical analyses using real data from literature are provided in Section 4. Conclusion and discussion of this work are provided in Section 5.

2. Modeling adaptive trait evolution

2.1. Property of model

To provide an overview of the models used in trait evolution, I first describe the properties of SDEs for several models, including the ones developed later in this paper. The SDE in Eq. (1) is a linear SDE if $\mu(y_t, t) = a_1(t)y_t + a_2(t)$ and $\tau(y_t, t) = b_1(t)y_t + b_2(t)$ are linear function of y_t. That is, $dy_t = (a_1(t)y_t + a_2(t))dt + (b_1(t)y_t + b_2(t))dW_t^y$. A linear SDE is autonomous if all coefficients are constants; a linear SDE is homogenous if $a_2(t) = 0$ and $b_2(t) = 0$; and a linear SDE is linear in the additive sense if $b_1(t) = 0$. For instance, the SDE for y_t in the OUBM model (Hansen et al., 2008) with $\mu(y_t, t) = \alpha_y(\theta_t - y_t)$ and $\tau_t(y_t, t) = \tau$ is a linear additive non-autonomous SDE. Since both θ_t and W_t are BMs in the OUBM model, the solution y_t for the SDE in Eq. (1) is a linear combination of two BMs (normal distributed). Hence y_t is a normal distributed stochastic variable in the OUBM model. Similarly y_t in the OUOU model (Jhwueng and Maroulas, 2014) is a normal stochastic variable, too. For the OUOUCIR model, the SDE for y_t with coefficients $\mu(y_t, t) = \alpha_y(\theta_t - y_t)$ and $\tau_t(y_t, t) = \tau_t^y$ is also a linear additive non-autonomous SDE. Since θ_t^y is an OU process, it is a normal distributed variable. However, the rate τ_t^y is a non-central chi-squared stochastic variable which leads to an intractable distribution for the solution y_t. See detailed explanation in Section 2.2. The properties of SDEs in model of trait evolution are categorized in Table 1.

One realization of the trajectories of OUBMBM, OUOUBM, OUBMCIR and OUOUCIR model on a single branch of tree is shown in Fig. 1. Trajectories for the optimum parameters $\theta_{it}, i = 1, 2$ are the same across models where $i = 1$ assumes that θ_{1t} is a BM, and $i = 2$ assumes that θ_{2t} is an OU process. Trajectories for the response traits y_t under each model are then generated and displayed according to the corresponding stochastic processes in each panel. We generally expect OUBMBM to have the greatest amount of variance, as there is the least amount of constraint on traits. OUOUBM, which has one further trait being pulled to an optimum with an OU process, should have less variance. With CIR models, given that there is a boundary, variance is expected to be further reduced. Whether this is true for a particular simulation depends on model parameters; for example, if a simulation occurs far from a boundary, a CIR model is indistinguishable from a BM model and should have the same properties. As an example of generally expected, I simulated 10,000 replicates for all four models using as similar as possible parameter values. The average of 10,000 standard deviations of the trajectories for the OUBMBM model is 9.43, the OUOUBM model is 6.24, the OUBMCIR model is 8.41 and the OUOUCIR model is 5.01. Overall, the response trait y_t with θ_t^y modeled by BM has higher variation than y_t with θ_t^y modeled by OU process (9.43 vs. 6.24; 8.41 vs. 5.01); while the response trait y_t with the rate τ_t^y modeled by a CIR process has lower variation than y_t with τ_t^y modeled by a BM (8.41 vs. 9.43; 5.01 vs. 6.24).

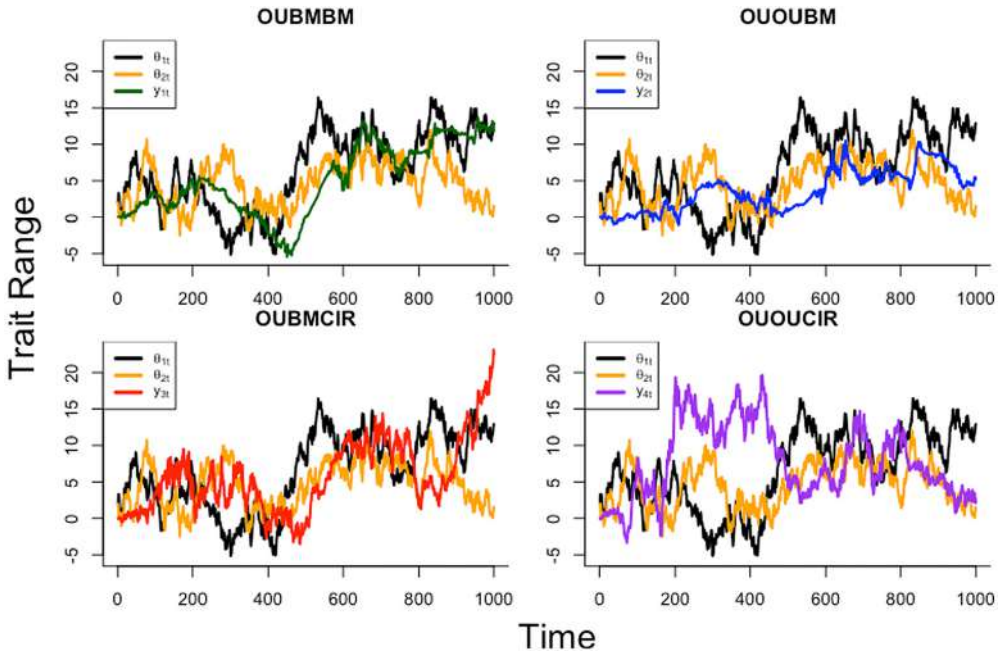


Fig. 1. Simulation of trait change using 1000 time steps. The trajectory of the covariates trait x_t (not display here) was first simulated under either Brownian motion or OU process; and the optimum paths $\theta_{it}, i = 1, 2$ were then computed using the linear relationship $\theta_t = b_0 + b_1 x_{1,t} + b_2 x_{2,t}$; finally, the path for the response traits $y_{it}, i = 1, 2, 3, 4$ was simulated based on models using Eqs. (2), (3), (5) and (6). θ_{1t} (in black) assumes that covariates $x_{1,t}$ and $x_{2,t}$ follow BM, and θ_{2t} (in orange) assumes that covariates $x_{1,t}$ and $x_{2,t}$ follow OU. On the top left panel, the trajectories of response trait y_{1t} (in green) are plotted for the OUBMBM model using parameter $(\alpha_y, \tau) = (0.01, 0.05)$; on the top right panel, the trajectories of response trait y_{2t} (in blue) are plotted for the OUOUBM model using parameter $(\alpha_y, \alpha_\theta, \tau) = (0.01, 0.02, 0.05)$; on the bottom left panel, the trajectories of response trait y_{3t} (in red) are plotted for the OUBMCIR model using parameter $(\alpha_y, \alpha_\tau, \theta_\tau, \sigma_\tau) = (0.01, 0.02, 0.75, 0.05)$; and on the bottom right panel, the trajectories of response trait y_{4t} (in purple) are plotted for the OUOUCIR model using parameter $(\alpha_y, \alpha_\theta, \alpha_\tau, \theta_\tau, \sigma_\tau) = (0.01, 0.02, 0.02, 0.75, 0.05)$. The regression coefficient vector (b_0, b_1, b_2) is set to $(2, 2.38, 1.42)$ for models without interaction. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

2.2. Distribution of model

In general by adopting Eqs. (2), (3), (5) and (6), the dynamic of $y_t, \theta_t^y, \tau_t^y$ can be jointly described by a random vector $\boldsymbol{Z}_t = (y_t, \theta_t^y, \tau_t^y)'$ that solves the system of SDE $d\boldsymbol{Z}_t = \boldsymbol{\mu}_t dt + \boldsymbol{D}_t d\boldsymbol{W}_t$, where $\boldsymbol{\mu}_t = \big(\mu(y_t, \Theta, t), \mu(\theta_t^y, \Theta, t), \mu(\tau_t, \Theta, t)\big)'$ is a vector of pull coefficient, $\boldsymbol{D}_t = \text{diag}\big[\tau(y_t, \Theta, t), \tau(\theta_t^y, \Theta, t), \tau(\tau_t^y, \Theta, t)\big]$ is a diagonal matrix, and $\boldsymbol{W}_t = (W_t^y, W_t^\theta, W_t^\tau)'$ is the associated random vector of Brownian motion and v' is the transpose of a vector v. The system of SDE can be represented in Eq. (7)

$$d\boldsymbol{Z}_t = (\boldsymbol{A}\boldsymbol{Z}_t + \boldsymbol{b}_t)dt + \boldsymbol{D}_t d\boldsymbol{W}_t, \tag{7}$$

where $\boldsymbol{A}$ is a 3×3 constant matrix of force parameters α_y, α_θ, and α_τ.

For a *homogeneous* model where the rate of evolution τ_t^y is a time invariant constant (e.g. $\boldsymbol{b}_t = 0$ and $\tau_t^y = \tau$ in the OUBM model and in the OUOU model), the $\boldsymbol{D}_t = \text{diag}[\tau, \sigma_\theta, 0]$ is a constant diagonal matrix. In this case, the system of SDE described in Eq. (7) given the initial condition $\boldsymbol{Z}_0 = (y_0, \theta_0, \tau_0^y)'$ at $t = 0$ has a unique solution $\boldsymbol{Z}_t = e^{-\boldsymbol{A}t}\boldsymbol{Z}_0 + \int_0^t e^{-\boldsymbol{A}(t-s)}\boldsymbol{D}_s d\boldsymbol{W}_s$. The expected value of $\boldsymbol{Z}_t$, defined by $\mathbb{E}[\boldsymbol{Z}_t] = \boldsymbol{Z}_0 e^{-\boldsymbol{A}t}$, can be calculated straightforwardly. The second moment of $\boldsymbol{Z}_t$, denoted by $\boldsymbol{P}_t = \mathbb{E}[\boldsymbol{Z}_t \boldsymbol{Z}_t']$, is uniquely determined by solving the system of an ordinary differential equation $\frac{d}{dt}\boldsymbol{P}_t = \boldsymbol{A}\boldsymbol{P}_t + \boldsymbol{P}_t\boldsymbol{A}' + \mathbb{E}[\boldsymbol{C}_t]$ where $\mathbb{E}[\boldsymbol{C}_t] = \boldsymbol{D}_t\boldsymbol{D}_t$ (Hansen et al., 2008; Jhwueng and Maroulas, 2014, 2016). Because $\boldsymbol{Z}_t$ is a normal random vector, its first component y_t is a normal random variable specified by its first and second moments of $\boldsymbol{Z}_t$. The solution $y_t = y_0 e^{-\alpha_y t} + \alpha_y e^{-\alpha_y t}\int_0^t e^{\alpha_y s}\theta_s^y ds + \tau\int_0^t e^{-\alpha_y(t-s)} dW_s^y$ is a linear combination of normal random variable and therefore y_t is again a normal random variable under the assumption that either θ_t^y is a BM (Hansen et al., 2008) or θ_t^y is an OU process (Jhwueng and Maroulas, 2014). However, due to the properties of a CIR process, it is not straightforward to specify the distribution of $\boldsymbol{Z}_t$ in the OUBMCIR model and in the OUOUCIR model. Taking the OUOUCIR model as a case where the drift and diffusion vector in the OUOUCIR model are

$$\boldsymbol{\mu}_t = \boldsymbol{A}\boldsymbol{Z}_t + \boldsymbol{b}_t = \begin{pmatrix} -\alpha_y & \alpha_y & 0 \\ 0 & -\alpha_\theta & 0 \\ 0 & 0 & -\alpha_\tau \end{pmatrix}\boldsymbol{Z}_t + \begin{pmatrix} 0 \\ 0 \\ \alpha_\tau\theta_\tau \end{pmatrix} \text{ and } \boldsymbol{D}_t = \begin{pmatrix} \tau_t^y & 0 & 0 \\ 0 & \sigma_\theta & 0 \\ 0 & 0 & \sigma_\tau\sqrt{\tau_t} \end{pmatrix}.$$

Table 2
A three taxa example of trait values simulated under Brownian motion model by the postorder tree traversal algorithm according to Fig. 2. The first row of the table indicates that the ancestral node (Anc. node) D started at the root of tree with phenotypic value (Anc. value) 0 and evolved to the descendant node (Des. node) E with phenotypic (Des. value) −0.32 with time (length) 1.2. Similarly, the second row of the table shows that the species D started at the root of tree and evolved to node C with value 2.48 with a time period of length 2.8. The third and the fourth row of the table show that species E evolved to taxon B (with value 0.92) and taxon A (with value −0.64) over a time period of length 1.6.

Anc. node	Des. node	length	Anc. value	Des. value
D	E	1.20	0	−0.32
D	C	2.80	0	2.48
E	B	1.60	−0.32	0.92
E	A	1.60	−0.32	−0.64

Note that the distribution of future values $\sqrt{\tau_t^y}$ in $\boldsymbol{D}_t$ follows a non-central chi distribution. By integrating Eq. (1) with respect to t, the solution y_t for OUOUCIR model is represented in Eq. (8)

$$y_t = y_0 + e^{-\alpha_y t}\int_0^t \alpha_y e^{\alpha_y s}\theta_s^y ds + e^{-\alpha_y t}\int_0^t \tau_s e^{\alpha_y s} dW_s^y = y_0 + ① + ②. \tag{8}$$

For ① in Eq. (8), since θ_s^y is an OU process with $\theta_s^y = e^{-\alpha_\theta s}\theta_0^y + \tilde{\theta}_y(1-e^{-\alpha_\theta s}) + \sigma_\theta\int_0^s e^{\alpha_\theta(v-s)}dW_v^\theta$ where $\tilde{\theta}_y$ is optimum of θ_s^y and θ_0^y is the initial condition at $t=0$. The integral $\int_0^t \alpha_y e^{\alpha_y s}\theta_s^y ds$ is computed as

$$\int_0^t \alpha_y\theta_0^y e^{(\alpha_y-\alpha_\theta)s}ds + \int_0^t \alpha_y\tilde{\theta}_y e^{\alpha_y s}(1-e^{-\alpha_\theta s})ds + \int_0^t \sigma_\theta\alpha_y e^{(\alpha_y-\alpha_\theta)s}\left(\int_0^s e^{\alpha_\theta v}dW_v^\theta\right)ds = ⓐ + ⓑ + ⓒ. \tag{9}$$

Note that ⓐ and ⓑ are both definite integrals where $ⓐ = \frac{\alpha_y\theta_0^y}{\alpha_y-\alpha_\theta}(e^{(\alpha_y-\alpha_\theta)t}-1)$ and $ⓑ = \tilde{\theta}_y(e^{\alpha_y t}-1) - \frac{\alpha_y\tilde{\theta}_y}{\alpha_y-\alpha_\theta}(e^{(\alpha_y-\alpha_\theta)t}-1)$. For ⓒ, since the stochastic integral $\int_0^s e^{\alpha_\theta v}dW_v^\theta$ is a normal random variable with mean 0 and variance (by Itô Isometry) $\frac{e^{2\alpha_\theta s}-1}{2\alpha_\theta}$, the integrand in ⓒ defined as $f_s = \sigma_\theta\alpha_y e^{(\alpha_y-\alpha_\theta)s}\left(\int_0^s e^{\alpha_\theta v}dW_v^\theta\right)$ is again a normal random variable with mean 0 and variance (by Itô Isometry) $v(s) = \frac{\sigma_\theta^2\alpha_y^2}{2\alpha_\theta}e^{2\alpha_y s}(1-e^{-2\alpha_\theta s})$. Then ⓒ is a definite integral of a normal distributed random variable with respect to time t.

For ② in Eq. (8), since τ_s is a CIR random variable with

$$\tau_s = \theta_\tau + (\tau_0-\theta_\tau)e^{-\alpha_\tau s} + \sigma_\tau e^{-\alpha_\tau s}\int_0^s e^{\alpha_\tau u}\sqrt{\tau_u}dW_u, \tag{10}$$

$② = e^{-\alpha_y t}\int_0^t \tau_s e^{\alpha_y s}dW_s^y$ can be separated into sum of three stochastic integrals: ⓓ + ⓔ +ⓕ. For $ⓓ = e^{-\alpha_y t}\int_0^t \theta_\tau e^{\alpha_y s}dW_s^y$, it is a normal random variable with mean 0 and variance $\theta_\tau^2\frac{1-e^{-2\alpha_y t}}{2\alpha_y}$. For $ⓔ = e^{-\alpha_y t}(\tau_0-\theta_\tau)\int_0^t e^{(\alpha_y-\alpha_\tau)s}dW_s^y$, it is another normal random variable with mean 0 and variance $(\tau_0-\theta_\tau)^2(e^{-2\alpha_\tau t}-e^{-2\alpha_y t})/(2(\alpha_y-\alpha_\tau))$.

$$ⓕ = \sigma_\tau\int_0^t e^{-(\alpha_y+\alpha_\tau)s}\left(\int_0^s e^{\alpha_\tau u}\sqrt{\tau_u}dW_u^\tau\right)dW_s^y. \tag{11}$$

However, the stochastic integral ⓕ in Eq. (11) is likely intractable. ⓕ is a stochastic integral with respect to Wiener process where its integrand $e^{-(\alpha_y+\alpha_\tau)s}\left(\int_0^s e^{\alpha_\tau u}\sqrt{\tau_u}dW_u^\theta\right)$ involves another stochastic integral of a square root of CIR random variable with respect to another Wiener process. Although it is known that $\tau_u|\tau_0$ follows a scaled non-central chi-squared distribution, it is not straightforward to determine the analytic form of the integrand (i.e. $\int_0^s e^{\alpha_\tau u}\sqrt{\tau_u}dW_u^\tau$) and it may be not possible to specify the distribution of ⓕ when the stochastic integral is computed using the intractable integrand.

2.3. Trait evolution along the tree

When the distribution of trait variables x_t and y_t are known, traits of a species A at the tip of the tree denoted by (x, y) can be simulated using the tree traversal algorithm (Morris, 1979) along a given tree $\mathbb{T}$ with known topology and branch lengths (the evolutionary times t). For instance, under the Brownian motion model (Felsenstein, 1985) a trait variable y_t of a species A at time t conditioned on its ancestor y_a is a normal random variable $y_t|y_a$ with mean y_a and variance $\sigma_y^2 t$. (i.e. $y_t|y_a \sim \mathcal{N}(y_a, \sigma_y^2 t)$). Similarly, $x_t|x_a \sim \mathcal{N}(x_a, \sigma_x^2 t)$. One realization of trajectories along a tree of 3 species (A, B, C) of y_t under Brownian motion is shown in Fig. 2 where the corresponding trait value simulated at each node/tip by tree traversal algorithm is shown in Table 2.

The tree traversal algorithm can be generally applied for any model with a known process. For instance, under an OU process (Hansen, 1997), $y_t|y_a$ is a normal random variable with mean $y_a e^{-\alpha_y t} + \theta_y(1-e^{-\alpha_y t})$, and variance $\sigma_y^2(1-e^{-2\alpha_y t})/(2\alpha_y)$). Trait values on each node of a tree can be simulated accordingly.

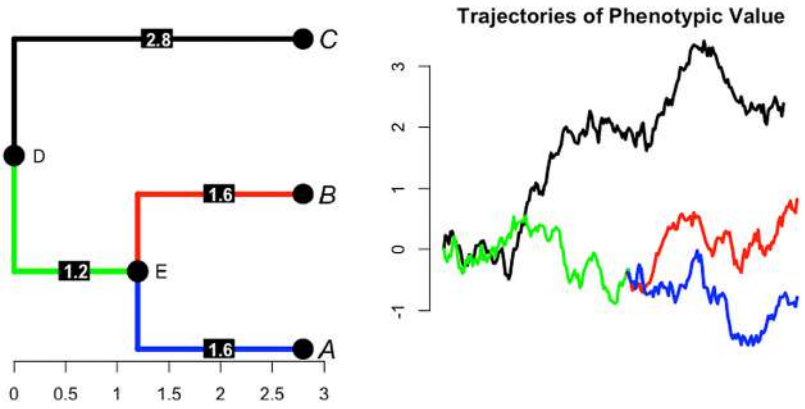


Fig. 2. Trait evolution on a rooted ultrametric phylogenetic tree. The plot on the left panel is a hypothetical phylogenetic species tree containing 3 taxa (A,B,C), an internal node (E) and the root node (D). Three species A, B, C evolve according to a Brownian motion with two speciation events at time 0 and time 1.2, respectively. The total evolutionary time is 2.8 from root D to taxon C, the evolutionary time is 1.2 from root D to node E, and 1.6 from node E to taxon A and to taxon B, respectively. The plot on the right panel shows the trajectories of phenotypic value of the three species, the trajectory in black shows the trajectories of phenotypic value evolved from species D (which is started at 0) to species C, the trajectory in green shows the trajectories of phenotypic value of species D evolved to species E, the trajectory in red shows the trajectories of phenotypic value of species E evolved to species B and the trajectory in blue shows the change of phenotypic value of species E evolved to species A. For this realization, the tips values and ancestral values are $\boldsymbol{Y} = (y_A, y_B, y_C)' = (-0.64, 0.92, 2.48)'$ and $(y_D, y_E) = (0, -0.32)'$, respectively. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Note that traits can be simulated without the tree traversal algorithm when models have known distributions. For instance, for the OU model in Hansen (1997), trait can be simulated directly under a multivariate normal distribution (i.e. $\boldsymbol{Y} \sim \mathcal{MVN}(\theta_y \mathbf{1}, \sigma_y^2 \boldsymbol{\Sigma}_{\alpha_y})$ where $\boldsymbol{Y} = (y_1, y_2, \ldots, y_n)'$ is the trait vector observed at the tips of the tree, θ_y is the overall mean, $\mathbf{1} = (1, 1, \ldots, 1)'$ is the vector of 1s and $\boldsymbol{\Sigma}_{\alpha_y}$ is the variance covariance structure for $\boldsymbol{Y}$ (which is based on the tree topology, branch lengths, and model parameters). Similarly, $\boldsymbol{X} \sim \mathcal{MVN}(\theta_x \mathbf{1}, \sigma_x^2 \boldsymbol{\Sigma}_{\alpha_x})$ where $\boldsymbol{X} = (x_1, x_2, \ldots, x_n)'$ is the trait vector observed at the tips of the tree (Jhwueng, 2013). Another two models are the OUBM model and the OUOU model which both are of multivariate normal distributions. Trait values $\boldsymbol{Y}$ can be simulated directly given the specified mean vector $\mathbb{E}[\boldsymbol{Y}]$ and variance structure $\mathrm{Var}[\boldsymbol{Y}]$ (Hansen et al., 2008; Jhwueng and Maroulas, 2014).

However, when distribution of a model is intractable (e.g. the OUBMCIR model and the OUOUCIR model), traits can only be simulated using the model. In the following subsection, the trait variable y_t for each of the models (OUBMBM, OUOUBM, OUBMCIR, OUOUCIR) included in this study is expressed. Traits are simulated under a specified model $\mathcal{M}$.

2.3.1. OUBMBM model

In the OUBMBM model with model parameters α_y, σ_x, τ and regression parameters $b_i, b_{ij}, i, j = 1, 2, \ldots, p$, covariate traits x_ts assuming a BM are first simulated on each node/tip of tree given σ_x. The optimal value θ_t^y on each node/tip is then computed via $\theta_t^y = \sum b_i x_{i,t} + \sum b_{ij} x_{i,t} x_{j,t}$ given b_i and b_{ij}. Simulating y_t at the nodes/tips of tree depends on the solution of y_t in Eq. (12) for y_t where

$$y_t = y_0 + e^{-\alpha_y t} \int_0^t \alpha_y e^{\alpha_y s} \theta_s^y ds + e^{-\alpha_y t} \int_0^t \tau_s e^{\alpha_y s} dW_s^y = y_0 + ① + ②. \tag{12}$$

For ①, by assuming the optimum θ_t^s follows Brownian motion (i.e. $\theta_s = \int_0^s \sigma_\theta dW_v^\theta = \sigma_\theta W_s^\theta \sim \mathcal{N}(0, \sigma_\theta^2 s)$), the term $\int_0^t \alpha_y e^{\alpha_y s} \theta_s^y ds = \int_0^t \alpha_y e^{\alpha_y s} \sigma_\theta W_s^\theta ds$ is a stochastic integral of Brownian motion with respect to time. Since $d(\theta_s^y e^{\alpha_y s}) = e^{\alpha_y s} d\theta_s^y + \theta_s^y de^{\alpha_y s}$, the integral $\int \theta_s^y de^{\alpha_y s} = \int_0^t d(\theta_s^y e^{\alpha_y s}) - \int_0^t e^{\alpha_y s} d\theta_s^y = \theta_t^y e^{\alpha_y t} - \int_0^t e^{\alpha_y s} d\theta_s^y$ is a normal random variable with mean θ_0 and variance $\sigma_\theta^2 \frac{e^{2\alpha_y t} - 1}{2\alpha_y}$. For ②, since the τ_s is a BM (i.e. $\tau_s = \int_0^s \tau dW_v^\tau = \tau W_s^\tau \sim \mathcal{N}(0, \tau^2 s)$), then ② $= e^{-\alpha_y t} \int_0^t \tau_s e^{\alpha_y s} dW_s^y = \int_0^t \tau W_s^\tau e^{\alpha_y (s-t)} dW_s^y$ is a stochastic integral that involves an integral of Brownian motion W_s^τ with respect to another Brownian motion W_s^y. Assuming that W_s^y and W_s^σ are two independent and identical processes, samples for ②are drawn by using the median of the trajectory of the stochastic integral $\int_0^t \tau W_s^\tau e^{\alpha_y (s-t)} dW_s^y$ which is simulated by R package `Sim.DiffProc`: `int.st` in Guidoum and Boukhetala (2017).

2.3.2. OUOUBM model

In the OUOUBM model with model parameters α_y, α_x(equivalent to α_θ), θ_x, σ_x, τ, and regression parameters $b_i, b_{ij}, i, j = 1, 2, \ldots, p$, covariate traits x_ts assuming an OU process are first simulated on each node/tip of tree given $\alpha_x, \theta_x, \sigma_x$. The optimum on each node and tip is then computed via $\theta_t^y = \sum b_i x_{i,t} + \sum b_{ij} x_{i,t} x_{j,t}$. Simulating y_t at the nodes/tips depends

on the solution of y_t in Eq. (13) for y_t where

$$y_t = y_0 + e^{-\alpha_y t}\int_0^t \alpha_y e^{\alpha_y s}\theta_s ds + e^{-\alpha_y t}\int_0^t \tau_s e^{\alpha_y s} dW_s^y = y_0 + ① + ②. \tag{13}$$

For ①, because θ_s^y is an OU process with $\theta_s^y = e^{-\alpha_\theta s}\theta_0^y + \tilde{\theta}_y(1-e^{-\alpha_\theta s}) + \sigma_\theta \int_0^s e^{\alpha_\theta(v-s)} dW_v^\theta$ where $\tilde{\theta}_y$ is optimum of θ_s^y and θ_0^y is the initial condition at $t=0$. The integral $\int_0^t \alpha_y e^{\alpha_y s}\theta_s^y ds$ equals to

$$\int_0^t \alpha_y \theta_0^y e^{(\alpha_y-\alpha_\theta)s} ds + \int_0^t \alpha_y \tilde{\theta}_y e^{\alpha_y s}(1-e^{-\alpha_\theta s}) ds + \int_0^t \sigma_\theta \alpha_y e^{(\alpha_y-\alpha_\theta)s}\left(\int_0^s e^{\alpha_\theta v} dW_v^\theta\right) ds = ⓐ + ⓑ + ⓒ. \tag{14}$$

Note that ⓐ and ⓑ are both definite integrals where $ⓐ = \frac{\alpha_y\theta_0^y}{\alpha_y-\alpha_\theta}(e^{(\alpha_y-\alpha_\theta)t}-1)$ and $ⓑ = \tilde{\theta}_y(e^{\alpha_y t}-1) - \frac{\alpha_y\tilde{\theta}_y}{\alpha_y-\alpha_\theta}(e^{(\alpha_y-\alpha_\theta)t}-1)$. For ⓒ, since the term $\int_0^s e^{\alpha_\theta v} dW_v^\theta$ is a normal random variable with mean 0 and variance $\frac{e^{2\alpha_\theta s}-1}{2\alpha_\theta}$, the integrand in ⓒ defined as $f_s = \sigma_\theta \alpha_y e^{(\alpha_y-\alpha_\theta)s}\int_0^s e^{\alpha_\theta v} dW_v^\theta$ is again a normal random variable with mean 0 and variance (by Itô Isometry) $v(s) = \frac{\sigma_\theta^2\alpha_y^2}{2\alpha_\theta} e^{2\alpha_y s}(1-e^{-2\alpha_\theta s})$. Hence $ⓒ = \int_0^t f_s ds = \int_0^t \frac{1}{\sqrt{2\pi v(s)}} e^{\frac{1}{2v(s)}s^2} ds$ can be numerically evaluated through a normal distribution given variance $v(\cdot)$ and a quantile value t (specified by a branch length of tree). For ②, as the rate τ_s is a BM, samples can be drawn under the method described in the OUBMBM model.

2.3.3. OUBMCIR model

In the OUBMCIR model with model parameters $\alpha_y, \sigma_x, \alpha_\tau, \theta_\tau, \sigma_\tau$, and regression parameters are $b_i s$, $b_{ij} s, i, j = 1, 2, \ldots, p$. Covariates traits x_ts assuming a BM variables are first simulated on each node/tip of tree given σ_x. The optimum on each node and tip is then computed via $\theta_t^y = \sum b_i x_{i,t} + \sum b_{ij} x_{i,t} x_{j,t}$. Simulating y_ts at the nodes/tips depends on the solution in Eq. (15) for y_t where

$$y_t = y_0 + e^{-\alpha_y t}\int_0^t \alpha_y e^{\alpha_y s}\theta_s ds + e^{-\alpha_y t}\int_0^t \tau_s e^{\alpha_y s} dW_s^y = y_0 + ① + ②. \tag{15}$$

For ①, since the optimum is a BM (i.e $\theta_s^y \sim \mathcal{N}(0, \sigma_\theta^2 s)$), samples are drawn using the method shown in ① in the OUBMBM model case.

For ②, it is a stochastic integral of a CIR random variable τ_s with respect to Brownian motion W_s^y. Note that $\tau_s|\tau_0$ follows a scaled non-central chi-squared distribution. Then the distribution of the random variable $\int_0^t \tau_s e^{\alpha s} dW_s^y$ conditioned on τ_0 can be seen as the sum of three independent random variables (see *prop.* 4 eqn 2.10 in Chan and Joshi, 2010). Moreover, Chan and Joshi (2010) and Glasserman and Kim (2011) showed that the exact distribution of $\int_0^t \tau_s e^{\alpha_y s} dW_s^y$, conditional on τ_0 can be represented by infinite sums and mixtures of gamma random variables (see prop 4. in Chan and Joshi (2010)). To simulate sample for ②, since the solution to the CIR SDE in Eq. (5) is given by

$$\tau_s = \theta_\tau + (\tau_0 - \theta_\tau)e^{-\alpha_\tau s} + \sigma_\tau e^{-\alpha_\tau s}\int_0^s e^{\alpha_\tau u}\sqrt{\tau_u} dW_u, \tag{16}$$

$② = e^{-\alpha_y t}\int_0^t \tau_s e^{\alpha_y s} dW_s^y$ can be separated into three integrals: ⓐ + ⓑ + ⓒ. For $ⓐ = e^{-\alpha_y t}\int_0^t \theta_\tau e^{\alpha_y s} dW_s^y$, it is a normal random variable with mean 0 and variance $\theta_\tau^2 \frac{1-e^{-2\alpha_y t}}{2\alpha_y}$. For $ⓑ = e^{-\alpha_y t}(\tau_0-\theta_\tau)\int_0^t e^{(\alpha_y-\alpha_\tau)s} dW_s^y$, it is another normal random variable with mean 0 and variance $(\tau_0-\theta_\tau)^2(e^{-2\alpha_\tau t}-e^{-2\alpha_y t})/(2(\alpha_y-\alpha_\tau))$. For $ⓒ = \sigma_\tau \int_0^t e^{-(\alpha_y+\alpha_\tau)s}\left(\int_0^s e^{\alpha_\tau u}\sqrt{\tau_u} dW_u^\tau\right) dW_s^y$, it is a stochastic integral with unknown distribution. Samples of ⓒ are drawn with the aid of numerical integration approach. First, $\int_0^s e^{\alpha_\tau u}\sqrt{\tau_u} dW_u^\tau$ is approximated by $\sum_{j=1}^{m_1} e^{\alpha_\tau j * s/m_1}\sqrt{\tau_j} W_j^\tau$ where τ_j is generated under a non-central chi-square random variable given s_j, and W_j^τ is drawn from a normal distribution with mean 0 and variance s/m_1. Next let $t = \cup_{i=0}^{m_2-1}\mathcal{I}_i$ where $\mathcal{I}_i = [s_i, s_{i+1}], i = 0, 1, 2, \ldots, m_2 - 1$. Then sample for ⓒ is drawn from the numerical approximation $ⓒ \approx \sigma_\tau \sum_{i=1}^{m_2}\left(e^{-(\alpha_y+\alpha_\tau)i*t/m_2}\left(\sum_{j=1}^{m_1} e^{\alpha_\tau j * s_i/m_1}\sqrt{\tau_j} W_j^\tau\right) W_i^\tau\right)$ where W_i^τ is a normal random variable with mean 0 and variance t/m_2. For efficient simulation, currently only 1 grid is used for analysis (i.e. $m_1 = m_2 = 1$). Accuracy of estimating the integral can be improved with finer grids at the cost of larger variation in trait values.

2.3.4. OUOUCIR model

In the OUOUCIR model, model parameters are $\alpha_y, \alpha_x, \theta_x, \sigma_x, \alpha_\tau, \theta_\tau, \sigma_\tau$ and regression parameters $b_i, b_{ij}, i, j = 1, 2, \ldots, p$, covariate traits x_ts assuming an OU process are first simulated on each node/tip of tree given $\alpha_x, \theta_x, \sigma_x$. The optimum on each node and tip can be calculated as $\theta_t^y = \sum b_i x_{i,t} + \sum b_{ij} x_{i,t} x_{j,t}$. Simulating y_t at the nodes/tips depends on the solution in Eq. (17) for y_t:

$$y_t = y_0 + e^{-\alpha t}\int_0^t \alpha e^{\alpha s}\theta_s^y ds + e^{-\alpha t}\int_0^t \tau_s e^{\alpha s} dW_s^y = y_0 + ① + ②. \tag{17}$$

For ①, since the optimum parameter θ_s^y evolved under an OU process, samples can be drawn using the method shown in ① in the OUOUBM model case. For ②, since the rate parameter τ_s evolves under a CIR model, samples can be drawn using the method shown in ② in the OUBMCIR model case.

2.4. Connection with optimal linear regression with interaction

To implement the model described in Eqs. (2), (3), (5) and (6) with interaction, it is necessary to determine the relation between the parameters in θ_t^y and parameters in the covariate $x_{i,t}$. Considering the case of two covariates $x_{1,t}, x_{2,t}$ and let $\theta_t^y = b_0 + b_1 x_{1,t} + b_2 x_{2,t} + b_{12} x_{1,t} x_{2,t}$. Differentiating both sides of the equation with respect to t yields to

$$d\theta_t^y = b_1 dx_{1,t} + b_2 dx_{2,t} + b_{12}(x_{2,t} dx_{1,t} + x_{1,t} dx_{2,t} + dx_{1,t} dx_{2,t}), \tag{18}$$

where $x_{i,t}, i = 1, 2$ are two diffusion processes that solve the following SDE $dx_{i,t} = \mu(x_{i,t}, \Theta, t)dt + \tau(x_{i,t}, \Theta, t)dW_t^{x_i}$, $i = 1, 2,\ t > 0$. By the property of stochastic calculus $dtdt \approx 0$, $dtdW_t^{x_i} \approx 0$, and $dW_t^{x_i} dW_t^{x_j} \approx \rho_{ij} dt$ where $-1 \le \rho_{ij} \le 1$, the square of $d\theta_t^y$ in Eq. (3) can be represented as $(d\theta_t^y)^2 = d\theta_t^y d\theta_t^y = \tau^2(\theta_t^y, \Theta, t)dt$.

Assuming that θ_t^y is either a BM or an OU process, one has $\tau(\theta_t^y, \Theta, t) = \sigma_\theta$. This yields to $(d\theta_t^y)^2 = \sigma_\theta^2 dt$. The relationship between σ_θ and $\sigma_{x_i}, i = 1, 2$ can be obtained by comparing the coefficients of dt after expanding the right hand side of Eq. (18) to get $\sigma_\theta^2 = \sigma_{x_1}^2(b_1^2 + 2b_1 b_{12} x_{2,t} + b_{12}^2 x_{2,t}^2) + \sigma_{x_2}^2(b_2^2 + 2b_2 b_{12} x_{1,t} + b_{12}^2 x_{1,t}^2) + 2\sigma_{x_1}\sigma_{x_2}\rho_{ij}(b_1 b_2 + b_{12} b_1 x_{1,t} + b_{12} b_2 x_{2,t} + b_{12}^2 x_{1,t} x_{2,t})$.

The general representation of the linear optimal regression with p covariates with interaction is $\theta_t^y = b_0 + \sum_{k=1}^p b_k x_{k,t} + \sum_{i=1}^p \sum_{j\neq i}^p b_{ij} x_{i,t} x_{j,t}$ as shown in Eq. (6).

Direct applying the same technique above implies

$$d\theta_t^y = \sum_{k=1}^p b_k \sigma_{x_k} dW_t^{x_k} + \sum_{i=1}^p \sum_{j\neq i}^p b_{ij}(x_{j,t}\sigma_{x_i} dW_t^{x_i} + x_{i,t}\sigma_{x_j} dW_t^{x_j} + \rho_{ij}\sigma_i\sigma_j x_{i,t} x_{j,t} dt). \tag{19}$$

Comparing $d\theta_t^y d\theta_t^y = \sigma_\theta^2 dt$ with the coefficient of dt in $dx_{i,t} dx_{i,t}$ yields

$$\begin{aligned}\sigma_\theta^2 = &\sum_{i=1}^n b_i^2 \sigma_{x_i}^2 + \sum_{i=1}^n \left(\sigma_{x_i}^2 \sum_{j\neq i}^n b_{ij} x_{j,t}^2 \right) + 2\sum_{i=1}^n \left(b_i \sigma_{x_i}^2 \sum_{j\neq i}^n b_{ij} x_{j,t} \right) \\ &+ \rho_{ij} \left(\sum_{j\neq i}^n \sigma_{x_i}\sigma_{x_j} b_i b_j + 2\sum_{j\neq i}^n \sigma_{x_i}\sigma_{x_j} b_{ij} b_i x_{i,t} + \sum_{j\neq i}^n \sigma_{x_i}\sigma_{x_j} b_{ij}^2 x_{i,t} x_{j,t} \right).\end{aligned} \tag{20}$$

For a case without interaction ($b_{ij} = 0$) and assuming independence of Wiener processes ($\rho_{ij} = 0$), Eq. (20) equals $\sigma_\theta^2 = \sum_{i=1}^n b_i^2 \sigma_{x_i}^2$. In general, Eq. (20) suggests that σ_θ^2 depends on the stochastic variable $x_{i,t}, i = 1, 2, \ldots, p$. In order to quantify σ_θ^2, the expected value of σ_θ^2 is computed for empirical data analysis. When $x_{i,t}$ is a Brownian motion random variable (i.e. $E[x_{i,t}] = 0$ and $E[x_{i,t}^2] = \sigma_{x_i}^2 t$), one has

$$E[\sigma_\theta^2] = \sum_{i=1}^p b_i^2 \sigma_{x_i}^2 + \sum_{i=1}^p \sigma_{x_i}^2 \sum_{j\neq i}^p b_{ij}^2 \sigma_{x_j}^2 t. \tag{21}$$

When $x_{i,t}$ is an OU process random variable, one has

$$E[\sigma_\theta^2] = \sum_{i=1}^p b_i^2 \sigma_{x_i}^2 + \sum_{i=1}^p \sigma_{x_i}^2 \sum_{j\neq i}^p b_{ij}^2 E[x_{j,t}^2] + 2\sum_{i=1}^p b_i \sigma_{x_i}^2 \sum_{j\neq i}^p b_{ij} E[x_{j,t}], \tag{22}$$

where $E[x_{j,t}] = x_0 \exp(-\alpha_x t) + \theta_x(1 - \exp(-\alpha_x t))$ and $E[x_{j,t}^2] = \sigma_{x_j}^2[1 - \exp(-2\alpha_x t)]/(2\alpha_x) + [x_0 \exp(-\alpha_x t) + \theta_x(1 - \exp(-\alpha_x t))]^2$.

A conceptional simulation scheme for variables $(y, \theta_y, \tau_y, x_1, x_2)$ at each node using a 3 species tree is shown in Fig. 3 for illustration.

3. Inference

3.1. Approximate Bayesian computation

As mentioned in Section 2.2, the closed form likelihood equations for y_t for the OUOUCIR model and the OUBMCIR model are intractable. To conduct statistical inference on the parameters of interest, I propose to use the Approximate Bayesian Computation (ABC) approach (Sisson et al., 2019). To initiate the ABC approach, samples of a parameter vector Θ are drawn from a specified joint prior probability distribution. Then Θ is used to simulate L replicates of trait $\boldsymbol{Y}_l, \mathcal{X}_l, l = 1, 2, \ldots, L$ under the model $\mathcal{M}$ where $\mathcal{X}_l = (\boldsymbol{X}_1, \boldsymbol{X}_1, \ldots, \boldsymbol{X}_m)$ is the set of m covariate traits. A set of summary statistics $S(\boldsymbol{Y}_l, \mathcal{X}_l), l = 1, 2, \ldots, L$ are computed from the simulated data and each is compared to the summary statistics of the raw data $S(\boldsymbol{Y}, \mathcal{X})$ using a Euclidean distance measure d. If the distance between $S(\boldsymbol{Y}_l, \mathcal{X}_l)$ and $S(\boldsymbol{Y}, \mathcal{X})$ is less than a given threshold δ (i.e. $d(S(\boldsymbol{Y}_l, \mathcal{X}_l), S(\boldsymbol{Y}, \mathcal{X})) < \delta$), then the sample Θ_l is accepted. Note that before computing the distance, I followed the suggestions in Blum (2010) and scaled each summary statistic by a robust estimate of the standard deviation (the median absolute deviation).

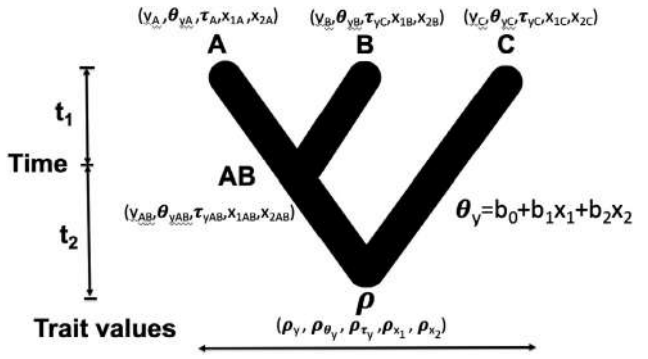


Fig. 3. A rooted phylogenetic tree of three species with tips A,B,C, root node ρ and an internal node AB. The response y_t, its optimum θ_t^y and rate τ_t^y and two covariates $x_{1,t}, x_{2,t}$ are variables of interest. Given the root values $(\rho_y, \rho_{\theta_y}, \rho_{\tau_y}, \rho_{x_1}, \rho_{x_2})$, the values at internal node AB $(y_{AB}, \theta_{AB}, \tau_{AB}, x_{1AB}, x_{2AB})$ as well as the values at the tips A, B, C $(y_i, \theta_i, \tau_i, x_{1i}, x_{2i})$, $i = A, B, C$ are simulated using postorder tree traversal algorithm under the model of adaptive trait evolution described in Section 2.3.

Table 3
Explicit formula of the summary statistics. Summary statistics are computed using tree $\mathbb{T}$ and trait data $\mathcal{X} = (\boldsymbol{X}_1, \boldsymbol{X}_1, \ldots, \boldsymbol{X}_m)$, $\boldsymbol{Y} = (y_1, y_2, \ldots, y_n)'$, formulae for summary statistics are presented using $\boldsymbol{Y}$. In Blomberg's K statistics, $\boldsymbol{C}$ is the phylogenetic covariance matrix transformed from the tree (Jhwueng, 2013; Adams, 2014). There is no explicit formula for λ in Pagel's λ model. The maximum likelihood estimate $\hat{\lambda}$ is reported to determine whether data exhibit significant phylogenetic dependence or not (Freckleton, 2012). Note that in the median statistics # is the count frequency, $\lfloor\cdot\rfloor$ is the floor function and $\lceil\cdot\rceil$ is the ceiling function.

Summary statistics	Formula
Mean	$\sum_i y_i/n$
Variance	$\sum_i (y_i - \bar{y})^2/(n-1)$
Median	$\frac{\boldsymbol{Y}\lfloor(\#\boldsymbol{Y}+1)\div 2\rfloor + \boldsymbol{Y}\lceil(\#\boldsymbol{Y}+1)\div 2\rceil}{2}$
Skewness	$\frac{\frac{1}{n}\sum_i (y_i-\bar{y})^3}{(\frac{1}{n}\sum_i (y_i-\bar{y})^2)^{3/2}}$
Kurtosis	$\frac{\frac{1}{n}\sum_i (y_i-\bar{y})^4}{(\frac{1}{n}\sum_i (y_i-\bar{y})^2)^2} - 3$
K	$K = \frac{(\boldsymbol{Y}-E(\boldsymbol{Y}))'(\boldsymbol{Y}-E(\boldsymbol{Y}))}{(\boldsymbol{Y}-E(\boldsymbol{Y}))'\boldsymbol{C}^{-1}(\boldsymbol{Y}-E(\boldsymbol{Y}))} \Big/ \frac{tr(\boldsymbol{C})-n(\mathbf{1}'\boldsymbol{C}^{-1}\mathbf{1})^{-1}}{n-1}$
λ	$\hat{\lambda}$

3.1.1. Summary statistics for models

As the OUBMCIR model and the OUOUCIR model fall out of the exponential family of distributions, it is theoretically infeasible to quantify all finite dimensional sufficient statistics. Nevertheless, it is still possible to implement non-sufficient statistics when inference is under the ABC framework. The inference could be more accurate with high efficiency if the summary statistics utilize all the information in the data. However, ABC fails to be accurate when using too many summary statistics because the distance increases with the number of summary statistics. Therefore to establish a procedure for choosing good summary statistics for ABC, I focus on choosing summary statistic on a pragmatic basis by making use of tree $\mathbb{T}$ and trait $\boldsymbol{Y}$, $\mathcal{X}$ so that the statistics could capture the relevant behavior of the model. Clarke et al. (2017) used the mean and the variance of the differences between each species and its closest neighbor in trait space for BM and OU model as the summary statistics (Csilléry et al., 2010). Bartoszek and Liò (2018) proposed to look at the sample mean and variance of the trait values in their R package `pcmabc`. O'Meara (2019) uses twenty-five summary statistics where eight summary statistics includes (i) the raw mean, (ii) the raw max, (iii) the raw min, (iv) the raw variance, (v) the raw median, (vi) the phylogenetic independent contrasts, (vii) the ancestral state reconstruction values and (viii) the range of ancestral state reconstruction confidence interval and seventeen from the `geiger::fitContinuous.hacked` function (Harmon et al., 2008), but these are then weighted for utility.

By considering the characteristics of the models, 12 summary statistics are used: the raw mean, the raw variance, the raw median, the raw skewness and the raw kurtosis of the given traits, and the mean, variance, median, skewness and kurtosis of the differences between each species and its closet neighbor suggested in Clarke et al. (2017) and the overall phylogenetic signal as measured by Bloomberg's K statistics which is a scale ratio of the variance among species over the contrasts variance (Blomberg et al., 2003) and Pagel's λ model which is a scaling parameter of the correlations between species, relative to the correlation expected under Brownian motion (Pagel, 1999). Note that smaller value of K indicates higher phylogenetic signal and larger value of λ indicates that traits are more similar amongst species than expected from their phylogenetic relationship. The explicit formula of the summary statistics is presented in Table 3.

The K statistics and Pagel's λ are computed by R package `phytools:phylosig` (Revell, 2012) while the phylogenetic independent contrast is computed in R package `ape:pic` (Paradis and Schliep, 2018). The twelve summary statistics should capture the central tendency, the overall dispersion, the level of asymmetry, and the tailedness of trait values, the signal embedded in the phylogenetic structuring of the trait values, and an even overall distribution of traits across the full resolved phylogenetic tree. After choosing appropriate summary statistics, a tolerance rate (accepted percentage of simulations) is set by default of 1 percent ($\delta = 0.01$) to get reliable results.

Then the posterior distribution of the parameters can be approximated using the accepted Θ_is. Algorithm 1 describes the ABC procedure used via `abc` package.

Algorithm 1 Approximate Bayesian Computation for $\mathcal{M}_1$: OUBMBM [$\Theta_1 = (\alpha_y, \sigma_x, \tau)$], $\mathcal{M}_2$: OUOUBM [$\Theta_2 = (\alpha_y, \alpha_x, \theta_x, \sigma_x, \tau)$], $\mathcal{M}_3$: OUBMCIR [$\Theta_3 = (\alpha_y, \sigma_x, \alpha_\tau, \theta_\tau, \sigma_\tau)$] and $\mathcal{M}_4$: OUOUCIR [$\Theta_4 = (\alpha_y, \alpha_x, \theta_x, \sigma_x, \alpha_\tau, \theta_\tau, \sigma_\tau)$] models.

Input: Tree $\mathbb{T}$ with known topology and branch lengths, starting value Θ_0, trait data set $\mathcal{D} = (\boldsymbol{Y}, \mathcal{X})$, root state $(\rho_y, \rho_{\theta_y}, \rho_{\tau_y}, \rho_{\mathcal{X}})$, prior distribution $\pi_i(\Theta)$, $i = 1, 2, 3, 4$ and a tolerance δ.

Output: Posterior sample for building up posterior distribution.

```
Compute summary statistics S(D) for raw data.
for each model M do
    for l = 1, ..., L do
        simulate sample Θ_l from prior π(Θ_0).
        simulate trait D_l = (Y_l, X_l) from Θ_l under model M.
        Compute summary statistics S(D_l) for simulated data.
    end for
end for
Apply ABC method using data S(D), S(D_l) and tolerate rate δ.
return Accepted posterior samples Θ_l for each model M, l = 1, 2, ..., ⌊Lδ⌋ where ⌊x⌋ returns the greatest integer less than or equal to x.
```

3.2. Model selection

The posterior probability $\Pr(\mathcal{M}|\mathcal{D})$ of a model $\mathcal{M}$ given data $\mathcal{D}$ is given by Bayes' theorem:

$$\Pr(\mathcal{M}|\mathcal{D}) = \frac{\Pr(\mathcal{D}|\mathcal{M})\Pr(\mathcal{M})}{\Pr(\mathcal{D})}. \tag{23}$$

I then compute the Bayes factors (BF) defined as a ratio of the likelihood probability of two different models $\mathcal{M}_i$ and $\mathcal{M}_j$, parameterized by model parameter vectors Θ_i and Θ_j. The Bayes factor BF is given by

$$\mathrm{BF}_{ij} = \frac{\Pr(\mathcal{D}|\mathcal{M}_i)}{\Pr(\mathcal{D}|\mathcal{M}_j)} = \frac{\int \Pr(\Theta_i|\mathcal{M}_i)\Pr(\mathcal{D}|\Theta_i, \mathcal{M}_i)\mathrm{d}\Theta_i}{\int \Pr(\Theta_j|\mathcal{M}_j)\Pr(\mathcal{D}|\Theta_j, \mathcal{M}_j)\mathrm{d}\Theta_j} = \frac{\Pr(\mathcal{M}_i|\mathcal{D})}{\Pr(\mathcal{M}_j|\mathcal{D})}\frac{\Pr(\mathcal{M}_j)}{\Pr(\mathcal{M}_i)}. \tag{24}$$

Currently for the posterior samples under the multinomial logistic regression method, I use the function `postpr` in R: `abc` package (Csilléry et al., 2012) to compute the posterior model probabilities where the posterior probabilities are estimated using multinomial logistic regression. This approximation holds when the different models are a priori equally likely (i.e. $\mathrm{BF}_{ij} = \Pr(\mathcal{M}_i|\mathcal{D})/\Pr(\mathcal{M}_j|\mathcal{D})$ is the ratio of the two posterior probabilities), and the same number of simulations are performed for each model. The analysis starts with four models, each with 50,000 replicates. The Euclidean distance d is computed for each replicate with respect to the realization (true data): rather than using d for each model, it is d for all models, so some models will have many more replicates within this bound than others. The Bayes factor between two models is computed as the ratio of frequencies of samples from each of these models that are below the threshold. With the acceptance rate of 1 percent ($\delta = 0.01$), 500 replicates are selected out of 50,000 replicates for each model. With four models there are $50,000 \times 4 = 200,000$ distance measures. The 200,000 distances are sorted and the cutoff is determined at the 500th position. The frequency is counted for each model that has the distance smaller than the cutoff. For example, suppose that among the 500 distances, OUBMBM accounts for 20, OUOUBM accounts for 40, OUBMCIR accounts for 200, and OUOUCIR accounts for 240 times. Then the Bayes factor of OUOUCIR with respect to OUBMBM is computed by $240 \div 20 = 12$. Kass and Raftery (1995) suggested that a value BF more than 100 would show very strong support for $\mathcal{M}_1$ over $\mathcal{M}_2$, between 10 and 100 would show strong support for $\mathcal{M}_1$ over $\mathcal{M}_2$, between 3.2 and 10 show substantial support for $\mathcal{M}_1$ over $\mathcal{M}_2$, finally a value between 1 and 3.2 is barely mentioning for $\mathcal{M}_1$ over $\mathcal{M}_2$.

4. Results

4.1. Simulation

I used extensive simulations to assess the performance and validity of the models. The parameters used are independent in their prior distributions. Without a likelihood function, one cannot determine a conjugate prior, but some

Table 4
Simulation results for assessing the model performance under parameter estimation for model parameter using uniform prior in Section 4.1.1. Four models (OUBMBM, OUOUBM, OUBMCIR and OUOUCIR) were investigated. Four different taxa sizes 16, 32, 64, 128 are used. On each column, the mean and 95 percent credible interval using 2500 posterior samples are reported for each model parameter.

Model	Taxa	α_y	α_x	θ_x	σ_x^2	τ	α_τ	θ_τ	σ_τ
	True value	0.5	0.75	0	2.5	1	0.65	1.5	1
OUOUCIR	16	0.55 (0.04,0.97)	0.86 (0.06,1.46)	−0.04 (−2.83,2.83)	2.31 (0.35,4.75)		0.62 (0.04,1.26)	1.45 (0.08,2.9)	0.97 (0.05,1.94)
	32	0.57 (0.04,0.98)	0.75 (0.04,1.46)	−0.07 (−2.87,2.82)	2.63 (0.52,4.8)		0.66 (0.03,1.27)	1.61 (0.11,2.95)	0.93 (0.06,1.95)
	64	0.58 (0.04,0.98)	0.95 (0.08,1.48)	0.18 (−2.82,2.85)	2.71 (0.78,4.73)		0.62 (0.03,1.26)	1.59 (0.1,2.92)	0.87 (0.05,1.92)
	128	0.47 (0.02,0.97)	0.84 (0.06,1.47)	0.04 (−2.83,2.81)	2.34 (0.52,4.67)		0.63 (0.04,1.26)	1.5 (0.13,2.93)	1.08 (0.17,1.96)
OUBMCIR	16	0.51 (0.03,0.97)			2.15 (0.24,4.79)		0.66 (0.03,1.27)	1.48 (0.07,2.92)	0.97 (0.04,1.95)
	32	0.53 (0.03,0.98)			2.33 (0.42,4.75)		0.65 (0.03,1.27)	1.48 (0.07,2.93)	0.95 (0.05,1.95)
	64	0.54 (0.03,0.98)			2.49 (0.6,4.79)		0.66 (0.03,1.27)	1.43 (0.07,2.92)	0.88 (0.04,1.94)
	128	0.56 (0.04,0.98)			2.49 (0.64,4.75)		0.65 (0.03,1.27)	1.4 (0.07,2.91)	0.83 (0.05,1.92)
OUOUBM	16	0.52 (0.09,0.96)	0.66 (0.03,1.46)	0.03 (−2.85,2.85)	2.46 (0.56,4.8)	0.99 (0.05,1.93)			
	32	0.46 (0.03,0.97)	0.82 (0.06,1.47)	−0.09 (−2.82,2.84)	2.31 (0.38,4.68)	0.98 (0.05,1.95)			
	64	0.48 (0.09,0.97)	0.6 (0.04,1.44)	0.01 (−2.84,2.86)	2.38 (0.5,4.83)	0.86 (0.03,1.93)			
	128	0.46 (0.05,0.96)	0.74 (0.04,1.46)	0.09 (−2.83,2.83)	2.51 (0.65,4.72)	0.98 (0.04,1.95)			
OUBMBM	16	0.48 (0.03,0.98)			2.45 (0.39,4.77)	1.01 (0.05,1.96)			
	32	0.47 (0.02,0.98)			2.19 (0.46,4.76)	1.01 (0.05,1.95)			
	64	0.51 (0.03,0.97)			2.36 (0.29,4.76)	1.02 (0.05,1.95)			
	128	0.5 (0.02,0.98)			2.31 (0.45,4.69)	0.98 (0.05,1.95)			

reasonable prior distribution must still be selected (Gelman et al., 2017). For parameter estimation, models are assessed where two set of informative priors are used. The tree for simulations is a balanced tree of 16, 32, 64 and 128 taxa with a height of 1 with Grafen branch lengths computed with R package ape (Paradis and Schliep, 2018; Grafen, 1989). Five runs are performed, where for each run $L = 50{,}000$ samples were simulated and tolerance rate ($\delta = 0.01$) is used, resulting in $5 \times 500 = 2500$ posterior samples to get reliable estimates. See Sections 4.1.1 and 4.1.2. The regression with interaction ($x_i x_j$) for this simulation is not included and the focus is on new model with modeling rate parameter with CIR process. In addition, a simulation using a 32 taxon and a 64 taxon balanced tree was performed to estimate the Bayesian coverage probability of parameters in each model. See Section 4.1.3.

4.1.1. Uniform prior

With the assumption that the prior of parameter follows uniform distribution, the following shows the range for each parameter. For the OUOUCIR model, the prior distributions for the model parameters are $\alpha_y \sim \mathcal{U}(0, 1)$, $\alpha_x \sim \mathcal{U}(0, 1.5)$, $\theta_x \sim \mathcal{U}(-3, 3)$, $\sigma_x^2 \sim \mathcal{U}(0, 5)$, $\alpha_\tau \sim U(0, 1.3)$, $\theta_\tau \sim \mathcal{U}(0, 3)$, $\sigma_\tau^2 \sim \mathcal{U}(0, 2)$. For the OUBMCIR model, the prior distributions for the model parameters are $\alpha_y \sim \mathcal{U}(0, 1)$, $\sigma_x^2 \sim \mathcal{U}(0, 5)$, $\alpha_\tau \sim \mathcal{U}(0, 1.3)$, $\theta_\tau \sim \mathcal{U}(0, 3)$, $\sigma_\tau \sim \mathcal{U}(0, 2)$. For the OUOUBM model, the prior distributions for the model parameters are $\alpha_y \sim \mathcal{U}(0, 1)$, $\alpha_x \sim \mathcal{U}(0, 1.5)$, $\theta_x \sim \mathcal{U}(-3, 3)$, $\sigma_x^2 \sim \mathcal{U}(0, 5)$, $\tau \sim \mathcal{U}(0, 2)$. For the OUBMBM model, the prior distributions for the model parameters are $\alpha_y \sim \mathcal{U}(0, 1)$, $\sigma_x^2 \sim \mathcal{U}(0, 5)$, $\tau \sim \mathcal{U}(0, 2)$. The prior information for the regression parameter is set to $b_0 \sim \mathcal{U}(5, 15)$, $b_1 \sim \mathcal{U}(-1, 9)$, $b_2 \sim \mathcal{U}(-8, 2)$. The root $\boldsymbol{\rho} = (\rho_y, \rho_x, \rho_\theta, \rho_\tau)$ is set to $(0, 0, 0, 1)$ for all models. For each unique combination of model and taxon size, I simulated evolution of one trait. The true parameter is set to $\Theta = (\alpha_y, \alpha_x, \theta_x, \sigma_x^2, \alpha_\tau, \theta_\tau, \sigma_\tau) = (0.5, 0.75, 0, 2.5, 0.65, 1.5, 1)$ for the OUOUCIR model,$\Theta = (\alpha_y, \sigma_x^2, \alpha_\tau, \theta_\tau, \sigma_\tau) = (0.5, 2.5, 0.65, 1.5, 1)$ for the OUBMCIR model, $\Theta = (\alpha_y, \alpha_x, \theta_x, \sigma_x^2, \tau) = (0.5, 0.75, 0, 2.5, 1)$ for the OUOUBM model and $\Theta = (\alpha_y, \sigma_x^2, \tau) = (0.5, 2.5, 1)$ for the OUBMBM model. The true value for regression parameters is set to $\beta = (\beta_0, \beta_1, \beta_2) = (10, 4, -3)$.

Simulation results of model parameters are shown in Table 4. Overall, parameters can be estimated more accurately. The posterior mean for each parameter is close to the true parameter value.

Simulation result for regression parameters is shown in Table 5. Most models on each taxa report satisfactory mean estimates for β_0 (true 10), β_1 (true 4) and β_2 (true -3).

The density plots of posterior samples of parameter are shown in Table 6. Overall, most parameters have their posterior mean close to the true value. This result assures that Algorithm 1 provides reliable procedure for parameter estimation. From the posterior density plots, trait data affect the posterior distribution on the parameter $\alpha_y, \alpha_x, \sigma_x^2, \sigma_\tau^2$, rate parameter τ, σ_τ^2 and regression parameter β_1, β_2 while fewer impact on the mean parameter $\theta_x, \theta_\tau, \beta_0$ is observed (posterior and prior look more similar). Noted that unlike model based likelihood approach where the posterior mean is a function of the mean of the prior and likelihood, model with intractable likelihood may have its posterior distribution similarly to the prior.

4.1.2. Informative priors

The second set of priors are set to specific distributions based on intuitive beliefs about the true values of the parameters. Following Uyeda and Harmon (2014), the prior distribution for θ is normal, which keeps the optimal peaks in a biologically reasonable part of space. One example of using an empirical prior in this way is from a study of bird tail length evolution (Jhwueng and Maroulas, 2014), where a reasonable prior for the adaptive optima θ of tail length used the distribution of tail length across all birds. Uyeda and Harmon (2014) used a log normal for the force parameter α and for the rate parameter σ. Here I set α to be the exponential distribution and σ to be the inverse gamma distribution. For the OUOUCIR model, the prior distributions for the model parameters are $\alpha_y \sim \mathcal{E}(\text{rate} = 5), \alpha_x \sim \mathcal{E}(\text{rate} = 8), \theta_x \sim \mathcal{N}(\text{mean} = 0, \text{sd} = 1), \sigma_x^2 \sim \mathcal{IG}(\text{shape} = 2, \text{scale} = 0.5), \alpha_\tau \sim \mathcal{E}(\text{rate} = 4),$

Table 5
Simulation results for assessing the model performance under parameter estimation for regression parameter using informative prior. Four models (OUBMBM, OUOUBM, OUBMCIR and OUOUCIR) were investigated. Four different taxa sizes 16, 32, 64,128 are used. On each column, the mean and 95 percent credible interval using 2500 posterior samples are reported for each regression parameter.

Model	Taxa True value	b_0 10	b_1 4	b_2 −3
OUOUCIR	16	10.06 (5.2,14.69)	3.78 (−0.73,8.71)	−3.15 (−7.73,1.78)
	32	10.05 (5.28,14.74)	4.38 (−0.73,8.8)	−2.9 (−7.76,1.66)
	64	10.05 (5.21,14.76)	3.88 (−0.75,8.75)	−3.16 (−7.77,1.7)
	128	9.94 (5.25,14.76)	3.69 (−0.78,8.71)	−2.8 (−7.64,1.76)
OUBMCIR	16	10.01 (5.25,14.72)	3.62 (−0.76,8.67)	−2.71 (−7.7,1.76)
	32	10.1 (5.25,14.76)	3.88 (−0.75,8.67)	−2.94 (−7.7,1.76)
	64	9.89 (5.23,14.77)	3.89 (−0.73,8.67)	−3 (−7.69,1.75)
	128	9.99 (5.26,14.77)	3.89 (−0.71,8.66)	−3.23 (−7.68,1.74)
OUOUBM	16	9.93 (5.28,14.73)	4.41 (−0.6,8.71)	−2.94 (−7.67,1.72)
	32	9.91 (5.31,14.79)	4.26 (−0.6,8.74)	−2.24 (−7.68,1.8)
	64	10.06 (5.3,14.79)	4.31 (−0.74,8.76)	−2.58 (−7.58,1.84)
	128	9.92 (5.29,14.74)	3.85 (−0.8,8.68)	−2.78 (−7.67,1.77)
OUBMBM	16	9.88 (5.26,14.81)	3.3 (−0.78,8.58)	−2.51 (−7.56,1.79)
	32	9.9 (5.21,14.79)	3.33 (−0.8,8.63)	−2.65 (−7.71,1.74)
	64	9.97 (5.22,14.73)	3.68 (−0.7,8.64)	−2.81 (−7.65,1.74)
	128	10.04 (5.21,14.73)	3.52 (−0.83,8.63)	−2.55 (−7.56,1.77)

$\theta_\tau \sim \log\mathcal{N}(\text{meanlog} = 1.5, \text{sdlog} = 0.2)$, $\sigma^2_\tau \sim \mathcal{IG}(\text{shape} = 2, \text{scale} = 1)$. For the OUBMCIR model, the prior distributions for the model parameters are $\alpha_y \sim \mathcal{E}(\text{rate} = 5)$, $\sigma^2_x \sim \mathcal{IG}(\text{shape} = 2, \text{scale} = 2)$, $\alpha_\tau \sim \mathcal{E}(\text{rate} = 4)$, $\theta_\tau \sim \log\mathcal{N}(\text{mean} = 0, \text{sd} = 1)$, $\sigma_\tau \sim \mathcal{IG}(\text{shape} = 2, \text{scale} = 1)$. For the OUOUBM model, the prior distributions for the model parameters are $\alpha_y \sim \mathcal{E}(\text{rate} = 8)$, $\alpha_x \sim \mathcal{E}(\text{rate} = 5)$, $\theta_x \sim \mathcal{N}(\text{mean} = 0, \text{sd} = 1)$, $\sigma^2_x \sim \mathcal{IG}(\text{shape} = 2, \text{scale} = 2)$, $\tau \sim \mathcal{IG}(\text{shape} = 2, \text{scale} = 1)$ where $\mathcal{N}$ is normal distribution. For the OUBMBM model, the prior distributions for the model parameters are $\alpha_y \sim \mathcal{E}(\text{rate} = 5)$, $\sigma^2_x \sim \mathcal{IG}(\text{shape} = 2, \text{scale} = 2)$, $\tau \sim \mathcal{IG}(\text{shape} = 2, \text{scale} = 1)$ where $\mathcal{E}$ is an exponential distribution and $\mathcal{IG}$ is an inverse gamma distribution. The prior information for the regression parameter is set to $b_0 \sim \mathcal{U}(-5, 5)$, $b_1 \sim \mathcal{U}(-7, 3)$, $b_2 \sim \mathcal{U}(-2, 8)$. Others recommend using weakly informative priors from the t distribution with 1 degree of freedom (a Cauchy distribution) (Gelman et al., 2008). The root $\boldsymbol{\rho} = (\rho_y, \rho_x, \rho_\theta, \rho_\tau)$ is set to (0, 0, 0, 1) for all models. For each unique combination of model and taxon size, I simulated evolution of one trait. For this set simulation, true parameter values are set to $\Theta = (\alpha_y, \alpha_x, \theta_x, \sigma^2_x, \alpha_\tau, \theta_\tau, \sigma_\tau) = (0.2, 0.125, 0, 0.5, 0.25, 1.5, 1)$ for the OUOUCIR model, $\Theta = (\alpha_y, \sigma^2_x, \alpha_\tau, \theta_\tau, \sigma_\tau) = (0.2, 0.5, 0.25, 1.5, 1)$ for the OUBMCIR model, $\Theta = (\alpha_y, \alpha_x, \theta_x, \sigma^2_x, \tau) = (0.2, 0.125, 0, 0.5, 1)$ for the OUOUBM model and $\Theta = (\alpha_y, \sigma^2_x, \tau) = (0.2, 0.5, 1)$ for the OUBMBM model. The true value for regression parameters is set to $\beta = (\beta_0, \beta_1, \beta_2) = (0, 2, -3)$.

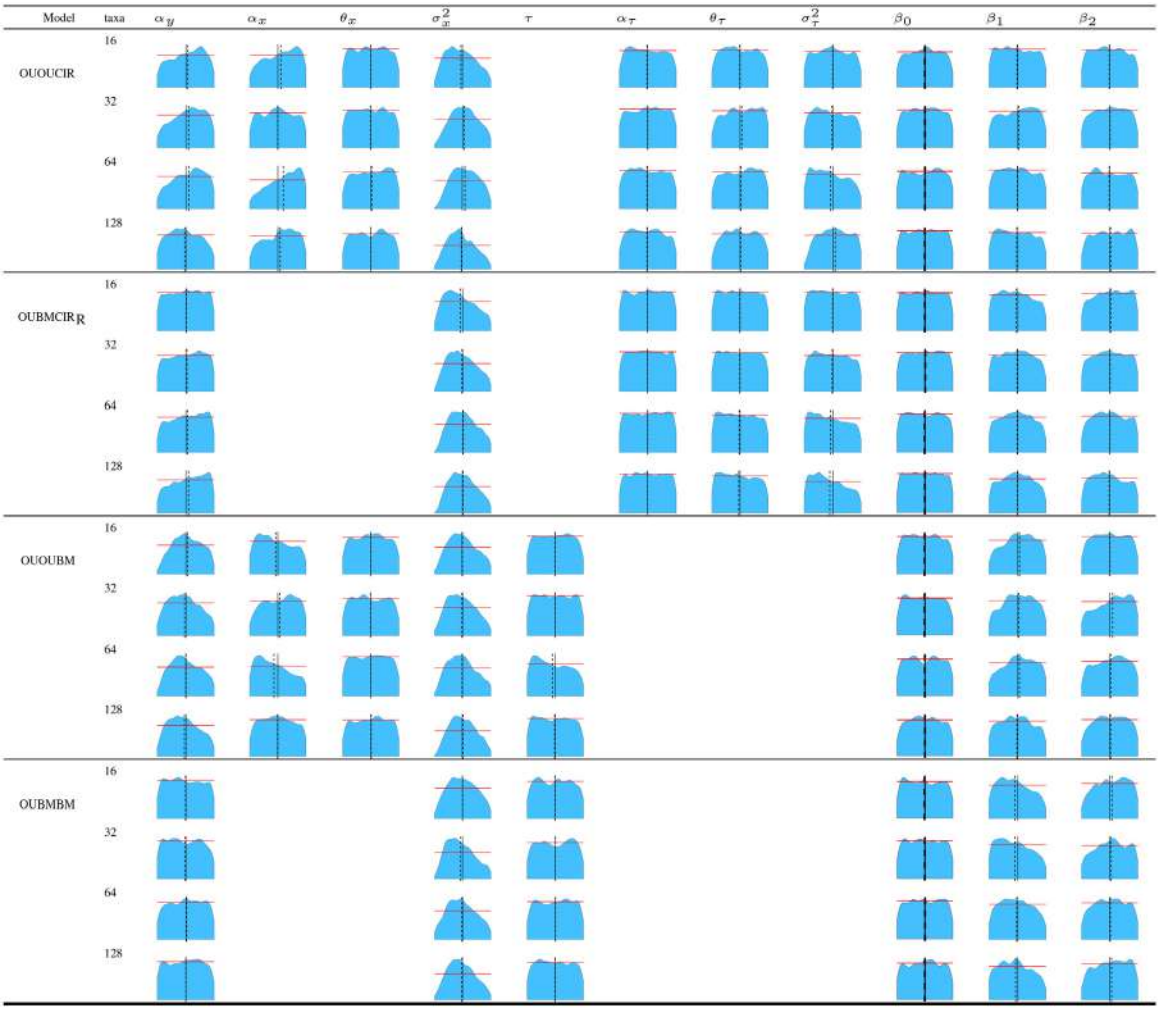
Results can be accessed in online supplemental material Table S1, Table S2 and Table S3. For this set of simulation, the posterior mean for each parameter is closed to its true value while the constructed 95% credible interval from the posterior sample covers the true parameter.

4.1.3. Coverage probability for credible interval

A p percent credible interval $\mathcal{I}_p = [\mathcal{I}_l, \mathcal{I}_r]$ in Bayesian context represents the shortest interval that covers a p percentage of the probability distribution on the domain. The probability densities within the credible interval are typically greater than those outside the interval even for skewed distribution (Baek et al., 2016). To estimate the coverage probability, simulations using a 32 and a 64 taxon balanced trees are performed separately with 100 runs where each run starts with $L = 20{,}000$ replicates for each model with tolerance rate $\delta = 0.01$ resulting 200 posterior samples. All priors are set to uniform distribution, and samples are simulated using the set up in Section 4.1.1.

A p percent Bayesian credible interval is computed using the high density interval (HDI) method implemented in R package `BayestestR` (Makowski et al., 2019) to guarantee all points within the interval have higher probability density than points outside the interval in the context of uncertainty characteristic of posterior distribution. The coverage probability is estimated using the 100 intervals from the 100 runs. It is expected that given a p percent credible interval, the true value across sims falls in there say around p percent of the time. For instance, for a 60% credible interval, the true value across sims falls in there say 66% of the time, at the 80% it falls in there 75% of the time, etc. The coverage probability for each model parameter in the OUOUCIR model is shown in Fig. 4 where the estimated coverage probability from 5 percent to 95 percent is reported by dot plots. Overall, for both 32 taxa and 64 taxa balanced tree cases simulations result show that most parameters have higher estimated coverage probability than the expected coverage probability. This may due to the asymmetric posterior distributions where the p percent credible interval built by the high density intervals (HDI) method is wider on one side of the interval and hence could have higher chance to cover the true parameter values. Results of the coverage probability for the OUBMCIR model are similar to the result here (see online supplemental material Figure S1).

Table 6
Distribution plots using uniform priors are reported for four models OUBMBM, OUOUBM, OUBMCIR and OUOUCIR. On each plot, the prior distribution (red curve), posterior distribution (blue density plot) from simulation results are generated and the true parameter value is plotted with a vertical line. Four different taxa sizes 16, 32, 64 and 128 are used. Note that due to each model lack likelihood, the posterior mean is not a combination of the prior mean and the likelihood mean as shown in the conventional likelihood based approach.



4.2. Cross validation

To address whether the correct model can be chosen, and how much data is required for this, an analysis study using cross validation is performed with varying the number of taxa. This is done using the R: `abc` package, which can do a leave-one-out cross validation for ABC (Csilléry et al., 2012). Two type of trees are used. For trees with 64, 128, 256 and 512 taxa, a balanced tree is simulated using the R: `ape` package. For trees with 50, 100, 200 and 500 taxa, 10,000 birth–death trees are simulated using the R: `TreeSim` package (Stadler, 2010) with speciation rate 2, extinction rate 0.5, probability of each tip is included into the final tree 0.6 and the time since origin 2. Given a tree, a trait data is simulated using parameters $\Theta = (\alpha_y, \sigma_x^2, \tau) = (0.2, 2, 1)$ for the OUBMBM model, $\Theta = (\alpha_y, \alpha_x, \theta_x, \sigma_x^2, \tau) = (0.2, 0.125, 0, 2, 1)$ for the OUOUBM model, $\Theta = (\alpha_y, \sigma_x^2, \alpha_\tau, \theta_\tau, \sigma_\tau^2) = (0.2, 2, 0.25, 0.5, 1)$ for the OUBMCIR model and for the OUOUCIR model $\Theta = (\alpha_y, \alpha_x, \theta_x, \sigma_x^2, \alpha_\tau, \theta_\tau, \sigma_\tau^2) = (0.2, 0.125, 0, 2, 0.25, 0.5, 1)$. The regression parameters are set to $(\beta_0, \beta_1, \beta_2) = (0, 1, 0.5)$.

To evaluate if ABC can distinguish between the four models, the twelve summary statistics are calculated in each model and the size of the cross validation sample for each model is set to 100.

Results of the confusion matrices based on 100 accepted samples for each model using the balanced tree are reported using bar plots in Fig. 5. For example, for sixty four taxa case in the upper left panel: taxa size 64 of Fig. 5, the right most bar plot set shows that among 100 samples, ABC model choice will identify the OUOUCIR model(green bar) for 65 times, the OUBMCIR model (blue bar) for 22 times, the OUBMBM model (pink bar) for 9 times and the OUOUBM model (orange

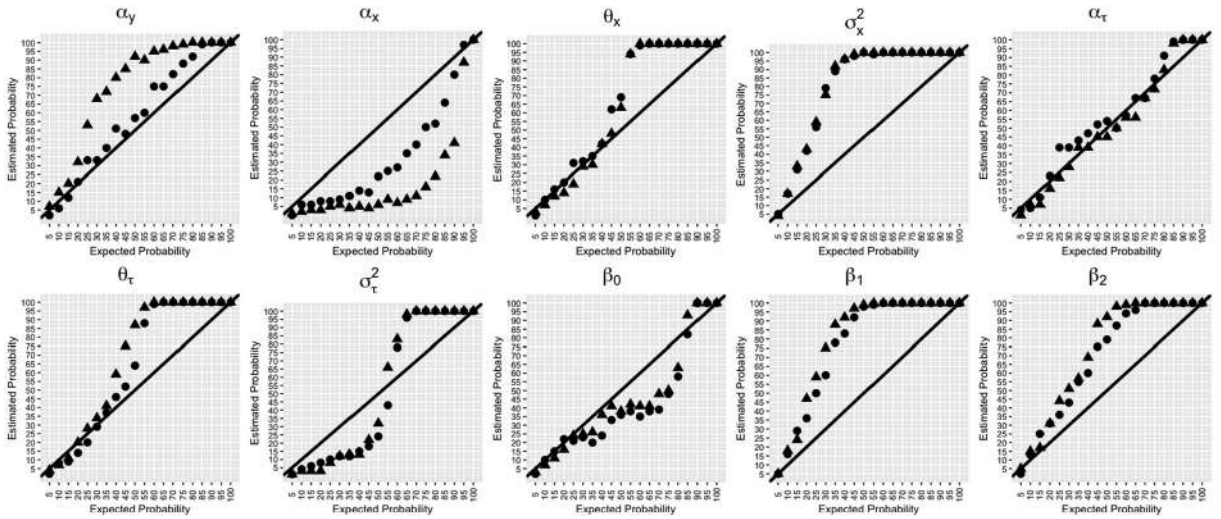


Fig. 4. Bayesian coverage probability for the OUOUCIR model. The horizontal axis shows the expected interval size (5% credible interval to 95% credible interval) and the vertical axis represents the frequency across the sims where the true parameter value falls within that credible interval. For 32 taxa the coverage probabilities are shown with circles while for 64 taxa the coverage probabilities are shown with triangles.

bar) for 4 times which accounts for misclassification proportions of $(1-\frac{65}{100})\times 100\% = 35\%$ for the OUOUCIR model while in the lower left panel of Fig. 5, the right most bar plot set of the taxa size of 512 case shows that for among 100 samples ABC model choice will identify the OUOUCIR model for 84 times, the OUBMCIR model for 15 times the OUBMBM model for 1 times and the OUOUBM for 0 times, which account for misclassification proportions of $1-\frac{84}{100}\times 100\% = 16\%$ for the OUOUCIR model.

In general, the frequency for ABC identifying the correct models increases with the taxa size. It is clear from this analysis that it is easier to distinguish at each taxa size. In particular, the OUOUCIR model (green bars in the right most panel across taxa sizes) constitutes a better fit for the data as supported with enough power in model choice. while the OUBMCIR model requires more taxa to identify the correct model (e.g. at taxa size 128 the blue bar in the OUBMCIR category is over 60% while other models are correctly identified more than 80% of the time). Note that factors beyond the number of taxa play a role: parameter choices that make the models create very different results can make them easier to distinguish, for example. All these factors are held constant here.

For the birth–death tree case, results for the OUBMBM, the OUOUBM and the OUOUCIR models are similar to the results in the balanced tree case. The OUBMCIR model requires more taxa to identify the correct model (see online supplemental material Figure S4). By comparing Fig. 5 and online supplemental material Figure S4, it is easier to distinguish models by using a balanced tree than using the birth–death tree. This is due to the death–birth tree is randomly simulated for each replication while the balanced tree is simulated once and used throughout the analysis.

4.3. Empirical data analysis

A compelling example to use these new models on is the branched coral morphology dataset of Sanchez and Lasker (2003a). They measured five traits of coral structure and used independent contrasts (Felsenstein, 1985) to examine correlations between these traits after correcting for a phylogeny, which they also inferred. Sanchez and Lasker (2003a) report a parsimony topology without meaningful branch lengths. For the models in this paper, branch lengths can have a substantial effect on results; for this example, I constructed arbitrary branch lengths (recovered from the arbitrary lengths used in the original paper figure), but a biologist using this software would be well-advised to construct a chronogram with branch lengths proportional to time. The phylogeny and trait data are shown in Fig. 6.

Corals are cnidarians (the same clade containing jellyfish and sea anemones). The ones in this paper form branched structures, with each branch formed by many coral polyps (which resemble tiny anemones) forming cells around them. While, Sanchez and Lasker (2003a) examine five traits of coral morphology, for simplicity I focus on just three: the thickness of the coral branches (T), the distance between polyps (D) and the polyp aperture (P): the hole through which the polyp sticks its tentacles. Sanchez and Lasker (2003a) found significant positive pairwise correlations between all these traits. Here, I use a model where optimal branch thickness (θ_y: T) is dependent on inter-polyp distance (x_1: D) and coral aperture (x_2: P) represented in Eq. (6). Inter-polyp distance and branch thickness evolve under OU in an OUOU** model and under BM under an OUBM** model. The gain of variance of branch thickness per increment of time can be either due to a Brownian model under OU**BM models or under a CIR model under OU**CIR. Traits are logarithm transformed into ratio scale for the rational of interpretation that the change of trait is measured in percentage for each species.

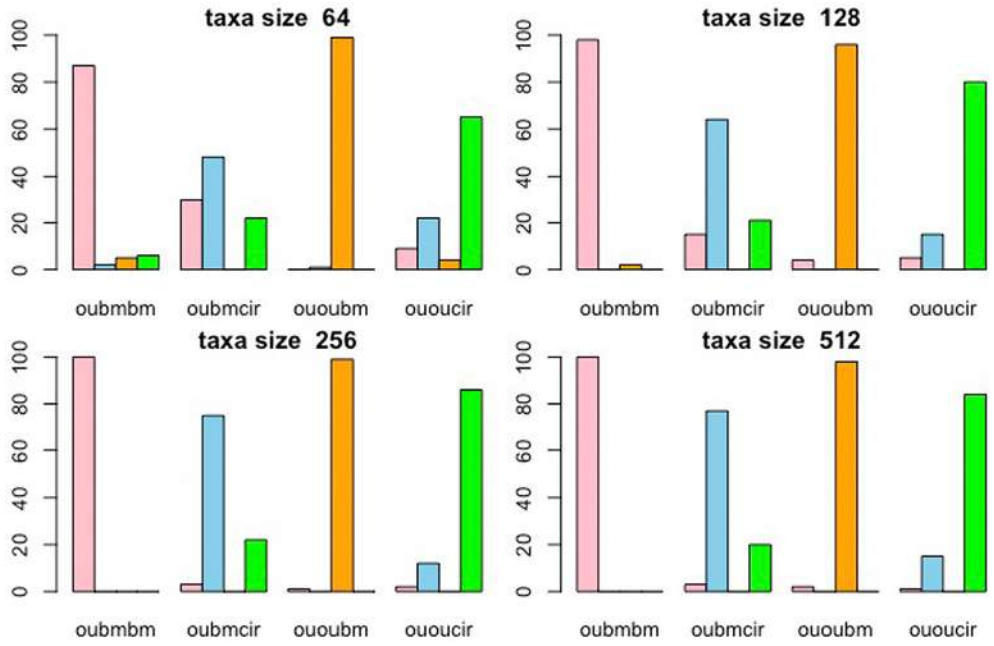


Fig. 5. Cross validation of models using balanced tree. Bar plots for confusion matrices from cross validation analysis under ABC multinomial logistic regression method for models of adaptive trait evolution. Four taxa sizes (64, 128, 256, 512) are considered for cross-validation. On each panel, the horizontal label for each bar plot represents the actual model, and the height of the bar plot represents frequency of correctly identifying the models (OUBMBM in pink, OUBMCIR in sky blue, OUOUBM in orange and OUOUCIR in green) for 100 samples. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

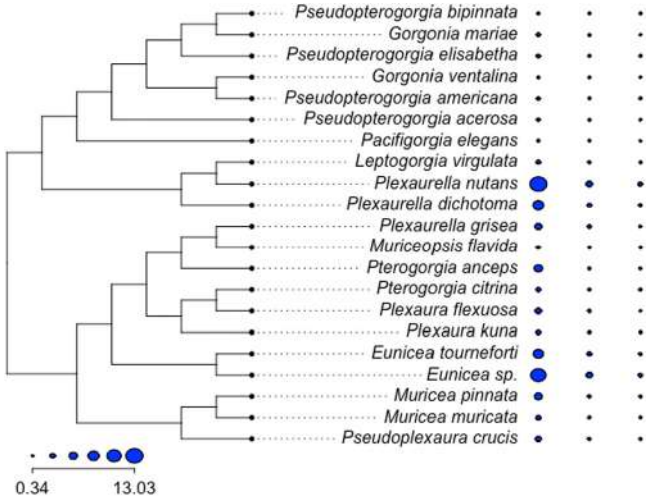


Fig. 6. Coral dataset in Sanchez and Lasker (2003a) used for this study. Traits data and their size were displayed using blue circles (from left to right: thickness (T), distance (D) and polyp (P)). Species names are shown in the center. The phylogenetic tree is adopted from Sanchez and Lasker (2003a) where the author used bootstrap majority-rule consensus tree. The branch lengths of the tree are measured using a ruler. To make it truly ultrametric, these distances are used to create a distance matrix and then converted it into a tree object under unweighted pair group method with arithmetic mean (UPGMA) method encoded in R: `phangorn` (Schliep, 2011). Plot was generated by R: `phytool` (Revell, 2012).

For choosing starting points for parameters, the MLE estimate $\hat{\Theta} = (\hat{\theta}, \hat{\sigma}^2)$ of BM model is used when covariate $x_i, i = 1, 2$ follows BM and the MLE estimate $\hat{\Theta} = (\hat{\alpha}, \hat{\theta}, \hat{\sigma}^2)$ of OU model is used when covariate $x_i, i = 1, 2$ follows OU from R: Geiger package `fitContinuous` (Harmon et al., 2008).

Since trait data are only observed on the tip with tree of known topology and branch length, the ancestral trait on internal nodes is simulated under the method described in Algorithm 1. In this situation, traits may vary with larger deviation on the internal states. Therefore, a wide range values for priors are used for empirical study.

Table 7
Bayes factor table for the coral dataset in Sanchez and Lasker (2003a). The posterior probability $P(\mathcal{M}|\mathcal{D})$ for each model is shown in the second row. The first row shows the models. The Bayes factor $\mathbf{BF}_{ij}$ for model $\mathcal{M}_i$ vs. model $\mathcal{M}_j$ is shown in the ith row and the jth column. For instance, the Bayes factor of the OUOUCIR model vs. the OUBMBM model is shown in the third row and the sixth column with value 7.610 which is computed by the ratio of the posterior model probability of the OUOUCIR model over the posterior model probability of the OUOUBM model (0.636/0.177 ≈ 3.589 after rounding).

Rank	$P(\mathcal{M}\|\mathcal{D})$	**0.636**	**0.177**	**0.103**	**0.084**
	Model $\mathcal{M}$	OUOUCIR	OUOUBM	OUBMCIR	OUBMBM
1st	OUOUCIR	1.000	**3.589**	**6.119**	**7.610**
2nd	OUOUBM	0.279	1.000	1.705	2.121
3rd	OUBMCIR	0.163	0.587	1.000	1.244
4th	OUBMBM	0.131	0.472	0.804	1.000

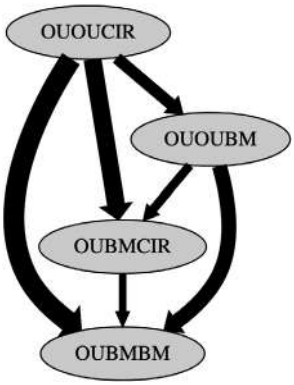


Fig. 7. Graphical display for comparing models of adaptive trait evolution with dynamic rate parameter $\tau(y_t, \Theta, t)$. The best model listed on top of the graph is the OUOUCIR model with posterior model probability ($P(\mathcal{M}|\mathcal{D}) = 0.636$, the second best model is the OUOUBM model ($P(\mathcal{M}|\mathcal{D}) = 0.177$), the third model is the OUBMCIR model ($P(\mathcal{M}|\mathcal{D}) = 0.103$) and the last model at the bottom of the graph is the OUBMBM model ($P(\mathcal{M}|\mathcal{D}) = 0.084$). Their Bayes factors are listed in Table 7.

All priors are set to uniform distribution. The corresponding range for each parameter including the natural lower bound 0 for nonnegative parameters or 3 times smaller than the parameter estimates for MLE estimates, and the upper bound is set to 3 times larger than the MLE estimates. For the OUBMBM model, $\alpha_y \sim \mathcal{U}(0, 3\hat{\alpha}_y)$, $\sigma_x^2 \sim \mathcal{U}(0, 3\sum_{i=1}^p \hat{\sigma}_{x_i}^2/p)$, $\tau \sim \mathcal{U}(0, 3(\hat{\sigma}_y^2 + \sum_{i=1}^p \hat{\sigma}_{x_i}^2)/(p+1))$. For the OUOUBM model, $\alpha_y \sim \mathcal{U}(0, 3\hat{\alpha}_y)$, $\alpha_x \sim \mathcal{U}(0, 3\hat{\alpha}_x)$, $\theta_x \sim \mathcal{U}(-3\sum_{i=1}^p \hat{\theta}_{x_i}/p, 3\sum_{i=1}^p \hat{\theta}_{x_i}/p)$, $\sigma_x^2 \sim \mathcal{U}(0, 3\sum_{i=1}^p \hat{\sigma}_{x_i}^2/p)$, $\tau \sim \mathcal{U}(0, 3(\hat{\sigma}_y^2 + \sum_{i=1}^p \hat{\sigma}_{x_i}^2)/(p+1))$. For the OUBMCIR model, $\alpha_y \sim \mathcal{U}(0, 3\hat{\alpha}_y)$, $\sigma_x^2 \sim \mathcal{U}(0, 3\sum_{i=1}^p \hat{\sigma}_{x_i}^2/p)$, $\alpha_\tau \sim \mathcal{U}(0, 3(\hat{\alpha}_y + \sum_{i=1}^p \hat{\alpha}_{x_i})/(p+1))$, $\theta_\tau \sim \mathcal{U}(0, 3(|\hat{\theta}_y + \sum_{i=1}^p \hat{\theta}_{x_i})|/(p+1))$. For the OUOUCIR model, $\alpha_y \sim \mathcal{U}(0, 3\hat{\alpha}_y)$, $\alpha_x \sim \mathcal{U}(0, 2\hat{\alpha}_x)$, $\theta_x \sim \mathcal{U}(-3\sum_{i=1}^p \hat{\theta}_{x_i}/p, 3\sum_{i=1}^p \hat{\theta}_{x_i}/p)$, $\sigma_x^2 \sim \mathcal{U}(0, 3\sum_{i=1}^p \hat{\sigma}_{x_i}^2/p)$, $\alpha_\tau \sim \mathcal{U}(0, 3(\hat{\alpha}_y + \sum_{i=1}^p \hat{\alpha}_{x_i})/(p+1))$, $\theta_\tau \sim \mathcal{U}(0, 3(|\hat{\theta}_y + \sum_{i=1}^p \hat{\theta}_{x_i})|/(p+1))$. For regression parameters, I first perform the generalized least squares (GLS) to obtain the estimates b_0, b_1, b_2 and their standard deviations, the lower and upper bounds for the priors distribution are set to the estimate plus/minus 3 standard deviations. I use the MLE estimates of OU for the root ρ_y, ρ_{θ_y} and ρ_{τ_y}), and use the MLE estimates of BM when x_1 and x_2 follow BM or use OU when x_1 and x_2 follow OU. When estimate $\hat{\alpha}_x \approx 0$ from OU in the case of assuming θ_t^y following OU, traits x_1, x_2 are instead simulated from BM model. I performed 50,000 simulations each under the OUBMBM, the OUOUBM, the OUBMCIR, and the OUOUCIR models. The posterior model probability of a given model is estimated using a multinomial logistic regression implemented in the function of `multinom` from R the package `nnet` (Venables and Ripley, 2002).

For Coral data in Sanchez and Lasker (2003a), the best model is the OUOUCIR model, followed by the OUBMBM model, the third model is the OUOUBM model and the last model is the OUBMCIR model. Their pairwise Bayes factors are shown in Table 7 and can be visualized in Fig. 7 using open source graph visualization software: Graphviz (Ellson et al., 2001). The best model (Rank = 1st) is the OUOUCIR model. This data set provides substantial support on the OUOUCIR model over the OUOUBM model with Bayes factor 3.589 which may show that the evolution of the rate τ_t^y is more appropriate described by the CIR process when comparing with the same model but with rate modeled by BM. This data set has substantial support on the OUOUCIR model over the OUBMCIR model with Bayes factor 6.119 which may indicate the optimum of thickness of coral evolved via a stabilized process (described by the OU process) rather than evolved with larger fluctuation (described by the BM) when modeling the rate τ_t^y under the CIR process. The dataset OUOUCIR model over the OUBMBM model with Bayes factor 7.610 for this dataset which may imply that the optima is more appropriate modeled by the OU process than BM and the rate is more appropriate modeled by CIR than BM.

Table 8 shows posterior mean for parameter of each models. For model parameter estimates, all models report large estimates of the force parameter α. For response trait, stronger forces (2.801, 3.585, 2.677 and 3.022 in the α_y column)

Table 8
Parameter estimates for model of adaptive trait evolution using coral data in Sanchez and Lasker (2003b). Model parameters are estimated under ABC-rejection methods where the posterior mean of parameters is reported. Parameters in GLS row show the regression estimates under generalized least square method.

Model	α_y	α_x	α_τ	θ_x	θ_τ	σ_x	τ	σ_τ	b_0	b_1	b_2
OUOUCIR	2.801	4.073	2.950	−0.121	1.639	1.242		1.884	0.825	1.165	0.253
OUBMCIR	3.585		3.097		1.577	0.951		1.823	0.899	1.153	0.150
OUOUBM	2.677	4.430		−0.150		1.241	4.609		0.836	1.216	0.230
OUBMBM	3.022					0.904	3.490		0.890	1.147	0.179
GLS									0.900	1.206	0.201

were detected to pull trait back to its optimum during evolution while stronger forces (4.073 and 4.430 in the α_x column) were found for the covariate traits assumed followed the OU process. The force parameter α_τ in the OUOUCIR model has estimates of value 2.950 which is slightly smaller than the value 3.097 for the force parameter α_τ in the OUBMCIR model. The optimum parameter for the covariate trait θ_x in the OUOUCIR model has value -0.121 which is close to the value -0.150 of the OUOUBM model. The optimum parameter θ_τ in the OUOUCIR model has value 1.639 while slightly larger value of α_τ 1.577 is estimated under the OUOUCIR model. For rate parameter σ_x of the covariate trait, the OUOU** models report moderate larger values (OUOUCIR: 1.242, OUOUBM: 1.241) than the OUBM** models (OUBMCIR: 0.951, OUBMBM: 0.904) while the rate of evolution τ in OUOUBM model for response trait y_t has moderate larger value estimate 4.609 than the estimate τ of value 3.490 in the OUBMBM model. For parameter σ_τ, the OUOUCIR model has estimate of 1.884 which is slightly larger than the estimate 1.823 in the OUBMCIR model.

For regression parameters, all models report positive regression slopes ($b_1, b_2 > 0$) which are consistent with the study in Sanchez and Lasker (2003b) where positive correlations among traits are reported. For intercept and slope coefficient β_0 and β_1, all modes report similar estimates, for β_2 the OUOU** models (OUOUCIR 0.253 and OUOUBM 0.230) report slightly higher coefficient for than the OUBM** models (OUBMCIR 0.150 and OUBMBM 0.179) while the estimate for β_2 using GLS estimates falls between the two types of models. Both the inter-polyp distance (D) and the polyp aperture (P) have positive effect on thicker coral branches (T). Since traits are log transformed, the regression effect is interpreted using ratio scale. Given two measurements (y', x_1', x_2'), (y, x_1, x_2), the difference between two predicted values is $\log\theta_{y'} - \log\theta_y = b_1(\log x_1' - \log x_1) + b_2(\log x_2' - \log x_2)$ which yields to $y'/y = (x_1'/x_1)^{b_1} \cdot (x_2'/x_2)^{b_2}$. Hence under the OUOUCIR model a 5 percent increase of the inter-poly distance (x_1) and 5 percent decrease of polyp aperture (x_2) would yield to 4.5 percent expected increase for the branch thickness (i.e. $1.05^{1.165} \cdot 0.95^{0.253} \approx 1.045$ and $(1.045 - 1)100\% = 4.5\%$); while for a 3 percent decrease of the inter-poly distance (x_1) and 7 percent increase of the polyp aperture, there could be 2.5 percent expected decrease for the branch thickness under the OUBMCIR model (i.e. $0.97^{1.153} \cdot 1.07^{0.150} \approx 0.975$ and $(0.975 - 1)100\% = -2.5\%$).

The same study was performed on seven additional datasets where for each dataset traits are log-transformed into a ratio scale. The results are reported in online supplemental information where one dataset in Fig (Ficus carica) study in Weiblen (2004) the OUBMCIR model provides the best fit among the four models. For 33 percents of the studies out of the eight study, one of the new CIR models (the OUBMCIR model and the OUOUCIR model) was the best (or second best) model, suggesting its potential utility. (See online supplemental material Table S4.)

5. Conclusion

Two models of adaptive trait evolution called the OUBMCIR model and the OUOUCIR model are developed in this framework. The new models were built to allow the optimal regression with interactions, and to especially incorporate the reduction of variance near zero that comes from CIR models, which may better model biological reality. Due to the intractable likelihood function of models, the trait variable y_t that solves the corresponding SDE is derived and clearly expressed. A heuristic algorithm combining tree traversal and Approximate Bayesian Computation is proposed here for inference. Current algorithm uses rejection method implemented in the R package `abc` (Blum, 2010) to obtain the posterior samples, one of future research directions is to develop advanced algorithm such as ABC-MCMC (Wegmann et al., 2009) to replace the unobserved variables by simulated variable for further inference.

The Euclidean measure as the distance metric using twelves summary statistics are used for assessing models performance, other distance measures may be more informative for trait evolving under a non-normal process. Simulations show that the performance of models is overall satisfactory and model parameters can be estimated well with acceptable values. The fit of the models to several empirical data sets is assessed using Bayes factors. The power analysis through cross validation analysis for the four models shows that models are adequate. Results from empirical study show that the new models (the OUOUCIR and the OUBMCIR models) can be as useful as the existed models (the OUBMBM and the OUOUBM models).

Several future research directions are suggested here (i) implement models that involve interaction effect using Eq. (6) with expressing the evolutionary rate of the optima σ_θ using Eq. (20). (ii) develop more efficient algorithm for drawing trait sample y_t under the models, in particular for computing the stochastic integrals with respect to time or Wiener process. (iii) design appropriate model selection criteria for selecting model of adaptive trait evolution.

Many interesting biological and ecological phenomena such as selection force, evolutionary optimum and evolutionary speed occurred during trait evolution, models of adaptive trait evolution built upon the optimal regression are useful tools to study them.

Acknowledgments

I would like to thank Brian O'Meara, the co-Editor, the associate editor and two anonymous reviewers for their tremendous help and constructive suggestions on improving this manuscript. I would also like to thank Yen-Suo Lai, Chih-Ping Wang and Chiu-Min Chou for useful discussions. Jhwueng's research was supported by the Ministry of Science and Technology [Grant #MOST-108- 2118-M-035-001], Taiwan, ROC and was assisted by attendance as a Short-term Visitor at the National Institute for Mathematical and Biological Synthesis, an Institute supported by the National Science Foundation through NSF Award #DBI-1300426, with additional support from The University of Tennessee, Knoxville.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.csda.2020.106924.

References

Adams, D.C., 2014. A generalized k statistic for estimating phylogenetic signal from shape and other high-dimensional multivariate data. Syst. Biol. 63 (5), 685–697.
Aguirre, Luis F, Herrel, Anthony, Van Damme, R, Matthysen, E, 2002. Ecomorphological analysis of trophic niche partitioning in a tropical savannah bat community. Proceedings of the Royal Society of London. Series B: Biological Sciences 269 (1497), 1271–1278.
Baek, J., Lesmes, L.A., Lu, Z.-L., 2016. qpr: An adaptive partial-report procedure based on Bayesian inference. J. Vis. 16 (10), 25.
Bartoszek, K., Liò, P., 2018. Modelling trait dependent speciation with approximate bayesian computation. arXiv preprint arXiv:1812.03715.
Beaulieu, J., Jhwueng, D.-C., Boettiger, C., OMeara, B., 2012. Modeling stabilizing selection: expanding the ornstein-uhlenbeck model of adaptive evolution. Evolution 66 (8), 2369–2383.
Blomberg, S.P., (0000). Beyond brownian motion and the ornstein-uhlenbeck process: Stochastic diffusion models for the evolution of quantitative characters. bioRxiv: http://dx.doi.org/10.1101/067363.
Blomberg, S.P., Garland, T., Ives, A.R., 2003. Testing for phylogenetic signal in comparative data: behavioral traits are more labile. Evolution 57 (4), 717–745.
Blum, M.G.B., 2010. Proceedings of COMPSTAT'2010: 19th International Conference on Computational Statistics. Paris France, August 22-27, 2010, Physica-Verlag HD, Heidelberg, pp. 47–56, Keynote, Invited and Contributed Papers, Ch. Choosing the Summary Statistics and the Acceptance Rate in Approximate Bayesian Computation.
Blum, M.G.B., François, O., 2010. Non-linear regression models for approximate Bayesian computation. Stat. Comput. 20 (1), 63–73. http://dx.doi.org/10.1007/s11222-009-9116-0.
Bonine, Kevin E, Gleeson, Todd T, Garland, Theodore, 2005. Muscle fiber-type variation in lizards (squamata) and phylogenetic reconstruction of hypothesized ancestral states. Journal of Experimental Biology 208 (23), 4529–4547.
Butler, M., King, A., 2004. Phylogenetic comparative analysis: a modeling approach for adaptive evolution. Amer. Nat. 164, 683–695.
Chan, J.H., Joshi, M., 2010. Fast and accurate long stepping simulation of the heston stochastic volatility model. ssrn.
Clarke, M., Thomas, G., Freckleton, R., 2017. Trait evolution in adaptive radiations: Modeling and measuring interspecific competition on phylogenies. Amer. Nat. 189 (2), 121–137. http://dx.doi.org/10.1086/689819, pMID: 28107052.
Cox, J., Ingersoll, J., Ross, S., 1985. A theory of the term structure of interest rates. Econometrica 53, 385–407.
Crespi, Bernard J, Teo, Roy, 2002. Comparative phylogenetic analysis of the evolution of semelparity and life history in salmonid fishes. Evolution 56 (5), 1008–1020.
Csilléry, K., Blum, M.G., Gaggiotti, O.E., François, O., 2010. Approximate Bayesian computation (abc) in practice. Trends Ecol. Evol. 25 (7), 410–418.
Csilléry, K., François, O., Blum, M.G., 2012. abc: an r package for approximate Bayesian computation (abc). Methods Ecol. Evol. 3 (3), 475–479.
Csilléry, K., Francois, O., Blum, M.G.B., 2012. abc: an r package for approximate Bayesian computation (abc). Methods Ecol. Evol. http://dx.doi.org/10.1111/j.2041-210X.2011.00179.x.
Ellson, J., Gansner, E., Koutsofios, L., North, S.C., Woodhull, G., 2001. Graphvizxopen source graph drawing tools. In: International Symposium on Graph Drawing. Springer, pp. 483–484.
Felsenstein, J., 1985. Phylogeny and the comparative method. Am. Nat. 125 (1), 1–15.
Freckleton, R.P., 2012. Fast likelihood calculations for comparative analyses. Methods Ecol. Evol. 3 (5), 940–947.
Gelman, A., Jakulin, A., Pittau, M.G., Su, Y.-S., et al., 2008. A weakly informative default prior distribution for logistic and other regression models. Ann. Appl. Stat. 2 (4), 1360–1383.
Gelman, A., Simpson, D., Betancourt, M., 2017. The prior can often only be understood in the context of the likelihood. Entropy 19 (10), 555.
Glasserman, P., Kim, K.-K., 2011. Gamma expansion of the heston stochastic volatility model. Finance Stoch. 15 (2), 267–296.
Grafen, A., 1989. The phylogenetic regression, philosophical transactions of the royal society of London. B. Biol. Sci. 326 (1233), 119–157.
Guidoum, A.C., Boukhetala, K., 2017. Sim.DiffProc: Simulation of Diffusion Processes. r package version 4.0.
Hansen, T.F., 1997. Stabilizing selection and the comparative analysis of adaptation. Evolution 51, 1341–1351.
Hansen, T., Martins, E., 1996. Translating between microevolutionary process and macroevolutionary patterns: the correlation structure of interspecific data. Evolution 50, 1404–1417.
Hansen, T., Pienaar, J., Orzack, S., 2008. A comparative method for studying adaptation to a randomly evolving environment. Evolution 62, 1965–1977.
Harmon, L., Weir, J., Brock, C., Glor, R., Challenger, W., 2008. Geiger: investigating evolutionary radiations. Bioinformatics 24, 129–131.
Jabot, F., Faure, T., Dumoulin, N., 2013. Easyabc: performing efficient approximate Bayesian computation sampling schemes using r. Methods Ecol. Evol. 4 (7), 684–687.
Jhwueng, D.-C., 2013. Assessing the goodness of fit of phylogenetic comparative methods: A meta-analysis and simulation study. PLoS One 8 (6), 1–12. http://dx.doi.org/10.1371/journal.pone.0067001.
Jhwueng, D.-C., Ho, L., Suchard, M., 2020. (2020). direct bayesian inference methods for analyzing positive traits under the cox-ingersoll-ross model. (in preparation).
Jhwueng, D.-C., Maroulas, V., 2014. Phylogenetic ornstein-uhlenbeck regression curves. Stat. Probab. Lett. 89, 110–117.

Jhwueng, D.-C., Maroulas, V., 2016. Adaptive trait evolution in random environment. J. Appl. Stat. 43 (12), 2310–2324.
Kass, R., Raftery, A., 1995. Bayes factors. J. Amer. Statist. Assoc. 90 (430), 773–795.
Kostikova, A., Silvestro, D., Pearman, P.B., Salamin, N., 2016. Bridging inter-and intraspecific trait evolution with a hierarchical Bayesian approach. Syst. Biol. 65 (3), 417–431.
Makowski, D., Ben-Shachar, M.S., Ludecke, D., 2019. bayestestr: Describing effects and their uncertainty, existence and significance within the Bayesian framework. J. Open Source Softw. 4 (40), 1541. http://dx.doi.org/10.21105/joss.01541, URL https://joss.theoj.org/papers/10.21105/joss.01541.
Molina-Borja, M, Rodríguez-Domínguez, MA, 2004. Evolution of biometric and life-history traits in lizards (gallotia) from the canary islands. Journal of Zoological Systematics and Evolutionary Research 42 (1), 44–53.
Morris, J., 1979. Traversing binary trees simply and cheaply. Inform. Process. Lett. 9 (5), 197–200.
Niewiarowski, Peter H, Angilletta, Michael J, Leaché, Adam D, 2004. Phylogenetic comparative analysis of life-history variation among populations of the lizard sceloporus undulatus: an example and prognosis. Evolution 58 (3), 619–633.
O'Meara, B., 2019. Treevo. https://github.com/bomeara/treevo.
OMeara, B., Ané, C., Sanderson, M., Wainwright, P., 2006. Testing different rates of continuous trait evolution using likelihood. Evolution 60, 922–933.
Pagel, M., 1999. Inferring the historical patterns of biological evolution. Nature 401 (6756), 877.
Paradis, E., Schliep, K., 2018. ape 5.0: an environment for modern phylogenetics and evolutionary analyses in r. Bioinformatics 35 (3), 526–528.
Revell, L.J., 2012. phytools: an r package for phylogenetic comparative biology (and other things). Methods Ecol. Evol. 3 (2), 217–223.
Sanchez, J., Lasker, H., 2003a. Patterns of morphological integration in marine modular organisms: supra-module organization in branching octocoral colonies. Proc. Biol. Sci. 270 (1528), 2039–2044.
Sanchez, J., Lasker, H., 2003b. Patterns of morphological integration in marine modular organisms: supra-module organization in branching octocoral colonies. Evolution 270, 2039–2044.
Schliep, K., 2011. phangorn: phylogenetic analysis in r. Bioinformatics 27 (4), 592–593. http://dx.doi.org/10.1093/bioinformatics/btq706.
Sisson, S., Fan, Y., Beaumont, M., 2019. Overview of approximate bayesian computation. arXiv preprint arXiv:1802.09720.
Stadler, T., 2010. Treesim in r-simulating trees under the birth-death model, r package.
Uyeda, J.C., Harmon, L.J., 2014. A novel Bayesian method for inferring and interpreting the dynamics of adaptive landscapes from phylogenetic comparative data. Syst. Biol. 63 (6), 902–918.
Venables, W.N., Ripley, B.D., 2002. Modern Applied Statistics with S, fourth ed. Springer, New York, ISBN: 0-387-95457-0, URL http://www.stats.ox.ac.uk/pub/MASS4.
Watanabe, Y.Y., Goldman, K.J., Caselle, J.E., Chapman, D.D., Papastamatiou, Y.P., 2015. Comparative analyses of animal-tracking data reveal ecological significance of endothermy in fishes. Proc. Natl. Acad. Sci. 112 (19), 6104–6109. http://dx.doi.org/10.1073/pnas.1500316112.
Webster, Andrea J, Purvis, Andy, 2002. Testing the accuracy of methods for reconstructing ancestral states of continuous characters. Proceedings of the Royal Society of London. Series B: Biological Sciences 269 (1487), 143–149.
Wegmann, D., Leuenberger, C., Excoffier, L., 2009. Efficient approximate bayesian computation coupled with markov chain monte carlo without likelihood. Genetics.
Weiblen, G., 2004. Correlated evolution in fig pollination. Syst. Biol. 53 (1), 128–139.

Modeling Rate of Adaptive Trait Evolution using Cox-Ingersoll-Ross Process: an Approximate Bayesian Computation Approach

Dwueng-Chwuan Jhwueng*

Department of Statistics, Feng-Chia University
No 100, Seatwen, Taichung Taiwan

1. Supplemental Material

1.1. List of pptx, R scripts and RData files

Below lists the click-able links for accessing the R script and data file to the figures and tables in this manuscript. All relevant files can be accessed at `https://tonyjhwueng.info/ououcir`

1. Tables S1, S2: `http://www.tonyjhwueng.info/ououcir/nonunifsimtable.html`
2. Table S3: `http://www.tonyjhwueng.info/ououcir/nonunifdistplot`.
3. Figure S1 : `http://www.tonyjhwueng.info/ououcir/summaryplotHDI.html`.
4. Figure S2: `http://www.tonyjhwueng.info/ououcir/cvbirthdeathtree.html`.
5. Table S4: `http://www.tonyjhwueng.info/ououcir/EmpiricalMaincodeV2/treetraitV2/`.

*Tel.:011-886-4-24517250 ext 4402
Email address: `dcjhwueng@fcu.edu.tw` (Dwueng-Chwuan Jhwueng)

1.2. Simulation using informative prior

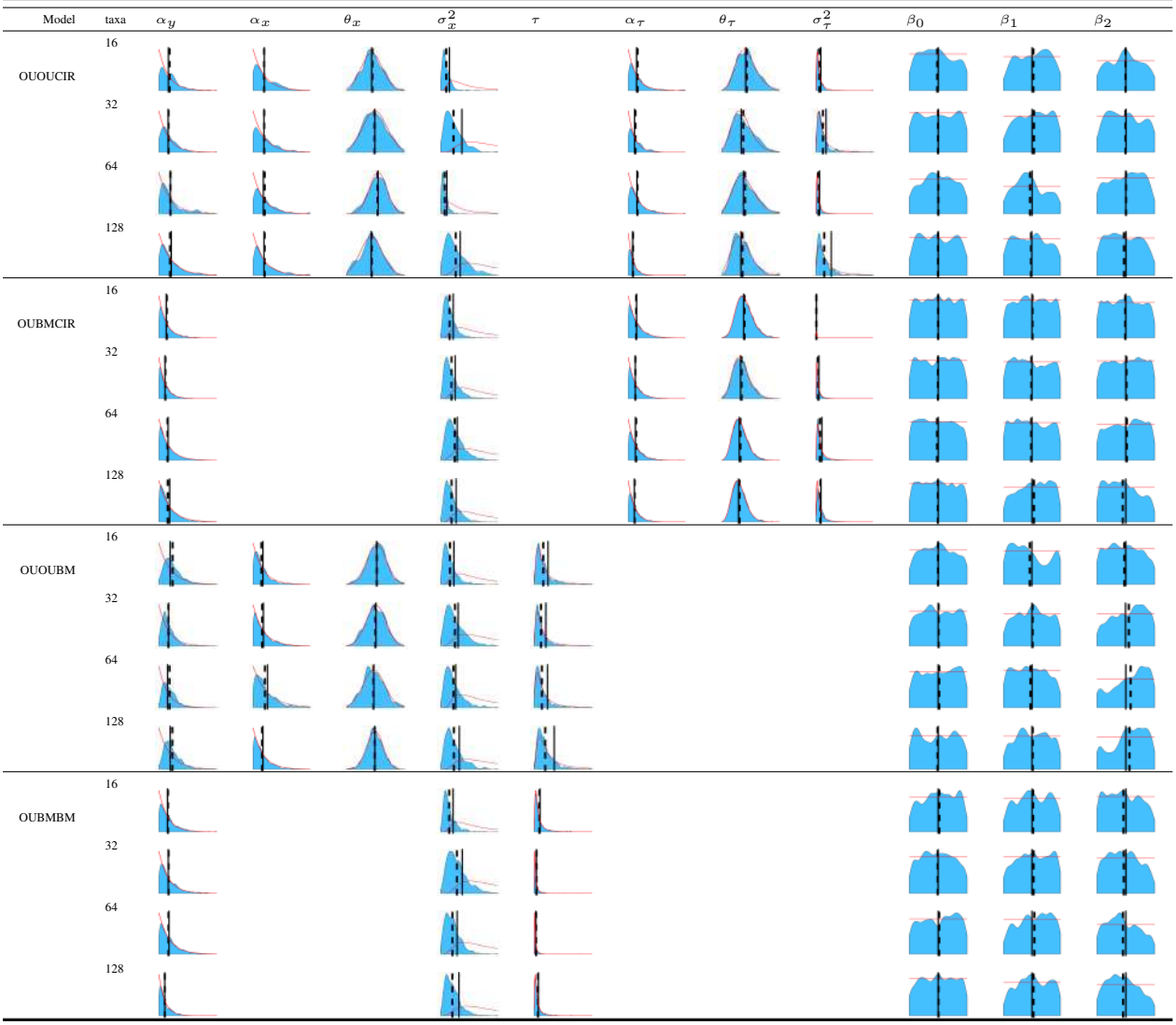
Table S1: Simulation results for assessing the model performance under parameter estimation using informative prior.Four models (OUBMBM, OUOUBM, OUBMCIR and OUOUCIR) are under investigation. Four different taxa sizes $16, 32, 64, 128$ are used. On each column, the mean and 95 percent credible interval using 2500 posterior samples are reported for each model parameter. The true parameters are $\alpha_y = 0.2, \alpha_x = 0.125, \theta_x = 0, \sigma_x^2 = 2, \tau = 1, \alpha_\tau = 0.25, \theta_\tau = 1.5$ and $\sigma_\tau = 1$. Note that $\theta_\tau \sim \log\mathcal{N}(\mu, \sigma^2)$ has expected value $E[\theta_\tau] = \exp(\mu+\sigma^2/2) = \exp(1/2) \approx 1.65$.

Model	taxa	α_y	α_x	θ_x	σ_x^2	τ	α_τ	θ_τ	σ_τ
True	value	0.2	0.125	0	0.5	1	0.25	1.5	1
OUOUCIR	16	0.23 (0.01,0.74)	0.12 (0,0.38)	0.07 (-1.63,1.95)	0.35 (0.1,0.89)		0.28 (0.01,1.21)	1.53 (1.06,2.17)	0.73 (0.16,1.96)
	32	0.21 (0,0.6)	0.13 (0,0.46)	-0.02 (-1.85,1.86)	0.33 (0.1,0.75)		0.29 (0.01,0.97)	1.57 (1.05,2.25)	0.73 (0.19,3.22)
	64	0.2 (0.01,0.69)	0.14 (0,0.5)	-0.07 (-2.35,2.01)	0.35 (0.1,0.89)		0.26 (0.01,0.94)	1.54 (1.05,2.22)	0.74 (0.16,2.43)
	128	0.18 (0.01,0.65)	0.13 (0.01,0.45)	0.05 (-1.85,1.89)	0.4 (0.12,0.98)		0.26 (0.01,0.82)	1.55 (1.13,2.2)	0.61 (0.19,2.2)
OUBMCIR	16	0.21 (0,0.78)			0.38 (0.13,0.88)		0.24 (0.01,0.82)	1.53 (1.01,2.23)	0.92 (0.17,3.04)
	32	0.21 (0,0.77)			0.39 (0.11,0.96)		0.25 (0.01,0.89)	1.54 (1,2.23)	0.78 (0.19,2.7)
	64	0.19 (0.01,0.69)			0.44 (0.16,0.98)		0.28 (0.01,0.96)	1.54 (1.03,2.29)	0.7 (0.19,2.1)
	128	0.16 (0,0.56)			0.37 (0.12,0.86)		0.26 (0.01,0.95)	1.54 (1.01,2.26)	0.92 (0.24,2.74)
OUOUBM	16	0.24 (0.03,0.64)	0.11 (0,0.4)	0.01 (-2.08,2)	0.37 (0.1,1.04)	0.69 (0.17,1.99)			
	32	0.2 (0.02,0.56)	0.11 (0,0.38)	-0.05 (-1.96,1.87)	0.43 (0.14,1.02)	0.63 (0.16,1.88)			
	64	0.25 (0.04,0.66)	0.1 (0,0.36)	-0.04 (-1.81,2.09)	0.43 (0.13,1.13)	0.64 (0.18,2.05)			
	128	0.23 (0.02,0.61)	0.11 (0,0.44)	-0.02 (-1.96,1.99)	0.38 (0.14,0.83)	0.62 (0.18,1.71)			
OUBMBM	16	0.21 (0,0.76)			0.36 (0.1,0.99)	0.9 (0.17,3.72)			
	32	0.21 (0.01,0.8)			0.39 (0.11,0.88)	1.04 (0.19,4.11)			
	64	0.19 (0.01,0.64)			0.38 (0.12,0.86)	1.11 (0.17,4.55)			
	128	0.2 (0.01,0.71)			0.35 (0.13,0.76)	0.95 (0.18,3.68)			

Table S2: Simulation results for assessing the model performance under parameter estimation for regression parameter using informative prior.Four models (OUBMBM, OUOUBM, OUBMCIR and OUOUCIR) are under investigation. Four different taxa sizes 16, 32, 64,128 are used. On each column, the mean and 95 percent credible interval using 2500 posterior samples are reported for each regression parameter. The true parameters have values $\beta_0 = 0, \beta_1 = -2$ and $\beta_2 = 3$.

Model	Taxa	b_0	b_1	b_2
True	value	0	-2	3
OUOUCIR	16	-0.17 (-4.7,4.71)	-1.77 (-6.58,2.63)	2.98 (-1.74,7.74)
	32	-0.03 (-4.83,4.67)	-1.68 (-6.63,2.72)	2.9 (-1.56,7.59)
	64	0.03 (-4.82,4.77)	-2.36 (-6.78,2.5)	3.04 (-1.81,7.51)
	128	-0.02 (-4.52,4.7)	-2.13 (-6.62,2.69)	2.66 (-1.78,7.5)
OUBMCIR	16	-0.01 (-4.72,4.73)	-1.91 (-6.71,2.78)	2.91 (-1.74,7.69)
	32	-0.04 (-4.76,4.7)	-2.11 (-6.77,2.76)	3.11 (-1.76,7.77)
	64	-0.18 (-4.8,4.7)	-2.14 (-6.78,2.73)	3.18 (-1.72,7.72)
	128	-0.06 (-4.78,4.78)	-1.67 (-6.67,2.7)	2.48 (-1.79,7.59)
OUOUBM	16	-0.02 (-4.76,4.82)	-2.37 (-6.8,2.85)	2.76 (-1.77,7.76)
	32	0.07 (-4.8,4.74)	-1.88 (-6.59,2.7)	3.56 (-1.54,7.78)
	64	0.18 (-4.69,4.77)	-2.29 (-6.81,2.66)	3.83 (-1.81,7.79)
	128	-0.08 (-4.81,4.8)	-1.85 (-6.59,2.7)	3.57 (-1.81,7.75)
OUBMBM	16	0.21 (-4.67,4.66)	-1.8 (-6.58,2.71)	2.58 (-1.79,7.55)
	32	-0.07 (-4.69,4.7)	-1.72 (-6.79,2.74)	2.63 (-1.76,7.47)
	64	0.16 (-4.81,4.64)	-1.58 (-6.68,2.83)	2.51 (-1.84,7.6)
	128	0.05 (-4.83,4.73)	-1.73 (-6.52,2.75)	2.5 (-1.76,7.54)

Table S3: Distribution plots using informative priors are reported for four models (OUBMBM, OUOUBM, OUBMCIR and OUOUCIR). On each plot, the prior distribution (red curve), the posterior distribution (in blue) from simulation results are generated. The true parameter value is plotted with a vertical line and the posterior mean is plotted with vertical dash line. Four different taxa sizes $16, 32, 64$ and 128 are used.



1.3. Coverage probability for credible intervals

Below list the bars plot for the Bayesian coverage probability for the OUBMCIR model, the OUOUBM model and the OUBMBM model.

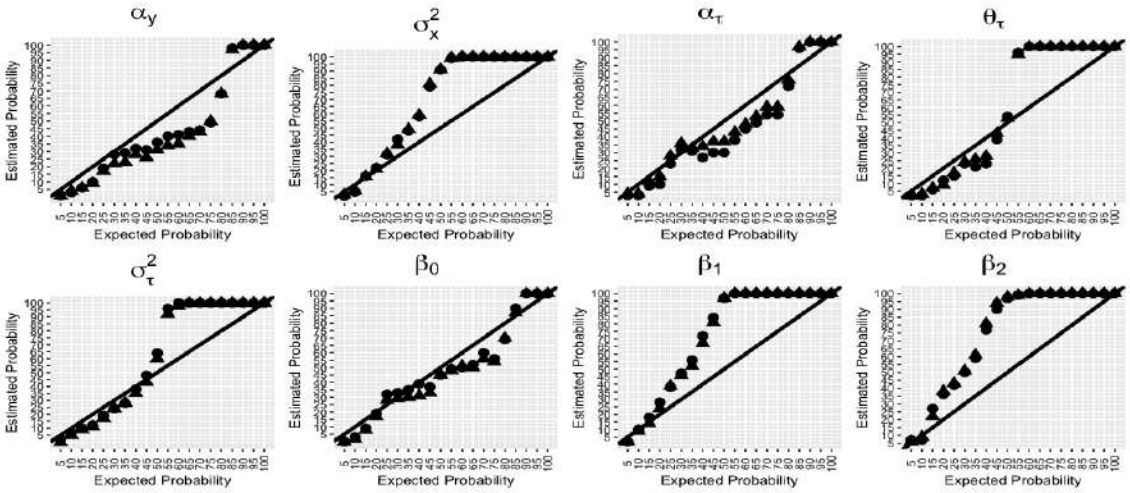


Figure S1: Bayesian coverage probability for the **OUBMCIR** model. The horizontal axis shows the expected interval size (5 % credible interval to 95% credible interval) and the vertical axis represents the frequency across the sims where the true parameter value falls within that credible interval. The circle is for 32 taxa and triangle is for 64 taxa. **summaryplotHDIv2.r**

1.4. Cross validation result using birth-death tree

Results of the confusion matrices based on 100 samples for each model using the birth-death tree are reported using bar plots in Figure 2 . For instance, for fifty taxa (taxa size 50) case in the upper left panel of Figure 2, the right most bar plot shows that for OUOUCIR being the true model, among 100 samples ABC model choice will identify the OUBMBM model (pink bar) for 6 times, the OUOUBM (orange bar) model for 3 times, the OUBMCIR model (blue bar) for 18 times, and the OUOUCIR model (green bar) 73 times which accounts for misclassification proportions of $(1 - \frac{73}{100}) \times 100\% = 27\%$ while in the lower left panel of Figure 2, the right most bar plot of the taxa size of 500 case shows that for the OUOUCIR model being the true model, among 100 samples ABC model choice will identify the OUBMBM model 1 times, the OUOUBM model 0 times, the OUBMCIR model 10 times, and the OUOUCIR model 89 times which account for misclassification proportions of $1-\frac{89}{100} \times 100\% = 11\%$. In general, the frequency for ABC identifying the correct models increases with the taxa size. It is clear from this analysis that it is easier to distinguish the OUOU** models (orange bar and green bar) at each taxa size. In particular, the OUOUCIR model (green bars in the right most panel across taxa sizes) constitutes a better fit for the data as supported with enough power in model choice. On the other hand, the OUBMCIR model requires more taxa to identify the correct model (e.g. at taxa size 500 the blue bar under the OUBMCIR category).

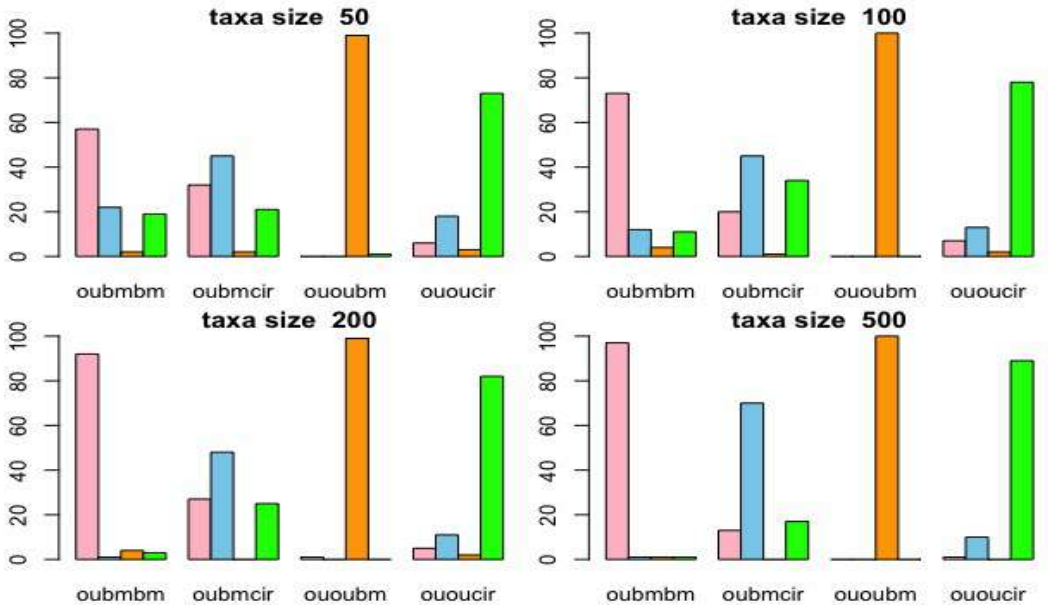


Figure S2: Bar plots for confusion matrices from cross validation analysis under ABC multinomial logistic regression method for models of adaptive trait evolution. Four taxa size (50, 100, 200, 500) of birth-death trees are used for cross-validation. On each panel, the labels in horizontal axis for each bar plot represents the true model, and the height of the bar plot represents frequency of correctly identifying the models (OUBMBM in pink, OUBMCIR in sky blue, OUOUBM in orange and OUOUCIR in green) for 100 samples.

1.5. Empirical data analysis

In the table 4, the OUBMCIR, and the OUOUCIR models accounts for one third population being the best models or the second best model for traits used in this study.

Table S4: The model selection by Bayes factor in Empirical Data. All data are log transformed. In coral dataset [1], the optimal regression model is set to polyp aperture $\sim$ distance between polyps + branch thickness. In fig dataset [2], the optimal regression model is set to gall width $\sim$ wall width + style length. In bat dataset [3], the optimal regression model is set to body mass $\sim$ head height + head length. In lizard dataset [4], the optimal regression model is set to body mass $\sim$ snout length + thigh muscle. In fish dataset [5], the optimal regression model is set to fecundity $\sim$ egg weight + body length. In lizard dataset [6], the optimal regression model is set to body length $\sim$ hatchling mass + hatchling mass. In lizard dataset [7], the optimal regression model is set to averaged mass $\sim$ clutch mass + clutch size. In foram dataset [8], the optimal regression model is set to area $\sim$ width + length.

Species	1^{st}	2^{nd}	3^{rd}	4^{th}	References
Coral	**OUOUCIR**	OUOUBM	OUBMBM	OUBMCIR	[1]
Fig	**OUBMCIR**	OUBMBM	OUOUBM	OUOUCIR	[2]
Bat	OUBMBM	OUOUBM	OUBMCIR	OUOUCIR	[3]
Lizard	OUBMBM	OUOUBM	OUBMCIR	OUOUCIR	[4]
Fish	OUBMBM	OUOUBM	OUBMCIR	OUOUCIR	[5]
Lizard	OUOUBM	**OUBMCIR**	OUBMBM	OUBMCIR	[6]
Lizard	OUOUBM	OUBMBM	OUOUCIR	OUBMCIR	[7]
Foram	OUBMBM	OUOUBM	OUOUCIR	OUBMCIR	[8]

References

[1] J. Sanchez, H. Lasker, Patterns of morphological integration in marine modular organisms: supra-module organization in branching octocoral colonies., Proceedings: Biological Sciences 270 (1528) (2003) 2039–2044.

[2] G. Weiblen, Correlated evolution in fig pollination., Syst. Biol. 53 (1) (2004) 128–139.

[3] L. F. Aguirre, A. Herrel, R. van Damme, E. Matthysen, Ecomorphological analysis of trophic niche partitioning in a tropical savannah bat community, Proceedings of the Royal Society of London B: Biological Sciences 269 (1497) (2002) 1271–1278. doi:10.1098/rspb.2002.2011.

[4] K. Bonine, T. Gleeson, T. J. Garland, Muscle fiber-type variation in lizards (squamata) and phylogenetic reconstruction of hypothesized ancestral states, The Journal of Experimental Biology 298 (2005) 4529–4547.

[5] B. J. Crespi, R. Teo, Comparative phylogenetic analysis of the evolution of semelparity and life history in salmonid fishes, Evolution 56 (5) (2002) 1008–1020.

[6] M. Molina-Borja, M. Rodríguez-Domínguez, Evolution of biometric and life-history traits in lizards (gallotia) from the canary islands, Journal of Zoological Systematics and Evolutionary Research 42 (1) (2004) 44–53.

[7] P. Niewiarowski, M. J. Angilletta, A. Leache, Phylogenetic comparative analysis of life-history variation among populations of the lizard sceloporus undulates: an example and prognosis, Evolution 58 (2004) 619–633.

[8] A. J. Webster, A. Purvis, Testing the accuracy of methods for reconstructing ancestral states of continuous characters, Proceedings of the Royal Society of London B: Biological Sciences 269 (1487) (2002) 143–149.

Modeling Rate of Adaptive Trait Evolution using Cox-Ingersoll-Ross Process: an Approximate Bayesian Computation Approach

Dwueng-Chwuan Jhwueng*

Department of Statistics, Feng-Chia University
No 100, Seatwen, Taichung Taiwan

1. Supplemental Material

1.1. List of pptx, R scripts and RData files

Below lists the click-able links for accessing the R script and data file to the figures and tables in this manuscript. All relevant files can be accessed at `https://tonyjhwueng.info/ououcir`

1. Tables S1, S2: `http://www.tonyjhwueng.info/ououcir/nonunifsimtable.html`
2. Table S3: `http://www.tonyjhwueng.info/ououcir/nonunifdistplot`.
3. Figure S1 : `http://www.tonyjhwueng.info/ououcir/summaryplotHDI.html`.
4. Figure S2: `http://www.tonyjhwueng.info/ououcir/cvbirthdeathtree.html`.
5. Table S4: `http://www.tonyjhwueng.info/ououcir/EmpiricalMaincodeV2/treetraitV2/`.

*Tel.:011-886-4-24517250 ext 4402
Email address: `dcjhwueng@fcu.edu.tw` (Dwueng-Chwuan Jhwueng)

1.2. Simulation using informative prior

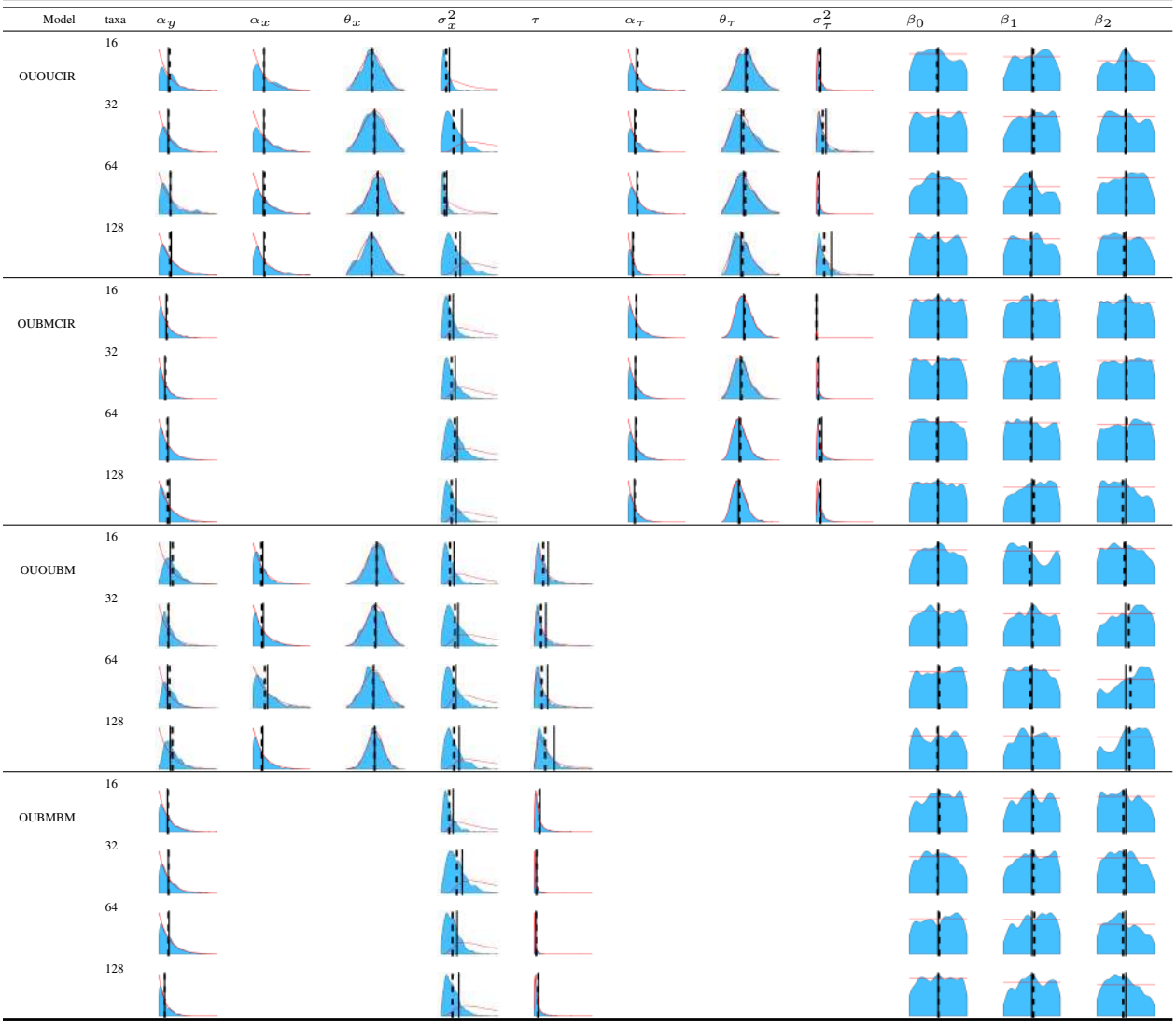
Table S1: Simulation results for assessing the model performance under parameter estimation using informative prior.Four models (OUBMBM, OUOUBM, OUBMCIR and OUOUCIR) are under investigation. Four different taxa sizes $16, 32, 64, 128$ are used. On each column, the mean and 95 percent credible interval using 2500 posterior samples are reported for each model parameter. The true parameters are $\alpha_y = 0.2, \alpha_x = 0.125, \theta_x = 0, \sigma_x^2 = 2, \tau = 1, \alpha_\tau = 0.25, \theta_\tau = 1.5$ and $\sigma_\tau = 1$. Note that $\theta_\tau \sim \log\mathcal{N}(\mu, \sigma^2)$ has expected value $E[\theta_\tau] = \exp(\mu+\sigma^2/2) = \exp(1/2) \approx 1.65$.

Model	taxa	α_y	α_x	θ_x	σ_x^2	τ	α_τ	θ_τ	σ_τ
True	value	0.2	0.125	0	0.5	1	0.25	1.5	1
OUOUCIR	16	0.23 (0.01,0.74)	0.12 (0,0.38)	0.07 (-1.63,1.95)	0.35 (0.1,0.89)		0.28 (0.01,1.21)	1.53 (1.06,2.17)	0.73 (0.16,1.96)
	32	0.21 (0,0.6)	0.13 (0,0.46)	-0.02 (-1.85,1.86)	0.33 (0.1,0.75)		0.29 (0.01,0.97)	1.57 (1.05,2.25)	0.73 (0.19,3.22)
	64	0.2 (0.01,0.69)	0.14 (0,0.5)	-0.07 (-2.35,2.01)	0.35 (0.1,0.89)		0.26 (0.01,0.94)	1.54 (1.05,2.22)	0.74 (0.16,2.43)
	128	0.18 (0.01,0.65)	0.13 (0.01,0.45)	0.05 (-1.85,1.89)	0.4 (0.12,0.98)		0.26 (0.01,0.82)	1.55 (1.13,2.2)	0.61 (0.19,2.2)
OUBMCIR	16	0.21 (0,0.78)			0.38 (0.13,0.88)		0.24 (0.01,0.82)	1.53 (1.01,2.23)	0.92 (0.17,3.04)
	32	0.21 (0,0.77)			0.39 (0.11,0.96)		0.25 (0.01,0.89)	1.54 (1,2.23)	0.78 (0.19,2.7)
	64	0.19 (0.01,0.69)			0.44 (0.16,0.98)		0.28 (0.01,0.96)	1.54 (1.03,2.29)	0.7 (0.19,2.1)
	128	0.16 (0,0.56)			0.37 (0.12,0.86)		0.26 (0.01,0.95)	1.54 (1.01,2.26)	0.92 (0.24,2.74)
OUOUBM	16	0.24 (0.03,0.64)	0.11 (0,0.4)	0.01 (-2.08,2)	0.37 (0.1,1.04)	0.69 (0.17,1.99)			
	32	0.2 (0.02,0.56)	0.11 (0,0.38)	-0.05 (-1.96,1.87)	0.43 (0.14,1.02)	0.63 (0.16,1.88)			
	64	0.25 (0.04,0.66)	0.1 (0,0.36)	-0.04 (-1.81,2.09)	0.43 (0.13,1.13)	0.64 (0.18,2.05)			
	128	0.23 (0.02,0.61)	0.11 (0,0.44)	-0.02 (-1.96,1.99)	0.38 (0.14,0.83)	0.62 (0.18,1.71)			
OUBMBM	16	0.21 (0,0.76)			0.36 (0.1,0.99)	0.9 (0.17,3.72)			
	32	0.21 (0.01,0.8)			0.39 (0.11,0.88)	1.04 (0.19,4.11)			
	64	0.19 (0.01,0.64)			0.38 (0.12,0.86)	1.11 (0.17,4.55)			
	128	0.2 (0.01,0.71)			0.35 (0.13,0.76)	0.95 (0.18,3.68)			

Table S2: Simulation results for assessing the model performance under parameter estimation for regression parameter using informative prior.Four models (OUBMBM, OUOUBM, OUBMCIR and OUOUCIR) are under investigation. Four different taxa sizes 16, 32, 64,128 are used. On each column, the mean and 95 percent credible interval using 2500 posterior samples are reported for each regression parameter. The true parameters have values $\beta_0 = 0, \beta_1 = -2$ and $\beta_2 = 3$.

Model	Taxa	b_0	b_1	b_2
True	value	0	-2	3
OUOUCIR	16	-0.17 (-4.7,4.71)	-1.77 (-6.58,2.63)	2.98 (-1.74,7.74)
	32	-0.03 (-4.83,4.67)	-1.68 (-6.63,2.72)	2.9 (-1.56,7.59)
	64	0.03 (-4.82,4.77)	-2.36 (-6.78,2.5)	3.04 (-1.81,7.51)
	128	-0.02 (-4.52,4.7)	-2.13 (-6.62,2.69)	2.66 (-1.78,7.5)
OUBMCIR	16	-0.01 (-4.72,4.73)	-1.91 (-6.71,2.78)	2.91 (-1.74,7.69)
	32	-0.04 (-4.76,4.7)	-2.11 (-6.77,2.76)	3.11 (-1.76,7.77)
	64	-0.18 (-4.8,4.7)	-2.14 (-6.78,2.73)	3.18 (-1.72,7.72)
	128	-0.06 (-4.78,4.78)	-1.67 (-6.67,2.7)	2.48 (-1.79,7.59)
OUOUBM	16	-0.02 (-4.76,4.82)	-2.37 (-6.8,2.85)	2.76 (-1.77,7.76)
	32	0.07 (-4.8,4.74)	-1.88 (-6.59,2.7)	3.56 (-1.54,7.78)
	64	0.18 (-4.69,4.77)	-2.29 (-6.81,2.66)	3.83 (-1.81,7.79)
	128	-0.08 (-4.81,4.8)	-1.85 (-6.59,2.7)	3.57 (-1.81,7.75)
OUBMBM	16	0.21 (-4.67,4.66)	-1.8 (-6.58,2.71)	2.58 (-1.79,7.55)
	32	-0.07 (-4.69,4.7)	-1.72 (-6.79,2.74)	2.63 (-1.76,7.47)
	64	0.16 (-4.81,4.64)	-1.58 (-6.68,2.83)	2.51 (-1.84,7.6)
	128	0.05 (-4.83,4.73)	-1.73 (-6.52,2.75)	2.5 (-1.76,7.54)

Table S3: Distribution plots using informative priors are reported for four models (OUBMBM, OUOUBM, OUBMCIR and OUOUCIR). On each plot, the prior distribution (red curve), the posterior distribution (in blue) from simulation results are generated. The true parameter value is plotted with a vertical line and the posterior mean is plotted with vertical dash line. Four different taxa sizes $16, 32, 64$ and 128 are used.



1.3. Coverage probability for credible intervals

Below list the bars plot for the Bayesian coverage probability for the OUBMCIR model, the OUOUBM model and the OUBMBM model.

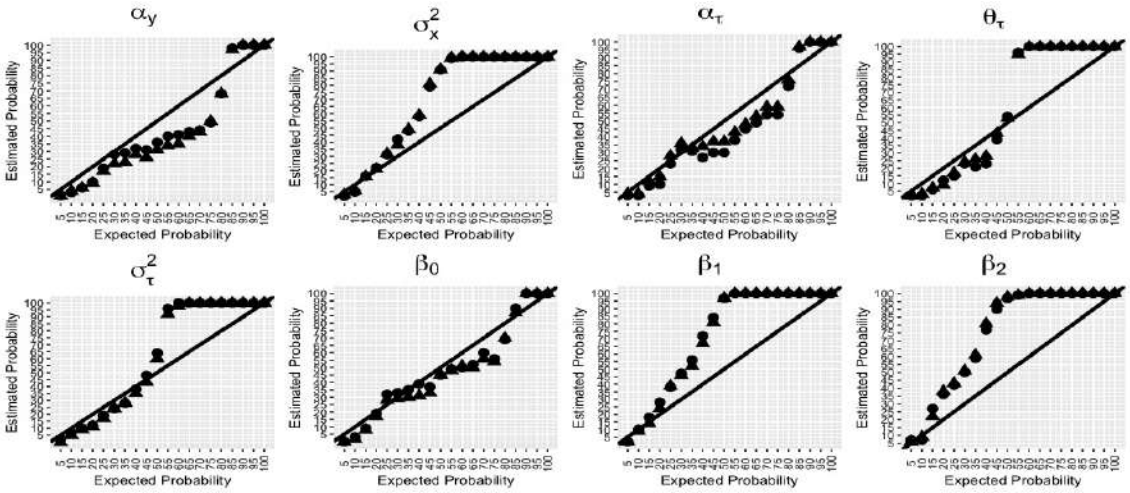


Figure S1: Bayesian coverage probability for the **OUBMCIR** model. The horizontal axis shows the expected interval size (5 % credible interval to 95% credible interval) and the vertical axis represents the frequency across the sims where the true parameter value falls within that credible interval. The circle is for 32 taxa and triangle is for 64 taxa. **summaryplotHDIv2.r**

1.4. Cross validation result using birth-death tree

Results of the confusion matrices based on 100 samples for each model using the birth-death tree are reported using bar plots in Figure 2 . For instance, for fifty taxa (taxa size 50) case in the upper left panel of Figure 2, the right most bar plot shows that for OUOUCIR being the true model, among 100 samples ABC model choice will identify the OUBMBM model (pink bar) for 6 times, the OUOUBM (orange bar) model for 3 times, the OUBMCIR model (blue bar) for 18 times, and the OUOUCIR model (green bar) 73 times which accounts for misclassification proportions of $(1 - \frac{73}{100}) \times 100\% = 27\%$ while in the lower left panel of Figure 2, the right most bar plot of the taxa size of 500 case shows that for the OUOUCIR model being the true model, among 100 samples ABC model choice will identify the OUBMBM model 1 times, the OUOUBM model 0 times, the OUBMCIR model 10 times, and the OUOUCIR model 89 times which account for misclassification proportions of $1-\frac{89}{100} \times 100\% = 11\%$. In general, the frequency for ABC identifying the correct models increases with the taxa size. It is clear from this analysis that it is easier to distinguish the OUOU** models (orange bar and green bar) at each taxa size. In particular, the OUOUCIR model (green bars in the right most panel across taxa sizes) constitutes a better fit for the data as supported with enough power in model choice. On the other hand, the OUBMCIR model requires more taxa to identify the correct model (e.g. at taxa size 500 the blue bar under the OUBMCIR category).

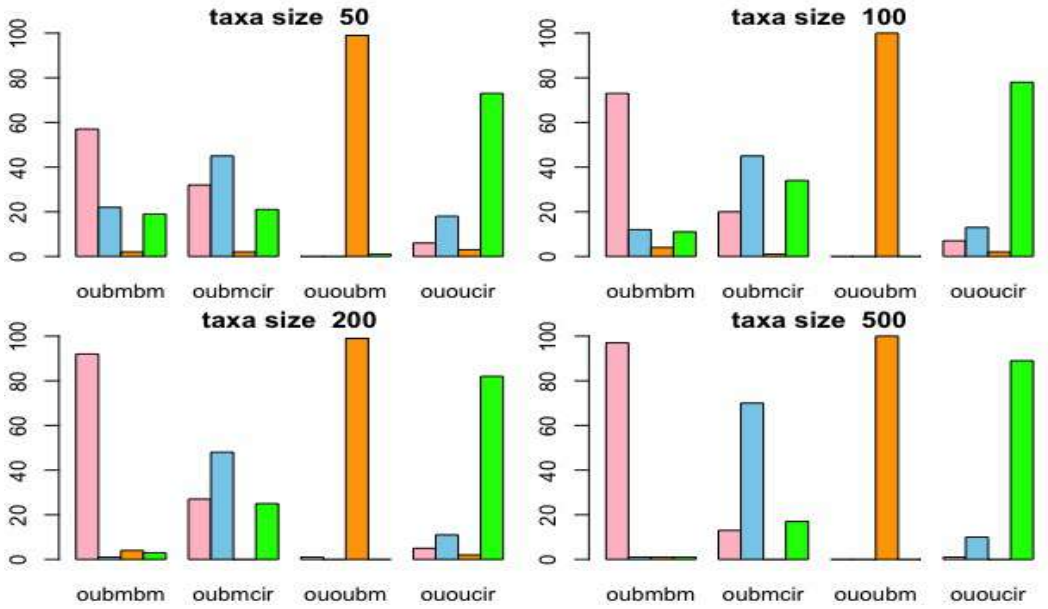


Figure S2: Bar plots for confusion matrices from cross validation analysis under ABC multinomial logistic regression method for models of adaptive trait evolution. Four taxa size (50, 100, 200, 500) of birth-death trees are used for cross-validation. On each panel, the labels in horizontal axis for each bar plot represents the true model, and the height of the bar plot represents frequency of correctly identifying the models (OUBMBM in pink, OUBMCIR in sky blue, OUOUBM in orange and OUOUCIR in green) for 100 samples.

1.5. Empirical data analysis

In the table 4, the OUBMCIR, and the OUOUCIR models accounts for one third population being the best models or the second best model for traits used in this study.

Table S4: The model selection by Bayes factor in Empirical Data. All data are log transformed. In coral dataset [1], the optimal regression model is set to polyp aperture $\sim$ distance between polyps + branch thickness. In fig dataset [2], the optimal regression model is set to gall width $\sim$ wall width + style length. In bat dataset [3], the optimal regression model is set to body mass $\sim$ head height + head length. In lizard dataset [4], the optimal regression model is set to body mass $\sim$ snout length + thigh muscle. In fish dataset [5], the optimal regression model is set to fecundity $\sim$ egg weight + body length. In lizard dataset [6], the optimal regression model is set to body length $\sim$ hatchling mass + hatchling mass. In lizard dataset [7], the optimal regression model is set to averaged mass $\sim$ clutch mass + clutch size. In foram dataset [8], the optimal regression model is set to area $\sim$ width + length.

Species	1^{st}	2^{nd}	3^{rd}	4^{th}	References
Coral	**OUOUCIR**	OUOUBM	OUBMBM	OUBMCIR	[1]
Fig	**OUBMCIR**	OUBMBM	OUOUBM	OUOUCIR	[2]
Bat	OUBMBM	OUOUBM	OUBMCIR	OUOUCIR	[3]
Lizard	OUBMBM	OUOUBM	OUBMCIR	OUOUCIR	[4]
Fish	OUBMBM	OUOUBM	OUBMCIR	OUOUCIR	[5]
Lizard	OUOUBM	**OUBMCIR**	OUBMBM	OUBMCIR	[6]
Lizard	OUOUBM	OUBMBM	OUOUCIR	OUBMCIR	[7]
Foram	OUBMBM	OUOUBM	OUOUCIR	OUBMCIR	[8]

References

[1] J. Sanchez, H. Lasker, Patterns of morphological integration in marine modular organisms: supra-module organization in branching octocoral colonies., Proceedings: Biological Sciences 270 (1528) (2003) 2039–2044.

[2] G. Weiblen, Correlated evolution in fig pollination., Syst. Biol. 53 (1) (2004) 128–139.

[3] L. F. Aguirre, A. Herrel, R. van Damme, E. Matthysen, Ecomorphological analysis of trophic niche partitioning in a tropical savannah bat community, Proceedings of the Royal Society of London B: Biological Sciences 269 (1497) (2002) 1271–1278. doi:10.1098/rspb.2002.2011.

[4] K. Bonine, T. Gleeson, T. J. Garland, Muscle fiber-type variation in lizards (squamata) and phylogenetic reconstruction of hypothesized ancestral states, The Journal of Experimental Biology 298 (2005) 4529–4547.

[5] B. J. Crespi, R. Teo, Comparative phylogenetic analysis of the evolution of semelparity and life history in salmonid fishes, Evolution 56 (5) (2002) 1008–1020.

[6] M. Molina-Borja, M. Rodríguez-Domínguez, Evolution of biometric and life-history traits in lizards (gallotia) from the canary islands, Journal of Zoological Systematics and Evolutionary Research 42 (1) (2004) 44–53.

[7] P. Niewiarowski, M. J. Angilletta, A. Leache, Phylogenetic comparative analysis of life-history variation among populations of the lizard sceloporus undulates: an example and prognosis, Evolution 58 (2004) 619–633.

[8] A. J. Webster, A. Purvis, Testing the accuracy of methods for reconstructing ancestral states of continuous characters, Proceedings of the Royal Society of London B: Biological Sciences 269 (1487) (2002) 143–149.