Phylogenetic confidence intervals for the optimal trait value

KRZYSZTOF BARTOSZEK and SERIK SAGITOV

May 19, 2014

Abstract

We consider a stochastic evolutionary model for a phenotype developing amongst n related species with unknown phylogeny. The unknown tree is modelled by a Yule process conditioned on n contemporary nodes. The trait value is assumed to evolve along lineages as an Ornstein–Uhlenbeck process. As a result, the trait values of the n species form a sample with dependent observations. We establish three limit theorems for the sample mean corresponding to three domains for the adaptation rate. In the case of fast adaptation, we show that for large n the normalized sample mean is approximately normally distributed. Using these limit theorems, we develop novel confidence interval formulae for the optimal trait value.

Keywords : Central limit theorem, Conditioned Yule process, macroevolution, martingales, Ornstein–Uhlenbeck process, phylogenetics

1 Introduction

Phylogenetic comparative methods deal with multi-species trait value data. This is an established and rapidly expanding area of research concerning evolution of phenotypes in groups of related species living under various environmental conditions. An important feature of such data is the branching structure of evolution causing dependence among the observed trait values. For this reason the usual starting point for phylogenetic comparative studies is an inferred phylogeny describing the evolutionary relationships. The likelihood can be computed by assuming a model for trait evolution along the branches of this fixed tree, such as the Ornstein-Uhlenbeck process.

The one-dimensional Ornstein-Uhlenbeck model is characterized by four parameters: the optimal value θ, the adaptation rate $\alpha > 0$, the ancestral value X_0, and the noise size σ. The classical Brownian motion model [11] can be viewed as a special case with $\alpha = 0$ and θ being irrelevant. As with any statistical procedure, it is important to be able to compute confidence intervals for these parameters. However, confidence intervals are often not mentioned in phylogenetic comparative studies [5].

There are a number of possible numerical ways of calculating such confidence intervals when the underlying phylogenetic tree is known. Using a regression framework one can apply standard regression theory methods to compute confidence intervals for (θ, X_0) conditionally on (α, σ^2) [12, 16, 21, 24]. Notably in [13] the authors derive analytical formulae for confidence intervals for X_0 under the Brownian motion model. In more complicated situations a parametric bootstrap is a (computationally very demanding) way out [5, 8, 20]. Another approach is to report a support surface [16, 17], or consider the curvature of the likelihood surface [4].

All of the above methods have in common that they assume that the phylogeny describing the evolutionary relationships is fully resolved. Possible errors in the topology can cause problems – the closer to the tips they occur, the more problematic they can be [34]. On the other hand, the regression estimators will remain unbiased even with a misspecified tree [25] and also seem to be robust with respect to errors in the phylogeny at least for the Brownian motion model [33]. There are only few papers addressing the issue of phylogenetic uncertainty. An MCMC procedure to jointly estimate the phylogeny and parameters of the Brownian model of trait evolution was suggested in [19, 18]. Recently, [27] develops an Approximate Bayesian Computation framework to estimate Brownian motion parameters in the case of an incomplete tree.

Our paper studies a situation when nothing is known about the phylogeny. The simplest stochastic model addressing this case is a combination of a Yule tree and the Brownian motion on top of it: already in the 1970s, a joint maximum likelihood estimation procedure of a Yule tree and Brownian motion on top of it was proposed in [10]. This basic evolutionary model allows for far reaching analytical analysis [9, 26]. A more realistic stochastic model of this kind combines the Brownian motion with a birth-death tree allowing for extinction of species [7]. For the latter model [26] explicitly compute the so-called interspecies correlation coefficient. Such "tree-free" models are appropriate for working with fossil data when there may be available rich fossilized phenotypic information but the molecular material might have degraded so much that it is impossible to infer evolutionary relationships. In [9] the usefulness of the tree-free approach for contemporary species is demonstrated in an Carnivora order case study.

Conditioned birth–death processes as stochastic models for species trees, have received significant attention in the last decade [3, 14, 22, 29, 30, 31]. In this work the unknown tree is modeled by the Yule process conditioned on n extant species while the evolution of a trait along a lineage is viewed as the Ornstein-Uhlenbeck process, see Fig. 1. We study the properties of the sample mean and sample variance computed from the vector of n trait values. Our main results are three asymptotic confidence interval formulae for the optimal trait value θ. These three formulae represent three asymptotic regimes for different values of the adaptation rate α.

In the discussion in [9] it is pointed out that "as evolutionary biologists further refine our knowledge of the tree of life, the number of clades whose phylogeny is truly unknown may diminish, along with interest in tree-free estimation methods." In our opinion, the easy-to-compute tree-free predictions will always play an important role of a sanity check to see whether the conclusions based on the inferred phylogeny deviate much from those from a "typical" phylogeny. Moreover, results like those presented here can also be used as a method of testing software for phylogenetic comparative models.

A detailed description of the evolutionary model along with our main results are presented in Section 2. Section 3 contains new formulae for the Laplace transforms of important characteristics of the conditioned Yule species tree: the time to origin U_n and the time $\tau^{(n)}$ to the most recent common ancestor for a pair of two species chosen at random out of n extant species. In Section 4 we calculate the interspecies correlation coefficient for the Yule–Ornstein–Uhlenbeck model and Section 5 contains the proof of our limit theorems. In Section 6 we establish the consistency of the stationary variance estimator, which is needed for our confidence interval formulae, cf [16] where the residual sum of squares was suggested to estimate the stationary variance. In Appendix A we calculate all the joint moments of U_n and $\tau^{(n)}$.

Our main result, Theorem 2.1, should be compared with the limit theorems obtained in [1, 2]. They also revealed three asymptotic regimes in a related, though different setting, dealing with a branching Ornstein–Uhlenbeck process. In their case the time of observation is deterministic and the number of the tree tips is random, while in our case the observation time is random and the number of the tips is deterministic. Although it is possible (with some effort) to deduce our results from [1, 2], our proof provides a much more elementary derivation. We believe that our approach will be useful in addressing other biologically relevant

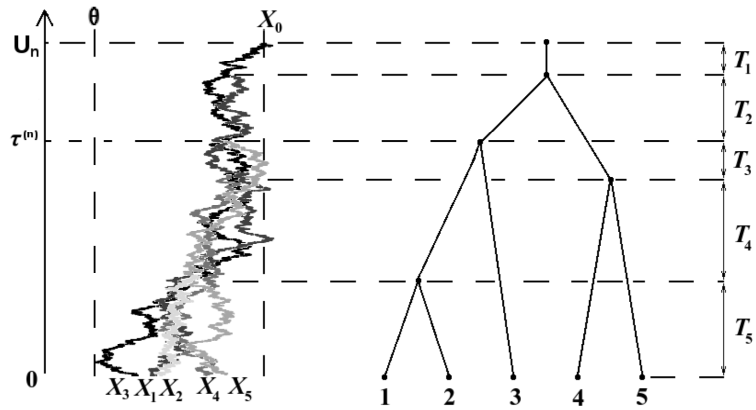

Figure 1: On the left: a branching Ornstein–Uhlenbeck process simulated on a realization of the Yule n-tree with $n = 5$ tips using the TreeSim [29, 30] and mvSLOUCH [4] R [23] packages. Parameters used are $\alpha = 1$, $\sigma = 1$, $X_0 - \theta = 2$, after the tree height U_n was scaled to 1. On the right: the species tree disregarding the trait values supplied with the notation for the inter–speciation times. For the pair of tips (2,3) the time $\tau^{(n)}$ to their most recent common ancestor is marked on the time axis (starting at present and going back to the time of origin).

issues like the formulae for the higher moments given in Appendix A.

2 The model and main results

This work deals with what we call the Yule–Ornstein–Uhlenbeck model which is characterized by four parameters $(X_0, \alpha, \sigma, \theta)$ and consists of two ingredients

1. the species tree connecting n extant species is modeled by the pure birth Yule process [35] with a unit speciation rate $\lambda = 1$ and conditioned on having n tips [14],

2. the observed trait values $(X_1^{(n)}, \ldots, X_n^{(n)})$ on the tips of the tree evolved from the ancestral state X_0 following the Ornstein–Uhlenbeck process with parameters (α, σ, θ).

Definition 2.1 *Let $(T_1, \ldots, T_n)$ be independent exponential random variables with parameters $(1, \ldots, n)$. We define the Yule n-tree as a random tree with n tips which* *is constructed using a bottom-up algorithm based on the following two simple rules.*

(1) During the time period T_k the tree has k branches.

(2) For $k \in [2,n]$ the reduction from k to $k-1$ branches occurs as two randomly chosen branches coalesce into one branch.

The height the Yule n-tree is now $U_n = T_1 + \ldots + T_n$.

As shown in [14], this definition corresponds to the standard Yule tree conditioned on having n tips at the moment of observation, assuming that the time to the origin has the improper uniform prior.

Following [8, 16], we model trait evolution along a lineage using the Ornstein–Uhlenbeck process $X(t)$ given by the stochastic differential equation

$$\mathrm{d}X(t) = -\alpha(X(t)-\theta)\mathrm{d}t + \sigma \mathrm{d}B(t), \quad X(0) = X_0. \tag{1}$$

Here $\alpha > 0$ is the adaptation rate, θ is the optimal trait value, σ^2 is the noise variance, and $B(t)$ is the standard Wiener process. The distribution of $X(t)$ is normal with,

$$\mathrm{E}[X(t)] = \theta + e^{-\alpha t}(X_0-\theta), \quad \mathrm{Var}[X(t)] = \frac{\sigma^2}{2\alpha}(1-e^{-2\alpha t}), \tag{2}$$

implying that $X(t)$ looses the effect of the ancestral state X_0 at an exponential rate. In the long run the Ornstein–Uhlenbeck process acquires a stationary normal distribution with mean θ and variance $\sigma^2/2\alpha$.

We propose asymptotic confidence interval formulae for the optimal value θ which take into account phylogenetic uncertainty. To this end we study properties of the sample mean and sample variance

$$\overline{X}_n = \frac{X_1^{(n)} + \ldots + X_n^{(n)}}{n}, \quad S_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i^{(n)} - \overline{X}_n)^2.$$

Using the properties of the Yule–Ornstein–Uhlenbeck model we find explicit expressions for $\mathrm{E}\left[\overline{X}_n\right]$, $\mathrm{Var}\left[\overline{X}_n\right]$, $\mathrm{E}\left[S_n^2\right]$, study the asymptotics of $\mathrm{Var}\left[S_n^2\right]$, and prove the following limit theorem revealing three different asymptotic regimes.

Theorem 2.1 *Let $\delta = \frac{X_0-\theta}{\sqrt{\sigma^2/2\alpha}}$ be a normalized difference between the ancestral and optimal values. Consider the normalized sample mean $\overline{Y}_n = \frac{\overline{X}_n-\theta}{\sqrt{\sigma^2/2\alpha}}$ of the*

Yule-Ornstein-Uhlenbeck process with $\overline{Y}_0 = \delta$. As $n \to \infty$ the process $\overline{Y}_n$ has the following limit behavior.

(i) If $\alpha > 0.5$, then $\sqrt{n} \cdot \overline{Y}_n$ is asymptotically normally distributed with zero mean and variance $\frac{2\alpha+1}{2\alpha-1}$.

(ii) If $\alpha = 0.5$, then $\sqrt{n/\ln n} \cdot \overline{Y}_n$ is asymptotically normally distributed with zero mean and variance 2.

(iii) If $\alpha < 0.5$, then $n^{\alpha} \cdot \overline{Y}_n$ converges a.s. and in L^2 to a random variable $Y_{\alpha,\delta}$ with $\mathrm{E}\left[Y_{\alpha,\delta}\right] = \delta\Gamma(1+\alpha)$ and $\mathrm{E}\left[Y_{\alpha,\delta}^2\right] = \left(\delta^2 + \frac{4\alpha}{1-2\alpha}\right)\Gamma(1+2\alpha)$.

Let z_x be the x-quantile of the standard normal distribution, and q_x be the x-quantile of the limit $Y_{\alpha,\delta}$. Denote by S_n the sample standard deviation defined as the square root of S_n^2. As it will be shown in Section 6, the sample variance S_n^2 is a consistent estimator of $\frac{\sigma^2}{2\alpha}$. This fact together with Theorem 2.1 allows us to state the following three approximate $(1-x)$-level confidence intervals for θ assuming that we know the value of α:

$$
\begin{aligned}
&\text{for } \alpha > 0.5 \qquad \overline{X}_n \pm z_{1-x/2} \cdot S_n \cdot \frac{K_\alpha}{\sqrt{n}}, \quad K_\alpha = \sqrt{\frac{2\alpha+1}{2\alpha-1}},\\
&\text{for } \alpha = 0.5 \qquad \overline{X}_n \pm z_{1-x/2} \cdot S_n \cdot \frac{\sqrt{2\ln n}}{\sqrt{n}},\\
&\text{for } \alpha < 0.5 \qquad (\overline{X}_n - q_{x/2} \cdot S_n \cdot n^{-\alpha},\ \overline{X}_n + q_{1-x/2} \cdot S_n \cdot n^{-\alpha}).
\end{aligned}
$$

Notably, the first of these confidence intervals differs from the classical confidence interval for the mean $(\overline{X}_n \pm z_{1-x/2} S_n/\sqrt{n})$ just by a factor K_α. The latter is larger than 1, as it should, in view of a positive correlation among the sample observations. The correction factor K_α becomes negligible in the case of a very strong adaptation, $\alpha \gg 1$, when the dependence due to common ancestry can be neglected.

Remark 2.1 *Observe that our standing assumption $\lambda = 1$, see Definition 2.1, of having one speciation event per unit of time causes no loss of generality. To incorporate an arbitrary speciation rate λ one has to replace in our formulae parameters α and σ^2 by α/λ and σ^2/λ. This transformation corresponds to the time scaling by factor λ in Eq. (1), it changes neither the optimal value θ nor the stationary variance $\sigma^2/(2\alpha)$.*

3 Sampling m leaves from the Yule n-tree

Here we consider the Yule n-tree, see Definition 2.1 and study some properties of its subtree joining m randomly (without replacement) chosen tips, where $m \in [2,n]$. In particular, we compute the joint Laplace transform of the height of the Yule n-tree $U_n = T_1 + \ldots + T_n$ and $\tau^{(n)}$, the height of the most recent common ancestor for two randomly sampled tips, see Fig. 1. For other results concerning the distribution of $\tau^{(n)}$ and U_n see also [14, 15, 22, 26, 28, 29, 31, 32].

Lemma 3.1 *Consider a random m-subtree of the conditioned Yule n-tree. It has* $m-1$ *bifurcating events. Let* $K_1^{(n,m)} < \ldots < K_{m-1}^{(n,m)}$ *be the consecutive numbers of the bifurcation events in the Yule n-tree (counted from the root toward the leaves) corresponding to the* $m-1$ *bifurcating events of the m-subtree. Put* $K_0^{(n,m)} = 0$ *and* $K_m^{(n,m)} = n$. *The sequence* $(K_m^{(n,m)}, \ldots, K_0^{(n,m)})$ *forms a time inhomogeneous Markov chain with transition probabilities*

$$\mathrm{P}(K_{j-1}^{(n,m)} = k | K_j^{(n,m)} = i) = p_{ik}^{(j)}, \quad 1 \le j < k < i \le n,$$

where $p_{i,0}^{(1)} = 1$ *for all* $i \ge 1$, *and*

$$p_{ik}^{(j)} = \frac{j(j-1)}{ik} \prod_{l=k+1}^{i-1} \frac{(l+1-j)(l+j)}{l^2}, \quad j = 2, \ldots, m.$$

PROOF Tracing the lineages of m randomly sampled tips of the Yule n-tree towards the root, the first coalescent event can be viewed as the success in a sequence of independent Bernoulli trials. This argument leads to the expression cf [29]

$$\begin{aligned} \mathrm{P}(K_{m-1}^{(n,m)} = k | K_m^{(n,m)} = n) &= \left(1 - \frac{\binom{m}{2}}{\binom{n}{2}}\right) \cdots \left(1 - \frac{\binom{m}{2}}{\binom{k+2}{2}}\right) \frac{\binom{m}{2}}{\binom{k+1}{2}} \\ &= \frac{m(m-1)}{nk} \prod_{i=k+1}^{n-1} \frac{(i+1-m)(i+m)}{i^2}, \quad k = m-1, \ldots, n-1, \end{aligned}$$

confirming the formula stated for the transition probabilities $p_{nk}^{(m)}$. The transition probabilities $p_{ik}^{(j)}$ for $j = 2, \ldots, m-1$ are obtained similarly. Notice, as a check, that $p_{j,j-1}^{(j)} = 1$.

$\square$

Lemma 3.2 *Consider the inter-bifurcation times for the m-subtree of the Yule n-tree*

$$\chi_j^{(n,m)} = T_{K_{j-1}^{(n,m)}+1} + \ldots + T_{K_j^{(n,m)}}, \quad j = 1, \ldots, m,$$

so that $U_n = \chi_1^{(n,m)} + \ldots + \chi_m^{(n,m)}$ *for any* $m \le n$, *and* $\tau^{(n)} = \chi_2^{(n,2)}$. *Then for* $x_j > -1$ *we have*

$$\mathrm{E}\left[\exp\left\{-\sum_{j=1}^{m} x_j \chi_j^{(n,m)}\right\}\right] = \sum_{k_1=1}^{n-1} \sum_{k_2=k_1+1}^{n-1} \cdots \sum_{k_{m-1}=k_{m-2}+1}^{n-1} \prod_{j=1}^{m} p_{k_j,k_{j-1}}^{(j)} \frac{b_{k_j,x_{j-1}}}{b_{k_{j-1},x_{j-1}}},$$

where $k_m = n$, $k_0 = 0$, *and*

$$b_{n,x} = \frac{1}{1+x} \cdot \frac{2}{2+x} \cdot \ldots \cdot \frac{n}{n+x} = \frac{\Gamma(n+1)\Gamma(x+1)}{\Gamma(n+x+1)}, \quad x > -1.$$

PROOF The Laplace transform of the sum of independent exponentials:

$$\begin{aligned}\mathrm{E}\left[\exp\{-x_1\chi_0^{(n,m)} - \ldots - x_m\chi_m^{(n,m)}\} | (K_{m-1}^{(n,m)}, \ldots K_1^{(n,m)}) = (k_{m-1}, \ldots k_1)\right] \\ = \prod_{j=1}^{m} \frac{k_{j-1}+1}{x_{j-1}+k_{j-1}+1} \cdots \frac{k_j}{x_{j-1}+k_j}\end{aligned}$$

together with Lemma 3.1 implies the stated equality

$$\begin{aligned}&\mathrm{E}\left[\exp\{-\sum_{j=1}^{m} x_j \chi_j^{(n,m)}\}\right] \\ &\quad = \sum_{k_1=1}^{n-1} \sum_{k_2=k_1+1}^{n-1} \cdots \sum_{k_{m-1}=k_{m-2}+1}^{n-1} \prod_{j=1}^{m} p_{k_j,k_{j-1}}^{(j)} \frac{k_{j-1}+1}{x_{j-1}+k_{j-1}+1} \cdots \frac{k_j}{x_{j-1}+k_j}.\end{aligned}$$

$\square$

Lemma 3.3 *The joint Laplace transform of the height of the Yule n-tree and the height of the most recent common ancestor for two randomly sampled tips is given by*

$$\mathrm{E}\left[e^{-xU_n-y\tau^{(n)}}\right]=\frac{2(n+1)b_{n,x+y}}{(n-1)}\sum_{k=1}^{n-1}\frac{b_{k,x}}{(k+2)(k+1)b_{k,x+y}}.$$

In particular,

$$\mathrm{E}\left[e^{-xU_n}\right]=b_{n,x}, \tag{3}$$

$$\mathrm{Var}\left[e^{-xU_n}\right]=b_{n,2x}-b_{n,x}^2, \tag{4}$$

and, denoting the harmonic number $h_n := 1+1/2+\ldots+1/n$*,*

$$\mathrm{E}\left[e^{-y\tau^{(n)}}\right]=\begin{cases}\frac{2-(n+1)(y+1)b_{n,y}}{(n-1)(y-1)}, & \text{for } y\neq 1,\\ \frac{2}{n-1}(h_n-1)-\frac{1}{n+1}, & \text{for } y=1.\end{cases} \tag{5}$$

PROOF Turning to Lemma 3.1 with $m=2$ we get

$$p_{n,k}^{(2)}=\frac{2}{nk}\prod_{i=k+1}^{n-1}\frac{(i-1)(i+2)}{i^2}=\frac{2(n+1)}{(n-1)(k+2)(k+1)},\quad k=1,\ldots,n-1,$$

and according to Lemma 3.2

$$\begin{aligned}\mathrm{E}\left[e^{-x(U_n-\tau^{(n)})-y\tau^{(n)}}\right]&=\sum_{k=1}^{n-1}p_{n,k}^{(2)}\frac{1}{x+1}\cdots\frac{k}{x+k}\frac{k+1}{y+k+1}\cdots\frac{n}{y+n}\\&=\frac{2(n+1)b_{n,y}}{(n-1)}\sum_{k=1}^{n-1}\frac{b_{k,x}}{(k+2)(k+1)b_{k,y}}.\end{aligned} \tag{6}$$

This implies the main formula claimed by Lemma 3.3 giving $\mathrm{E}\left[e^{-xU_n}\right]=b_{n,x}$ after putting $y=0$. With $x=0$,

$$\begin{aligned}\mathrm{E}\left[e^{-y\tau^{(n)}}\right]&=\frac{2(n+1)b_{n,y}}{(n-1)}\sum_{k=1}^{n-1}\frac{1}{(k+2)(k+1)b_{k,y}}\\&=\frac{2(n+1)!}{(n-1)\Gamma(y+n+1)}\sum_{k=1}^{n-1}\frac{\Gamma(k+1+y)}{\Gamma(k+3)}.\end{aligned}$$

When $y = 1$ this directly becomes

$$\mathrm{E}\left[e^{-\tau^{(n)}}\right] = \frac{2}{n-1}(h_n - 1) - \frac{1}{n+1}.$$

In the case of $y \neq 1$ we use the following relation (easily verified by induction when $z \neq y$)

$$\sum_{k=1}^{n-1} \frac{\Gamma(k+y)}{\Gamma(k+z+1)} = \frac{\Gamma(n+z)\Gamma(y+1) - \Gamma(z+1)\Gamma(n+y)}{\Gamma(z+1)\Gamma(n+z)(z-y)} \tag{7}$$

to derive

$$\mathrm{E}\left[e^{-y\tau^{(n)}}\right] = \frac{2\Gamma(n+1+y) - \Gamma(n+2)\Gamma(y+2)}{(n-1)\Gamma(y+n+1)(y-1)} = \frac{2-(n+1)(y+1)b_{n,y}}{(n-1)(y-1)}.$$

□

Lemma 3.4 *As $n \to \infty$ for positive x and y we have the following asymptotic results*

$$\begin{aligned}
\mathrm{E}\left[e^{-xU_n}\right] &\sim \Gamma(x+1)n^{-x}, \\
\mathrm{E}\left[e^{-y\tau^{(n)}}\right] &\sim \begin{cases} \frac{1+y}{1-y}\Gamma(y+1)\cdot n^{-y}, & \text{if } 0<y<1, \\ 2n^{-1}\ln n, & \text{if } y=1, \\ \frac{2}{y-1}n^{-1}, & \text{if } y>1, \end{cases} \\
\mathrm{E}\left[e^{-xU_n - y\tau^{(n)}}\right] &\sim \begin{cases} C_{x,y}n^{-x-y}, & \text{if } 0<y<1, \\ 2\Gamma(x+1)n^{-x-1}\ln n, & \text{if } y=1, \\ \frac{2\Gamma(x+1)}{y-1}\cdot n^{-x-1}, & \text{if } y>1, \end{cases}
\end{aligned}$$

where

$$C_{x,y} = 2\Gamma(x+y+1)\sum_{k=1}^{\infty} \frac{b_{k,x}}{(k+2)(k+1)b_{k,x+y}}.$$

PROOF The stated results are obtained from Lemma 3.3 using the first of the following three asymptotic properties of the function $b_{n,x}$

$$b_{n,x} \sim \Gamma(x+1)n^{-x},\ n \to \infty,$$
$$\frac{1-(n+1)b_{n,x}}{x-1} \to h_n - 1,\ x \to 1,$$
$$x^{-1}(1-b_{n,x}) \to h_n,\ x \to 0.$$

These three relations will often be used tacitly in what follows.

□

4 Interspecies correlation

Denote by $\mathscr{Y}_n$ the σ–algebra containing all information on the Yule n-tree. The scaled trait values $Y_i^{(n)} := \frac{X_i^{(n)}-\theta}{\sigma/\sqrt{2\alpha}}$, in view of Eq. (2), are conditionally normal with

$$\mathrm{E}\left[Y_i^{(n)}|\mathscr{Y}_n\right] = \delta e^{-\alpha U_n},$$
$$\mathrm{Var}\left[Y_i^{(n)}|\mathscr{Y}_n\right] = 1 - e^{-2\alpha U_n},$$

which together with the results from Section 3 entails

$$\mathrm{E}\left[Y_i^{(n)}\right] = \delta b_{n,\alpha},$$
$$\mathrm{Var}\left[Y_i^{(n)}\right] = 1 - \delta^2 b_{n,\alpha}^2 + (\delta^2 - 1) b_{n,2\alpha}.$$

Lemma 4.1 *In the framework of the Yule-Ornstein-Uhlenbeck model, for an arbitrary pair of traits we have*

$$\mathrm{Cov}\left[Y_i^{(n)}, Y_j^{(n)}|\mathscr{Y}_n\right] = e^{-2\alpha\tau_{ij}^{(n)}} - e^{-2\alpha U_n},$$

where $\tau_{ij}^{(n)}$ is the backward time to the most recent common ancestor of the tips (i,j).

PROOF Denote by $Y_{ij}^{(n)}$ the normalized trait value of the most recent common ancestor of the tips (i,j). Let $\mathscr{Y}_{ij}^{(n)}$ stand for the σ–algebra generated by $(\mathscr{Y}_n, Y_{ij}^{(n)})$, then using Eq. (2) we get

$$\mathrm{Cov}\left[Y_i^{(n)}, Y_j^{(n)} | \mathscr{Y}_{ij}^{(n)}\right] = 0, \qquad \mathrm{E}\left[Y_i | \mathscr{Y}_{ij}^{(n)}\right] = \mathrm{E}\left[Y_j | \mathscr{Y}_{ij}^{(n)}\right] = e^{-\alpha\tau_{ij}^{(n)}} Y_{ij}^{(n)},$$

implying the statement of this lemma

$$\begin{aligned}\mathrm{Cov}\left[Y_i^{(n)}, Y_j^{(n)} | \mathscr{Y}_n\right] &= \mathrm{Var}\left[e^{-\alpha\tau_{ij}^{(n)}} Y_{ij}^{(n)} | \mathscr{Y}_n\right] \\ &= e^{-2\alpha\tau_{ij}^{(n)}}(1 - e^{-2\alpha(U_n - \tau_{ij}^{(n)})}) = e^{-2\alpha\tau_{ij}^{(n)}} - e^{-2\alpha U_n}.\end{aligned}$$

□

Lemma 4.2 *Consider the interspecies correlation coefficient, the unconditioned correlation between two randomly sampled trait values*

$$\rho_n = \frac{1}{n(n-1)} \sum_i \sum_{j \neq j} \frac{\mathrm{Cov}\left[X_i^{(n)}, X_j^{(n)}\right]}{\mathrm{Var}\left[X_1^{(n)}\right]} = \frac{1}{n(n-1)} \sum_i \sum_{j \neq j} \frac{\mathrm{Cov}\left[Y_i^{(n)}, Y_j^{(n)}\right]}{\mathrm{Var}\left[Y_1^{(n)}\right]}.$$

If $\alpha \neq 0.5$, *then*

$$\rho_n = 1 - \frac{2\alpha(n-1) + (n+1)((1+2\alpha)b_{n,2\alpha} - 1)}{(n-1)(2\alpha - 1)(1 + (\delta^2 - 1)b_{n,2\alpha} - \delta^2 b_{n,\alpha}^2)},$$

and in the case of $\alpha = 0.5$

$$\rho_n = 1 - \frac{n+1}{n-1} \frac{n+2-2h_n}{n + \delta^2(1 - (n+1)b_{n,0.5}^2)}.$$

PROOF According to Lemma 4.1 we have,

$$\frac{2}{n(n-1)} \sum_{i<j} \mathrm{Cov}\left[Y_i^{(n)}, Y_j^{(n)}\right] = \mathrm{E}\left[e^{-2\alpha\tau^{(n)}} - e^{-2\alpha U_n}\right] + \delta^2\,\mathrm{Var}\left[e^{-\alpha U_n}\right]$$

leading to

$$\rho_n = 1 - \frac{1 - \mathrm{E}\left[e^{-2\alpha\tau^{(n)}}\right]}{1 - \mathrm{E}\left[e^{-2\alpha U_n}\right] + \delta^2 \operatorname{Var}\left[e^{-\alpha U_n}\right]}.$$

Applying the results of Section 3 we arrive at the asserted relations for ρ_n. Observe that asymptotically as $n \to \infty$ the interspecies correlation coefficient decays to 0 as

$$\rho_n \sim \begin{cases} \frac{2}{1-2\alpha}\Gamma(1+2\alpha) + \delta^2\left(\Gamma(1+2\alpha) - \Gamma^2(\alpha+1)\right) n^{-2\alpha} & 0 < \alpha < 0.5, \\ 2n^{-1}\ln n & \alpha = 0.5, \\ \frac{2}{2\alpha-1} n^{-1} & \alpha > 0.5. \end{cases}$$

$\square$

Lemma 4.3 *Consider the sample mean $\overline{Y}_n = n^{-1}(Y_1^{(n)} + \ldots + Y_n^{(n)})$ and the sample variance*

$$D_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(Y_i^{(n)} - \overline{Y}_n)^2,$$

of the scaled trait values. For all $\alpha > 0$ we have $\mathrm{E}\left[\overline{Y}_n\right] = \delta b_{n,\alpha}$. For $\alpha \neq 0.5$

$$\operatorname{Var}\left[\overline{Y}_n\right] = \frac{1 + 2\alpha - (4\alpha n + 1 + 2\alpha) b_{n,2\alpha}}{(2\alpha - 1)n} + \delta^2 (b_{n,2\alpha} - b_{n,\alpha}^2),$$
$$\mathrm{E}\left[D_n^2\right] = 1 + \frac{(1+2\alpha)(n+1) b_{n,2\alpha} - 2}{(2\alpha-1)(n-1)},$$

and in the singular case $\alpha = 0.5$

$$\operatorname{Var}\left[\overline{Y}_n\right] = \frac{2(h_n - 1)}{n} + \frac{\delta^2 - 1}{n+1} - \delta^2 b_{n,0.5}^2,$$
$$\mathrm{E}\left[D_n^2\right] = \frac{n - 2h_n}{n-1}.$$

PROOF Obviously, $\mathrm{E}\left[\overline{Y}_n\right] = \mathrm{E}\left[Y_i^{(n)}\right] = \delta b_{n,\alpha}$. To prove the other assertions we turn to [26], where the concept of interspecies correlation was originally introduced. It was shown there that the variance of the sample average and the expectation of the sample variance can be compactly expressed as

$$\begin{aligned}\operatorname{Var}\left[\overline{X}_n\right] &= \left(\frac{1}{n}+\frac{n-1}{n}\rho_n\right)\operatorname{Var}\left[X_1^{(n)}\right],\\ \mathrm{E}\left[S_n^2\right] &= (1-\rho_n)\operatorname{Var}\left[X_1^{(n)}\right].\end{aligned}$$

Since $\operatorname{Var}\left[Y_1^{(n)}\right] = \frac{2\alpha}{\sigma^2}\operatorname{Var}\left[X_1^{(n)}\right]$, $\operatorname{Var}\left[\overline{Y}_n\right] = \frac{2\alpha}{\sigma^2}\operatorname{Var}\left[\overline{X}_n\right]$, and $\mathrm{E}\left[D_n^2\right] = \frac{2\alpha}{\sigma^2}\mathrm{E}\left[S_n^2\right]$ it remains to combine Lemma 4.2 with the known expression for $\operatorname{Var}\left[Y_1^{(n)}\right]$.

A more direct proof of Lemma 4.3 can be obtained using the following result on conditional expectations.

□

Lemma 4.4 *We have*

$$\begin{aligned}\mathrm{E}\left[\overline{Y}_n|\mathscr{Y}_n\right] &= \delta e^{-\alpha U_n},\\ \mathrm{E}\left[\overline{Y}_n^2|\mathscr{Y}_n\right] &= n^{-1}+(1-n^{-1})\,\mathrm{E}\left[e^{-2\alpha\tau^{(n)}}|\mathscr{Y}_n\right]-e^{-2\alpha U_n}+\delta^2 e^{-2\alpha U_n},\\ \operatorname{Var}\left[\overline{Y}_n^2|\mathscr{Y}_n\right] &= n^{-1}+(1-n^{-1})\,\mathrm{E}\left[e^{-2\alpha\tau^{(n)}}|\mathscr{Y}_n\right]-e^{-2\alpha U_n}.\end{aligned}$$

PROOF The main assertion follows from

$$\begin{aligned}\operatorname{Var}\left[Y_1^{(n)}+\ldots+Y_n^{(n)}|\mathscr{Y}_n\right] &= n(1-e^{-2\alpha U_n})+2\sum_{i<j}(e^{-2\alpha\tau_{ij}^{(n)}}-e^{-2\alpha U_n})\\ &= n-n^2e^{-2\alpha U_n}+n(n-1)\,\mathrm{E}\left[e^{-2\alpha\tau^{(n)}}|\mathscr{Y}_n\right].\end{aligned}$$

□

5 Proof of Theorem 2.1

Lemma 5.1 *Put* $V_n^{(x)} := b_{n,x}^{-1}\cdot e^{-xU_n}$ *with* $\mathrm{E}\left[V_n^{(x)}\right]=1$. *For any* $x>-1$ *the sequence* $\{V_n^{(x)},\mathscr{Y}_n\}_{n\geq 0}$ *forms a martingale converging a.s. and in* L^2. *Moreover,* $(U_n-\log n)$ *converges in distribution to a random variable having the standard Gumbel distribution.*

PROOF The martingale property is obvious

$$\mathrm{E}\left[V_{n+1}^{(x)}|\mathscr{Y}_n\right] = b_{n+1,x}^{-1} \cdot e^{-xU_n}\,\mathrm{E}\left[e^{-xT_{n+1}}\right] = b_{n,x}^{-1} \cdot e^{-xU_n} = V_n^{(x)}.$$

Since the second moments

$$\mathrm{E}\left[(V_n^{(x)})^2\right] = b_{n,x}^{-2} \cdot \mathrm{E}\left[e^{-2xU_n}\right] = \frac{b_{n,2x}}{(b_{n,x})^2}$$

are uniformly bounded over n, we may conclude that $V_n^{(x)} \to V^{(x)}$ a.s. and in L^2 with $\mathrm{E}\left[V^{(x)}\right] = 1$. It follows that $\mathrm{E}\left[V_n^{(x)}\right] \to 1$, and therefore, $\mathrm{E}\left[e^{-x(U_n - \log n)}\right] \to \Gamma(x+1)$. The latter is a convergence of Laplace transforms confirming the stated convergence in distribution.

□

Observe that the Gumbel limit for $U_n - \log n$ can be obtained using the classical extreme value theory, in view of the representation

$$U_n \stackrel{d}{=} \sum_{i=1}^{n} i^{-1} E_i \stackrel{\mathscr{D}}{=} \max(E_1, \ldots, E_n)$$

in terms of independent exponentials with parameter 1. Notice also that $U_{n+1}/2$ has the same distribution as the total branch length of Kingman's n-coalescent.

Lemma 5.2 *Denote by $\mathscr{F}_n$ the σ–algebra containing information on the Yule n-tree realization as well as the corresponding information on the evolution of trait values. Set*

$$H_n := (n+1)e^{(\alpha-1)U_n}\overline{Y}_n, \quad n \geq 0.$$

The sequence $\{H_n, \mathscr{F}_n\}_{n\geq 0}$ forms a martingale with $\mathrm{E}[H_n] = H_0 = \delta$.

PROOF Notice that,

$$\begin{aligned}\mathrm{E}\left[e^{(\alpha-1)T_{n+1}} \sum_{i=1}^{n+1} Y_i^{(n+1)} | \mathscr{F}_n\right] &= \mathrm{E}\left[e^{-T_{n+1}}\right] \left(\sum_{i=1}^{n} Y_i^{(n)} + n^{-1} \sum_{j=1}^{n} Y_j^{(n)}\right) \\ &= \frac{n+1}{n+2}\frac{n+1}{n} \sum_{i=1}^{n} Y_i^{(n)} = \frac{(n+1)^2}{n+2}\overline{Y}_n.\end{aligned}$$

Hence

$$\mathrm{E}\left[H_{n+1}|\mathscr{F}_n\right] = \frac{n+2}{n+1}e^{(\alpha-1)U_n}\,\mathrm{E}\left[e^{(\alpha-1)T_{n+1}}\sum_{i=1}^{n+1}Y_i^{(n+1)}|\mathscr{F}_n\right] = H_n.$$

$\square$

Lemma 5.3 *For all positive* α *we have* $\mathrm{Var}\left[\mathrm{E}\left[e^{-2\alpha\tau^{(n)}}|\mathscr{Y}_n\right]\right] = O(n^{-3})$ *as* $n \to \infty$.

PROOF For a given realization of the Yule n-tree we denote by $\tau_1^{(n)}$ and $\tau_2^{(n)}$ two independent versions of $\tau^{(n)}$ corresponding to two independent choices of pairs of tips out of n available. We have,

$$\mathrm{E}\left[\left(\mathrm{E}\left[e^{-2\alpha\tau^{(n)}}|\mathscr{Y}_n\right]\right)^2\right] = \mathrm{E}\left[\mathrm{E}\left[e^{-2\alpha(\tau_1^{(n)}+\tau_2^{(n)})}|\mathscr{Y}_n\right]\right] = \mathrm{E}\left[e^{-2\alpha(\tau_1^{(n)}+\tau_2^{(n)})}\right].$$

Writing

$$\pi_{n,k} := p_{n,k}^{(2)}, \qquad f(a,k,n) = \frac{k+1}{a+k+1}\cdots\frac{n}{a+n}$$

and using the ideas of Section 3 we obtain

$$\begin{aligned}\mathrm{E}\left[\left(\mathrm{E}\left[e^{-2\alpha\tau^{(n)}}|\mathscr{Y}_n\right]\right)^2\right] &= \sum_{k=1}^{n-1} f_{4\alpha}(k,n)\pi_{n,k}^2 \\ &\quad + 2\sum_{k_1=1}^{n-1}\sum_{k_2=k_1+1}^{n-1} f_{2\alpha}(k_1,k_2)f_{4\alpha}(k_2,n)\pi_{n,k_1}\pi_{n,k_2}.\end{aligned}$$

On the other hand,

$$\left(\mathrm{E}\left[e^{-2\alpha\tau^{(n)}}\right]\right)^2 = \left(\sum_{k_1} f_{2\alpha}(k_1,n)\pi_{n,k_1}\right)\left(\sum_{k_2} f_{2\alpha}(k_2,n)\pi_{n,k_2}\right).$$

Taking the difference between the last two expressions we find

$$\mathrm{Var}\left[\mathrm{E}\left[e^{-2\alpha\tau^{(n)}}|\mathscr{Y}_n\right]\right] = \sum_k \left(f_{4\alpha}(k,n) - f_{2\alpha}(k,n)^2\right)\pi_{n,k}^2$$
$$+2\sum_{k_1=1}^{n-1}\sum_{k_2=k_1+1}^{n-1} f_{2\alpha}(k_1,k_2)\left(f_{4\alpha}(k_2,n) - f_{2\alpha}(k_2,n)^2\right)\pi_{n,k_1}\pi_{n,k_2}.$$

Using the simple equality

$$a_1\cdots a_n - b_1\cdots b_n = \sum_{i=1}^{n} b_1\cdots b_{i-1}(a_i - b_i)a_{i+1}\cdots a_n$$

we see that it suffices to prove that,

$$\sum_{k=1}^{n-1} A_{n,k}\pi_{n,k}^2 = O(n^{-4}),$$
$$\sum_{k_1=1}^{n-1}\sum_{k_2=k_1+1}^{n-1} f_{2\alpha}(k_1,k_2)A_{n,k_2}\pi_{n,k_1}\pi_{n,k_2} = O(n^{-3}),$$

where

$$A_{n,k} := \sum_{j=k+1}^{n-1} f_{2\alpha}(k,j)^2\left(\frac{2\alpha}{2\alpha+j+1}\right)^2 f_{4\alpha}(j,n).$$

To verify these two asymptotic relations observe that

$$A_{n,k} < \frac{k+1}{4\alpha+k+1}\cdots\frac{n}{4\alpha+n}\sum_{i=k+1}^{n}\frac{4\alpha^2}{(2\alpha+i)^2} < \frac{4\alpha^2 b_{n,4\alpha}}{b_{k,4\alpha}}\sum_{i=k+1}^{n}\frac{1}{i(i-1)} < \frac{4\alpha^2 b_{n,4\alpha}}{k b_{k,4\alpha}}.$$

Since $\pi_{n,k} = \frac{2(n+1)}{(n-1)(k+2)(k+1)}$, it follows

$$\sum_{k=1}^{n-1} A_{n,k}\pi_{n,k}^2 < c_1 b_{n,4\alpha}\sum_{k=1}^{n-1}\frac{1}{k^5 b_{k,4\alpha}} < c_2 n^{-4\alpha}\sum_{k=1}^{n} n^{4\alpha-5} < c_2 n^{-4},$$

and

$$\sum_{k_1=1}^{n-1}\sum_{k_2=k_1+1}^{n-1} f_{2\alpha}(k_1,k_2)A_{n,k_2}\pi_{n,k_1}\pi_{n,k_2} < c_3 b_{n,4\alpha}\sum_{k_1=1}^{n-1}\sum_{k_2=k_1+1}^{n-1}\frac{b_{k_2,2\alpha}}{b_{k_1,2\alpha}b_{k_2,4\alpha}}\frac{1}{k_1^2 k_2^3}$$
$$< c_4 n^{-4\alpha}\sum_{k_2=2}^{n} k_2^{2\alpha-3}\sum_{k_1=1}^{k_2} k_1^{2\alpha-2} < c_4 n^{-4\alpha}\sum_{k_2=2}^{n} k_2^{4\alpha-4} < c_4 n^{-3}.$$

$\square$

PROOF OF THEOREM 2.1 (I) AND (II). Let $\alpha > 0.5$. To establish the stated normal approximation it is enough to prove the convergence in probability of the first two conditional moments

$$(\mu_n, \sigma_n^2) := \left(\sqrt{n}\,\mathrm{E}\left[\overline{Y}_n|\mathscr{Y}_n\right],\ n\,\mathrm{Var}\left[\overline{Y}_n|\mathscr{Y}_n\right]\right) \xrightarrow{P} \left(0, \frac{2\alpha+1}{2\alpha-1}\right), \quad n \to \infty,$$

since then, due to the conditional normality of $\overline{Y}_n$, we will get the following convergence of characteristic functions

$$\mathrm{E}\left[e^{i\gamma\sqrt{n}\cdot\overline{Y}_n}\right] = \mathrm{E}\left[e^{i\mu_n\gamma-\sigma_n^2\gamma^2/2}\right] \to e^{-\frac{2\alpha+1}{2(2\alpha-1)}\gamma^2}.$$

Now, due to Lemma 4.4 we can write

$$(\mu_n, \sigma_n^2) = \left(\sqrt{n}\delta e^{-\alpha U_n},\ 1+(n-1)\,\mathrm{E}\left[e^{-2\alpha\tau^{(n)}}|\mathscr{Y}_n\right] - ne^{-2\alpha U_n}\right).$$

Using relations from Section 4 we see that

$$\mathrm{E}\left[\sigma_n^2\right] = 1 - nb_{n,2\alpha} + \frac{2-(n+1)(2\alpha+1)b_{n,2\alpha}}{2\alpha-1} \to \frac{2\alpha+1}{2\alpha-1}.$$

It remains to observe that on one hand, according to Lemma 5.3

$$1+(n-1)\,\mathrm{E}\left[e^{-2\alpha\tau^{(n)}}|\mathscr{Y}_n\right] \xrightarrow{P} \frac{2\alpha+1}{2\alpha-1},$$

and on the other hand, $ne^{-2\alpha U_n} \xrightarrow{P} 0$, implying that $\sigma_n^2 \xrightarrow{P} \frac{2\alpha+1}{2\alpha-1}$. This together with $\mu_n \to 0$ holding in L^2 and therefore in probability, entails $(\mu_n, \sigma_n^2) \xrightarrow{P} (0, \frac{2\alpha+1}{2\alpha-1})$, finishing the proof of part (i). Part (ii) is proven similarly.

PROOF OF THEOREM 2.1 (III). Let $0 < \alpha < 0.5$. Turning to Lemma 5.2 observe that the martingale $H_n = (n+1)e^{(\alpha-1)U_n}\overline{Y}_n$ has uniformly bounded second moments. Indeed, due to Lemma 4.4

$$\begin{aligned}\mathrm{E}\left[H_n^2\right] &= (n+1)^2\,\mathrm{E}\left[e^{2(\alpha-1)U_n}\,\mathrm{E}\left[\overline{Y}_n^2|\mathscr{Y}_n\right]\right] \\ &< c_1 n\,\mathrm{E}\left[e^{-2(1-\alpha)U_n}\right] + c_2 n^2\,\mathrm{E}\left[e^{-2(1-\alpha)U_n-2\alpha\tau^{(n)}}\right] + c_3 n^2\,\mathrm{E}\left[e^{-2\alpha U_n}\right].\end{aligned}$$

Thus, according to Lemma 3.4 we have $\sup_n \mathrm{E}\left[H_n^2\right] < \infty$. Referring to the martingale L^2-convergence theorem we conclude that $H_n \to H_\infty$ almost surely and in L^2. Due to Lemma 5.1 it follows that

$$n^\alpha \overline{Y}_n = \frac{n^\alpha b_{n,\alpha-1}}{n+1} V_n^{(\alpha-1)} H_n \to V^{(\alpha-1)} H_\infty =: Y_{\alpha,\delta} \quad \text{a.s. and in } L^2.$$

Finally, as $n \to \infty$

$$\begin{aligned} n^\alpha \mathrm{E}\left[\overline{Y}_n\right] &= \delta n^\alpha b_{n,\alpha} \to \delta \Gamma(1+\alpha), \\ n^{2\alpha} \mathrm{E}\left[\overline{Y}_n^2\right] &= n^{2\alpha-1} + n^{2\alpha}(1-n^{-1}) \mathrm{E}\left[e^{-2\alpha\tau^{(n)}}\right] + n^{2\alpha}(\delta^2-1) \mathrm{E}\left[e^{-2\alpha U_n}\right] \\ &\to \left(\delta^2 + \frac{4\alpha}{1-2\alpha}\right) \Gamma(1+2\alpha). \end{aligned}$$

6 Consistency of the sample variance

Recall that $\mathrm{E}\left[S_n^2\right] = \frac{\sigma^2}{2\alpha} \mathrm{E}\left[D_n^2\right]$, and according to Lemma 4.3 we have $\mathrm{E}\left[D_n^2\right] \to 1$. The aim of this section is to show that $\mathrm{Var}\left[D_n^2\right] \to 0$ as $n \to \infty$ which is equivalent to

$$\mathrm{E}\left[D_n^4\right] \to 1, \quad n \to \infty. \tag{8}$$

To this end we will need the following formula, see Eq. (13) in [6] valid for any normally distributed vector (Z_1, Z_2, Z_3, Z_4) with means (m_1, m_2, m_3, m_4) and covariances $\mathrm{Cov}\left[Z_i, Z_j\right] = c_{ij}$:

$$\mathrm{Cov}\left[Z_1 Z_2, Z_3 Z_4\right] = m_1 m_3 c_{24} + m_1 m_4 c_{23} + m_2 m_3 c_{14} + m_2 m_4 c_{13} + c_{13} c_{24} + c_{14} c_{23}.$$

In the special case with $m_i = m$ it follows

$$\mathrm{E}\left[Z_1 Z_2 Z_3 Z_4\right] = m^4 + m^2(c_{12} + c_{13} + c_{14} + c_{23} + c_{24} + c_{34}) + c_{12} c_{34} + c_{13} c_{24} + c_{14} c_{23}. \tag{9}$$

Writing Y_i instead of $Y_i^{(n)}$ we use the representation

$$D_n^2 = \frac{n}{n-1} \left(\frac{1}{n} \sum_{i=1}^n Y_i^2 - \overline{Y}_n^2 \right) = \frac{1}{n} \sum_i Y_i^2 - \frac{2}{n(n-1)} \sum_i \sum_{j>i} Y_i Y_j$$

to find out that

$$\begin{aligned}\mathrm{E}\left[D_n^4\right] = &\frac{1}{n^2}\Big(\sum_i \mathrm{E}\left[Y_i^4\right] + 2\sum_i\sum_{j>i}\mathrm{E}\left[Y_i^2Y_j^2\right]\Big)\\
&-\frac{4}{n^2(n-1)}\Big(\sum_i\sum_{j>i}\mathrm{E}\left[Y_i^3Y_j\right]+\sum_i\sum_{j>i}\mathrm{E}\left[Y_iY_j^3\right]+\sum_i\sum_{j>i}\sum_{k\neq i,j}\mathrm{E}\left[Y_i^2Y_jY_k\right]\Big)\\
&+\frac{4}{n^2(n-1)^2}\Big(\sum_i\sum_{j>i}\mathrm{E}\left[Y_i^2Y_j^2\right]+\sum_i\sum_{j>i}\sum_{k\neq i,j}\mathrm{E}\left[Y_i^2Y_jY_k\right]+\sum_i\sum_{j>i}\sum_{k\neq i,j}\mathrm{E}\left[Y_iY_j^2Y_k\right]\\
&\qquad+\sum_i\sum_{j>i}\sum_{k\neq i,j}\sum_{m>k;\, m\neq i,j}\mathrm{E}\left[Y_iY_jY_kY_m\right]\Big).\end{aligned}$$

Denoting by (W_1,W_2,W_3,W_4) a random sample without replacement of four trait values out of n available, so that

$$\begin{aligned}\mathrm{E}\left[W_1^4\right] &= n^{-1}\sum_i \mathrm{E}\left[Y_i^4\right],\\
\mathrm{E}\left[W_1^3W_2\right] &= \frac{1}{n(n-1)}\sum_i\sum_{j\neq i}\mathrm{E}\left[Y_i^3Y_j\right],\\
\mathrm{E}\left[W_1^2W_2^2\right] &= \frac{1}{n(n-1)}\sum_i\sum_{j\neq i}\mathrm{E}\left[Y_i^2Y_j^2\right],\\
\mathrm{E}\left[W_1^2W_2W_3\right] &= \frac{1}{n(n-1)(n-2)}\sum_i\sum_{j\neq i}\sum_{k\neq i,j}\mathrm{E}\left[Y_i^2Y_jY_k\right],\\
\mathrm{E}\left[W_1W_2W_3W_4\right] &= \frac{1}{n(n-1)(n-2)(n-3)}\sum_i\sum_{j\neq i}\sum_{k\neq i,j}\sum_{m\neq i,j,k}\mathrm{E}\left[Y_iY_jY_kY_m\right],\end{aligned}$$

we derive

$$\begin{aligned}\mathrm{E}\left[D_n^4\right] = &n^{-1}\mathrm{E}\left[W_1^4\right] - 4n^{-1}\mathrm{E}\left[W_1^3W_2\right] + \frac{n^2-2n+3}{n(n-1)}\mathrm{E}\left[W_1^2W_2^2\right]\\
&-\frac{2(n-2)(n-3)}{n(n-1)}\mathrm{E}\left[W_1^2W_2W_3\right] + \frac{(n-2)(n-3)}{n(n-1)}\mathrm{E}\left[W_1W_2W_3W_4\right].\end{aligned} \tag{10}$$

We compute the five fourth-order moments in the last expression using the conditional normality of the random quadruple (W_1,W_2,W_3,W_4) with conditional moments given by

$$
\begin{aligned}
\mathrm{E}\left[W_i|\mathscr{Y}_n\right] &= \delta e^{-\alpha U_n},\\
\mathrm{E}\left[W_i^2|\mathscr{Y}_n\right] &= 1+(\delta^2-1)e^{-2\alpha U_n},\\
\mathrm{Var}\left[W_i|\mathscr{Y}_n\right] &= 1-e^{-2\alpha U_n},\\
\mathrm{Cov}\left[W_i,W_j|\mathscr{Y}_n\right] &= \mathrm{E}\left[e^{-2\alpha\tau_{ij}^{(n,4)}}|\mathscr{Y}_n\right]-e^{-2\alpha U_n}, \quad i,j\in\{1,2,3,4\}, \quad i\neq j,
\end{aligned}
$$

where $\tau_{ij}^{(n,m)}$ is the time to the most recent ancestor for the pair of tips (i,j) among m randomly chosen tips of the Yule n-tree. Clearly, all $\tau_{ij}^{(n,4)}$ have the same distribution as $\tau^{(n)}$, and for

$$
v_{ij}^{(n)} := \mathrm{E}\left[e^{-2\alpha\tau_{ij}^{(n,4)}}|\mathscr{Y}_n\right]
$$

we can find the asymptotics of

$$
\mathrm{E}\left[v_{ij}^{(n)}\right]=\mathrm{E}\left[e^{-2\alpha\tau^{(n)}}\right], \quad \mathrm{E}\left[v_{ij}^{(n)}e^{-2\alpha U_n}\right]=\mathrm{E}\left[e^{-2\alpha U_n-2\alpha\tau^{(n)}}\right]
$$

using Lemma 3.4. Notice also that

$$
\mathrm{E}\left[(v_{ij}^{(n)})^2\right]\sim\left(\mathrm{E}\left[e^{-2\alpha\tau^{(n)}}\right]\right)^2, \quad n\to\infty.
$$

This follows from Lemma 5.3 and Lemma 3.4 as

$$
\mathrm{E}\left[(v_{ij}^{(n)})^2\right]=\mathrm{E}\left[\left(\mathrm{E}\left[e^{-2\alpha\tau^{(n)}}|\mathscr{Y}_n\right]\right)^2\right]=\mathrm{Var}\left[\mathrm{E}\left[e^{-2\alpha\tau^{(n)}}|\mathscr{Y}_n\right]\right]+\left(\mathrm{E}\left[e^{-2\alpha\tau^{(n)}}\right]\right)^2.
$$

(i) With $Z_1=Z_2=Z_3=Z_4=W_1$ in Eq. (9), we obtain

$$
\mathrm{E}\left[W_1^4|\mathscr{Y}_n\right]=\delta^4e^{-4\alpha U_n}+6\delta^2e^{-2\alpha U_n}(1-e^{-2\alpha U_n})+3(1-e^{-2\alpha U_n})^2,
$$

and therefore $\mathrm{E}\left[W_1^4\right]\to 3$ as $n\to\infty$.

(ii) Using Eq. (9) with $Z_1=Z_2=Z_3=W_1$ and $Z_4=W_2$ we obtain

$$
\begin{aligned}
\mathrm{E}\left[W_1^3W_2|\mathscr{Y}_n\right] &= \delta^4e^{-4\alpha U_n}+3\delta^2e^{-2\alpha U_n}(1-e^{-2\alpha U_n})\\
&\quad +3\delta^2e^{-2\alpha U_n}(v_{12}^{(n)}-e^{-2\alpha U_n})+3(1-e^{-2\alpha U_n})(v_{12}^{(n)}-e^{-2\alpha U_n})\\
&= 3v_{12}^{(n)}-3(\delta^2-1)(1-v_{12}^{(n)})e^{-2\alpha U_n}+(\delta^4-6\delta^2+3)e^{-4\alpha U_n},
\end{aligned}
$$

resulting in $\mathrm{E}\left[W_1^3 W_2\right] \to 0$ as $n \to \infty$.

(iii) Eq. (9) with $Z_1 = Z_2 = W_1$ and $Z_3 = Z_4 = W_2$ gives

$$\begin{aligned}
\mathrm{E}\left[W_1^2 W_2^2 | \mathscr{Y}_n\right] &= \delta^4 e^{-4\alpha U_n} + 2\delta^2 e^{-2\alpha U_n}(1 - e^{-2\alpha U_n}) \\
&\quad + 4\delta^2 e^{-2\alpha U_n}(\nu_{12}^{(n)} - e^{-2\alpha U_n}) + (1 - e^{-2\alpha U_n})^2 + 2(\nu_{12}^{(n)} - e^{-2\alpha U_n})^2 \\
&= 1 + 2(\delta^2 - 1)e^{-2\alpha U_n} \\
&\quad + (\delta^4 - 6\delta^2 + 5)e^{-4\alpha U_n} + 4(\delta^2 - 1)\nu_{12}^{(n)} e^{-2\alpha U_n} + 2(\nu_{12}^{(n)})^2,
\end{aligned}$$

so that $\mathrm{E}\left[W_1^2 W_2^2\right] \to 1$ as $n \to \infty$.

(iv) Using a consequence of Eq. (9),

$$\mathrm{E}\left[Z_1^2 Z_2 Z_3\right] = m^4 + m^2(c_{11} + 2c_{12} + 2c_{13} + c_{23}) + c_{11}c_{23} + 2c_{12}c_{13},$$

we get

$$\begin{aligned}
\mathrm{E}\left[W_1^2 W_2 W_3 | \mathscr{Y}_n\right] &= \delta^4 e^{-4\alpha U_n} + \delta^2 e^{-2\alpha U_n}(1 - e^{-2\alpha U_n}) \\
&\quad + \delta^2 e^{-2\alpha U_n}(2\nu_{12}^{(n)} + 2\nu_{13}^{(n)} + \nu_{23}^{(n)} - 5e^{-2\alpha U_n}) \\
&\quad + (1 - e^{-2\alpha U_n})(\nu_{23}^{(n)} - e^{-2\alpha U_n}) + 2(\nu_{12}^{(n)} - e^{-2\alpha U_n})(\nu_{13}^{(n)} - e^{-2\alpha U_n}) \\
&= (\delta^2 - 1)e^{-2\alpha U_n} + (\delta^4 - 6\delta^2 + 3)e^{-4\alpha U_n} + 2\nu_{12}^{(n)}\nu_{13}^{(n)} \\
&\quad + 2(\delta^2 - 1)e^{-2\alpha U_n}(\nu_{12}^{(n)} + \nu_{13}^{(n)}) + (1 + (\delta^2 - 1)e^{-2\alpha U_n})\nu_{23}^{(n)}.
\end{aligned}$$

Using the Cauchy-Schwarz inequality

$$0 \leq \mathrm{E}\left[\nu_{12}^{(n)}\nu_{13}^{(n)}\right] \leq \left(\mathrm{E}\left[\nu_{12}^{(n)}\right]\right)^2 = \left(\mathrm{E}\left[e^{-2\alpha\tau^{(n)}}\right]\right)^2 = o(1),$$

we obtain $\mathrm{E}\left[W_1^2 W_2 W_3\right] \to 0$ as $n \to \infty$.

(v) According to Eq. (9) we have

$$\begin{aligned}
\mathrm{E}\left[W_1W_2W_3W_4|\mathscr{Y}_n\right] = {} & \delta^4 e^{-4\alpha U_n} \\
& + \delta^2 e^{-2\alpha U_n}(\nu_{12}^{(n)} + \nu_{13}^{(n)} + \nu_{14}^{(n)} + \nu_{23}^{(n)} + \nu_{24}^{(n)} + \nu_{34}^{(n)} - 6e^{-2\alpha U_n}) \\
& + (\nu_{12}^{(n)} - e^{-2\alpha U_n})(\nu_{34}^{(n)} - e^{-2\alpha U_n}) \\
& + (\nu_{13}^{(n)} - e^{-2\alpha U_n})(\nu_{24}^{(n)} - e^{-2\alpha U_n}) \\
& + (\nu_{14}^{(n)} - e^{-2\alpha U_n})(\nu_{23}^{(n)} - e^{-2\alpha U_n}),
\end{aligned}$$

implying

$$\begin{aligned}
\mathrm{E}\left[W_1W_2W_3W_4|\mathscr{Y}_n\right] = {} & (\delta^4 - 6\delta^2 + 3)e^{-4\alpha U_n} + \nu_{12}^{(n)}\nu_{34}^{(n)} + \nu_{12}^{(n)}\nu_{34}^{(n)} + \nu_{12}^{(n)}\nu_{34}^{(n)} \\
& + (\delta^2 - 1)e^{-2\alpha U_n}(\nu_{12}^{(n)} + \nu_{13}^{(n)} + \nu_{14}^{(n)} + \nu_{23}^{(n)} + \nu_{24}^{(n)} + \nu_{34}^{(n)}).
\end{aligned}$$

Using an estimate for $\mathrm{E}\left[\nu_{12}^{(n)}\nu_{34}^{(n)}\right] = \mathrm{E}\left[\nu_{12}^{(n)}\nu_{34}^{(n)}\right] = \mathrm{E}\left[\nu_{12}^{(n)}\nu_{34}^{(n)}\right]$ similar to that we used in (iv), we find $\mathrm{E}\left[W_1W_2W_3W_4\right] \to 0$ as $n \to \infty$.

Finally, putting the above results (i) - (v) into Eq. (10) we arrive at Eq. (8).

Acknowledgments

We are grateful to Thomas F. Hansen and Anna Stokowska for helpful suggestions and comments. Special thanks to an anonymous referee for the constructive suggestions to the earlier version of the paper, in particular, for the remark added after Lemma 5.1.

The research of Serik Sagitov was supported by the Swedish Research Council grant 621-2010-5623. Krzysztof Bartoszek was supported by the Centre for Theoretical Biology at the University of Gothenburg, Svenska Institutets Östersjösamarbete scholarship nr. 11142/2013, Stiftelsen för Vetenskaplig Forskning och Utbildning i Matematik (Foundation for Scientific Research and Education in Mathematics), Knut and Alice Wallenbergs travel fund, Paul and Marie Berghaus fund, the Royal Swedish Academy of Sciences, and Wilhelm and Martina Lundgrens research fund.

A All moments of U_n and $\tau^{(n)}$

Eq. (3) for the Laplace transforms of the random variable U_n can be used to calculate the moments of U_n using,

$$\mathrm{E}\left[U_n^m\right] = (-1)^m(\partial^m \mathrm{E}\left[e^{-xU_n}\right]/\partial x^m)|_{x=0}. \tag{11}$$

For a fixed n we introduce the following notation,

$$\begin{aligned}
A(x) &= \frac{1}{x+1}\cdot\ldots\cdot\frac{1}{x+n},\\
b_m(x) &= \frac{1}{(x+1)^m}+\ldots+\frac{1}{(x+n)^m},\\
\mathbf{b}_m(x) &= (b_1(x),\ldots,b_m(x)).
\end{aligned}$$

Notice that $A(0) = 1/n!$ and $b_m(0) = H_{n,m}$ is the n–th generalized harmonic number of order m,

$$H_{n,m} = \sum_{i=1}^{n}\frac{1}{i^m}. \tag{12}$$

We can write Eq. (3) as $\mathrm{E}\left[e^{-xU_n}\right] = n!A(x)$. Its first derivative with respect to x is $-n!A(x)b_1(x)$, and the second derivative is $n!A(x)(b_1(x)^2+b_2(x))$. For the general recursive formula we introduce the following notation. We will denote by $\mathbf{k} = (k_1,k_2,\ldots)$ infinite dimensional vectors with integer–valued components, and write $\mathbf{k}\in\mathscr{A}_m$ if all $k_i\geq 0$ and $|\mathbf{k}| := \sum_{i=1}^{m}k_i i = m$. Therefore $\mathscr{A}_m$ represents the set of all possible ways to represent m as a sum of positive integers. We will also use the multi–index notation $\mathbf{b}_m(x)^{\mathbf{k}} = b_1(x)^{k_1}\cdot\ldots\cdot b_m(x)^{k_m}$.

Since $A'(x) = -A(x)b_1(x)$, and $b'_m(x) = -mb_{m+1}(x)$, we can show by induction that,

$$\frac{\partial^m}{\partial x^m}\mathrm{E}\left[e^{-xU_n}\right] = (-1)^m n!A(x)\sum_{\mathbf{k}\in\mathscr{A}_m}c_{\mathbf{k}}\mathbf{b}_m(x)^{\mathbf{k}}, \tag{13}$$

where coefficients $c_{\mathbf{k}}$ are defined for all vectors $\mathbf{k} = (k_1,k_2,\ldots)$ with integer–valued components using the recursion,

$$c_{\mathbf{k}} = \sum_{j=0}^{m}(jk_j+1)c_{\mathbf{k},j}, \tag{14}$$

with $m = |\mathbf{k}|$ and

$$\begin{aligned}
c_{\mathbf{k},0} &= c_{(k_1-1,k_2,k_3,\dots)}, \\
c_{\mathbf{k},j} &= c_{(k_1,\dots,k_j+1,k_{j+1}-1,\dots)},\ j \geq 1.
\end{aligned}$$

The boundary conditions for the recursion of Eq. (14) consist of two parts:

- $c_{\mathbf{k}} = 0$, if all $k_i = 0$, or one of the coordinates of the vector $\mathbf{k}$ is negative,
- $c_{\mathbf{k}} = 1$ if $k_1 \geq 1$ and all other $k_i = 0$.

We conclude from Eq. (13) that,

$$\mathrm{E}\left[U_n^m\right] = \sum_{\mathbf{k}\in\mathscr{A}_m} c_{\mathbf{k}} \prod_{i=1}^{m} H_{n,i}^{k_i}.$$

The technique for calculating the m–th derivative of the Laplace transform of $\tau^{(n)}$ given by Eq. (5) is the same but requires new notation

$$\begin{aligned}
\hat{A}(y) &= \frac{1}{y-1}\cdot\frac{1}{y+2}\cdot\ldots\cdot\frac{1}{y+n}, \\
\hat{b}_m(y) &= \frac{1}{(y-1)^m} + \frac{1}{(y+2)^m} + \ldots + \frac{1}{(y+n)^m}.
\end{aligned}$$

Notice that $\hat{A}'(y) = -\hat{A}(y)\hat{b}_1(y)$, $\hat{b}_m'(y) = -m\hat{b}_{m+1}(y)$, $\hat{A}(0) = -n!$ and $\hat{b}_m(0) = H_{n,m}$ if m is even or $\hat{b}_m(0) = H_{n,m} - 2$ if m is odd. One can then inductively show that,

$$\begin{aligned}
\frac{\partial^m}{\partial y^m}\mathrm{E}\left[e^{-y\tau^{(n)}}\right] = \frac{(-1)^m 2m!}{(n-1)(y-1)^{m-1}} &- \frac{(-1)^{m+1}(n+1)!}{n-1}\hat{A}(y)(\hat{b}_1(y)^m \\
&+ \sum_{\substack{\mathbf{k}\in\mathscr{A}_m \\ k_1<m}} c_{\mathbf{k}}\hat{\mathbf{b}}_m(y)^{\mathbf{k}}),
\end{aligned}$$

with the coefficients $c_{\mathbf{k}}$ defined as previously by Eq. (14). Therefore, we get,

$$\mathrm{E}\left[\tau^{(n)^m}\right] = \frac{2m!}{n-1} - (H_{n,1}-2)^m + \sum_{\substack{\mathbf{k}\in\mathscr{A}_m \\ k_1<m}} c_{\mathbf{k}} \prod_{\substack{i=1 \\ i \text{ odd}}}^{m} (H_{n,i}-2)^{k_i} \prod_{\substack{i=1 \\ i \text{ even}}}^{m} H_{n,i}^{k_i}.$$

Similarly we can use Eq. (6) to calculate the joint moments for $U_n - \tau^{(n)}$ and $\tau^{(n)}$ in terms of,

$$\begin{aligned} A^{(i,j)}(x) &= \frac{1}{x+i+1} \cdot \ldots \cdot \frac{1}{x+j}, \\ b_m^{(i,j)}(x) &= \frac{1}{(x+i+1)^m} + \ldots + \frac{1}{(x+j)^m}. \end{aligned}$$

For $m \geq 1$ and $r \geq 1$ we first get,

$$\begin{aligned} \frac{\partial^{m+r}}{\partial x^m \partial y^r} \mathrm{E}\left[e^{-x(U_n-\tau^{(n)})-y\tau^{(n)}}\right] &= (-1)^{m+r} \frac{2(n+1)!}{n-1} \\ &\times \sum_{j=1}^{n-1} \frac{A^{(0,j)}(x)A^{(j,n)}(y)}{(j+1)(j+2)} \left(\sum_{\mathbf{k}\in\mathscr{A}_m} c_{\mathbf{k}} \mathbf{b}_m^{(0,j)}(x)^{\mathbf{k}}\right) \left(\sum_{\mathbf{k}\in\mathscr{A}_r} c_{\mathbf{k}} \mathbf{b}_r^{(j,n)}(y)^{\mathbf{k}}\right), \end{aligned}$$

and then from the above,

$$\begin{aligned} \mathrm{E}\left[(U_n - \tau^{(n)})^m \tau^{(n)^r}\right] &= (-1)^{m+r} \frac{2(n+1)}{n-1} \\ &\times \sum_{j=1}^{n-1} \frac{1}{(j+1)(j+2)} \left(\sum_{\mathbf{k}\in\mathscr{A}_m} c_{\mathbf{k}} \prod_{i=1}^{m} H_{j,i}^{k_i}\right) \left(\sum_{\mathbf{k}\in\mathscr{A}_r} c_{\mathbf{k}} \prod_{i=1}^{r} (H_{n,i} - H_{j,i})^{k_i}\right). \end{aligned}$$

References

[1] R. Adamczak and P. Miłoś. CLT for Ornstein–Uhlenbeck branching particle system. *ArXiv e-prints*, 2011.

[2] R. Adamczak and P. Miłoś. U–statistics of Ornstein–Uhlenbeck branching particle system. *J. Th. Pobab.*, in press.

[3] D. Aldous and L. Popovic. A critical branching process model for biodiversity. *Adv. Appl. Probab.*, 37(4):1094–1115, 2005.

[4] K. Bartoszek, J. Pienaar, P. Mostad, S. Andersson, and T. F. Hansen. A phylogenetic comparative method for studying multivariate adaptation. *J. Theor. Biol.*, 314:204–215, 2012.

[5] C. Boettiger, G. Coop, and P. Ralph. Is your phylogeny informative? Measuring the power of comparative methods. *Evolution*, 2012.

[6] G. W. Bohrnstedt and A. S. Goldberger. On the exact covariance of products of random variables. *J. Am. Stat. Assoc.*, 64:1439–1442, 1969.

[7] F. Bokma. Time, species and seperating their effects on trait variance in clades. *Syst. Biol.*, 59(5):602–607, 2010.

[8] M. A. Butler and A. A. King. Phylogenetic comparative analysis: a modelling approach for adaptive evolution. *Am. Nat.*, 164(6):683–695, 2004.

[9] F. W. Crawford and M. A. Suchard. Diversity, disparity, and evolutionary rate estimation for unresolved Yule trees. *Syst. Biol.*, 62(3):439–455, 2013.

[10] A. W. F. Edwards. Estimation of the branch points of a branching diffusion process. *J. Roy. Stat. Soc. B*, 32(2):155–174, 1970.

[11] J. Felsenstein. Phylogenies and the comparative method. *Am. Nat.*, 125(1): 1–15, 1985.

[12] T. Garland and A. R. Ives. Using the past to predict the present: Confidence intervals for regression equations in phylogenetic comparative methods. *Am. Nat.*, 155(3):346–364, 2000.

[13] T. Garland, P. E. Midford, and A. R. Ives. An introduction to phylogenetically based statistical methods, with a new method for confidence intervals on ancestral values. *Amer. Zool.*, 39:374–388, 1999.

[14] T. Gernhard. The conditioned reconstructed process. *J. Theor. Biol.*, 253: 769–778, 2008.

[15] T. Gernhard. New analytic results for speciation times in neutral models. *B. Math. Biol.*, 70:1082–1097, 2008.

[16] T. F. Hansen. Stabilizing selection and the comparative analysis of adaptation. *Evolution*, 51(5):1341–1351, 1997.

[17] T. F. Hansen, J. Pienaar, and S. H. Orzack. A comparative method for studying adaptation to a randomly evolving environment. *Evolution*, 62:1965–1977, 2008.

[18] J. P. Huelsenbeck and B. Rannala. Detecting correlation between characters in a comparative analysis with uncertain phylogeny. *Evolution*, 57(6):1237–1247, 2003.

[19] J.P. Huelsenbeck, B. Rannala, and J.P. Masly. Accommodating phylogenetic uncertainty in evolutionary studies. *Science*, 88:2349–2350, 2000.

[20] A. R. Ives, P. E. Midford, and T. Garland. Within–species variation and measurement error in phylogenetic comparative methods. *Syst. Biol.*, 56(2): 252–270, 2007.

[21] E. P. Martins and T. F. Hansen. Phylogenies and the comparative method: a general approach to incorporating phylogenetic information into the analysis of interspecific data. *Am. Nat.*, 149(4):1341–1351, 1997.

[22] A. Mooers, O. Gascuel, T. Stadler, H. Li, and M. Steel. Branch lengths on birth-death trees and the expected loss of phylogenetic diversity. *Syst. Biol.*, 61(2):195–203, 2012.

[23] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2013. URL `http://www.R-project.org`.

[24] F. J. Rohlf. Comparative methods for the analysis of continuous variables: geometric interpretations. *Evolution*, 55(11):2143–2160, 2001.

[25] F. J. Rohlf. A comment on phylogenetic correction. *Evolution*, 60(7):1509–1515, 2006.

[26] S. Sagitov and K. Bartoszek. Interspecies correlation for neutrally evolving traits. *J. Theor. Biol.*, 309:11–19, 2012.

[27] G. J. Slater, L. J. Harmon, D. Wegmann, P. Joyce, L. J. Revell, and M. E. Alfaro. Fitting models of continuous trait evolution to incompletely sampled comparative data using approximate bayesian computation. *Evolution*, 66 (3):752–762, 2012.

[28] T. Stadler. Lineages–through–time plots of neutral models for speciation. *Math. Biosci.*, 216:163–171, 2008.

[29] T. Stadler. On incomplete sampling under birth-death models and connections to the sampling-based coalescent. *J. Theor. Biol.*, 261(1):58–68, 2009.

[30] T. Stadler. Simulating trees with a fixed number of extant species. *Syst. Biol.*, 60(5):676–684, 2011.

[31] T. Stadler and M. Steel. Distribution of branch lengths and phylogenetic diversity under homogeneous speciation models. *J. Theor. Biol.*, 297:33–40, 2012.

[32] M. Steel and A. McKenzie. Properties of phylogenetic trees generated by Yule–type speciation models. *Math. Biosci.*, 170:91–112, 2001.

[33] E. A. Stone. Why the phylogenetic regression appears robust to tree misspecification. *Syst. Biol.*, 60(3):245–260, 2011.

[34] M. R. E. Symonds. The effects of topological inaccuracy in evolutionary trees on the phylogenetic comparative method of independent contrasts. *Syst. Biol.*, 51(4):541–553, 2002.

[35] G. U. Yule. A mathematical theory of evolution: based on the conclusions of Dr. J. C. Willis. *Philos. T. Roy. Soc. B*, 213:21–87, 1924.