

Article

# Phylogenetic Curved Optimal Regression for Adaptive Trait Evolution

Dwueng-Chwuan Jhwueng \*  and Chih-Ping Wang

Department of Statistics, Feng-Chia University, Taichung 40724, Taiwan; a09760381100@gmail.com

\* Correspondence: dcjhwueng@fcu.edu.tw

**Abstract:** Regression analysis using line equations has been broadly applied in studying the evolutionary relationship between the response trait and its covariates. However, the characteristics among closely related species in nature present abundant diversities where the nonlinear relationship between traits have been frequently observed. By treating the evolution of quantitative traits along a phylogenetic tree as a set of continuous stochastic variables, statistical models for describing the dynamics of the optimum of the response trait and its covariates are built herein. Analytical representations for the response trait variables, as well as their optima among a group of related species, are derived. Due to the models' lack of tractable likelihood, a procedure that implements the Approximate Bayesian Computation (ABC) technique is applied for statistical inference. Simulation results show that the new models perform well where the posterior means of the parameters are close to the true parameters. Empirical analysis supports the new models when analyzing the trait relationship among kangaroo species.

**Keywords:** adaptive trait evolution; approximate Bayesian computation; geometric Brownian motion; geometric Ornstein–Uhlenbeck process; phylogenetic comparative analysis



**Citation:** Jhwueng, D.-C.; Wang, C.-P. Phylogenetic Curved Optimal Regression for Adaptive Trait Evolution. *Entropy* **2021**, *23*, 218. <https://doi.org/10.3390/e23020218>

Academic Editor: Jacinto Martín  
Received: 15 November 2020  
Accepted: 8 February 2021  
Published: 10 February 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



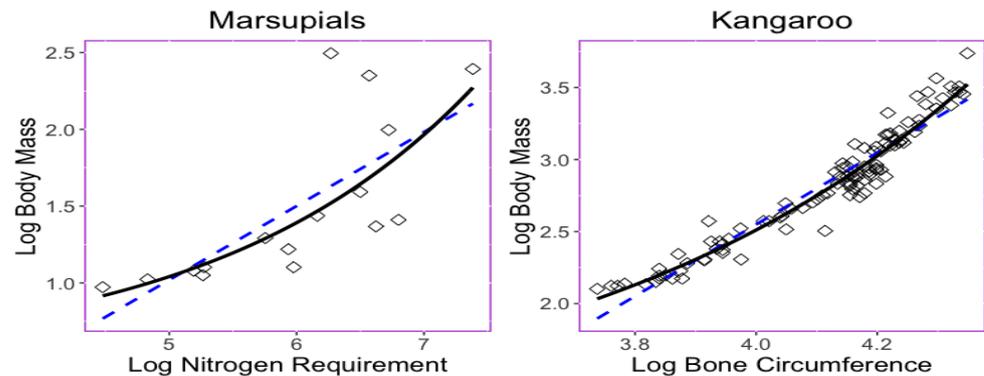
**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Species evolve across generations. For quantitative-trait evolution, scientists apply phylogenetic comparative methods (PCMs) to study the evolutionary relationship of a group of related species where a phylogenetic tree is incorporated for describing affinity among species [1–8]. Most current regression models in PCMs assume that the response trait variable  $y$  is linear with its covariates  $x$ s where the estimated line equation (e.g.,  $y = b_0 + \sum_{i=1}^n b_i x_i$ ) is used to predict the response trait [8–10]. However, the allometric relationship between body mass and other organisms is also often observed in nonlinear form (i.e.,  $y = kx^a$ ). Logarithm transformation ( $\log y = \log k + a \log x$ ) is usually considered as a regular procedure prior to analysis [11]. From a statistical perspective, log transformation on the data reduces skewness, decreasing the variability, conforming data close to the normal distribution, and placing dependent variable and covariates in a linear-like relationship [12–14]. From an evolutionary perspective, because most traits of particular species fall within a certain range, interpreting trait changes using raw scales may produce unreasonable results. Hence, convex transformation by the logarithm function is often applied to convert the raw data of the interval type into the ratio type. This has particular advantages, for example, a change in body mass of 0.2 kg might not be important for a male red kangaroo with a weight from 55 to 90 kg, but probably matters substantially for a wallaby with a weight of about 1.6 kg; a 1.36% change in body mass for both species is interpretable under log-transformed data.

Nevertheless, even the log transformation helps to convert the trait relationship from nonlinear into a moderate linear type, and there exists a nonlinear relationship among some log-transformed data [15]. The trait relationship shown in Figure 1 provides two examples in which nonlinear exponential regressions could provide a better fit with less predicted errors than those obtained when using linear regression. The left panel in Figure 1

displays the bivariate relationship between the body mass ( $x$ ) and the maintenance nitrogen requirement ( $y$ ) in the log scale of the marsupial species [16,17]. The exponential equation  $y = 0.486 + 0.047 \exp(0.490x)$  has a root mean square deviation with a value of 0.341, while the linear regression model  $y = -1.389 + 0.482x$  has a root mean square deviation 0.344.



**Figure 1.** Scatter plots and the relationship of the bivariate trait dataset described by the least squares regression lines or curves. **(left)** Relationship between maintenance nitrogen requirement and body mass in marsupials [16,17]. Exponential curve  $y = 0.486 + 0.047 \exp(0.490x)$  (RMSD = 0.341) and line equation  $y = -1.389 + 0.482x$  (RMSD = 0.344) shown. **(right)** Relationship between bone circumference and the body mass in kangaroos [18]. Exponential curve  $y = 1.051 + 0.003 \exp(1.510x)$  (RMSD = 0.092) and line  $y = -7.407 + 2.489x$  (RMSD = 0.112) shown.

The right panel in Figure 1 displays relationship between the thigh-bone (femoral) circumference ( $x$ ) and body mass ( $y$ ) of the kangaroo species [18] is shown Figure 1. The exponential equation  $y = 1.051 + 0.003 \exp(1.510x)$  has a root mean square deviation with a value of 0.092, while the linear regression model  $y = -7.407 + 2.489x$  has a root mean square deviation of 0.112. Parameters  $\beta_1, \beta_2, \beta_3$  in the exponential curve  $y = \beta_1 + \beta_2 \exp(\beta_3 x)$  were estimated under a least-squares method performed using the following step. Since  $\beta_1$  adds to the complexity of the model, an estimate of  $\beta_1$  is established by using the half value of the minimum of the responses; then, traits are subtracted from this value, which yields the model  $y - \beta_1 = \beta_2 \exp(\beta_3 x)$ . Parameters  $\beta_2$  and  $\beta_3$  are estimated through the least-squares method for the model  $\log(y - \beta_1) = \log(\beta_2) + \beta_3 x$ .

In the marsupial and kangaroo datasets, exponential regression models yield to smaller RMSD than those in linear regression models. Conceiving that the potential use of exponential regression models to study phylogenetic-traits relationships, the empirical datasets in Figure 1 call for novel phylogenetic comparative methods. In this framework, we developed models for adaptive trait evolution where the optimum of the trait undergoes stabilizing selection and has an exponential relationship with the predictor trait. Our work is distinguished from the work in [19], which mainly makes use of multiple linear regression. Our ultimate goal was to provide feasible models for scientists to analyze their valuable data for research.

Prior to developing new models, the background of phylogenetic adaptive trait evolution was introduced follows. Hansen et al. [20] developed a popular model (OUBM model) for phylogenetic adaptive trait evolution where the response trait variable is assumed following an Ornstein–Uhlenbeck (OU) process dynamic where the optimum of the response trait is assumed with a linear relationship with Brownian motion (BM) covariates. Later, various scientists made further efforts to expand the OUBM model of Hansen et al. via considering an Ornstein–Uhlenbeck process covariates (OUOU model) [21,22], a Cox–Ingersoll–Ross process for rate evolution [19], or extending the OUBM model to the multivariate case [23–25].

In general, the generalized model for phylogenetic adaptive trait evolution assumes that trait variable  $y_t$  solves stochastic differential equation (SDE) in Equation (1):

$$dy_t = \alpha_t^y (\theta_t^y - y_t) dt + \sigma_t^y dW_t^y, \quad (1)$$

where parameter  $\alpha_t^y$  is the force that pulled the trait back to its optimum  $\theta_t^y$ , parameter  $\sigma_t^y$  is called the evolutionary rate for the trait variable  $y_t$ , and  $W_t^y$  is a Wiener process with independent Gaussian increment, with mean 0 and variance  $t$ . Let  $\alpha_t^y = \alpha_y$  and  $\sigma_t^y = \sigma_y$  be constants. By multiplying the integrating factor  $\exp(\alpha_y t)$  and then integrating on both sides of Equation (1),  $y_t$  can be expressed explicitly, as shown in Equation (2)

$$y_t = \exp(-\alpha_y t) y_0 + \int_0^t \alpha_y \exp(-(\alpha_y t - \alpha_y s)) \theta_s^y ds + \sigma_y \int_0^t \exp(-(\alpha_y t - \alpha_y s)) dW_s^y, \quad (2)$$

where  $\exp(-\alpha_y t) y_0$  is a deterministic term with initial condition  $y_0$  at  $t = 0$ , and term  $\sigma_y \int_0^t \exp(-(\alpha_y t - \alpha_y s)) dW_s^y$  is a stochastic integral with respect to  $W_s^y$ , and is, again, a Gaussian variable with mean 0 and variance  $\sigma_y^2 (1 - \exp(-2\alpha_y t)) / (2\alpha_y)$  (obtained by applying Itô isometry [26]) and

$$\textcircled{A} = \exp(-\alpha_y t) \int_0^t \alpha_y \exp(\alpha_y s) \theta_s^y ds \quad (3)$$

is an integral with respect to time.

Optimal  $\theta_t^y$ , has a functional relationship with the covariate  $x_t$  represented in Equation (4)

$$\theta_t^y = f(\beta, x_t), \quad (4)$$

where  $\beta$  is the vector of regression parameters.

In Equation (4), when optimum  $\theta_t^y$  and covariate trait variable  $x_t$  are in a linear relationship (i.e.,  $\theta_t^y = \beta_0 + \sum_{i=1}^p \beta_i x_{i,t}$  where  $x_{i,t}, i = 1, \dots, p$  are identical independently distributed continuous stochastic random variables), the dynamics of  $\theta_t^y$  can be characterized through identifying the dynamics of the linear combination of identical independent distributed covariates  $x_{i,t}, i = 1, 2, \dots, p$ . For Gaussian process covariates  $x_{i,t}$ , optimal  $\theta_t^y$  follows a Brownian motion if covariates  $x_{i,t}$  follows Brownian motion (i.e.,  $dx_{i,t} = \sigma_{x_i} dW_t^{x_i}$ ), called the OUBM model [20]. On the other hand,  $\theta_t^y$  is an OU process if  $x_{i,t}$  are OU processes (i.e.,  $dx_{i,t} = \alpha_{x_i} (\theta_{x_i} - x_{i,t}) dt + \sigma_{x_i} dW_t^{x_i}$ ), called the OUOU model [21].

In this work, we assumed that an exponential relationship existed between trait optimum  $\theta_t^y$  and its covariate  $x_t$ . The development of the new models is described as follows. When assuming an exponential relationship between the optimum  $\theta_t^y$  and a Brownian motion covariate  $x_t$ , the optimum  $\theta_t^y$  follows a well known geometric Brownian motion [27]. By assuming an exponential relationship between optimum  $\theta_t^y$  and its Ornstein–Uhlenbeck process-type covariate  $x_t$ , the optimum  $\theta_t^y$  follows a geometric Ornstein–Uhlenbeck process [28].

We assumed that the covariate trait variable  $x_t$  evolved under Gaussian processes (e.g., Brownian motion or Ornstein–Uhlenbeck process); hence, the analytic expression of  $\textcircled{A}$  in Equation (3) depends on the expression between  $\theta_t^y$  and its covariate  $x_t$ . Both evolutionary rate ( $\sigma_t^y$ ) and force  $\alpha_t^y$  in Equation (1) are assumed to be positive constants throughout this work (i.e.,  $\alpha_t^y = \alpha_y > 0, \sigma_t^y = \sigma_y > 0$ ). Hence, we focused on developing of models by implementing the curved relationship between optimum  $\theta_t^y$  and its covariate  $x_t$ . The new model is named OUGBM (see Section 2.1.1) when trait  $y_t$  represented in Equation (2) admits a generalized OU process dynamic, and its optimum  $\theta_t^y$  has an exponential relationship with Brownian motion covariates  $x_t$ . The new model is named OUGOU (see Section 2.1.2) if  $y_t$  admits a generalized OU process, and  $\theta_t^y$  has an exponential relationship with OU process covariates  $x_t$ . We also implemented the OUBM (see Section 2.2.1) and OUOU (see Section 2.2.2) models for comparison with the new models. Since species are evolutionarily related, the models were developed with the assumption that evolutionary dependency among a group of species is along a given root phylogenetic tree (see Section 2.3). Due to those new models' lack of model likelihood, we propose the use of the approximate Bayesian computation procedure for model inference (see Section 2.4).

## 2. Materials and Methods

### 2.1. Optimal Exponential Regression

Consider an exponential relationship between the optimum and its covariate as follows

$$\theta_t^y = f(\beta, x_t) = \beta_1 + \beta_2 \exp(\beta_3 x_t). \tag{5}$$

The relationship in Equation (5) is commonly applied in growth/decay studies with  $\beta_1$  representing the value of maximal growth (if  $\beta_3 < 0$ ) or minimal decay (if  $\beta_3 > 0$ ). By using Equation (5), two models (OUGBM and OUGOU) were developed, as reported in Sections 2.1.1 and 2.1.2, respectively.

#### 2.1.1. OUGBM Model

Let  $x_t$  be a Brownian motion random variable that solves the SDE  $dx_t = \sigma_x dW_t^x$  (i.e.,  $\mu_t = 0$  and  $\sigma_t = \sigma_x$  in the SDE  $dx_t = \mu_t dt + \sigma_t dW_t^x$ ). Suppose the optimum of the response trait  $\theta_t^y$  has an exponential relationship with  $x_t$ , as shown in Equation (5). The first step is to express  $\theta_t^y$  in terms of model parameters  $\sigma_x$  in  $x_t$ . By taking a derivative in Equation (5) with respect to  $t$ , one has  $d\theta_t^y = \beta_2 d \exp(\beta_3 x_t)$ . Let  $f(t, x) = \exp(\beta_3 x)$  with partial derivative  $f_t = 0, f_x = \beta_3 \exp(\beta_3 x)$  and  $f_{xx} = \beta_3^2 \exp(\beta_3 x)$ . By applying Itô's lemma [26]  $df = (f_t + \mu_t f_x + \sigma_t^2 f_{xx}/2)dt + \sigma_t f_x dW_t$ , one has

$$d \exp(\beta_3 x_t) = \left( \sigma_x^2 \beta_3^2 \exp(\beta_3 x_t) / 2 \right) dt + (\sigma_x \beta_3 \exp(\beta_3 x_t)) dW_t,$$

which is known as the SDE for a geometric Brownian motion random variable  $f(x_t)$  with constant of percentage drift parameter  $\mu = \sigma_x^2 \beta_3^2 / 2$  and a constant of percentage volatility parameter  $\sigma = \sigma_x \beta_3$ . The analytical solution is

$$f(x_t) = f(x_0) \exp\left(\left(\mu - \sigma^2 / 2\right) + \sigma_x W_t\right).$$

Plugging  $f(x_t) = \exp(\beta_3 x_t)$  into Equation (5) and then simplifying the equation yields to an explicit representation for the optimum as follows.

$$\theta_t^y = \beta_1 + \beta_2 \exp(\beta_3(x_0 + \sigma_x W_t^x)). \tag{6}$$

To draw a sample for trait  $y_t$  considering the expression of  $y_t$  in Equation (2), it suffices to recognize the dynamics of  $\textcircled{A}$  in Equation (3), where  $\textcircled{A} = \int_0^t \alpha_y \exp(-\alpha_y(t-s)) \theta_s^y ds$ . This can be performed by replacing  $\theta_t^y$  with  $\beta_1 + \beta_2 \exp(\beta_3(x_0 + \sigma_x W_t^x))$  in Equation (6), which yields to

$$\textcircled{A} = \beta_1(1 - \exp(-\alpha_y t)) + \beta_2 \alpha_y \exp(-\alpha_y t + \beta_3 x_0) \int_0^t \exp(\alpha_y s + \beta_3 \sigma_x W_s^x) ds,$$

where  $\int_0^t \exp(\alpha_y s + \beta_3 \sigma_x W_s^x) ds := S_t$  is a definite integral with respect to time, and the integrand  $\exp(\alpha_y s + \beta_3 \sigma_x W_s^x)$  is a geometric Brownian motion variable [29–31].

Currently there is no analytical expression for  $S_t$ . The authors in [31,32] extensively studied the problem and provided a numerical solution through the Laplace transform. In particular, when  $t$  approaches to  $\infty$ , and the reciprocal of  $S_t$  has a limit distribution of gamma type with shape parameter  $(2\alpha_y) / (\beta_3^2 \sigma_x^2)$  and scale parameter  $(\beta_3^2 \sigma_x^2) / 2$  for  $\alpha_y > 0$  at  $t \rightarrow \infty$  (see Prop. 4.4.4 in [30]). In our modeling framework, since  $t$  represented evolutionary time and was of finite value (i.e.,  $0 < t < 1$  after scaling tree in the models), samples of  $S_t$  were drawn from the definite integral of a geometric Brownian motion variable  $\exp(\alpha_y t + \beta_3 \sigma_x W_t^x)$  with respect to time on time domain  $[0, t]$  using Simpson's rule [33]. Hence, given  $t, \alpha_y, \sigma_x, \sigma_y$ , samples of trait variables  $y_t, x_t$  and  $\theta_t^y$  were accordingly drawn with the aid of R package pracma [34] to compute the stochastic integral.

### 2.1.2. OUGOU Model

Let  $x_t$  be the Ornstein–Uhlenbeck process variable that solves the SDE.

$$dx_t = \alpha_x(\theta_x - x_t)dt + \sigma_x dW_t^x. \quad (7)$$

Given the exponential relationship between  $\theta_t^y$  and  $x_t$  as  $\theta_t^y = \beta_1 + \beta_2 \exp(\beta_3 x_t)$ , by taking differentials with respect to  $t$  on both sides yields to  $d\theta_t^y = \beta_2 d \exp(\beta_3 x_t)$ . Let  $z_t = \exp(\beta_3 x_t)$ , again by Itô's lemma and use Equation (7), one has

$$dz_t = \alpha_x \beta_3 \left( \theta_x - x_t + (\sigma_x^2 \beta_3) / (2\alpha_x) \right) \exp(\beta_3 x_t) dt + \sigma_x \beta_3 \exp(\beta_3 x_t) dW_t^x,$$

which implies that  $z_t$  is a geometric Ornstein–Uhlenbeck process [28].  $z_t$  can be expressed as

$$\log z_t = \log z_0 \exp(-\alpha_x \beta_3 t) + \theta_x (1 - \exp(-\alpha_x \beta_3 t)) + \beta_3 \sigma_x \exp(-\alpha_x \beta_3 t) \int_0^t \exp(\alpha_x \beta_3 s) dW_s^x.$$

Hence,  $\theta_t^y = \beta_1 + \beta_2 \exp(\beta_3 x_t)$  can be expressed as

$$\theta_t^y = \beta_1 + \beta_2 \exp \left( \beta_3 x_0 \exp(-\alpha_x \beta_3 t) + \theta_x (1 - \exp(-\alpha_x \beta_3 t)) + \sigma_x \beta_3 \exp(-\alpha_x \beta_3 t) \int_0^t \exp(\alpha_x \beta_3 s) dW_s^x \right). \quad (8)$$

To draw a sample for trait  $y_t$ , considering the expression of  $y_t$  in Equation (2), it suffices to recognize the dynamics of  $\textcircled{A}$  in Equation (3) where  $\textcircled{A} = \int_0^t \alpha_y \exp(-\alpha_y(t-s)) \theta_s^y ds$ . By using Equation (8) for  $\theta_t^y$ , one has

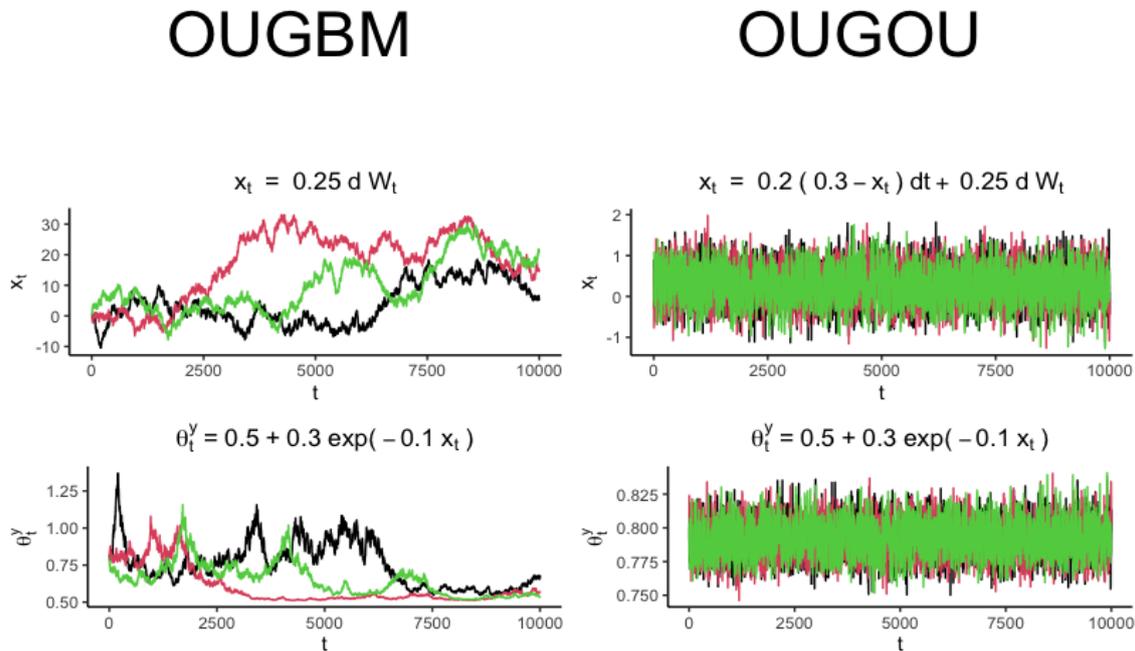
$$\textcircled{A} = \beta_1 (1 - \exp(-\alpha_y t)) + \beta_2 \alpha_y \exp(-\alpha_y t) A_t$$

where

$$A_t = \int_0^t \exp \left( -\alpha_y s + \theta_x + \exp(-\alpha_x \beta_3 s) \left( \beta_3 x_0 - 1 + \sigma_x \beta_3 \int_0^s \exp(\alpha_x \beta_3 u) dW_u^x \right) \right) ds$$

is a definite integral of geometric OU process with respect to time. Currently, there is no analytical expression for  $A_t$ , so we used R package pracma [34] to draw samples of  $A_t$  where the definite integral was computed over a finite grid by Simpson's rule. On each grid sample of  $\int_0^s \exp(\alpha_x \beta_3 u) dW_u^x$  were generated by a normal variable with mean 0 and variance  $(\exp(2\alpha_x \beta_3 s) - 1) / (2\alpha_x \beta_3)$ .

Sections 2.1.1 and 2.1.2 provide the fundamental framework for phylogenetic exponential optimal regression for adaptive trait evolution. Once  $\textcircled{A}$  in Equation (3) was fully recognized, samples of trait variable  $y_t$  could be drawn accordingly by using the  $y_t$  expressed in Equation (2). Trajectories for optimal response  $\theta_t^y$  and the covariate  $x_t$  for the OUGBM and OUGOU models are shown in Figure 2.



**Figure 2.** Trajectory simulation for the OUGBM and the OUGOU models. Each plot contains three realizations generated from the corresponding model. Trajectories of optimum  $\theta_t^y$  were generated by evaluating the exponential relationship  $\theta_t^y = 0.5 + 0.3 \exp(-0.1x_t)$  from the realization of covariate  $x_t$ .

2.2. Optimal Linear Regression

Optimal  $\theta_t^y$  of response trait  $y_t$  and its covariate  $x_t$  has a linear relationship, as follows:

$$\theta_t^y = \beta_1 + \beta_2 x_t. \tag{9}$$

Two optimal linear regression models, OUBM [20] using a BM covariate, and OUOU [21] using an OU process covariate were developed in the literature. We included both models in this study for comparison with the optimal exponential regression model.

2.2.1. OUBM Model

When the dynamic of optimum  $\theta_t^y$  was assumed with a linear relationship with the BM covariate  $x_t = \sigma_x W_t^x$ , then

$$\theta_t^y = \beta_1 + \beta_2 \sigma_x W_t^x. \tag{10}$$

To draw a sample for trait  $y_t$ , considering the expression of  $y_t$  in Equation (2), it suffices to recognize the dynamics of  $\textcircled{A}$  in Equation (3).  $\textcircled{A} = \int_0^t \alpha_y \exp(-\alpha_y(t-s)) \theta_s^y ds$  is computed by replacing  $\theta_t^y$  with the right-hand side of Equation (10). Hence,

$$\textcircled{A} = \beta_1(1 - \exp(-\alpha_y t)) + \beta_2 \sigma_x \exp(-\alpha_y t) \left( W_t^x - \int_0^t \exp(\alpha_y s) dW_s^x \right)$$

which is a normal variable with mean  $\beta_1(1 - \exp(-\alpha_y t))$  and variance  $\beta_2^2 \sigma_x^2 (t \exp(-2\alpha_y t) - 2(\exp(-\alpha_y t) - \exp(-2\alpha_y t))/\alpha_y + (1 - \exp(-2\alpha_y t))/(2\alpha_y))$ .

2.2.2. OUOU Model

Let  $\theta_t^y = \beta_1 + \beta_2 x_t$  where  $x_t = \theta_x + \exp(-\alpha_x t)(x_0 - \theta_x + \sigma_x \int_0^t \exp(\alpha_x s) dW_s^x)$  is a random OU process variable. By replacing  $x_t$  in terms of  $t, \alpha_x, \sigma_x, W_t^x$  to Equation (9), one has

$$\theta_t^y = \beta_1 + \beta_2 \theta_x + \beta_2 \exp(-\alpha_x t) \left( x_0 - \theta_x + \sigma_x \int_0^t \exp(\alpha_x s) dW_s^x \right). \quad (11)$$

To draw a sample for trait variable  $y_t$ , considering the expression of  $y_t$  in Equation (2), it suffices to recognize the dynamics of  $\textcircled{A}$  in Equation (3), where  $\textcircled{A} = \int_0^t \alpha_y \exp(-\alpha_y(t-s)) \theta_s^y ds$ . By expressing  $\theta_t^y$  in Equation (11),

$$\textcircled{A} = \textcircled{1} + \textcircled{2} + \textcircled{3}$$

where

$$\textcircled{1} = \int_0^t \alpha_y \exp(-\alpha_y(t-s)) (\beta_1 + \beta_2 \theta_x) ds = (\beta_1 + \beta_2 \theta_x) (1 - \exp(-\alpha_y t)),$$

$$\textcircled{2} = \alpha_y \beta_2 (x_0 - \theta_x) \int_0^t \exp(-\alpha_y t + (\alpha_y - \alpha_x) s) ds = -\alpha_y \beta_2 (x_0 - \theta_x) (\exp(-\alpha_y t) - \exp(-\alpha_x t)) / (\alpha_y - \alpha_x),$$

and

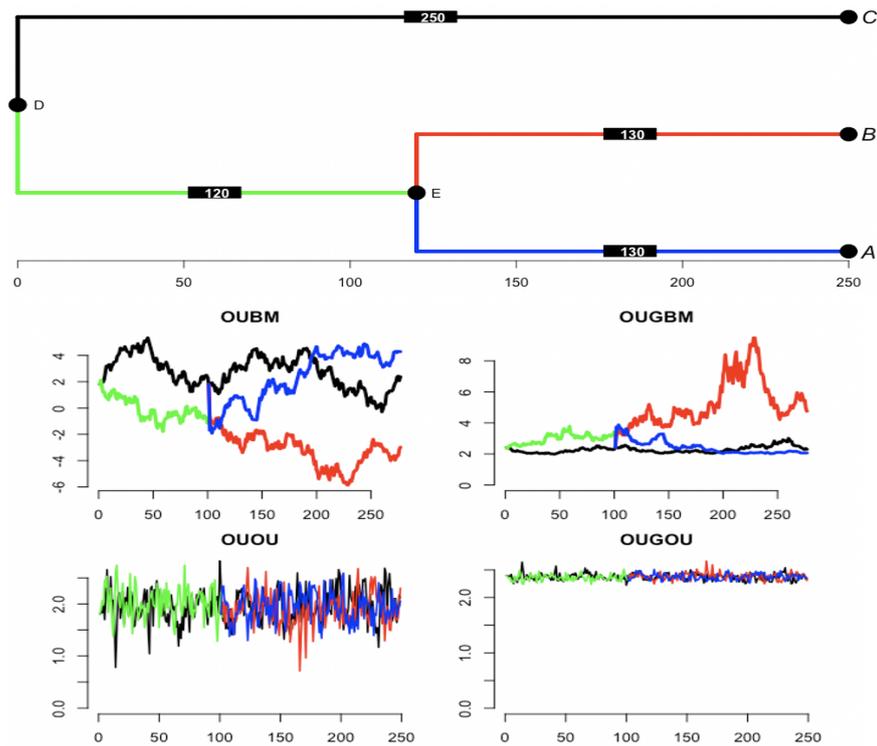
$$\begin{aligned} \textcircled{3} &= \alpha_y \sigma_x \beta_2 \int_0^t \exp(-\alpha_y t + (\alpha_y - \alpha_x) s) \int_0^s \exp(\alpha_x u) dW_u^x ds \\ &= \alpha_y \sigma_x \beta_2 \exp(-\alpha_y t) / (\alpha_y - \alpha_x) \left( \exp(-\alpha_x t) \int_0^t \exp(\alpha_x s) dW_s^x - \exp(-\alpha_y t) \int_0^t \exp(\alpha_y s) dW_s^x \right) \end{aligned}$$

which is a normal variable with mean 0 and variance  $\alpha_y^2 \sigma_x^2 \beta_2^2 ((1 - \exp(-2\alpha_x t)) / (2\alpha_x) + (1 - \exp(-2\alpha_y t)) / (2\alpha_y)) / (\alpha_y - \alpha_x)^2$ .

### 2.3. Optimal Adaptive-Trait Evolution along Phylogenetic Tree

A phylogenetic tree provides evidence of the summary of evolutionary history of living species [35]. For a mutation occurring in an individual identified on a lineage of the tree where the mutation changed the phenotype of the organism such as kangaroos, that mutation may change the moving style from bipedal walking to bipedal hopping. Such a mutation may need many generations to be achieved. However, the trait may be difficult to predict when a lineage is fixed for a derived trait; descendants would inherit the trait until a subsequent evolution change occurs. For a clade that contains marsupials such as kangaroos, wallabies, koalas, and possums, their differences are the results of changes after their common ancestor begins to diversify. Here, a phylogenetic tree provides information to organize this biological diversity where internal nodes depict a common ancestry and contain the formation of the degree of relatedness that is relative to the entire evolutionary history. As adopting tree thinking that living species share a common ancestor is broadly accepted in evolutionary biology, the tree provides evidence in how to conceptualize the broad sweep of biological diversity.

For trait evolution, a group of currently observed species has beautifully expressed affinity by the evolutionary tree. From the mathematical side, changes in trait value among a group of species along a phylogenetic tree can be realized by the relevant stochastic process. One realization of  $y_t$  using a BM predictor in the OUBM and OUGBM models, and one realization of trajectories  $y_t$  for the OU process-based predictor in the OUOU and OUGOU models using a 3-species phylogenetic tree are shown in Figure 3. Box plots of 100 simulated optimal-trait and response trait samples under the tree in Figure 3 using the tree traversal algorithm can be accessed in Figure S1 in the online Supplemental Material, displaying the spread of traits across models.



**Figure 3.** Simulation of optimal trajectories along the tree using Gaussian process covariates. (top) Tree of 3 taxa is simulated from coalescent process using R package's `ape` function `rcoal` [36]. Original tree has branch of length 250 from root node D to tip C, 120 from node D to node E, and 130 from E to B and from E to A. Edge length is increased by multiplying a constant, and trajectories are simulated at each unit under relevant processes. Trajectories of predictor  $x_t$  assume Brownian motion with rate  $\sigma_x = 0.5$  first simulated along the tree with  $x_0 = 0$  at root  $\rho$ . For the covariate  $x_t$  under an Ornstein–Uhlenbeck process dynamics with parameters  $\alpha_x = 0.625$ , optimum  $\theta_x = 0.25$ , and rate  $\sigma_x = 0.5$ . Trait is first simulated along the tree with starting point  $x_0 = 0$  at the root  $\rho$ . Optimum  $\theta_t^y$  as a function of  $x_t$  is computed under each model using the regression parameters  $(\beta_1, \beta_2, \beta_3) = (1.8, 0.6, -0.2)$ . For linear model  $\theta_t^y = 1.8 + 0.6x_t$ , and for exponential model  $\theta_t^y = 1.8 + 0.6 \exp(-0.2x_t)$ .

#### 2.4. Approximate Bayesian Computation

Due to the exponential relationship between  $\theta_t^y$  and its covariate  $x_t$ , the stochastic variable  $y_t$  shown in Equation (2) includes a definite integral of stochastic variable  $\theta_t^y$  with respect to time  $t$ . The distribution for the definite integral of the geometric OU process with respect to time is currently not known. Hence, the OUGOU model lack of closed-form likelihood as stochastic variable (A) in Equation (3) embedded in  $y_t$  is intractable. Approximate Bayesian computation (ABC) was used for statistical inference herein. The ABC procedure is a likelihood free based method used for model inference. To start an ABC algorithm, data are first simulated from the model using parameters drawn from prior distributions. Then, a set of the summary statistics for samples and raw data are calculated. For the ABC rejection method, a distance function  $d(\cdot)$  and a threshold  $\delta$  are used to determine posterior samples by comparing summary statistics of observed data and simulated data [37].

To determine posterior samples, we adopted the 12 summary statistics from [19], and used the mean, median, standard deviation, skewness, kurtosis, and the phylogenetic tree based statistics: the contrast mean, the contrast standard deviation, the contrast skewness, the contrast kurtosis [1], and two phylogenetics-related statistics: (i) the Bloomberg's  $\mathcal{K}$  statistic (measures the relatedness of species in a clade when compared to randomly selected species from the same tree  $\mathcal{T}$ ) [38] and (ii) the Pagel's  $\lambda$  statistic (measures the

strength of trait heritability from the ancestor) [39]. For  $\mathcal{K} = \text{obs}\left(\frac{\text{MSE}_0}{\text{MSE}}\right) / \exp\left(\frac{\text{MSE}_0}{\text{MSE}}\right)$  where  $\text{MSE}_0$  is the mean square root of the observed tip data measured from phylogenetic correct mean and  $\text{MSE}$  is the mean squared error of the observed data calculated using the variance covariance matrix derived from the candidate tree. For trait vector  $\mathbf{Y}$  following a Brownian motion model (i.e.,  $\mathbf{Y} \sim \text{MVN}(\mu\mathbf{1}_n, \sigma^2\mathbf{C})$ ), one has

$$\text{MSE}_0 = (\mathbf{Y} - \hat{\mu}\mathbf{1}_n)^t(\mathbf{Y} - \hat{\mu}\mathbf{1}_n)/(n-1) \text{ and } \text{MSE} = (\mathbf{Z} - \hat{\mu}\mathbf{1}_n)^t(\mathbf{Z} - \hat{\mu}\mathbf{1}_n)/(n-1),$$

where  $\hat{\mu} = \mathbf{1}^t\mathbf{C}^{-1}\mathbf{Y}/\mathbf{1}^t\mathbf{C}^{-1}\mathbf{1}$  is the phylogenetic corrected mean, and  $\mathbf{Z} = \mathbf{P}\mathbf{Y}$  is the transformed  $\mathbf{Y}$  vector obtained from the generalized least-square procedure. Matrix  $\mathbf{P}$  satisfies equation  $\mathbf{D}\mathbf{V}\mathbf{D} = \mathbf{I}$ , where  $\mathbf{V} = \sigma^2\mathbf{C}$  is the variance covariance matrix and  $\mathbf{I}$  is the identity matrix. Relatively small  $\text{MSE}_0$  occurs when there is little covariance within the tip data that is explained by the candidate tree, and it leads to a smaller value of the ratio of  $\text{MSE}_0/\text{MSE}$  (weaker phylogenetic signal). Conversely, while if the candidate tree precisely demonstrate the variance-covariance pattern observed in the data, then there is a small  $\text{MSE}$ , which results in large value of  $\text{MSE}_0/\text{MSE}$  (stronger phylogenetic signal) [38]. Pagel's  $\lambda$  statistic parameter  $\lambda$  is calculated by optimizing the likelihood function of the model, assuming that observed trait vector  $\mathbf{Y} = (y_1, \dots, y_n)^t$  follows multivariate normal distribution  $\mathbf{Y} \sim \text{MVN}(\mu\mathbf{1}_n, \sigma^2(\lambda\mathbf{C} + (1-\lambda)\mathbf{I}))$  where  $\mathbf{1}_n = (1, \dots, 1)$  is vector of 1s, and  $\mathbf{I}$  is an identity matrix,  $\mathbf{C}$  is phylogenetic affinity matrix transformed from the given phylogenetic tree [40]. Since both the MLE for mean  $\mu$  and variance  $\sigma^2$  can be written as a function of  $\lambda$ ,

$$\hat{\mu} = \frac{\mathbf{1}_n[\lambda\mathbf{C} + (1-\lambda)\mathbf{I}]^{-1}\mathbf{Y}}{\mathbf{1}_n^t[\lambda\mathbf{C} + (1-\lambda)\mathbf{I}]^{-1}\mathbf{1}_n} \text{ and } \hat{\sigma}^2 = \frac{(\mathbf{Y} - \hat{\mu}\mathbf{1}_n)^t[\lambda\mathbf{C} + (1-\lambda)\mathbf{I}]^{-1}(\mathbf{Y} - \hat{\mu}\mathbf{1}_n)}{n},$$

$\lambda$  can be estimated by optimizing the likelihood function over its domain  $\lambda \in [0, 1]$ . Those statistics resulted in a great interest in evolutionary-biology research [19,41,42]. Euclidean distance measure  $d = d(S(\mathbf{Y}, \mathbf{X}), S(\mathbf{Y}', \mathbf{X}')) = \|S(\mathbf{Y}, \mathbf{X}) - S(\mathbf{Y}', \mathbf{X}')\|_2$  corresponds to those statistics  $S$ , where  $S(\mathbf{Y}, \mathbf{X})$  and  $S(\mathbf{Y}', \mathbf{X}')$  are computed from observed and simulated-trait data, respectively. The procedure for parameter estimation under the ABC rejection method is shown in Algorithm 1.

---

**Algorithm 1** Approximate Bayesian computation for the models of adaptive trait evolution.

---

**Require:** Trait datasets: response  $\mathbf{Y} = (y_1, y_2, \dots, y_n)^t$ , covariate  $\mathbf{X} = (x_1, x_2, \dots, x_n)^t$ ; tree  $\mathcal{T}$ , a threshold  $\delta$ , and model  $\mathcal{M}_j$ , starting parameter value  $\Theta_{0j}$  and priors  $\pi_j(\cdot)$  for  $\Theta_j, j = 1, 2, \dots, m$ .

**Ensure:** Posterior samples  $\Theta_i, i = 1, 2, \dots, mL\delta$ .

- 1: Calculate the summary statistics  $S_0 = S(\mathbf{Y}, \mathbf{X})$ .
  - 2: **for**  $i = 1, 2, \dots, L$  **do**
  - 3:   **for**  $j = 1, 2, \dots, m$  **do**
  - 4:     Draw samples  $\Theta_{ij}$  from prior  $\pi_j$  under model  $\mathcal{M}_j$ .
  - 5:     Simulate trait set  $\mathbf{Y}_{ij}, \mathbf{X}_{ij}$  under model  $\mathcal{M}_j$  and its parameters  $\Theta_{ij}$ .
  - 6:     Evaluate summary statistics  $S_{ij} = S(\mathbf{Y}_{ij}, \mathbf{X}_{ij})$ .
  - 7:     Compute distance  $d_{ij}$  between  $S_0$  and  $S_{ij}$ .
  - 8:   **end for**
  - 9: **end for**
  - 10: Order the distance  $\{d_{ij}\}_{i,j=1}^{L,m}$  from the small least to the largest  $\{d_k, k = 1, 2, \dots, Lm\}$ .
  - 11: **return** Posterior samples  $\Theta_i : i = 1, 2, \dots, mL\delta$ .
-

### 2.5. Interpretation of Change of Optimum by Its Covariate

As traits are logarithm-transformed prior to analysis, the change in response traits is measured on a ratio scale under two types of regression methods: (i) optimal linear regression or (ii) optimal exponential regression. Below, we briefly describe the change in optimum by its covariate.

(i) In optimal linear regression: First, given  $\theta_{y_1} = \beta_1 + \beta_2 x_1$  and  $\theta_{y_2} = \beta_1 + \beta_2 x_2$ , the two equations in log scale are written as  $\log(\theta_{y_1}) = \beta_1 + \beta_2 \log(x_1)$  and  $\log(\theta_{y_2}) = \beta_1 + \beta_2 \log(x_2)$ . The difference between the two equations is  $\log(\theta_{y_2}) - \log(\theta_{y_1}) = \beta_2(\log(x_2) - \log(x_1))$  which implies that  $\log(\theta_{y_2}/\theta_{y_1}) = \beta_2 \log(x_2/x_1)$ . Hence,  $\theta_{y_2}/\theta_{y_1} = (x_2/x_1)^{\beta_2}$  depends on values of  $\beta_2$ ,  $x_1$  and  $x_2$ . Let  $x_1 = 1$ ,  $x_2 = 1.1x_1$  and  $\beta_2 = 0.5$ , then  $\theta_{y_2}/\theta_{y_1} = (1.1x_1/x_1)^{0.5} = (1.1)^{0.5} = 1.0488$ , which means that a 10% increase in the covariate  $x$  results in 4.88% increase in the optimum of response  $\theta_y$ .

(ii) In optimal exponential regression: First, given  $\theta_{y_1} = \beta_1 + \beta_2 \exp(\beta_3 x_1)$  and  $\theta_{y_2} = \beta_1 + \beta_2 \exp(\beta_3 x_2)$ , the two equations in log scale are written as  $\log(\theta_{y_1}) = \beta_1 + \beta_2 \exp(\beta_3 \log(x_1))$  and  $\log(\theta_{y_2}) = \beta_1 + \beta_2 \exp(\beta_3 \log(x_2))$ . The difference between the two equations is  $\log(\theta_{y_2}) - \log(\theta_{y_1}) = \beta_2 x_2^{\beta_3} - \beta_2 x_1^{\beta_3}$ , which implies that  $\log(\theta_{y_2}/\theta_{y_1}) = \beta_2(x_2^{\beta_3} - x_1^{\beta_3})$ . Hence,  $\theta_{y_2}/\theta_{y_1} = \exp(\beta_2(x_2^{\beta_3} - x_1^{\beta_3}))$  depends on covariate  $x_1, x_2$ . Let  $\beta_2 = 0.5, \beta_3 = 1$  and  $x_2 = 1.1x_1$ , then  $\theta_{y_2}/\theta_{y_1} = \exp(0.5(1.1x_1 - x_1))$ . Set  $x_1 = 1$ , then  $\theta_{y_2}/\theta_{y_1} = \exp(0.5((1.1) - 1)) = 1.0513$ . So, a 10% increase in covariate  $x$  would result in a 5.13% increase in optimal response  $\theta_t^y$ .

## 3. Results

### 3.1. Simulation

#### 3.1.1. Parameter Estimation

To validate the new models, their performance was assessed through extensive simulations. Prior parameter distributions were assumed to be independent. Some appropriate priors were selected because of the models' lack of tractable likelihood without a conjugate prior. A balanced tree of 16, 32, 64 and 128 with a height of 1, and Grafen branch length simulated by R: ape was used for the simulation. To obtain reliable estimates, 2000 ( $=500 \times 4$ ) posterior samples were obtained from four runs, in each run, 50,000 samples were generated, and a tolerance rate ( $\delta = 0.01$ ) was used to obtain 500 posterior samples. Two sets of true parameters and priors were used for simulation. For the first set, all priors used uniform distribution. For the second set, priors were set to a specific distribution by intuitive beliefs about the true values of the parameters [43,44]. For the nonuniform prior,  $\theta_x$  was assumed to be normal, as it was reasonable to assumed that the optimum remained at the peaks. An example of using the normal prior comes from a study of coral polyp evolution [19], where a suitable prior for the adaptive optimum  $\theta_x$  of polyp thickness used the normal distribution of polyp thickness across all corals. The exponential prior was used for force parameters  $\alpha_x$  and  $\alpha_y$ , and the inverse gamma was used for the rate parameters  $\sigma_x$  and  $\sigma_y$ . The setup of hyperparameters for priors is listed in Table 1.

**Table 1.** Simulation setup for true parameter values and prior distributions.  $\mathcal{U}$ , uniform distribution;  $\mathcal{E}$ , exponential distribution;  $\mathcal{IG}$  inverse gamma distribution; and  $\mathcal{N}$ , normal distribution. In inverse gamma distribution, sh = shape and sc = scale.

Par	True 1	Prior 1	True 2	Prior 2
$\alpha_y$	0.50	$\mathcal{U}(0, 1)$	0.20	$\mathcal{E}$ (rate = 5)
$\alpha_x$	0.125	$\mathcal{U}(0, 0.25)$	0.125	$\mathcal{E}$ (rate = 8)
$\theta_x$	0.00	$\mathcal{U}(-5, 5)$	1.00	$\mathcal{N}$ (mean = 1, sd = 1)
$\sigma_x$	2.50	$\mathcal{U}(0, 5)$	0.5	$\mathcal{IG}$ (sh = 2, sc = 0.5)
$\sigma_y$	1.00	$\mathcal{U}(0, 2)$	0.5	$\mathcal{IG}$ (sh = 2, sc = 0.5)
$\beta_1$	0.00	$\mathcal{U}(-1, 1)$	0.00	$\mathcal{U}(-5, 5)$
$\beta_2$	1.00	$\mathcal{U}(0, 2)$	-2.00	$\mathcal{U}(-7, 3)$
$\beta_3$	-0.50	$\mathcal{U}(-1, 0)$	-0.5	$\mathcal{U}(2, -3)$

Root state  $\rho = (\rho_y, \rho_x, \rho_\theta)$  was set to a trivial value of 0 for all models. For each taxon size, one trait was simulated under each model from the simulation.

The results for uniform priors from this simulation of model parameters are shown in Table 2. The results for the second set using informative priors from this simulation of model parameters are shown in Tables S2 and S3 in the online Supplemental Material.

**Table 2.** Simulation results of validating models through model parameter estimation using uniform prior. Four different taxon sizes of 16, 32, 64, and 128 were used for the four models (OUGBM, OUGOU, OUBM and OUOU). Means and 95% credible intervals using 2000 posterior samples from 4 individual runs on each model are reported for each model parameter on each column.

Model	Taxa	$\alpha_y$	$\alpha_x$	$\theta_x$	$\sigma_x$	$\sigma_y$
	True Value	0.5	0.125	0	2.5	1
OUGBM	16	0.52 (0.06, 0.96)			2.16 (0.26, 4.59)	0.89 (0.16, 1.83)
	32	0.53 (0.08, 0.95)			1.83 (0.25, 4.3)	0.95 (0.2, 1.82)
	64	0.54 (0.09, 0.95)			1.66 (0.2, 4.1)	0.93 (0.2, 1.78)
	128	0.52 (0.08, 0.95)			1.65 (0.21, 4.07)	0.91 (0.2, 1.78)
OUGOU	16	0.44 (0.04, 0.95)	0.12 (0.01, 0.24)	-1.14 (-4.49, 2.81)	2.25 (0.65, 4.28)	1.16 (0.38, 1.88)
	32	0.47 (0.04, 0.95)	0.12 (0.01, 0.24)	-1.22 (-4.59, 2.75)	2.52 (0.82, 4.56)	0.99 (0.19, 1.83)
	64	0.48 (0.04, 0.95)	0.12 (0.01, 0.24)	-1.16 (-4.58, 2.93)	2.61 (0.87, 4.59)	0.95 (0.18, 1.81)
	128	0.49 (0.04, 0.95)	0.12 (0.01, 0.24)	-1.16 (-4.58, 2.88)	2.57 (0.79, 4.57)	0.9 (0.16, 1.78)
OUBM	16	0.5 (0.05, 0.95)			2.14 (0.59, 4.22)	1.13 (0.11, 1.92)
	32	0.56 (0.07, 0.96)			2.05 (0.55, 4.21)	1.05 (0.1, 1.92)
	64	0.52 (0.06, 0.96)			1.95 (0.48, 4.12)	1.07 (0.11, 1.92)
	128	0.54 (0.06, 0.96)			1.92 (0.51, 4.05)	1.06 (0.11, 1.91)
OUOU	16	0.53 (0.05, 0.95)	0.12 (0.01, 0.24)	0.63 (-4.15, 4.49)	2.13 (0.64, 4.11)	1.08 (0.12, 1.92)
	32	0.55 (0.05, 0.95)	0.12 (0.01, 0.24)	0.94 (-4.15, 4.56)	1.9 (0.42, 4)	1.06 (0.13, 1.9)
	64	0.53 (0.05, 0.94)	0.12 (0.01, 0.24)	0.79 (-4.18, 4.54)	1.85 (0.42, 3.96)	1.06 (0.12, 1.91)
	128	0.55 (0.05, 0.95)	0.12 (0.01, 0.24)	0.79 (-4.26, 4.54)	1.81 (0.44, 3.92)	1.05 (0.11, 1.9)

Overall, parameters could be estimated reasonably well with acceptable accuracy. The posterior mean of each parameter was close to the true parameter value under uniform priors. Results for the uniform priors from this simulation of regression parameters are shown in Table 3. On each taxon, most models showed reasonable mean estimates for  $\beta_1$  (true 0),  $\beta_2$  (true 1),  $\beta_3$  (true -0.5). Results guaranteed that Algorithm 1 provided a reliable procedure for estimating parameters.

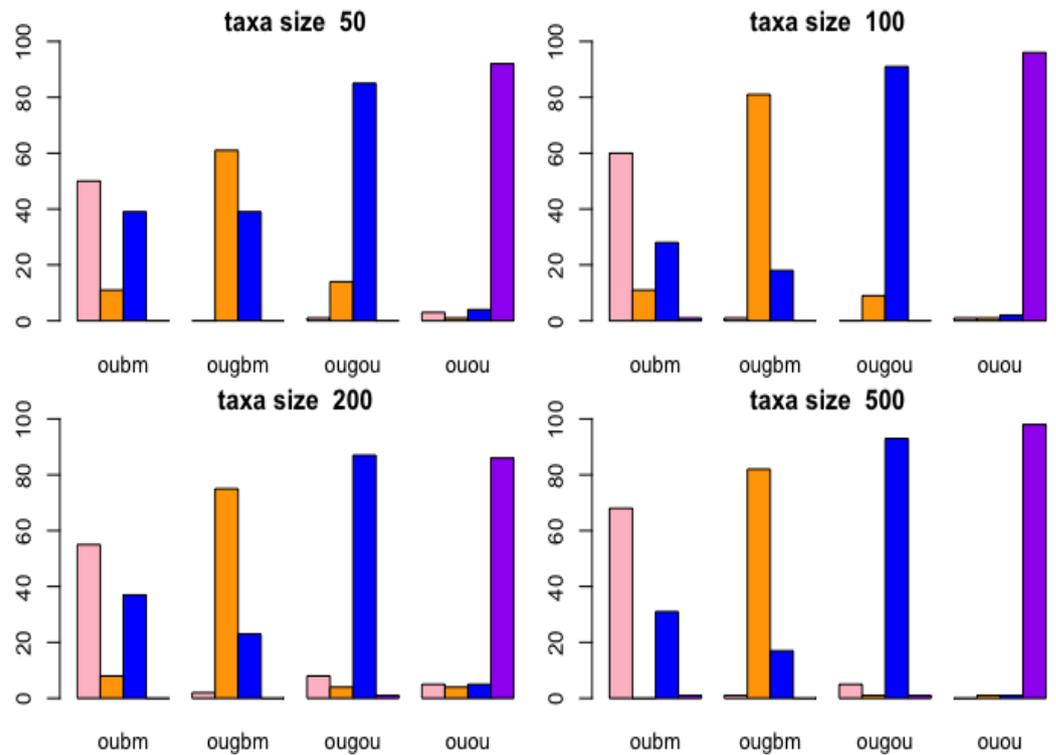
**Table 3.** Simulation results of validating models through regression parameter estimation using uniform prior. Four different taxon sizes of 16, 32, 64, and 128 were used for four models (OUGBM, OUGOU, OUBM and OUOU). Means and 95% credible intervals using 2000 posterior samples from 4 individual runs on each model were reported for each regression parameter on each column.

Model	Taxa	$\beta_1$	$\beta_2$	$\beta_3$
	True Value	0	1	−0.5
OUGBM	16	−0.08 (−0.91, 0.84)	1.01 (0.14, 1.89)	−0.46 (−0.94, −0.05)
	32	−0.02 (−0.9, 0.87)	0.95 (0.12, 1.87)	−0.47 (−0.95, −0.05)
	64	−0.02 (−0.91, 0.89)	0.96 (0.14, 1.86)	−0.48 (−0.95, −0.05)
	128	0.01 (−0.9, 0.9)	0.97 (0.14, 1.86)	−0.48 (−0.95, −0.04)
OUGOU	16	−0.01 (−0.92, 0.88)	0.88 (0.06, 1.89)	−0.47 (−0.92, −0.05)
	32	−0.03 (−0.92, 0.88)	0.89 (0.07, 1.89)	−0.48 (−0.94, −0.05)
	64	−0.05 (−0.92, 0.88)	0.88 (0.07, 1.89)	−0.48 (−0.93, −0.05)
	128	−0.05 (−0.92, 0.88)	0.91 (0.07, 1.89)	−0.49 (−0.94, −0.05)
OUBM	16	−0.03 (−0.88, 0.89)	0.8 (0.11, 1.81)	
	32	0.01 (−0.89, 0.9)	0.78 (0.09, 1.82)	
	64	−0.01 (−0.9, 0.89)	0.79 (0.09, 1.83)	
	128	−0.02 (−0.9, 0.89)	0.8 (0.09, 1.83)	
OUOU	16	−0.11 (−0.9, 0.88)	0.86 (0.11, 1.81)	
	32	−0.11 (−0.9, 0.88)	0.81 (0.09, 1.83)	
	64	−0.1 (−0.89, 0.88)	0.85 (0.1, 1.85)	
	128	−0.09 (−0.89, 0.88)	0.82 (0.1, 1.84)	

### 3.1.2. Cross-Validation

Cross-validation is used to investigate how many taxa are needed and whether the correct model can be chosen from a candidate set. Leave-one-out cross-validation was performed under ABC using the R: abc package [37]. The balanced trees of taxon sizes 64, 128, 256, and 512 taxa were simulated using R: ape package, while 10,000 birth–death trees of taxon sizes 50, 100, 200, and 500 with birth rate 2, death rate 0.5, the time since origin 2, and probability of 0.5 for each tip were included in the final tree and simulated using the R: TreeSim package [45]. One trait datum was simulated along a given tree using parameters with values set up in Table 1 using uniform distribution. To assess if ABC could distinguish between the models, the 12 summary statistics were calculated in each model. For each model, the size of the cross-validation samples was set to 100.

Results of the confusion matrix are reported with birth–death tree cases by bar plots in Figure 4. In the lower right panel (taxon size 500) in Figure 4, the bar plots in the OUGOU categories shows that for ABC model choice will identify the OUOU model 1 time, the OUGOU model 93 times, the OUGBM model 1 time, the OUBM model 5 times, which yielded to the misclassification proportion for the OUGOU model of  $(1 - 93/100) \times 100\% = 7\%$ ; in the upper left panel (taxon size 50), the rightmost bar plots in the OUOU categories shows that the ABC model identified the OUOU model (purple) 90 times, the OUGOU model (blue) 4 times, the OUGBM model (orange) 1 time and the OUBM model (pink) 3 times among the 100 samples, which yielded to the misclassification proportion for the OUOU model of  $(1 - 90/100) \times 100\% = 10\%$ .

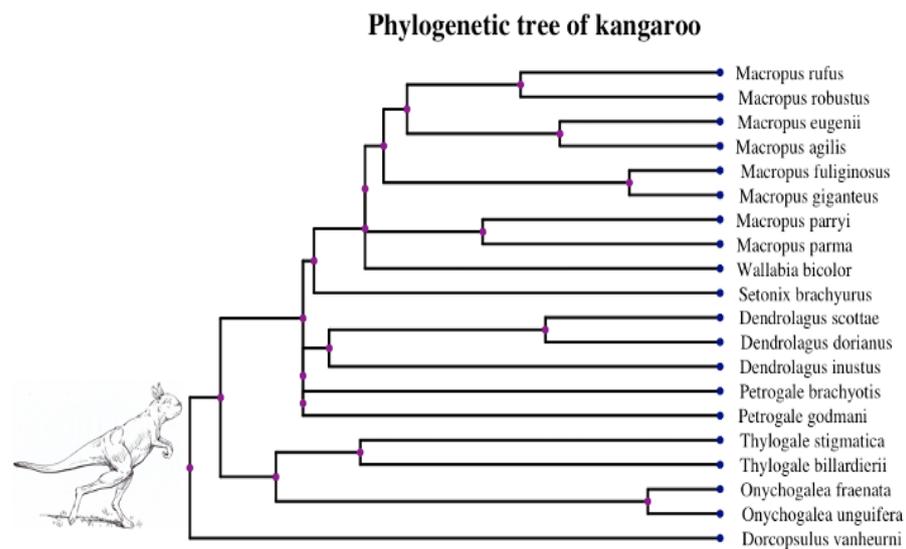


**Figure 4.** Cross-validation of models using birth-death tree. Bar plots show results of confusion matrices from cross-validation analysis under Approximate Bayesian Computatin (ABC) multinomial logistic-regression method for models of adaptive trait evolution. Four taxon sizes of 50, 100, 200, and 500 of birth-death trees were considered. The actual model is shown in the horizontal label for each bar plot on each panel, and the frequency of correctly identifying the models is represented by the height of the bar plots.

From this analysis, models are distinguishable at each taxon size. When taxa increase, ABC can more frequently identify the correct models. There are other factors, such as the choices of parameters and number of models in the candidates set, which may impact the power of correctly identifying the correct models. Here, we used constant factors. Results of the confusion matrix for each model reported with the balanced tree cases were similar to those of the case with the birth-death tree cases, and can be accessed in Figure S2 in the online Supplemental Material.

### 3.2. Empirical Analysis

Kangaroos are bipedal, and using their femoral midshaft circumference is especially suitable for predicting body mass. We used the trait datasets in [18] and applied our models by treating femoral bone circumference as the covariate to explore its impact on the optimum of body mass. The phylogeny of kangaroos is shown in Figure 5 and trait values corresponding to the species can be accessed in Table S3 in the Supplemental Material. Prior to log transformation, data were scaled by the feature-scaling method [46], while the curved relationship remains unchanged under this scaling. Our ABC algorithm worked properly for the dataset where traits were simulated within a reasonable range.



**Figure 5.** Phylogenetic tree of 20 giant kangaroo species in [18]. Image at the root of the tree is a reconstruction of *Sthenurus stirlingi* [47], an extinct giant kangaroo in walking pose.

Posterior means for the parameters of each model is shown in Table 4.

**Table 4.** Posterior means of parameters for kangaroo dataset.

Model	Parameter							
	$\alpha_x$	$\alpha_y$	$\sigma_x$	$\sigma_y$	$\theta_x$	$\beta_1$	$\beta_2$	$\beta_3$
EXP						0.5987	0.2946	0.4251
OUGBM		0.0016	1.3420	0.7888		0.6848	0.2985	0.4281
OUGOU	0.0014	0.0015	0.8034	0.5480	-1.2113	0.5208	0.3258	0.4293
LS						0.2078	0.5125	
OUBM		0.0015	1.4413	0.7392		0.1504	0.4996	
OUOU	0.0014	0.0014	0.9952	0.6931	-0.5732	0.2392	0.5713	

For the kangaroo dataset, all models reported relatively small estimates of force parameters  $\alpha_x$  and  $\alpha_y$ , which indicated that relative weak force was detected to pull the trait back to its optimum during evolution. For rate parameter,  $\sigma_x$ , the OU\*OU models reported a smaller value than that of OU\*BM models, while  $\sigma_y$  for the response trait in all model was between 0.5 and 0.8. For optimum parameter for the covariate trait,  $\theta_x$ , both exponential model (OUGOU) and linear model reported negative values. For regression parameters, linear models (OUBM and OUOU) reported a positive regression slope  $\beta_2$  which was consistent with [18], where positive correlation among traits was reported. Regression parameters  $\beta_i, i = 1, 2, 3$  reported relatively closed values across the same class of models. Overall, our results predicted that bone circumference has a positive effect on the optimum body mass, which is consistent with the result in [18] when using phylogenetic independent contrast as the response trait.

We used Bayes factors (BF) to compare the models. The posterior probability  $P(\mathcal{M}|\mathcal{D})$  of a model  $\mathcal{M}$  given data  $\mathcal{D}$  is given by Bayes' theorem:  $\Pr(\mathcal{M}|\mathcal{D}) = \Pr(\mathcal{D}|\mathcal{M})\Pr(\mathcal{M})/\Pr(\mathcal{D})$ . We adopted the method in [19] and computed the BF, defined as the ratio of the posterior model probabilities of two different models  $\mathcal{M}_i$  and  $\mathcal{M}_j$ , parameterized by model parameter vectors  $\Theta_i$  and  $\Theta_j$ . This is performed by using function `postpr` in the R package `abc` [37], where posterior model probabilities are estimated using the rejection method.

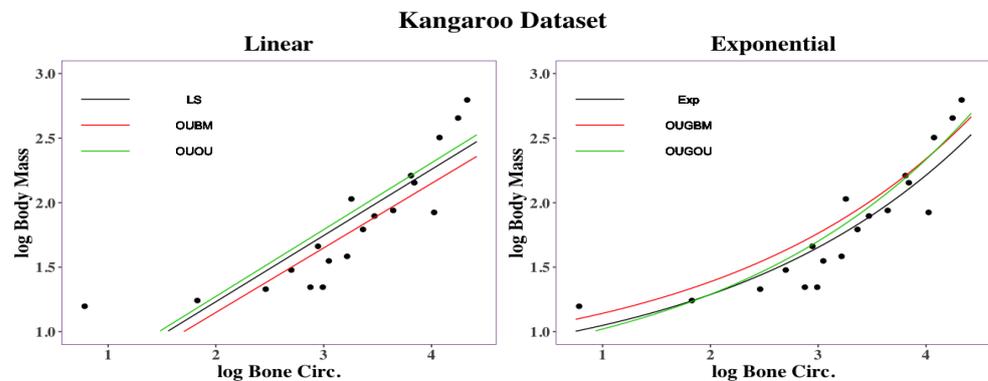
The model comparison under the Bayes factor is shown in Table 5. For the kangaroo data in [18], the best model was the OUGBM model, followed by the OUGOU, OUBM,

and OUOU models. Their pairwise Bayes factors are shown in Table 5. The best model (Rank = 1st) was the OUGBM model. This dataset provides relative equal support for all the exponential OUG\*\* models, a result which was slightly higher than the linear OU\*\* models with the Bayes factor 1.5000 for OUGBM model over OUBM, and 2.1132 for OUGBM over OUOU model. This indicates that the evolution of the optimum  $\theta_i^y$  was also more appropriately described by the geometric BM process predictor than that described by a linear predictor.

**Table 5.** Bayes factor table for kangaroo dataset. Posterior probability  $P(\mathcal{M}|\mathcal{D})$  for each model is shown in the first row; and models shown in second row. Bayes factor  $\mathbf{BF}_{ij}$  for model  $\mathcal{M}_i$  vs. model  $\mathcal{M}_j$  shown in  $i$ th row and  $j$ th column. Acceptance rate was set to 1% ( $\delta = 0.01$ ) for the kangaroo dataset.

	$P(\mathcal{M} \mathcal{D})$	0.3360	0.2810	0.2240	0.1590
Rank	Model $\mathcal{M}$	OUGBM	OUGOU	OUBM	OUOU
1st	OUGBM	1.0000	1.1957	1.5000	2.1132
2nd	OUGOU	0.8363	1.0000	1.2545	1.7673
3rd	OUBM	0.6667	0.7972	1.0000	1.4088
4th	OUOU	0.4732	0.5658	0.7098	1.0000

Regression curves are shown in Figure 6. Overall, the exponential models (EXP, OUGBM, and OUGOU) returned smaller RMSD values than those of the linear models (LS, OUBM, and OUOU) suggesting the utility of the new models. To interpret the impact on the optimal  $\theta_y$  by its covariate  $x$ , we again used the two transformation methods described in Section 2.5 and the posterior mean of parameters in Table 4.



**Figure 6.** Regression curves for kangaroo traits: femoral bone circumference vs. body mass. Linear regression had an RMSD = 0.2748 under the LS method, RMSD = 0.2924 under the OUBM model, and RMSD = 0.2788 under the OUOU model. Exponential regression had an RMSD = 0.2122 under the exponential method, RMSD = 0.2228 under the OUGBM model, and RMSD = 0.1982 under the OUGOU model.

For a 1% decrement of the covariate bone circumference across kangaroos, it was expected that there would be a 0.13% decrement of the body mass  $\theta_y$  under the EXP model, 0.13% decrement under the OUGBM model, 0.14% decrement under the OUGOU model, 0.51% decrement under the LS model, 0.50% decrement under the OUBM model and 0.52% decrement under the OUOU model. For a 5% increment of the covariate bone circumference across the kangaroos, it was expected that there would be a 0.62% increment of the body mass  $\theta_y$  under the EXP model, 0.63% increment under the OUGBM model, 0.69% decrement under the OUGOU model, 2.53% increment under the LS model, 2.47% decrement under the OUBM model, and 2.56% increment under the OUOU model. Overall, the exponential models predicted smaller optimum changes of the optima than the linear

models did for this dataset. A list of optimum changes corresponding to the covariate under those models can be seen in Table S4 in online Supplemental Material.

#### 4. Discussion

Two phylogenetic optimal exponential regression models, OUGBM and OUGOU, for adaptive trait evolution under stabilizing selection were developed. Simulations showed that the new models were validated where posterior means of parameters were close to their true parameter values. The utility of the new regression models in phylogenetic comparative analysis is accessed by analyzing the kangaroo dataset, and results showed that the new models could be appropriately used and are more competitive than the linear models.

Parameter estimation for regression parameters in the ABC procedure depends on several factors. While appropriate priors are required for simulating samples, the choice of the hyperparameters is also important. In this study, uniform distribution with bounds of regression estimates  $\pm 5$  times their standard deviations was used. As results showed the fit of the model, the choice of the parameters for ABC inference provides a reasonable range to cover the true parameters.

The OU process is applied to model stabilizing selection, but is currently criticised for simply being a trait-tracking movement process [48]. Our models assumed that the optimum was tracked by its covariates in a nonlinear functional manner. While our approach provides options for analyzing trait data from the aspect of adaptive trait evolution, it remains to be seen whether models can accurately estimate the adaptive optima from the stabilizing selection, as described in the literature [48]. Undoubtedly, it would be very interesting to investigate this open question for all OU process-based PCMs [8].

Phylogenetic comparative methods are very useful statistical methods to answer evolutionary questions. Those methods, which were developed on the basis of the property of stochastic process remains, require more improvement so that they are able to face the challenges of an intrinsic evolutionary process, which merely a simple Brownian motion model or an OU process model can solve [40,49]. Our models provide feasible options to users in the community to account for nonlinearity in the relationship between the trait optima undergoing stabilizing selection and predictor traits. The models and procedures included in this study were implemented into the R package *ouxy* [50].

**Supplementary Materials:** The following are available at <https://www.mdpi.com/1099-4300/23/2/218/s1>, Figure S1: Box plots of simulated trait values, Figure S2: Cross-validation of models using balanced tree, Table S1: Simulation results of validating models through model parameter estimation using informative priors, Table S2: Simulation results of validating models through regression parameter estimation using informative priors, Table S3: Body mass and bone circumference for Kangaroo species, Table S4: Percentage change of the optimal trait impacted by its covariate.

**Author Contributions:** The authors' individual contributions are provided as the following: Conceptualization, D.-C.J.; methodology, C.-P.W. and D.-C.J.; software, D.-C.J.; validation, D.-C.J.; formal analysis, C.-P.W. and DCJ; investigation, C.-P.W. and D.-C.J.; resources, D.-C.J.; data curation, C.-P.W. and D.-C.J.; writing—original draft preparation, D.-C.J.; writing—review and editing, D.-C.J.; visualization, C.-P.W. and D.-C.J.; supervision, D.-C.J.; project administration, D.-C.J.; funding acquisition, D.-C.J. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by Ministry of Science and Technology, Taiwan (grant no. MOST-109- 2118-M-035-003), 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.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** The data presented in this study are available on request from the corresponding author.

**Acknowledgments:** The authors thank the editors and two anonymous reviewers for their constructive suggestions for improving the early version of this manuscript.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Felsenstein, J. Phylogenies and the comparative method. *Am. Nat.* **1985**, *125*, 1–15. [[CrossRef](#)]
2. Lynch, M. Methods for the analysis of comparative data in evolutionary biology. *Evolution* **1991**, *45*, 1065–1080. [[CrossRef](#)]
3. Harvey, P.H.; Pagel, M.D. *The Comparative Method in Evolutionary Biology*; Oxford University Press: Oxford, UK, 1991; Volume 239.
4. Felsenstein, J. *Inferring Phylogenies*; Sinauer Associates: Sunderland, MA, USA, 2004; Volume 2.
5. O’Meara, B.C. Evolutionary inferences from phylogenies: A review of methods. *Annu. Rev. Ecol. Evol. Syst.* **2012**, *43*, 267–285. [[CrossRef](#)]
6. Hernández, C.E.; Rodríguez-Serrano, E.; Avaria-Llautureo, J.; Inostroza-Michael, O.; Morales-Pallero, B.; Boric-Bargetto, D.; Canales-Aguirre, C.B.; Marquet, P.A.; Meade, A. Using phylogenetic information and the comparative method to evaluate hypotheses in macroecology. *Methods Ecol. Evol.* **2013**, *4*, 401–415. [[CrossRef](#)]
7. Pennell, M.W.; Harmon, L.J. An integrative view of phylogenetic comparative methods: Connections to population genetics, community ecology, and paleobiology. *Ann. N. Y. Acad. Sci.* **2013**, *1289*, 90–105. [[CrossRef](#)]
8. Garamszegi, L.Z. *Modern Phylogenetic Comparative Methods and Their Application in Evolutionary Biology. Concepts and Practice*; Springer: London, UK, 2014.
9. Grafen, A. The phylogenetic regression. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **1989**, *326*, 119–157. [[PubMed](#)]
10. Freckleton, R.P.; Harvey, P.H.; Pagel, M. Phylogenetic analysis and comparative data: A test and review of evidence. *Am. Nat.* **2002**, *160*, 712–726. [[CrossRef](#)]
11. Xiao, X.; White, E.; Hooten, M.; Durham, S. On the use of log-transform vs. nonlinear regression for analyzing biological power laws. *Ecology* **2011**, *92*, 1887–1894. [[CrossRef](#)] [[PubMed](#)]
12. Harmon, L. *Phylogenetic Comparative Methods: Learning from Trees*; CreateSpace Independent Publishing Platform: Charleston, SC, USA, 2018.
13. Ives, A.R.; Garland, T., Jr. Phylogenetic logistic regression for binary dependent variables. *Syst. Biol.* **2009**, *59*, 9–26. [[CrossRef](#)] [[PubMed](#)]
14. Maddison, W.P.; Midford, P.E.; Otto, S.P. Estimating a binary character’s effect on speciation and extinction. *Syst. Biol.* **2007**, *56*, 701–710. [[CrossRef](#)] [[PubMed](#)]
15. Packard, G.C. On the use of log-transformation versus nonlinear regression for analyzing biological power laws. *Biol. J. Linn. Soc.* **2014**, *113*, 1167–1178. [[CrossRef](#)]
16. Klaassen, M.; Nolet, B.A. Stoichiometry of endothermy: Shifting the quest from nitrogen to carbon. *Ecol. Lett.* **2008**, *11*, 785–792. [[CrossRef](#)]
17. Hume, I.D. *Marsupial Nutrition*; Cambridge University Press: Cambridge, UK, 1999.
18. Helgen, K.M.; Wells, R.T.; Kear, B.P.; Gerdtz, W.R.; Flannery, T.F. Ecological and evolutionary significance of sizes of giant extinct kangaroos. *Aust. J. Zool.* **2006**, *54*, 293–303. [[CrossRef](#)]
19. Jhwueng, D.C. Modeling rate of adaptive trait evolution using Cox–Ingersoll–Ross process: An Approximate Bayesian Computation approach. *Comput. Stat. Data Anal.* **2020**, *145*, 106924. [[CrossRef](#)]
20. Hansen, T.F.; Pienaar, J.; Orzack, S.H. A comparative method for studying adaptation to a randomly evolving environment. *Evolution* **2008**, *62*, 1965–1977. [[CrossRef](#)] [[PubMed](#)]
21. Jhwueng, D.C.; Maroulas, V. Phylogenetic ornstein–uhlenbeck regression curves. *Stat. Probab. Lett.* **2014**, *89*, 110–117. [[CrossRef](#)]
22. Jhwueng, D.C.; Maroulas, V. Adaptive trait evolution in random environment. *J. Appl. Stat.* **2016**, *43*, 2310–2324. [[CrossRef](#)]
23. Bartoszek, K.; Pienaar, J.; Mostad, P.; Andersson, S.; Hansen, T.F. A phylogenetic comparative method for studying multivariate adaptation. *J. Theor. Biol.* **2012**, *314*, 204–215. [[CrossRef](#)]
24. Cressler, C.E.; Butler, M.A.; King, A.A. Detecting adaptive evolution in phylogenetic comparative analysis using the Ornstein–Uhlenbeck model. *Syst. Biol.* **2015**, *64*, 953–968. [[CrossRef](#)]
25. Marass, F.; Mouliere, F.; Yuan, K.; Rosenfeld, N.; Markowitz, F. A phylogenetic latent feature model for clonal deconvolution. *Ann. Appl. Stat.* **2016**, *10*, 2377–2404. [[CrossRef](#)]
26. Oksendal, B. *Stochastic Differential Equations: An Introduction with Applications*; Springer Science & Business Media: Berlin, Germany, 2013.
27. Øksendal, B. Stochastic differential equations. In *Stochastic Differential Equations*; Springer: Berlin, Germany, 2003; pp. 65–84.
28. Vega, C.A.M. Calibration of the exponential Ornstein–Uhlenbeck process when spot prices are visible through the maximum log-likelihood method. Example with gold prices. *Adv. Differ. Equations* **2018**, *2018*, 269. [[CrossRef](#)]
29. Lyasoff, A. Another look at the integral of exponential Brownian motion and the pricing of Asian options. *Financ. Stochastics* **2016**, *20*, 1061–1096. [[CrossRef](#)]

30. Dufresne, D. The distribution of a perpetuity, with applications to risk theory and pension funding. *Scand. Actuar. J.* **1990**, *1990*, 39–79. [[CrossRef](#)]
31. Dufresne, D. The integral of geometric Brownian motion. *Adv. Appl. Probab.* **2001**, *33*, 223–241. [[CrossRef](#)]
32. Yor, M. On some exponential functionals of Brownian motion. *Adv. Appl. Probab.* **1992**, *24*, 509–531. [[CrossRef](#)]
33. Burden, R.L.; Faires, J.D. *Numerical Analysis, 9th ed.*; Brooks Cole Publishing: Monterey, CA, USA, 2010.
34. Borchers, H.W. *Pracma: Practical Numerical Math Functions*. 2019. R package version 2.2.9. Available online: <https://CRAN.R-project.org/package=pracma> (accessed on 12 September 2020).
35. Baum, D. Trait evolution on a phylogenetic tree: Relatedness. *Nat. Educ.* **2008**, *1*, 191.
36. Paradis, E.; Schliep, K. ape 5.0: An environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics* **2019**, *35*, 526–528. [[CrossRef](#)] [[PubMed](#)]
37. Csillery, K.; Francois, O.; Blum, M.G.B. abc: An R package for approximate Bayesian computation (ABC). *Methods Ecol. Evol.* **2012**. [[CrossRef](#)]
38. Blomberg, S.P.; Garland, T., Jr.; Ives, A.R. Testing for phylogenetic signal in comparative data: Behavioral traits are more labile. *Evolution* **2003**, *57*, 717–745. [[CrossRef](#)] [[PubMed](#)]
39. Pagel, M. Inferring the theoretical patterns of biological evolution. *Nature* **1999**, *401*, 877–884. [[CrossRef](#)]
40. Adams, D.C.; Felice, R.N. Assessing trait covariation and morphological integration on phylogenies using evolutionary covariance matrices. *PLoS ONE* **2014**, *9*, e94335. [[CrossRef](#)]
41. Bartoszek, K.; Liò, P. Modelling trait dependent speciation with Approximate Bayesian Computation. *arXiv* **2018**, arXiv:1812.03715.
42. Lepers, C.; Billiard, S.; Porte, M.; Méléard, S.; Tran, V.C. Inference with selection, varying population size and evolving population structure: Application of ABC to a forward-backward coalescent process with interactions. *arXiv* **2019**, arXiv:1910.10201.
43. Uyeda, J.C.; Harmon, L.J. A novel Bayesian method for inferring and interpreting the dynamics of adaptive landscapes from phylogenetic comparative data. *Syst. Biol.* **2014**, *63*, 902–918. [[CrossRef](#)]
44. Bastide, P.; Ho, L.S.T.; Baele, G.; Lemey, P.; Suchard, M.A. Efficient Bayesian Inference of General Gaussian Models on Large Phylogenetic Trees. *arXiv* **2020**, arXiv:2003.10336.
45. Stadler, T. *TreeSim: Simulating Phylogenetic Trees*. 2019. R package version 2.4. Available online: <https://CRAN.R-project.org/package=TreeSim> (accessed on 28 August 2020).
46. Bo, L.; Wang, L.; Jiao, L. Feature scaling for kernel fisher discriminant analysis using leave-one-out cross validation. *Neural Comput.* **2006**, *18*, 961–978. [[CrossRef](#)] [[PubMed](#)]
47. Janis, C.M.; Buttrill, K.; Figueirido, B. Locomotion in extinct giant kangaroos: Were sthenurines hop-less monsters? *PLoS ONE* **2014**, *9*, e109888. [[CrossRef](#)] [[PubMed](#)]
48. Cooper, N.; Thomas, G.H.; Venditti, C.; Meade, A.; Freckleton, R.P. A cautionary note on the use of Ornstein Uhlenbeck models in macroevolutionary studies. *Biol. J. Linn. Soc.* **2016**, *118*, 64–77. [[CrossRef](#)] [[PubMed](#)]
49. Cornwell, W.; Nakagawa, S. Phylogenetic comparative methods. *Curr. Biol.* **2017**, *27*, R333–R336. [[CrossRef](#)] [[PubMed](#)]
50. Jhwueng, D.C. Building an adaptive trait simulator package to infer parametric diffusion model along phylogenetic tree. *MethodsX* **2020**, *7*, 100978. [[CrossRef](#)] [[PubMed](#)]