mathematics

MDPI

Article

Stochastic Modeling of Adaptive Trait Evolution in Phylogenetics: A Polynomial Regression and Approximate Bayesian Computation Approach

Dwueng-Chwuan Jhwueng * and Chia-Hua Chang

Department of Statistics, Feng-Chia University, Taichung 40724, Taiwan; m1119502@o365.fcu.edu.tw
* Correspondence: dcjhwueng@fcu.edu.tw

Abstract: In nature, closely related species often exhibit diverse characteristics, challenging simplistic line interpretations of trait evolution. For these species, the evolutionary dynamics of one trait may differ markedly from another, with some traits evolving at a slower pace and others rapidly diversifying. In light of this complexity and concerning the phenomenon of trait relationships that escape line measurement, we introduce a novel general adaptive optimal regression model, grounded on polynomial relationships. This approach seeks to capture intricate patterns in trait evolution by considering them as continuous stochastic variables along a phylogenetic tree. Using polynomial functions, the model offers a holistic and comprehensive description of the traits of the studied species, accounting for both decreasing and increasing trends over evolutionary time. We propose two sets of optimal adaptive evolutionary polynomial regression models of k^{th} order, named the Ornstein–Uhlenbeck Brownian Motion Polynomial (OUBMP$_k$) model and Ornstein–Uhlenbeck Ornstein–Uhlenbeck Polynomial (OUOUP$_k$) model, respectively. Assume that the main trait value y_t is a random variable of the Ornstein–Uhlenbeck (OU) process and that its optimal adaptive value θ_t^y has a polynomial relationship with other traits x_t for statistical modeling, where x_t can be a random variable of Brownian motion (BM) or OU process. As analytical representations for the likelihood of the models are not feasible, we implement an approximate Bayesian computation (ABC) technique to assess the performance through simulation. We also plan to apply models to the empirical study using the two datasets: the longevity vs. fecundity in the Mediterranean nekton group, and the trophic niche breadth vs. body mass in carnivores in a European forest region.

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

MSC: 60G15; 60H35; 92-10

check for updates

Received: 4 December 2024
Revised: 30 December 2024
Accepted: 31 December 2024
Published: 6 January 2025

Citation: Jhwueng, D.-C.; Chang, C.-H. Stochastic Modeling of Adaptive Trait Evolution in Phylogenetics: A Polynomial Regression and Approximate Bayesian Computation Approach. *Mathematics* **2025**, *13*, 170. https://doi.org/10.3390/math13010170

Copyright: © 2025 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

The tapestry of life is woven with rich diversity, with even closely related species showcasing a range of characteristics that often defy evolutionary interpretations [1]. In the realm of biology, statistical polynomial regression has been extensively employed to elucidate complex relationships. Ref. [2] used it to investigate the phenotypic response curves of organisms subjected to varying environmental gradients. Such mathematical models were also used to understand the correlations between morphological traits and performance metrics [3]. Ref. [4] adopted polynomial regression models to show the curvature of reaction norms in plant traits in response to environmental alterations. Ref. [5] explored

polynomial regression, among other methods, as key tools in ecological data analysis. Ref. [6] innovatively applied polynomial regression to analyze the thermal performance curves of ectotherms; and ref. [7] highlighted the nonlinear impacts of temperature on the trophic interactions in the food web. Refs. [8,9] further highlighted the value of polynomial regression in the representation of reaction norms, especially in evolutionary contexts and when species are exposed to novel environments.

Delving into the relationships between evolutionary traits reveals a nuanced landscape. Some traits, for instance, might display a declining trend up to a certain evolutionary threshold, only to increase beyond that point and then decrease after that point. In the left panel of Figure 1, the relationship between longevity and fecundity in Mediterranean nekton species [10] is better captured by a quadratic regression model ($y = 1.73 - 0.07x + 0.01x^2$, RMSD = 1.107, AIC = 81.970) compared to a linear model ($y = 1.93 + 0.03x$, RMSD = 1.192, AIC = 81.996) and a cubic regression ($y = 4.410 - 0.977x + 0.102x^2 - 0.003x^3$, RMSD = 1.151, AIC = 82.538) . The results suggest that incorporating quadratic terms into the regression model provides a better fit for the data, capturing the nonlinear relationship between longevity and fecundity in Mediterranean nekton species. Similarly, in the right panel of Figure 1, the relationship between trophic niche breadth and body mass in carnivores is best modeled with a cubic regression ($y = 13.282 - 4.409x + 0.488x^2 - 0.017x^3$, RMSD = 0.197, AIC = -5.311), significantly outperforming the linear model ($y = 1.555 - 0.121x$, RMSD = 0.408, AIC = 45.774) and the quadratic model ($y = 5.841 - 1.286x + 0.075x^2$, RMSD = 0.262, AIC = 14.120). This suggests that intermediate body sizes may optimize resource use and dietary diversity, reflecting ecological adaptations. These findings highlight the need for advanced polynomial modeling to better understand complex ecological relationships.

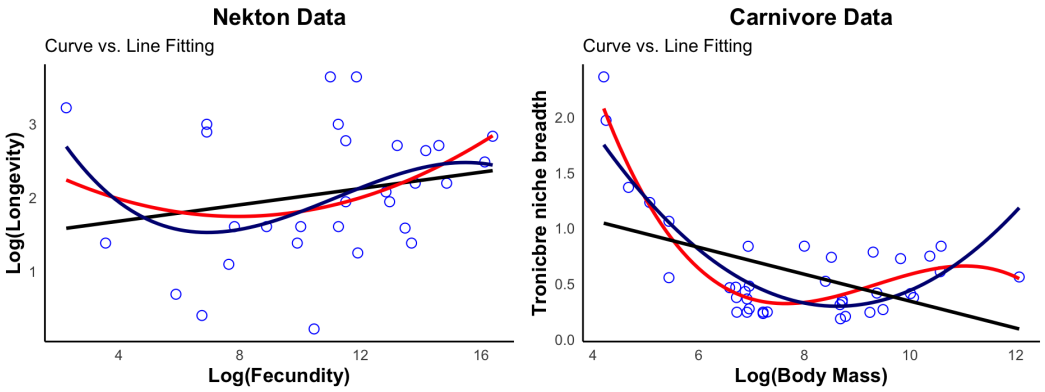


Figure 1. Comparison of regression models: (**Left**) Quadratic regression (red) vs. cubic regression (navy) and linear regression (black) for the relationship between longevity and fecundity in Mediterranean nekton data. (**Right**) Cubic regression (red) vs. quadratic regression (navy) and linear regression (black) for the relationship between trophic niche breadth and body mass in carnivore data.

A more severe situation could occur as shown in Figure 2, where the fit of the model under different regression curves would lead to inconsistent interpretations. Line regression, while powerful in its simplicity, assumes a straight-line relationship between variables, which may fail. For a clearer understanding, using the metaphor of men in a cave feeling an elephant: one feels the tusk and thinks it is a spear, another the tail and thinks it is a rope, and another the leg, thinking it is a tree trunk. Each forms a conclusion based only on the part they touch [11]. This metaphor parallels the longevity vs. fecundity data in Figure 2: (i) in the low-fecundity region (i.e., fecundity has values smaller than 8), where longevity and fecundity have a negative correlation, one might conclude that the more offspring an organism has, the shorter its lifespan. (ii) in the high-fecundity region (i.e.,

fecundity has values greater than 8), where longevity and fertilization have a positive correlation, it implies that having more offspring could be associated with a longer life.

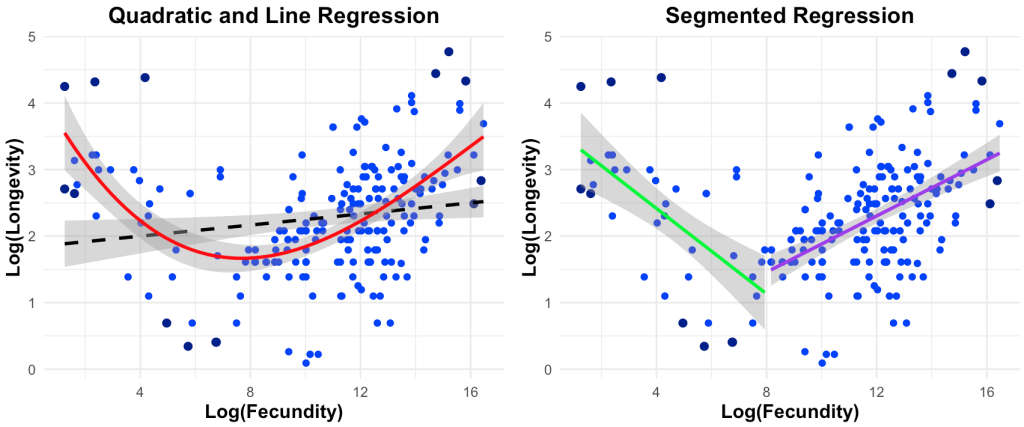


Figure 2. (**Left**): Quadratic (red) vs. linear (black dashed) regression for log-transformed longevity and fecundity. (**Right**): Segmented regression showing distinct trends for low-fecundity (green) and high-fecundity (purple) regions.

In studying the intricate relationships between longevity and fecundity, focusing solely on segmented data (e.g., high- or low-fecundity regions) or studying the complex relationship between trophic niche breadth and body mass, focusing solely on segmented data (e.g., high- or low-fecundity regions or high- or low-body mass) without appropriate analytical tools, can lead to similarly fragmented insights. While linear regression is effective for monotonic trends, it struggles with curvilinear patterns. Polynomial regression offers a more flexible approach, capturing nonmonotone correlations and providing a fuller, more accurate representation of these complex relationships. This highlights the need for comprehensive methods in adaptive evolutionary studies where linear assumptions are insufficient.

On the other hand, in closely allied species, species are assumed with evolutionary relationships, which stack up the complexities and demand innovative analytical tools that go beyond traditional line regression models. Although phylogenetically independent contrasts [12] remove the effects of phylogenetics, they are designed to take the value of the change between a pair of closed related species (for example, longevity vs. change in fertilization in Figure 2, left panel). Even the popular method with the software packages [13] fits a linear model for more general cases, missing the implementation of the polynomial case in the sense of multiple adaptive trait evolution where the optimum is built on other traits.

The empirical data presented in Figures 1 and 2 underscore the potential of polynomial regression models to study the relationship between traits for the evolution of adaptive traits that are phylogenetic, as shown in Figure 3. This observation motivates the development of essential phylogenetic comparative methods tailored to handle this new type of data. We describe the novel implementation of the polynomial regression model of adaptive phylogenetic traits in the following section.

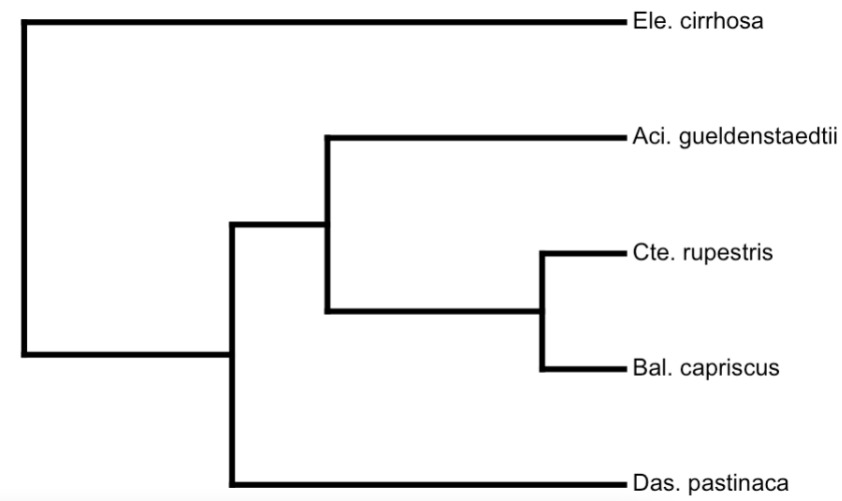


Figure 3. The Phylogenetic hierarchical relationship is established using the `TimeTree` [14], where five nekton species are presented as exemplary examples, as illustrated in Figure 1, left panel.

2. Methods

We first introduce the background of the phylogenetic regression model for adaptive trait evolution. The OUBM model, a prominent model in this field, was developed by [15]. This model assumes that the response trait variable follows an Ornstein–Uhlenbeck (OU) dynamic process. In this setup, the response trait optimum is modeled as having a line relationship with covariates that behave according to Brownian motion.

Subsequent research has built on and expanded the OUBM model. Notable advances include the introduction of the OUOU model, which incorporates covariates of the Ornstein–Uhlenbeck process [16,17]. In addition, efforts have been made to integrate a Cox–Ingersoll–Ross process to model the rate of evolution [18]. In addition, the OUBM model has been extended to analyze multivariate cases, as demonstrated in references [19–21]. There are non-linear exponential models such as [22,23] with application in [24]. We do not intend to include them here, as we focus on the polynomial case. The representation of the trait variable is shown in Section 2.1, where the BM polynomial covariate model with degree k, denoted as the OUBMP_k model (i.e., OUBMP_1 is the OUBM model from the literature [15]), is in Section 2.1.1, and the OU polynomial covariate model with degree k, denoted as the OUOUP_k model (i.e., OUOUP_1 is the OUOU model from the literature [16]), is introduced in Section 2.1.2.

2.1. Representation of the Trait Variable

In the generalized model for the evolution of adaptive phylogenetic traits, the trait variable, denoted as y_t, follows the stochastic differential equation (SDE) presented in Equation (1):

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

where α_t^y signifies the force pulling the trait towards its optimum θ_t^y, σ_t^y is the evolutionary rate of y_t, and W_t^y represents a Wiener process with Gaussian increments. Assuming that α_t^y and σ_t^y are constants (α_y and σ_y, respectively), we obtain Equation (2) by integrating Equation (1) with the integrating factor $\exp(\alpha_y t)$:

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

with $\exp(-\alpha_y t)y_0$ as a deterministic term (initial condition y_0 at $t = 0$) and the stochastic integral $\sigma_y \int_0^t \exp(-(\alpha_y t - \alpha_y s))dW_s^y$, which is Gaussian with a mean of 0 and a variance of $\sigma_y^2(1 - \exp(-2\alpha_y t))/(2\alpha_y)$.

Furthermore, let the integral in the middle of the right hand side of Equation (3) be as follows:

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

Equation (3) is an integral with respect to time, where the stochastic variable θ_s^y in included in the integrand function in Equation (6).

The focal variable, optimum θ_t^y, is assumed to be functionally related to the covariate x_t, as shown in Equation (4):

$$\theta_t^y = f(\beta, x_t). \tag{4}$$

Here, $\beta = (\beta_0, \beta_1, \ldots, \beta_p)^t$ denotes a vector of regression parameters.

In Equation (4), if θ_t^y and x_t have a linear relationship, the dynamics of θ_t^y are determined by the dynamics of a linear combination of continuously distributed identical stochastic random variables $x_{i,t}, i = 1, \ldots, p$. Since the covariates x_t follow Brownian motion, $\theta_t^y = a + bx_t$, which represents a linear relationship, also follows Brownian motion (denoted as the OUBM model [15]). If $x_{i,t}$s are OU processes, θ_t^y is an OU process (denoted the OUOU model [16]).

In this framework, we expand the model to allow for polynomial relationships between the optimum θ_t^y and covariate x_t regression model proposed in Equation (5)

$$\theta_t^y = \sum_{j=0}^{p} \beta_j x_t^j. \tag{5}$$

Plugging Equation (5) into Equation (3), one has the following:

$$A = \exp(-\alpha_y t)\alpha_y \sum_{j=0}^{p} \beta_j \int_0^t \exp(\alpha_y s)x_s^j ds. \tag{6}$$

This is further discussed in the BM polynomial covariate model (OUBMP$_k$) in Section 2.1.1 and OU polynomial covariate model (OUOUP$_k$) in Section 2.1.2.

2.1.1. The OUBMP$_k$ Model

The OUBMP$_k$ model is a polynomial model with x_t in Equation (6) as the BM covariate variable shown in Equation (9)

$$x_t = \sigma_x W_t^x. \tag{7}$$

One can then represent $\theta_t^y = \sum_{j=0}^{p} b_j x_t^j$ in Equation (5) as follows:

$$\theta_t^y = \sum_{j=0}^{p} b_j \sigma_x^j (W_t^x)^j. \tag{8}$$

Then, by incorporating Equation (9) into Equation (3), we have the following:

$$A = \alpha_y \exp(-\alpha_y t) \sum_{j=0}^{p} \beta_i \sigma_x^j \int_0^t \exp(\alpha_y s)(W_s^x)^j ds. \tag{9}$$

y_t in Equation (2) for the OUBMP$_k$ model can be represented as shown in Equation (10):

$$y_t = \exp(-\alpha_y t)y_0 + \alpha_y \exp(-\alpha_y t) \sum_{j=0}^{p} b_j \sigma_x^j \int_0^t \exp(\alpha_y s)(W_s^x)^j ds + \sigma_y \int_0^t \exp(-(\alpha_y t - \alpha_y s))dW_s^y. \tag{10}$$

Furthermore, let

$$I_j(t) = \int_0^t \exp(\alpha_y s)(W_s^x)^j ds \tag{11}$$

be the jth order integral of the stochastic variable $\exp(\alpha_y s)W_s^x$. It can be seen that $I_j(t), j = 1, 2, 3, 4, \ldots$ does not straightforwardly align with a distribution , largely due to the non-linear relationship with W_s; therefore, determining the precise distribution of this integral is a complex task. While using integration by parts, $I_1(t)$ and $I_2(t)$ may align with the normal distribution; however, $I_j(t), j \geq 3$ faces the challenge of specifying the exact distribution. For practical implementation, we compute its expected value $\mathbb{E}\left[I_j(t)\right]$ in Equation (12).

$$\mathbb{E}\left[I_j(t)\right] = \int_0^t \exp(\alpha_y s)\mathbb{E}\left[(W_s^x)^j\right] ds. \tag{12}$$

It is known that for $s \in [0, \infty)$,

$$\mathbb{E}\left[(W_s^x)^j\right] = \begin{cases} s^{j/2}(j-1)!!, & \text{if } j \text{ is even} \\ 0, & \text{if } j \text{ is odd} \end{cases} \tag{13}$$

where $\mathbb{E}[(W_s^x)^2] = s$, $\mathbb{E}[(W_s^x)^4] = 3s^2$, $\mathbb{E}[(W_s^x)^6] = 15s^3$, and $\mathbb{E}[(W_s^x)^8] = 105s^4$, for example.

Then, the expected value of $\mathbb{E}\left[I_j(t)\right]$ can be written as shown in Equation (14).

$$\mathbb{E}\left[I_j(t)\right] = \begin{cases} (j-1)!! \int_0^t \exp(\alpha_y s)s^{j/2} ds, , & \text{if } j \text{ is even} \\ 0, & \text{if } j \text{ is odd} \end{cases} \tag{14}$$

For $j = 2, 4, 6, 8$, $\mathbb{E}\left[I_j(t)\right]$ is represented as a definite integral in Table 1.

Table 1. Expressions for $\mathbb{E}\left[I_j(t)\right]$ for specific $j = 2, 4, 6, 8$.

j	$\mathbb{E}\left[I_j(t)\right]$
2	$\int_0^t \exp(2\alpha_y s)s\, ds$
4	$3\int_0^t \exp(2\alpha_y s)s^2\, ds$
6	$15\int_0^t \exp(2\alpha_y s)s^3\, ds$
8	$105\int_0^t \exp(2\alpha_y s)s^4\, ds$

Each of these integrals in Table 1 is a definite integral that can be directly evaluated using integration techniques that are appropriate for the specific form of the integrated. Therefore, the trait variable y_t in Equation (10) can be drawn accordingly.

2.1.2. The OUOUP$_k$ Model

The biological reasoning for assuming the Ornstein–Uhlenbeck (OU) process for traits in the OUOUP$_k$ model lies in its ability to represent adaptive evolutionary processes, such as stabilizing selection and movement toward an optimal trait value. Unlike Brownian motion (BM), which models random walks without bounds, the OU process introduces a tendency for traits to revert to a specific adaptive optimum over time. This is particularly relevant when studying traits under selective pressures or when there is a biological rationale for constraints on trait evolution.

The OUOUP$_k$ model is an extension of the polynomial model from the OUBMP$_k$ model with the assumption that x_t is the OU covariate variable that solves the SDE $dx_t = \alpha_x(\theta_x - x_t)dt + \sigma_x dW_t^x$, as expressed in Equation (15)

$$x_t = \theta_x + \exp(-\alpha_x t)(x_0 - \theta_x) + \sigma_x \exp(-\alpha_x t)\left(\int_0^t \exp(\alpha_x s) dW_s^x\right). \tag{15}$$

Let $p(t|\theta_x, \alpha_x, x_0) = 1 - \exp(-\alpha_x t)(x_0 - \theta_x)$, $q(t|\sigma_x, \alpha_x) = \sigma_x \exp(-\alpha_x t)$, and $J(t, W_s|\alpha_x) = \int_0^t \exp(\alpha_x s) dW_s^x$. One can represent x_t in Equation (15) as follows:

$$x_t = p(t|\theta_x, \alpha_x, x_0) + q(t|\sigma_x, \alpha_x) J(t, W_s|\alpha_x) := p_t + q_t J_t. \tag{16}$$

Then, the power of x_t in Equation (16) can be expressed using the binomial expansion as follows:

$$x_t^j = (p_t + q_t J_t)^j = \sum_{c=0}^{j} \binom{j}{c} p_t^{j-c} q_t^c J_t^c \tag{17}$$

We can then represent $\theta_t^y = \sum_{j=0}^{p} \beta_j x_t^j$ in Equation (5) as follows:

$$\theta_t^y = \sum_{j=0}^{p} \beta_j \left(\sum_{c=0}^{j} \binom{j}{c} p_t^{j-c} q_t^c J_t^c\right). \tag{18}$$

Plugging Equation (18) in Equation (3), one has the following:

$$A = \alpha_y \exp(-\alpha_y t) \sum_{j=0}^{p} \beta_j \int_0^t \exp(\alpha_y s) \sum_{c=0}^{j} \binom{j}{c} p_s^{j-c} q_s^c J_s^c ds. \tag{19}$$

y_t in Equation (2) for the OUOUP$_k$ model can be represented as shown in Equation (20):

$$\begin{aligned} y_t = {} & \exp(-\alpha_y t) y_0 + \alpha_y \exp(-\alpha_y t) \sum_{j=0}^{p} \beta_j \exp(\alpha_y s) \sum_{c=0}^{j} \binom{j}{c} \int_0^t p_s^{j-c} q_s^c J_s^c ds \\ & + \sigma_y \int_0^t \exp(-(\alpha_y t - \alpha_y s)) dW_s^y. \end{aligned} \tag{20}$$

Define

$$k_c(t) = \int_0^t p_s^{j-c} q_s^c J_s^c ds \tag{21}$$

as for a positive integer c the power term $J_s^c = \left(\int_0^t \exp(\alpha_x s) dW_s^x\right)^c$ in Equation (21) that has no explicit distribution. We instead create a numerical approximation to evaluate the stochastic integral J_s by calculating an integral using a weighted sum. The weighting comes from the exponential function $\exp(\alpha_x s)$, which applies to a sequence of values from 0 to s with 100 points. The differences between successive points in $W_0, W_1, \ldots, W_s$ are computed using $W_i - W_{i-1}, i = 1, 2, \ldots, 101$, and the product of these differences with the exponential term is summed up to yield the integral J_s. We use 100 replicates and evaluate the mean of J_s as d. The $k_c(t)$ is then computed using the definite integral using the composite rule for integral, taking $p_s^{j-c} q_s^c J_s^c$ as the integrand with 100 grid points on each interval (branch lengths) of the tree. The detailed computation can be accessed as an R function `oulscqouModel` in Appendix A.1.

Figure 4 shows the trajectory for a one-dimensional optimal covariate polynomial for the OUBMP$_3$ model and the OUOUP$_3$ model, respectively. In short, Figure 4 highlights the hierarchical interactions between these variables: x_t influences θ_t^y, which in turn drives the dynamics of y_t. By comparing the OUBMP$_3$ and OUOUP$_3$ models, the figure demonstrates how the different stochastic processes (Brownian Motion vs. Ornstein–Uhlenbeck) shape the behavior of x_t and consequently affect the dynamics of θ_t^y and y_t. This visualization underscores the importance of the underlying model structures in shaping the joint dynamics of the system.

2.2. Modeling Trait Evolution Along the Tree

To simulate the trait variable y_t as specified in Equation (2), we use previously developed techniques, particularly for simulating A in Equation (3), which involves integrating the power of the stochastic variable x_t in both Brownian Motion (BM) and Ornstein–Uhlenbeck (OU) processes. The procedure is described in the following.

Given a group of n species, denote $y = (y_1, y_2, \ldots, y_n)$ as the corresponding response trait, $\theta^y = (\theta_1^y, \theta_2^y, \ldots, \theta_n^y)$ as the optimum of y, and $x = (x_1, x_2, \ldots, x_n)$ as the covariate. To simulate trait data y, θ, x for inference using a specified phylogenetic tree, denoted as $\mathcal{T}$ and encompassing a set of branch lengths, $\mathcal{B} = \{b_1, b_2, \ldots, b_L\}$, a tree traversal algorithm [25] is employed to begin from the root to simulate traits at each node $\mathcal{I} = \{i_1, i_2, \ldots, i_L\}$ and at the tips of the tree. The algorithm starts by setting the root node with an initial value of (x_0, θ_0^y, y_0). At each subsequent node, the covariate x_t is simulated, conditioned on the previous node i_s, and θ_a is evaluated under the polynomial relationship.

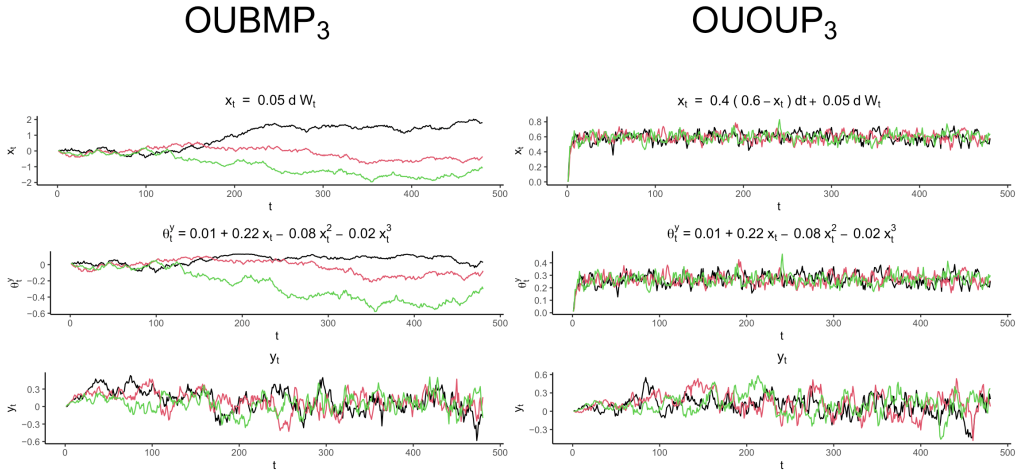


Figure 4. Trajectory simulations for two models—the OUBMP$_3$ model and the OUOUP$_3$ model—are conducted using degree 3 polynomial covariates. The corresponding trajectories of the covariates x_t, the optimum θ_t^y, is computed using these covariates, and the trait y_t is computed using the methods in Section 2.1. Each panel contain 3 independent trajectories.

Finally, y_t at the descendant node is simulated, conditioned on y_a (the ancestral node), using the branch length t_s between the ancestor-descendant nodes.

For the OUBMP$_k$ model, on the node $i_k \in \mathcal{I}$, the trait y_{i_k} is simulated in Equation (22)

$$y_{i_k} = \exp(-\alpha_y t_a) y_a + \alpha_y \exp(-\alpha_y t_a) \sum_{j=0}^{p} b_j \sigma_x^j \int_0^{t_a} \exp(\alpha_y s)(W_s^x)^j ds + \sigma_y \int_0^{t_a} \exp(-(\alpha_y t - \alpha_y s)) dW_s^y. \quad (22)$$

For the OUOUP$_k$ model, on the node $i_k \in \mathcal{I}$, y_{i_k} is simulated in Equation (23):

$$y_{i_k} = \exp(-\alpha_y t_a) y_a + \alpha_y \exp(-\alpha_y t_a) \sum_{j=0}^{p} b_j \int_0^{t_a} \exp(\alpha_y s) \sum_{c=0}^{j} \binom{j}{c} p_s^{j-c} q_s^c I_s^c ds + \sigma_y \int_0^{t_a} \exp(-(\alpha_y t - \alpha_y s)) dW_s^y. \quad (23)$$

For the OUBMP model, the model parameters are $(\alpha_y, \sigma_y, \sigma_x) = (0.5, 1, 1)$. For the OUOUP model, the model parameters are $(\alpha_y, \alpha_x, \theta_x, \sigma_x, \sigma_y) = (0.15, 0.12, 0, 0.1, 0.08$. For both models, the corresponding regression parameters are set to $(\beta_0, \beta_1, \beta_2, \beta_3) = (0.1, 1, -0.2, 0.1, -0.12)$ The roots for both models are set as $\rho_y = 0$ and $\rho_x = 1$, with $\theta_0^y = 1$. The parameter values were chosen based on conventional settings commonly used in similar studies [18,22], where these values reflect biologically meaningful constraints, as α

helps to set bounds (generally no more than 3) [25,26], and σ represents the rate, which is set to a reasonable scale to capture variability without being excessively large.

We illustrate the algorithm using an example in Figure 5 of the three taxa tree case to generate covariates x_i, optima θ_i, and traits y_i at the tip $i = A, B, C$, at the internal node E. The root node $D = x_0$ is set to be a constant. Figure 5 shows the trajectory of the optimal θ_t^y along a three-species phylogenetic tree using the optimal covariate polynomials using the same parameter in Figure 4.

For summarizing the simulation schemes of x, θ, y, we used 100 replicates and reported the boxplots shown in the subsequent Figure 6.

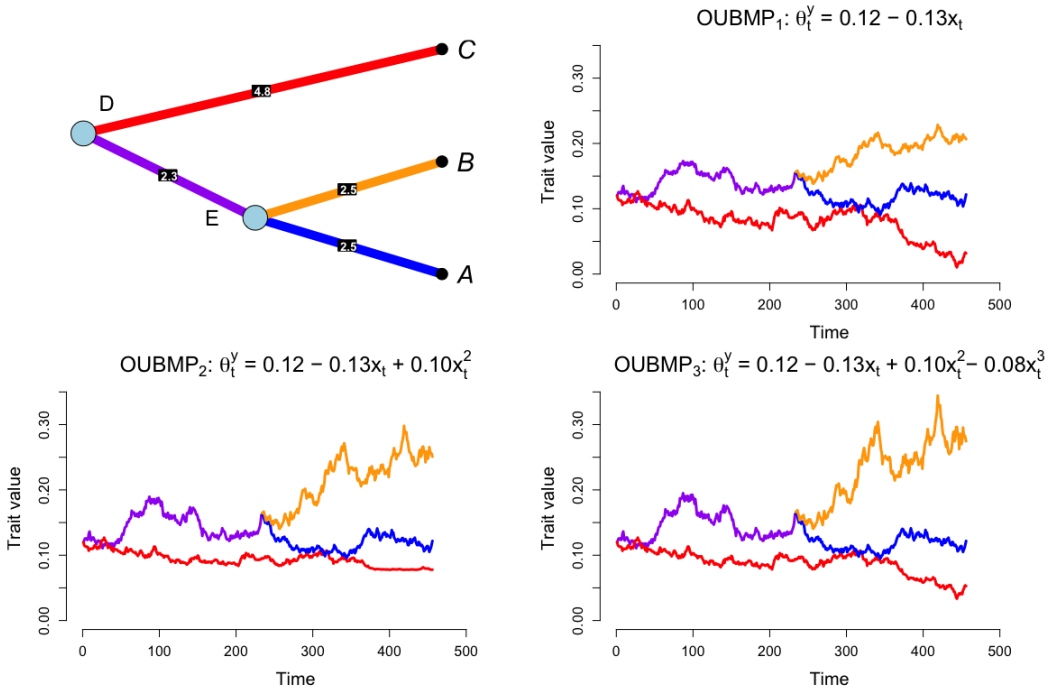


Figure 5. Simulations of trajectories for traits of three species A, B, C with trait variables $(\theta_{A,t}^y, \theta_{B,t}^y, \theta_{C,t}^y)$ along the given tree are depicted in various panels: the BM covariate model up to the linear ($OUBMP_1$) is shown in the upper-right panel, while the quadratic ($OUBMP_2$) and cubic ($OUBMP_3$) polynomial covariate models are displayed in the lower-left and lower-right panels, respectively.

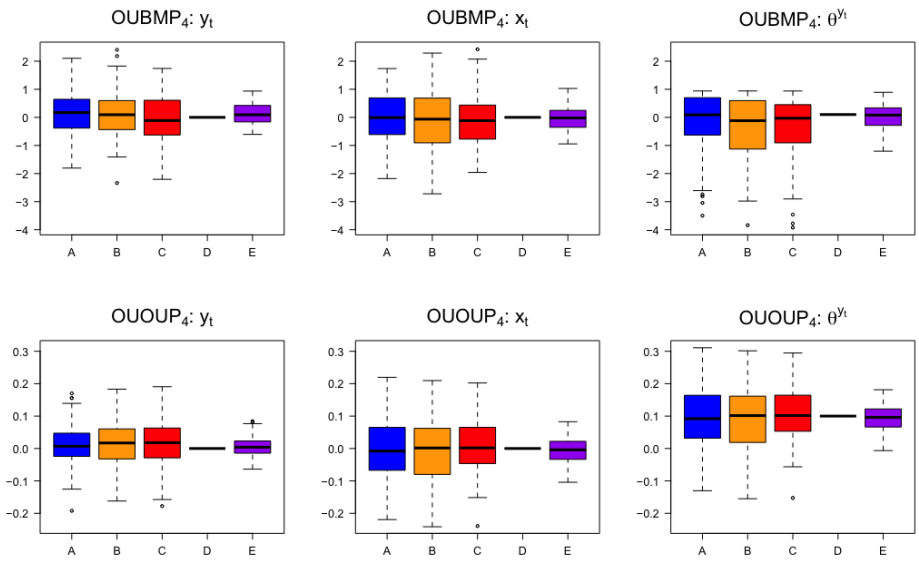


Figure 6. Boxplots representing both models for each node A, B, C, D, E of the tree are presented in Figure 5, illustrating the outcome of the simulations.

2.3. Approximate Bayesian Computation

Due to the polynomial characteristics, between θ_t^y and its covariate x_t, the stochastic variable y_t shown in Equation (2) includes a definite integral of the stochastic variable θ_t^y with respect to time t. To the best of our knowledge, the distribution of the definite integral of the power of the Gaussian process lacks an analytical formulation. For example, the distribution of the cubic term $\int_0^t [W_s^x]^3 ds$ for the OUBMP$_3$ model and $\int_0^t \int_0^s [\exp(\alpha_x u) dW_u^x]^3 ds$ for the OUOUP$_3$ model may not be exactly specified. Hence, the model's lack of closed-form likelihood might not be intractable. Approximate Bayesian Computation (ABC) is used herein for statistical inference. The ABC procedure is a likelihood-free method used for model inference. To start the ABC algorithm, the data are first simulated from the model using the parameters drawn from the prior distributions. Then, a set of summary statistics is calculated for the samples and raw data. For the ABC rejection method, a distance function $d(\cdot)$ and a threshold δ are used to determine posterior samples by comparing the summary statistics of the observed and simulated data [27–29].

We adopt the summary statistics from [18,22], including the mean, median, standard deviation, skewness, and kurtosis, as well as statistics based on phylogenetic trees, including the contrast mean, contrast variance, contrast skewness, and contrast standard deviation [30]. Since the polynomial feature is present, we also add the Lagrange polynomial [31] as a summary statistic, constructed from the selected median points of each quantile, providing an approximation of the data. Note that although it captures certain aspects of the data trend, the overall fit may not be as precise in regions where the data deviate from these median points. Those statistics have generated great interest in the research field of evolutionary biology [18,32,33]. We also employ phylogenetic contrasts for key statistical measures such as mean, variance, skewness, and kurtosis, calculated using the `pic` function of R's `ape` package [34]. In particular, we analyze these statistics through both the trait vector y and its phylogenetic contrast vector $\check{y}$. The summary statistics used in our ABC framework are shown in Table 2.

Table 2. Summary statistics used for the phylogenetic optimal polynomial regression model of trait evolution.

Summary Statistics	Notation
Mean	$\bar{y}, \bar{x}$
Median	y_m, x_m
SD	$\mathrm{sd}_y, \mathrm{sd}_x$
Skewness	$\mathrm{skew}_y, \mathrm{skew}_x$
Kurtosis	$\mathrm{Kurt}_y, \mathrm{Kurt}_x$
Contrast mean	$\bar{\check{y}}, \bar{\check{x}}$
Contrast SD	$\mathrm{sd}_{\check{y}}, \mathrm{sd}_{\check{x}}$
Contrast skewness	$\mathrm{skew}_{\check{y}}, \mathrm{skew}_{\check{x}}$
Contrast kurtosis	$\mathrm{Kurt}_{\check{y}}, \mathrm{Kurt}_{\check{x}}$
Lagrange polynomial	$L(x) = \sum_{i=0}^{n} y_i l_i(x), l_i(x) = \prod_{j=0, j\neq i}^{n} (x - x_j)/(x_i - x_j)$

The procedure for estimating parameters using the ABC rejection method is shown in Algorithm 1.

Algorithm 1 Approximate Bayesian Computation for the polynomial covariate model of adaptive trait evolution.

Require: Data on traits: outcome vector $y = (y_1, y_2, \ldots, y_n)^t$, predictor vector $X = (x_1, x_2, \ldots, x_p)$ where $x_j = (x_{1j}, x_{2j}, \ldots, x_{nj})^t$; phylogenetic tree $\mathcal{T}$ with branch lengths set $\mathcal{B} = \{b_1, b_2, \ldots, b_L\}$, initial parameter estimates Θ_0, prior distribution $\pi(\cdot)$ for Θ and a cut-off value δ.

Ensure: Sampled parameters Θ_i, for $i = 1, 2, \ldots, mN\delta$, alongside the corresponding model indices, where m denotes the total model count, N is the total number of simulations for each model.

1: Compute the initial summary metrics $S_0 = S(y, X)$.
2: **for** $i = 1$ to N **do**
3: Sample parameters $\{\Theta_i\}$ from the prior distribution π.
4: Generate simulated trait and predictor sets y_i, x_i using the model with parameters Θ_i.
5: Determine summary metrics $S_i = S(y_i, X_i)$.
6: Calculate all distances $\{d_i\}_{i=1}^{mN}$ between S_0 and S_i between all samples within models.
7: **end for**
8: **return** Return the sampled parameters $\{\Theta_i : i = 1, 2, \ldots, N\delta\}$ yield the smallest distance in summary statistics relative to the given trait data.

2.4. Interpret the Regression Effect

In the form of the optimal θ_t^y is a polynomial to the covariates x_t:

$$\log \theta_t^y = \beta_0 + \sum_{j=1}^{p} \beta_j \log x_t^j \tag{24}$$

For interpretation in the latter empirical analysis, one has the difference between the two measures $(\theta_2^y, x_2), (\theta_1^y, x_1)$ in the log scale, which can be written as follows:

$$\log \theta_2^y - \log \theta_1^y = \sum_{j=1}^{p} \beta_j \log x_2^j - \sum_{j=1}^{p} \beta_j \log x_1^j, \tag{25}$$

which yields Equation (26):

$$\frac{\theta_2^y}{\theta_1^y} = \prod_{j=1}^{p} \left(\frac{x_2}{x_1}\right)^{j\beta_j} \tag{26}$$

For example, given $x_2 = 1.1x_1$ and $j = 3$, it is the cubic polynomial regression, and with $\beta_1 = -0.8$, $\beta_2 = 1$, and $\beta_3 = 0.5$ one would expect that for a one percent increase in the covariate, the ratio for the optimal status would have a value of $1.01^{-0.8} \times 1.01^{2\times 1} \times 1.01^{3\times 0.5} = 0.9594 = 95.94\%$, which accounts for a 4.06 percent decrease in the optimal.

3. Results

The simulation study is presented in Section 3.1, and the empirical study in shown in Section 3.2.

3.1. Simulation

Below, we provide results that present the posterior distributions for the OUBMP_4 and OUOUP_4 models' parameters and regression coefficients, evaluated using different numbers of taxa (8, 16, 32, and 64) under the balanced tree.

For the OUBMP_4 model, the true model parameter values are set to $\alpha_y = 0.5$, $\sigma_x = 0.5, \sigma_y = 0.5$, where the prior distributions of the model parameters are set to $\alpha_y \sim \mathcal{E}(\lambda = 2)$ (exponential with rate 2), $\sigma_x \sim \text{Inv}\Gamma(\alpha = 2, \beta = 0.5)$ (inverse gamma), and $\sigma_y \sim$ Inverse Gamma$(\alpha = 2, \beta = 0.5)$ (inverse gamma). The true values of the regres-

sion parameters are $(\beta_0, \beta_1, \beta_2, \beta_3, \beta_4) = (0, -2, -0.5, 1, 0.8)$, and their prior distributions are assumed to be uniform: $\beta_0 \sim U(-5, 5)$, $\beta_1 \sim U(-7, 3)$, $\beta_2 \sim U(-3, 2)$, $\beta_3 \sim U(0, 2)$, $\beta_4 \sim U(0.6, 1.0)$. Since the simulation requires setting the parameter values at the root of the tree, the trait values are set to the following: $\rho = (\rho_y, \rho_x, \rho_{\sigma_y}, \rho_{\sigma_x}) = (0, 0, 1, 1)$.

For the OUOUP$_4$ model, the true parameter values are set to $\alpha_y = 0.5, \alpha_x = 0.125$, $\theta_x = 0, \sigma_x = 2.5, \sigma_y = 1$, where the prior distributions of the model parameters are uniform: $\alpha_y \sim U(0, 1)$, $\alpha_x \sim U(0, 0.25)$, $\theta_x \sim N(-2, 2)$, $\sigma_x \sim U(0, 5)$, and $\sigma_y \sim U(0, 2)$. The true values of the regression parameters are $(\beta_0, \beta_1, \beta_2, \beta_3, \beta_4) = (0, 1, 0.1, -0.1, 0.8)$, and their prior distributions are assumed to be uniform: $\beta_0 \sim U(-1, 1)$, $\beta_1 \sim U(0, 2)$, $\beta_2 \sim U(0, 0.2)$, $\beta_3 \sim U(-0.2, 0)$, $\beta_4 \sim U(0.6, 1.0)$. Since the simulation requires setting the parameter values at the root of the tree, the trait values are set to the following: $\rho = (\rho_y, \rho_x, \rho_{\sigma_y}, \rho_{\sigma_x}) = (0, 0, 1, 1)$.

A reasonable range of the parameters is set. The selection of model parameters is based on common practices in the field. One reason for our parameter choices is that we use a log transformation of the data, which is meaningful for biological traits that change on a logarithmic scale. This transformation ensures that the trait values are on a relatively closed scale. Parameters α_y and α_x are typically set below 3 to accurately model the stabilizing selection observed in biological phenomena. Similarly, the rates σ_x and σ_y are chosen based on standard practices. Optimal parameters θ_y and θ_x exhibit more variability, while regression parameters β are flexible and depend on the relationship between the covariate x and the optimal θ. We initialize β using regular regression estimates with some error, allowing for flexibility in their selection. Therefore, the chosen parameters and their priors are the most suitable on the basis of comparisons with the existing literature.

For the model parameters, the results are given in Table 3. We observe that the results indicate that the true values of α_y, σ_x, and σ_y are well captured by the median values and their 95% credible intervals.

Table 3. Simulation results for the model parameters for a balanced tree case.

Model	Taxa	α_y (0.5)	α_x (0.125)	θ_x (0)	σ_x (2.5)	σ_y (1)
OUOUP$_4$	8	0.53 (0.07, 0.98)	0.13 (0.01, 0.24)	−0.04 (−1.86, 1.89)	3.52 (1.92, 4.91)	0.95 (0.05, 1.94)
OUOUP$_4$	16	0.53 (0.08, 0.98)	0.13 (0, 0.25)	−0.11 (−1.86, 1.91)	3.06 (1.66, 4.79)	1.01 (0.1, 1.94)
OUOUP$_4$	32	0.53 (0.06, 0.98)	0.14 (0.01, 0.25)	−0.01 (−1.93, 1.87)	3.02 (1.73, 4.83)	0.95 (0.04, 1.94)
OUOUP$_4$	64	0.56 (0.07, 0.99)	0.12 (0.01, 0.25)	0.05 (−1.89, 1.88)	2.42 (1.14, 4.58)	0.68 (0.02, 1.81)
OUBMP$_4$	8	0.26 (0.01, 1.35)	–	–	0.42 (0.18, 0.92)	0.34 (0.12, 0.9)
OUBMP$_4$	16	0.37 (0.01, 1.92)	–	–	0.39 (0.19, 0.77)	0.33 (0.13, 0.77)
OUBMP$_4$	32	0.42 (0.02, 1.69)	–	–	0.34 (0.16, 0.66)	0.34 (0.15, 0.73)
OUBMP$_4$	64	0.32 (0.01, 1.77)	–	–	0.45 (0.23, 0.87)	0.34 (0.12, 0.84)

For the regression coefficients in Table 4, where β_0 through β_4 generally exhibit broad credible intervals, the true values are consistently covered.

These results highlight that while the model captures the general trends of parameter estimates, increasing the number of taxa improves the precision of parameter estimates, especially for the regression coefficients, reflecting a more robust inference with larger sample sizes. We also performed simulations for three other types of trees: a pure birth tree, birth–death tree, and coalescent tree (see Appendix B). The results were similar to those of the balanced tree, with the posterior parameter provided estimates close to the true values and the credible intervals covering the true values, which suggests that the parameter estimates under our framework are independent of the tree type.

Table 4. Simulation results for the regression parameters under the balanced tree case.

Model	Taxa	β_0 (0)	β_1 (1)	β_2 (0.1)	β_3 (-0.1)	β_4 (0.8)
OUOUP$_4$	8	-0.02 (-0.94, 0.94)	0.99 (0.04, 1.91)	0.1 (0.01, 0.19)	-0.1 (-0.2, -0.01)	0.8 (0.61, 0.99)
OUOUP$_4$	16	0.04 (-0.96, 0.96)	1 (0.06, 1.95)	0.1 (0, 0.2)	-0.1 (-0.19, -0.01)	0.82 (0.61, 0.99)
OUOUP$_4$	32	0 (-0.95, 0.95)	0.93 (0.03, 1.97)	0.1 (0.01, 0.19)	-0.1 (-0.2, 0)	0.81 (0.61, 0.99)
OUOUP$_4$	64	0.04 (-0.9, 0.95)	0.99 (0.08, 1.94)	0.1 (0.01, 0.2)	-0.1 (-0.2, -0.01)	0.82 (0.61, 0.99)
OUBMP$_4$	8	-0.2 (-4.42, 4.74)	-2.12 (-6.74, 2.71)	-0.82 (-2.91, 1.82)	0.98 (0.06, 1.93)	0.8 (0.61, 0.99)
OUBMP$_4$	16	-0.01 (-4.28, 4.64)	-2.07 (-6.73, 2.67)	-0.25 (-2.84, 1.82)	0.99 (0.04, 1.93)	0.81 (0.61, 0.99)
OUBMP$_4$	32	-0.02 (-4.48, 4.38)	-1.93 (-6.75, 2.79)	-0.63 (-2.86, 1.86)	0.94 (0.05, 1.95)	0.81 (0.61, 0.98)
OUBMP$_4$	64	0.05 (-4.44, 4.6)	-1.95 (-6.64, 2.68)	-0.31 (-2.88, 1.87)	0.91 (0.06, 1.94)	0.8 (0.61, 0.99)

3.2. Empirical Analysis

We apply our models to study the longevity (years) vs. fecundity in Mediterranean nekton in Section 3.2.1 and the trophic niche breadth vs. body mass in carnivores in Section 3.2.2.

3.2.1. Longevity (Years) vs. Fecundity in Mediterranean Nekton

We apply our model to analyze the relationship between longevity (in years) and mean fecundity (number of eggs) in a group of 205 Mediterranean nekton species [10]. Currently, the names of these 205 species, listed in [10], have been entered into the `TimeTree` database [14]. From this, we obtained a resolved phylogenetic tree for 184 species. After matching the trait data with the labeling of the tree tips, we were able to include 173 species in our analysis. The phylogenetic tree of the nekton is shown in Figure 7. We use 30 species case to present our analysis, where the 30 species were uniformly randomly selected from the 173 species, and their trait relationship remains present in the curve-like relationship. The phylgoenetic relationship is shown in Figure 7.

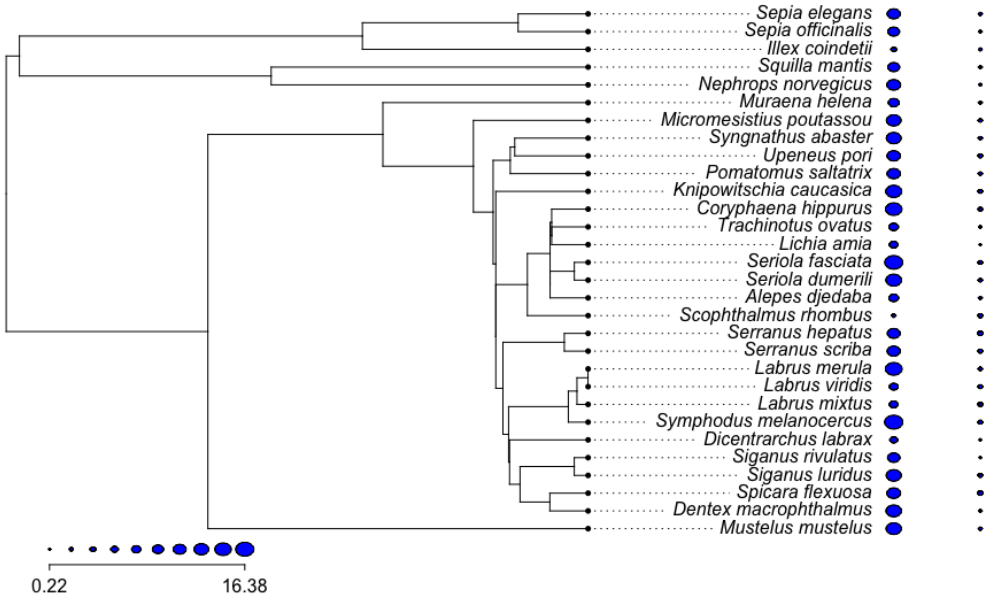


Figure 7. Phylogenetic tree and traits (blue dots) of 30 nekton species.

Table 5 shows the posterior parameter estimates.

Table 5. Posterior parameter estimate: 30 species cases.

Model	α_x	α_y	σ_x	σ_y	θ_x	β_0	β_1	β_2	β_3
LS						1.704	0.056		
OUBMP$_1$		2.087	21.832	1.035		1.728	0.057		
OUOUP$_1$	1.560	1.567	18.998	1.332	0.352	1.713	0.056		
SQ						3.921	−0.508	0.030	
OUBMP$_2$		1.522	19.990	1.326		3.924	−0.502	0.030	
OUOUP$_2$	1.462	1.155	17.783	1.356	0.877	3.916	−0.504	0.030	
CUBIC						4.761	−0.919	0.082	−0.002
OUBMP$_3$		1.344	18.642	1.466		4.721	−0.896	0.081	−0.002
OUOUP$_3$	1.582	1.441	19.174	1.635	0.712	4.745	−0.930	0.081	−0.002

The regression curve under the three categories is shown in Figure 8. By applying the interpretation in Section 2.4 and using the regression estimates from Table 5, the corresponding percent increases and decreases for the optimal θ^y associated with a 1% increase in the covariate within the nekton data set are shown in Table 6 where the linear models show a change of around 1.82–1.84%, the quadratic models around 4.29–4.31%, and the cubic models have the largest increase at 4.82–4.87%, indicating greater sensitivity in higher-order models.

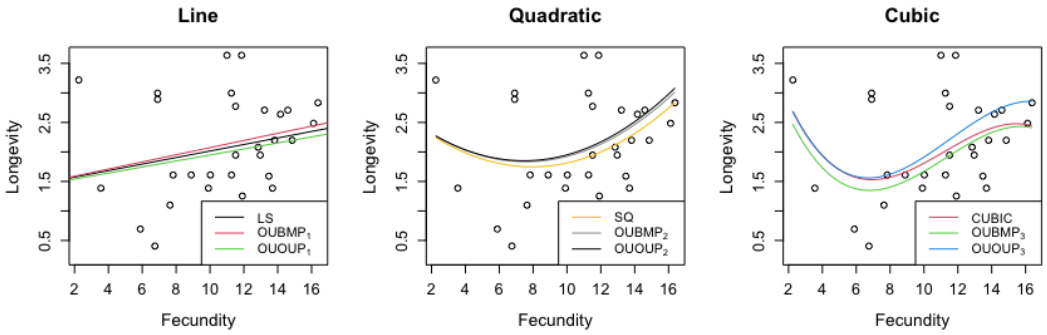


Figure 8. Regression curves for the 30 species cases. (**Left**): line regression including (**middle**) quadratic, (**right**) cubic.

Table 6. Change in the optimal θ^y value for each model when encountering a 1% covariate increase.

Linear Models	Quadratic Models	Cubic Models
LS (1.82%)	**SQ** (4.31%)	**CUBIC** (4.87%)
OUBMP$_1$ (1.84%)	**OUBMP**$_2$ (4.31%)	**OUBMP**$_3$ (4.82%)
OUOUP$_1$ (1.83%)	**OUOUP**$_2$(4.29%)	**OUOUP**$_3$ (4.85%)

We compare the estimation error with the root mean square deviation (RMSD), calculated as RMSD $= \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}$, where y_i is the observed value and$\hat{y}_i$ is the predicted value regression models. The results are show in Table 7.

Table 7. RMSD for 30 nekton species cases across 3 different types of models.

Line (RMSD)	Quadratic (RMSD)	Cubic (RMSD)
LS (0.8587)	**SQ** (0.8302)	**CUBIC** (0.8106)
OUBMP$_1$ (0.8612)	**OUBMP**$_2$ (0.8394)	**OUBMP**$_3$ (0.8214)
OUOUP$_1$ (0.8617)	**OUOUP**$_2$ (0.8456)	**OUOUP**$_3$ (0.8338)

The root mean square deviation (RMSD) values in Table 7 are given for different models applied to 30 nekton species. Three model types—linear, quadratic, and cubic—were compared. Overall, the quadratic and the cubic models achieved better fits than the line model. In particular, the cubic models across all categories had the lowest values.

3.2.2. Trophic Niche Breadth vs. Body Mass in Carnivores

We studied the breadth of trophic niches versus body mass in carnivores [1,35]. The original dataset was sourced from [1] and includes data on the food niche breadth and body mass of terrestrial carnivores across different forest regions in Europe. Various types of foods (insectivores and rodents, rabbits and muskrats, ungulates, livestock, birds, bird eggs, reptiles, amphibians, fish, earthworms, and fruits, seeds, and other plant materials) are given with their diet composition values. We used the Shannon–Wiener diversity index as the covariate trait x to evaluate the trophic niche breadth $H = -\sum_{i=1}^{n} p_i \ln(p_i)$, where p_i represents the proportion of each type of food in the diet. The response trait is the body mass of the carnivores. Table 8 shows the parameter estimate.

The phylgoenetic relationships and traits are shown in Figure 9.

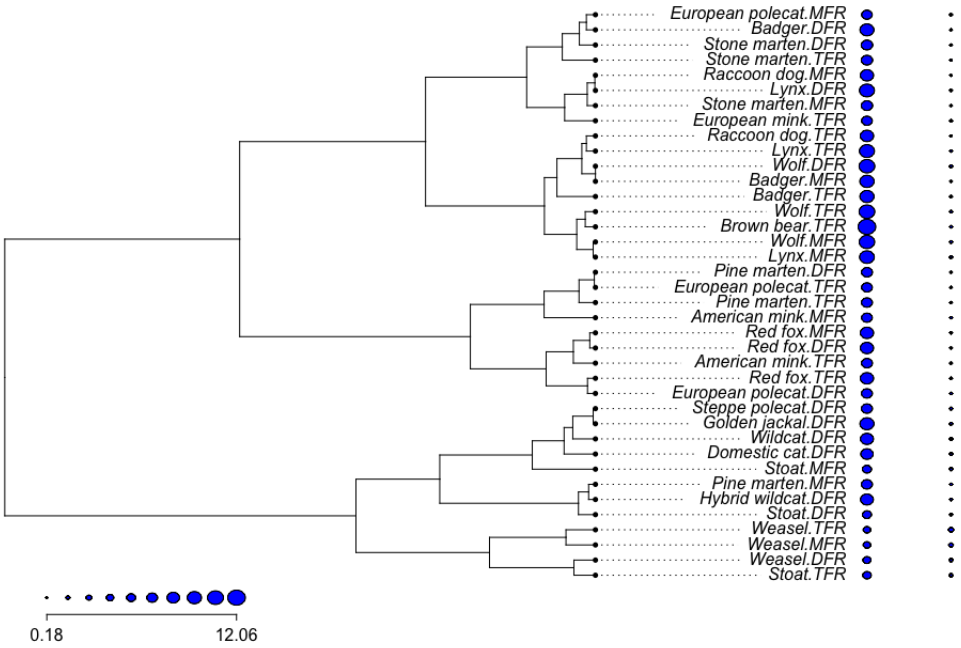


Figure 9. The phylogenetic relationships and traits (blue dots) of 38 carnivore species.

Table 8. Posterior parameter estimate.

Model	α_x	α_y	σ_x	σ_y	θ_x	β_0	β_1	β_2	β_3
LS						1.556	−0.121		
OUBMP$_1$		1.726	19.288	9.034		1.553	−0.122		
OUOUP$_1$	1.642	1.671	18.919	10.028	0.159	1.567	−0.120		
SQ						5.842	−1.286	0.075	
OUBMP$_2$		1.478	16.698	9.944		5.839	−1.281	0.074	
OUOUP$_2$	1.870	1.156	17.282	9.814	0.001	5.846	−1.287	0.075	
CUBIC						13.283	−4.410	0.488	−0.017
OUBMP$_3$		1.650	20.585	15.596		13.115	−4.416	0.489	−0.017
OUOUP$_3$	1.395	1.236	15.222	10.341	−0.079	13.310	−4.404	0.487	−0.017

The regression curves are shown in Figure 10.

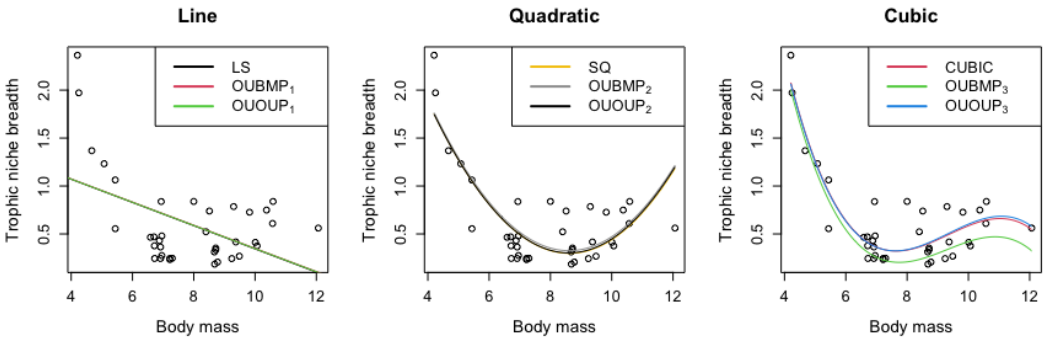


Figure 10. Regression curve for the carnivore trophic breadth nich vs. body mass.

By using the estimate β coefficients in Table 8, Table 9 summarizes the percentage increase or decrease in θ^y for each model based on a 1% increase in x.

Table 9. Percentage change in θ^y for a 1% increase in x.

Linear Models	Quadratic Models	Cubic Models
LS (−0.24%)	**SQ** (−2.32%)	**CUBIC** (−7.20%)
OUBMP$_1$ (−0.24%)	**OUBMP**$_2$ (−2.31%)	**OUBMP**$_3$ (−7.21%)
OUOUP$_1$ (−0.24%)	**OUOUP**$_2$ (−2.33%)	**OUOUP**$_3$ (−7.20%)

Under these models, it is shown that the relationship between range of food resources and the optimal value of body weight is an inverse relationship. This demonstrates that for the 38 carnivore species, smaller body sizes reflect lighter-weight species. These species consume a wider range of food, but larger and heavier species will eat a narrower range of food. This evidence supports the findings of ref. [36], demonstrating that smaller carnivores often have broader diets, as they can exploit a wider range of prey sizes, such as insects, rodents, and small birds. Species like foxes and weasels are opportunistic feeders with diverse food options. In contrast, larger carnivores like whales, lions, and polar bears tend to have narrower diets, specializing in large prey to meet their higher caloric needs, with limited dietary flexibility due to size and habitat constraints.

The values of the root mean square deviation (RMSD) are given in Table 10 for different models applied to the 38 carnivore species.

Table 10. RMSD for 38 carnivores species cases across 3 different types of models.

Line (RMSD)	Quadratic (RMSD)	Cubic (RMSD)
LS (0.4084)	**SQ** (0.2623)	**CUBIC** (0.1978)
OUBMP$_1$ (0.4084)	**OUBMP**$_2$ (0.2624)	**OUBMP**$_3$ (0.2432)
OUOUP$_1$ (0.4088)	**OUOUP**$_2$ (0.2623)	**OUOUP**$_3$ (0.1987)

The table compares three model types: linear, quadratic, and cubic. The quadratic and cubic models show a better fit than the linear model based on the RMSD values. Among them, the cubic models consistently provide the best fit across all categories, with the lowest RMSD values, indicating that they capture the data more accurately than the others.

4. Discussion

The proposed novel phylogenetic regression methods provide an alternative for analyzing comparative data with curve-like relationships and serve as a valuable option for researchers seeking to effectively analyze their data. Specifically, our general adaptive

optimal polynomial regression models, OUBMP_k and OUOUP_k, offer a novel approach to capturing curvilinear evolutionary relationships among traits. By employing polynomial functions within an Ornstein-Uhlenbeck framework, these models allow for a more nuanced understanding of trait evolution, accounting for both accelerating and decelerating trends over time. Our selection of model parameters follows common practices in the field. We apply a log transformation to the data, appropriate for biological traits that vary on a logarithmic scale, ensuring trait values remain on a relatively closed scale. We use of approximate Bayesian computation provides a feasible method for evaluating the models' performance when analytical solutions are not available.

In our current framework, we encounter challenges in evaluating the distribution of y_t, as the stochastic integral $I_j(t) = \int_0^t \exp(\alpha_y s)(W_s^x)^j ds$ has no unknown distribution for $j \geq 3$. Instead, we use the expected value due to the complexity of this process. There might be several approaches to approximate its distribution, such as Monte Carlo simulation, which simulates multiple paths of W_s^j and computes $I_j(t)$ for each, though this can be computationally intensive. Moment matching fits $I_j(t)$'s distribution based on theoretical moments but may struggle with asymmetry or heavy tails [37]. Edgeworth expansion approximates the distribution using series expansions and assumes near-normality, requiring higher moments [38]. Lastly, kernel density estimation is a nonparametric method that estimates the distribution from simulated samples, though it demands a large sample size for accuracy.

There is a concept of nonlinear phylogenetic regression using regularized kernels [39]. Following this theoretical approach, a research study [23] suggested a fractional polynomial method which involves selecting a transformation for a predictor variable x from possible functions (e.g., $x^{-2}, x^{-1}, x^{-0.5}, x^{0.5}$). These transformations serve as independent variables in a linear model and reflect polynomial powers $-2, -1, -0.5, 0.5$ in the Box-Tidwell transformation [40]. It might be worth to expand our models to this approach that effectively capture the nonlinear characteristics of polynomial power in adaptive trait evolution.

Author Contributions: Conceptualization, D.-C.J.; Methodology, D.-C.J. and C.-H.C.; Validation, C.-H.C.; Formal analysis, D.-C.J. and C.-H.C.; Investigation, D.-C.J.; Data curation, C.-H.C.; Writing—original draft, D.-C.J.; Visualization, D.-C.J.; Funding acquisition, D.-C.J. All authors have read and agreed to the published version of the manuscript.

Funding: This research and APC were funded by the Ministry of Science and Technology, Taiwan (grant No. MOST-112- 2118-M-035-003).

Data Availability Statement: The original contributions presented in the study are included in the article; Further inquiries can be directed to the corresponding author(s).

Acknowledgments: We sincerely thank the editors and the three anonymous reviewers for their valuable and constructive feedback, which has greatly improved this manuscript. Thanks to Bao-Yuan Huang for his technical support on an earlier version of this manuscript.

Conflicts of Interest: The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Appendix A. Code and Scripts

The scripts and relevant files developed for this project can be accessed at: http://tonyjhwueng.info/phypolyregopt (accessed on 30 December 2024).

Appendix A.1. The Models

1. The trait simulator under the phylogenetic optimal BM polynomial covariate model: http://tonyjhwueng.info/phypolyregopt/oubmnlregpolybm.html (accessed on 30 December 2024).
2. The trait simulator under the phylogenetic optimal OU polynomial covariate model: http://tonyjhwueng.info/phypolyregopt/oubmnlregpolyou.html (accessed on 30 December 2024).
3. Main script for running the empirical analysis: http://tonyjhwueng.info/phypolyregopt/polyempiricalanalysis.html (accessed on 30 December 2024).

Appendix A.2. Figures and Tables

- Figure 1: http://tonyjhwueng.info/phypolyregopt/nektonAcarnidata2.html (accessed on 30 December 2024).
- Figure 2: http://tonyjhwueng.info/phypolyregopt/tzanatos2020mediterraneanplot.html (accessed on 30 December 2024).
- Figure 3: http://tonyjhwueng.info/phypolyregopt/netkon5spe.html (accessed on 30 December 2024).
- Figure 4: http://tonyjhwueng.info/phypolyregopt/polyreg1Dplotv3.html (accessed on 30 December 2024).
- Figure 5: http://tonyjhwueng.info/phypolyregopt/outreepolybmouv2.html (accessed on 30 December 2024).
- Figure 6: http://tonyjhwueng.info/phypolyregopt/sumboxoubmp4ououp4v2.html (accessed on 30 December 2024).
- Figure 7: https://tonyjhwueng.info/phypolyregopt/30Nektontree.html (accessed on 30 December 2024).
- Figure 8: https://tonyjhwueng.info/phypolyregopt/regresultsplotfull_tzanatos2020mediterranean30spev2.html (accessed on 30 December 2024).
- Figure 9: https://tonyjhwueng.info/phypolyregopt/38carnivoretree.html (accessed on 30 December 2024).
- Figure 10: https://tonyjhwueng.info/phypolyregopt/regresultsplotfulllanszki2019nonv2.html (accessed on 30 December 2024).
- Tables 3 and 4: https://tonyjhwueng.info/phypolyregopt/summarynlregsimp4.html (accessed on 30 December 2024).
- Tables 5 and 6: http://tonyjhwueng.info/phypolyregopt/paramtable_tzanatos2020mediterranean.html (accessed on 30 December 2024).
- Table 7: https://tonyjhwueng.info/phypolyregopt/regresultsplotfulltzanatos2020mediterranean30spe.html (accessed on 30 December 2024).
- Tables 8 and 9: https://tonyjhwueng.info/phypolyregopt/paramtablelanszki2019non.html (accessed on 30 December 2024).
- Table 10: https://tonyjhwueng.info/phypolyregopt/regresultsplotfulllanszki2019nonTest5000.R (accessed on 30 December 2024).

Appendix B. Simulation

The simulation results for parameter estimation under different tree cases are reported below.

Table A1. Simulation results for the model parameters for the pure birth tree case.

Model	Taxa	α_y (0.5)	α_x (0.125)	θ_x (0)	σ_x (2.5)	σ_y (1)
OUOUP$_4$	8	0.57 (0.08, 0.97)	0.12 (0, 0.24)	0.14 (−1.89, 1.84)	3.83 (2.02, 4.94)	0.87 (0.05, 1.92)
OUOUP$_4$	16	0.52 (0.06, 0.96)	0.13 (0.01, 0.24)	0.08 (−1.9, 1.91)	3.06 (1.52, 4.82)	0.93 (0.04, 1.93)
OUOUP$_4$	32	0.53 (0.04, 0.97)	0.13 (0.01, 0.24)	−0.06 (−1.86, 1.84)	1.94 (0.72, 4.24)	0.81 (0.05, 1.92)
OUOUP$_4$	64	0.50 (0.03, 0.97)	0.13 (0.01, 0.24)	−0.04 (−1.89, 1.91)	1.88 (0.71, 3.91)	0.83 (0.04, 1.95)
OUBMP$_4$	8	0.58 (0.02, 2.32)	–	–	0.29 (0.10, 1.58)	0.69 (0.14, 6.33)
OUBMP$_4$	16	0.55 (0.03, 1.97)	–	–	0.25 (0.11, 0.70)	0.45 (0.14, 2.08)
OUBMP$_4$	32	0.37 (0.02, 1.49)	–	–	0.25 (0.09, 0.55)	0.40 (0.12, 1.70)
OUBMP$_4$	64	0.61 (0.07, 1.30)	–	–	0.34 (0.15, 0.74)	0.24 (0.09, 1.00)

Table A2. Simulation results for the regression parameters for the pure birth tree case.

Model	Taxa	β_0 (0)	β_1 (1)	β_2 (0.1)	β_3 (−0.1)	β_4 (0.8)
OUOUP$_4$	8	−0.06 (−0.94, 0.93)	0.97 (0.04, 1.93)	0.10 (0.00, 0.20)	−0.10 (−0.19, −0.01)	0.79 (0.61, 0.99)
OUOUP$_4$	16	0.01 (−0.92, 0.94)	1.00 (0.06, 1.93)	0.09 (0.01, 0.19)	−0.11 (−0.20, −0.01)	0.80 (0.61, 0.99)
OUOUP$_4$	32	−0.03 (−0.95, 0.96)	0.96 (0.05, 1.95)	0.11 (0.01, 0.19)	−0.11 (−0.20, −0.01)	0.79 (0.61, 0.99)
OUOUP$_4$	64	−0.10 (−0.94, 0.94)	1.09 (0.04, 1.93)	0.10 (0.01, 0.20)	−0.10 (−0.19, −0.01)	0.80 (0.61, 0.99)
OUBMP$_4$	8	0.00 (−2.93, 2.78)	−2.41 (−6.84, 2.77)	−0.51 (−2.92, 1.90)	0.97 (0.08, 1.95)	0.81 (0.61, 0.99)
OUBMP$_4$	16	−0.10 (−2.55, 1.54)	−2.12 (−6.84, 2.72)	−0.38 (−2.88, 1.89)	0.93 (0.04, 1.95)	0.82 (0.61, 0.99)
OUBMP$_4$	32	0.04 (−0.98, 1.35)	−2.04 (−6.68, 2.69)	−0.62 (−2.85, 1.87)	0.94 (0.03, 1.89)	0.79 (0.61, 0.99)
OUBMP$_4$	64	−0.55 (−3.37, 2.38)	−1.76 (−6.68, 2.72)	−0.43 (−2.85, 1.85)	0.98 (0.04, 1.95)	0.79 (0.61, 0.99)

Table A3. Simulation results for the model parameters for the birth–death tree case.

Model	Taxa	α_y (0.5)	α_x (0.125)	θ_x (0)	σ_x (2.5)	σ_y (1)
OUOUP$_4$	8	0.55 (0.08, 0.97)	0.15 (0.03, 0.24)	−0.13 (−1.75, 1.81)	1.21 (0.35, 3.22)	0.86 (0.03, 1.91)
OUOUP$_4$	16	0.52 (0.14, 0.98)	0.12 (0.03, 0.24)	−0.16 (−1.9, 1.92)	3.74 (1.82, 4.94)	1.08 (0.06, 1.95)
OUOUP$_4$	32	0.62 (0.16, 0.98)	0.16 (0.05, 0.24)	0 (−1.9, 1.86)	1.34 (0.25, 4.33)	0.91 (0.04, 1.94)
OUOUP$_4$	64	0.38 (0.06, 0.98)	0.1 (0.02, 0.24)	0.01 (−1.88, 1.88)	2.75 (0.97, 4.81)	1.1 (0.06, 1.94)
OUBMP$_4$	8	0.59 (0.07, 2.22)	–	–	0.27 (0.1, 0.75)	0.43 (0.12, 1.56)
OUBMP$_4$	16	0.44 (0.05, 1.52)	–	–	0.3 (0.1, 0.74)	0.27 (0.09, 1.06)
OUBMP$_4$	32	0.89 (0.31, 2.02)	–	–	0.3 (0.12, 0.65)	0.24 (0.09, 0.88)
OUBMP$_4$	64	0.52 (0.15, 1.84)	–	–	0.28 (0.11, 0.62)	0.38 (0.1, 1.43)

Table A4. Simulation results for the regression parameters for the birth–death tree case.

Model	Taxa	β_0 (0)	β_1 (1)	β_2 (0.1)	β_3 (−0.1)	β_4 (0.8)
OUOUP$_4$	8	−0.04 (−0.97, 0.94)	0.99 (0.06, 1.95)	0.1 (0.01, 0.19)	−0.1 (−0.2, 0)	0.79 (0.61, 0.99)
OUOUP$_4$	16	0.06 (−0.95, 0.96)	0.98 (0.06, 1.96)	0.1 (0, 0.2)	−0.1 (−0.19, 0)	0.81 (0.61, 0.99)
OUOUP$_4$	32	0.04 (−0.93, 0.95)	0.99 (0.03, 1.95)	0.1 (0, 0.2)	−0.09 (−0.2, 0)	0.81 (0.61, 0.99)
OUOUP$_4$	64	0.01 (−0.95, 0.95)	1.02 (0.06, 1.93)	0.1 (0, 0.2)	−0.1 (−0.19, 0)	0.79 (0.61, 0.99)
OUBMP$_4$	8	0.06 (−2.07, 2.33)	−2.07 (−6.75, 2.78)	−0.51 (−2.87, 1.87)	1.04 (0.06, 1.94)	0.8 (0.61, 0.99)
OUBMP$_4$	16	0.26 (−4.45, 4.59)	−1.61 (−6.88, 2.77)	−0.48 (−2.77, 1.92)	1.03 (0.07, 1.96)	0.81 (0.61, 0.99)
OUBMP$_4$	32	−2.4 (−4.65, 0.61)	−2.19 (−6.75, 2.73)	−0.55 (−2.84, 1.91)	0.91 (0.03, 1.93)	0.81 (0.61, 0.99)
OUBMP$_4$	64	0.01 (−1.2, 0.95)	−1.86 (−6.66, 2.65)	−0.6 (−2.9, 1.84)	1.06 (0.09, 1.95)	0.8 (0.61, 0.99)

Table A5. Simulation results for the model parameters for the coalescent tree case.

Model	Taxa	α_y (0.5)	α_x (0.125)	θ_x (0)	σ_x (2.5)	σ_y (1)
$\mathbf{OUOUP_4}$	8	0.42 (0.02, 0.99)	0.12 (0, 0.24)	0 (−1.84, 1.88)	1.52 (0.12, 4.26)	1.09 (0.09, 1.95)
$\mathbf{OUOUP_4}$	16	0.55 (0.07, 0.98)	0.12 (0, 0.25)	0.07 (−1.87, 1.87)	3.32 (1.86, 4.85)	0.93 (0.04, 1.93)
$\mathbf{OUOUP_4}$	32	0.52 (0.04, 0.98)	0.12 (0.01, 0.24)	0.05 (−1.9, 1.9)	2.47 (0.44, 4.55)	0.82 (0.03, 1.94)
$\mathbf{OUOUP_4}$	64	0.48 (0.03, 0.95)	0.13 (0.01, 0.24)	0.02 (−1.89, 1.92)	1.95 (0.22, 4.33)	0.88 (0.04, 1.93)
$\mathbf{OUBMP_4}$	8	0.41 (0.02, 1.71)	–	–	0.28 (0.1, 0.92)	0.34 (0.1, 1.47)
$\mathbf{OUBMP_4}$	16	0.58 (0.02, 2.43)	–	–	0.25 (0.08, 0.66)	0.46 (0.12, 2.47)
$\mathbf{OUBMP_4}$	32	0.27 (0.01, 1.26)	–	–	0.28 (0.09, 1.01)	0.57 (0.12, 3.64)
$\mathbf{OUBMP_4}$	64	0.2 (0.01, 0.93)	–	–	0.26 (0.09, 1.44)	0.64 (0.13, 5.33)

Table A6. Simulation results for the regression parameters for the coalescent tree case.

Model	Taxa	β_0 (0)	β_1 (1)	β_2 (0.1)	β_3 (−0.1)	β_4 (0.8)
$\mathbf{OUOUP_4}$	8	0.01 (−0.87, 0.91)	1.07 (0.13, 1.92)	0.11 (0.01, 0.19)	−0.1 (−0.19, −0.01)	0.8 (0.62, 0.98)
$\mathbf{OUOUP_4}$	16	−0.01 (−0.89, 0.89)	1.02 (0.12, 1.9)	0.1 (0.01, 0.19)	−0.1 (−0.19, −0.01)	0.82 (0.62, 0.98)
$\mathbf{OUOUP_4}$	32	−0.01 (−0.88, 0.88)	1.03 (0.1, 1.89)	0.1 (0.01, 0.19)	−0.1 (−0.19, −0.01)	0.8 (0.62, 0.97)
$\mathbf{OUOUP_4}$	64	0 (−0.91, 0.91)	1 (0.14, 1.9)	0.1 (0.01, 0.19)	−0.1 (−0.19, −0.01)	0.81 (0.62, 0.98)
$\mathbf{OUBMP_4}$	8	0.48 (−2.7, 3.4)	−1.53 (−6.45, 2.53)	−0.43 (−2.71, 1.83)	0.96 (0.1, 1.92)	0.79 (0.62, 0.98)
$\mathbf{OUBMP_4}$	16	0.01 (−1.21, 1.13)	−1.61 (−6.45, 2.64)	−0.3 (−2.73, 1.73)	1.09 (0.1, 1.92)	0.79 (0.62, 0.98)
$\mathbf{OUBMP_4}$	32	0 (−0.38, 0.41)	−1.58 (−6.43, 2.63)	−0.55 (−2.78, 1.78)	0.99 (0.1, 1.86)	0.8 (0.62, 0.98)
$\mathbf{OUBMP_4}$	64	−0.01 (−0.68, 0.33)	−1.99 (−6.32, 2.27)	−0.3 (−2.8, 1.85)	0.98 (0.08, 1.86)	0.79 (0.62, 0.98)

References

1. Lanszki, J.; Heltai, M.; Kövér, G.; Zalewski, A. Non-linear relationship between body size of terrestrial carnivores and their trophic niche breadth and overlap. *Basic Appl. Ecol.* **2019**, *38*, 36–46. [CrossRef]
2. Agrawal, A.A. Phenotypic plasticity in the interactions and evolution of species. *Science* **2001**, *294*, 321–326. [CrossRef]
3. Kingsolver, J.G.; Huey, R.B. Introduction: The evolution of morphology, performance, and fitness. *Integr. Comp. Biol.* **2003**, *43*, 361–366. [CrossRef]
4. Schlichting, C.D. The evolution of phenotypic plasticity in plants. *Annu. Rev. Ecol. Syst.* **1986**, *17*, 667–693. [CrossRef]
5. Bolker, B.M.; Brooks, M.E.; Clark, C.J.; Geange, S.W.; Poulsen, J.R.; Stevens, M.H.H.; White, J.S.S. Generalized linear mixed models: A practical guide for ecology and evolution. *Trends Ecol. Evol.* **2009**, *24*, 127–135. [CrossRef] [PubMed]
6. Huey, R.B.; Kingsolver, J.G. Evolution of thermal sensitivity of ectotherm performance. *Trends Ecol. Evol.* **1989**, *4*, 131–135. [CrossRef] [PubMed]
7. Barton, B.T.; Schmitz, O.J. Experimental warming transforms multiple predator effects in a grassland food web. *Ecol. Lett.* **2009**, *12*, 1317–1325. [CrossRef] [PubMed]
8. Murren, C.J.; Maclean, H.J.; Diamond, S.E.; Steiner, U.K.; Heskel, M.A.; Handelsman, C.A.; Ghalambor, C.K.; Auld, J.R.; Callahan, H.S.; Pfennig, D.W.; et al. Evolutionary change in continuous reaction norms. *Am. Nat.* **2014**, *183*, 453–467. [CrossRef] [PubMed]
9. Ghalambor, C.K.; McKay, J.K.; Carroll, S.P.; Reznick, D.N. Adaptive versus non-adaptive phenotypic plasticity and the potential for contemporary adaptation in new environments. *Funct. Ecol.* **2007**, *21*, 394–407. [CrossRef]
10. Tzanatos, E.; Moukas, C.; Koutsidi, M. Mediterranean nekton traits: Distribution, relationships and significance for marine ecology monitoring and management. *PeerJ* **2020**, *8*, e8494. [CrossRef]
11. Porter, W.F.; Underwood, H.B. Of elephants and blind men: Deer management in the US National Parks. *Ecol. Appl.* **1999**, *9*, 3–9. [CrossRef]
12. Quader, S.; Isvaran, K.; Hale, R.; Miner, B.; Seavy, N. Nonlinear relationships and phylogenetically independent contrasts. *J. Evol. Biol.* **2004**, *17*, 709–715. [CrossRef] [PubMed]
13. Ho, L.S.T.; Ane, C. A linear-time algorithm for Gaussian and non-Gaussian trait evolution models. *Syst. Biol.* **2014**, *63*, 397–408.
14. Kumar, S.; Suleski, M.; Craig, J.M.; Kasprowicz, A.E.; Sanderford, M.; Li, M.; Stecher, G.; Hedges, S.B. TimeTree 5: An expanded resource for species divergence times. *Mol. Biol. Evol.* **2022**, *39*, msac174. [CrossRef]
15. 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]
16. Jhwueng, D.C.; Maroulas, V. Phylogenetic ornstein–uhlenbeck regression curves. *Stat. Probab. Lett.* **2014**, *89*, 110–117. [CrossRef]
17. Jhwueng, D.C.; Maroulas, V. Adaptive trait evolution in random environment. *J. Appl. Stat.* **2016**, *43*, 2310–2324. [CrossRef]
18. 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]

19. 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] [PubMed]
20. 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]
21. Marass, F.; Mouliere, F.; Yuan, K.; Rosenfeld, N.; Markowetz, F. A phylogenetic latent feature model for clonal deconvolution. *Ann. Appl. Stat.* **2016**, *10*, 2377–2404. [CrossRef]
22. Jhwueng, D.C.; Wang, C.P. Phylogenetic Curved Optimal Regression for Adaptive Trait Evolution. *Entropy* **2021**, *23*, 218. [CrossRef]
23. Hubin, A.; Heinze, G.; De Bin, R. Fractional Polynomial Models as Special Cases of Bayesian Generalized Nonlinear Models. *Fractal Fract.* **2023**, *7*, 641. [CrossRef]
24. Panciroli, E.; Benson, R.B.; Fernandez, V.; Fraser, N.C.; Humpage, M.; Luo, Z.X.; Newham, E.; Walsh, S. Jurassic fossil juvenile reveals prolonged life history in early mammals. *Nature* **2024**, *632*, 815–822. [CrossRef]
25. Felsenstein, J. Evolutionary trees from DNA sequences: A maximum likelihood approach. *J. Mol. Evol.* **1981**, *17*, 368–376. [CrossRef]
26. Pennell, M.W.; Eastman, J.M.; Slater, G.J.; Brown, J.W.; Uyeda, J.C.; FitzJohn, R.G.; Alfaro, M.E.; Harmon, L.J. geiger v2.0: An expanded suite of methods for fitting macroevolutionary models to phylogenetic trees. *Bioinformatics* **2014**, *30*, 2216–2218. [CrossRef] [PubMed]
27. Minter, A.; Retkute, R. Approximate Bayesian Computation for infectious disease modelling. *Epidemics* **2019**, *29*, 100368. [CrossRef]
28. Raynal, L.; Marin, J.M.; Pudlo, P.; Ribatet, M.; Robert, C.P.; Estoup, A. ABC random forests for Bayesian parameter inference. *Bioinformatics* **2019**, *35*, 1720–1728. [CrossRef] [PubMed]
29. Csillery, K.; Francois, O.; Blum, M.G.B. abc: An R package for approximate Bayesian computation (ABC). *Methods Ecol. Evol.* **2012**, *3*, 475–479. [CrossRef]
30. Felsenstein, J. Phylogenies and the comparative method. *Am. Nat.* **1985**, *125*, 1–15. [CrossRef]
31. Atangana, A.; Araz, S.İ. *New Numerical Scheme with Newton Polynomial: Theory, Methods, and Applications*; Academic Press: Cambridge, MA, USA, 2021.
32. Bartoszek, K.; Liò, P. Modelling trait dependent speciation with Approximate Bayesian Computation. *arXiv* **2018**, arXiv:1812.03715. [CrossRef]
33. 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. [CrossRef]
34. Paradis, E.; Schliep, K. ape 5.0: An environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics* **2019**, *35*, 526–528. [CrossRef]
35. Chang, C.H. Application of Polynomial Regression Model to Study Phylogenetic Adaptation Trait Evolution. Master's Thesis, Feng-Chia University, Taichung, Taiwan, 2024.
36. Tucker, M.A.; Rogers, T.L. Examining the prey mass of terrestrial and aquatic carnivorous mammals: Minimum, maximum and range. *PLoS ONE* **2014**, *9*, e106402. [CrossRef] [PubMed]
37. Scarciotti, G.; Teel, A.R. On moment matching for stochastic systems. *IEEE Trans. Autom. Control* **2021**, *67*, 541–556. [CrossRef]
38. Decamps, M.; De Schepper, A. Edgeworth expansions of stochastic trading time. *Phys. A Stat. Mech. Its Appl.* **2010**, *389*, 3179–3192. [CrossRef]
39. Rosas-Puchuri, U.; Santaquiteria, A.; Khanmohammadi, S.; Solis-Lemus, C.; Betancur-R, R. Non-linear phylogenetic regression using regularized kernels. *Methods Ecol. Evol.* **2023**, *15*, 1611–1623. [CrossRef]
40. Thompson, W.K.; Xie, M.; White, H.R. Transformations of covariates for longitudinal data. *Biostatistics* **2003**, *4*, 353–364. [CrossRef] [PubMed]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.