



# **Modeling a trait-dependent diversification process coupled with molecular evolution on a random species tree**

Daniah Tahir, Sylvain Glémin, Martin Lascoux, Ingemar Kaj

## **► To cite this version:**

Daniah Tahir, Sylvain Glémin, Martin Lascoux, Ingemar Kaj. Modeling a trait-dependent diversification process coupled with molecular evolution on a random species tree. *Journal of Theoretical Biology*, 2019, 461, pp.189-203. <10.1016/j.jtbi.2018.10.032>. <hal-02370723>

**HAL Id: hal-02370723**

**<https://hal.science/hal-02370723v1>**

Submitted on 21 Nov 2019

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization

# Modeling a trait-dependent diversification process coupled with molecular evolution on a random species tree

Daniah Tahir<sup>a,\*</sup>, Sylvain Glémin<sup>b,c</sup>, Martin Lascoux<sup>b</sup>, Ingemar Kaj<sup>a</sup>

<sup>a</sup>Department of Mathematics, Uppsala University, Box 480, SE-751 06 Uppsala, Sweden

<sup>b</sup>Department of Plant Ecology and Evolution, Uppsala University, Norbyvägen 18D, SE-752 36 Uppsala, Sweden

<sup>c</sup>UMR 5554 ISEM (Université de Montpellier-CNRS-IRD-EPHE), FR-34095 Montpellier Cedex 5, France

---

## Abstract

Understanding the evolution of binary traits, which affects the birth and survival of species and also the rate of molecular evolution, remains challenging. In this work, we present a probabilistic modeling framework for binary trait, random species trees, in which the number of species and their traits are represented by an asymmetric, two-type, continuous time Markov branching process. The model involves a number of different parameters describing both character and molecular evolution on the so-called ‘reduced’ tree, consisting of only extant species at the time of observation. We expand our model by considering the impact of binary traits on  $dN/dS$ , the normalized ratio of nonsynonymous to synonymous substitutions. We also develop mechanisms which enable us to understand the substitution rates on a phylogenetic tree with regards to the observed traits. The properties obtained from the model are illustrated with a phylogeny of outcrossing and selfing plant species, which allows us to investigate not only the branching tree rates, but also the molecular rates and the intensity of selection.

**Keywords:** Branching processes, Irreversible transitions, Binary traits, Phylogenetic trees, Mutation rates

**MSC:** 60J85, 92D15

---

## 1. Introduction

One of the main challenges in modern evolutionary biology entails recognizing how different traits evolve on a phylogenetic tree, and how they can affect the evolution of other species characteristics. Of particular importance is the impact of life history or ecological traits on molecular evolutionary rates (Nikolaev et al., 2008; Smith and Donoghue, 2008; Mayrose and Otto, 2011; Figuet et al., 2016; Levy Karin et al., 2017). Among other things, comprehending the relationship between the two could help unravel the causes of molecular clock variation. Models have been developed to associate continuous or dis-

crete traits with molecular rates. For example, the program TraitRateProp (Levy Karin et al., 2017) allows the identification of sequence sites undergoing shifts in rate associated with phenotypic changes. Levy Karin et al. (2017) used TraitRateProp to identify shifts in site evolutionary rates within the RPS8 protein associated with transition to a heterotrophic lifestyle in orchids. However, existing methods do not allow for the phenotypic traits to also affect rates of speciation and extinction. A classical approach involves mapping traits on a phylogenetic tree, and then matching them with the evolutionary rates that have been measured on the corresponding branches. In such approaches, it is usually assumed that the phylogenetic tree is fixed, and is independent of the evolving traits. Yet, it is well known that some traits can affect the diversification process, and hence, the tree shape, which can bias the reconstruction of ancestral states as well as state transitions along the tree (Gold-

---

\*Corresponding author

Email addresses: [daniah.tahir@math.uu.se](mailto:daniah.tahir@math.uu.se) (Daniah Tahir),

[sylvain.glemin@univ-montp2.fr](mailto:sylvain.glemin@univ-montp2.fr) (Sylvain Glémin),

[martin.lascoux@ebc.uu.se](mailto:martin.lascoux@ebc.uu.se) (Martin Lascoux),

[ingemar.kaj@math.uu.se](mailto:ingemar.kaj@math.uu.se) (Ingemar Kaj)

berg and Igić, 2008).

A minimal model addressing the differences in speciation rates between groups, will need to consider both trait evolution and species diversification. The binary-state speciation and extinction (BiSSE) model (Maddison et al., 2007) was precisely developed for this purpose, and it also helped disentangle the two aspects. It was later extended to the ClaSSE (cladogenetic-state speciation and extinction) model (Goldberg and Igić, 2012). These models have been extensively used and extended to test for the effect of traits on species diversification (Mayrose et al., 2011; Beaulieu and O’Meara, 2016). However, to our knowledge, this approach has not been incorporated yet to assess for the possible effect of traits on molecular evolution.

Ideally, we would like to have a model which considers the impact of traits on not only the species diversification rates, but also on molecular evolutionary rates. A typical example where both approaches have been conducted independently, is the evolution of selfing from outcrossing in plant species, in the context of the so-called ‘dead-end’ hypothesis (Stebbins, 1957). The popular claim that the transition towards selfing is an evolutionary dead-end (Igić and Busch, 2013) relies on two assumptions: the transition from outcrossing to selfing is irreversible, and, selfing species have negative diversification rates (difference between speciation and extinction rates). In the short term, selfing offers reproductive assurance and transmission advantage over outcrossing (Fisher, 1941), since selfing species can contribute to outcross pollen while fertilizing their own ovules at the same time. Outcrossing also limits a plant’s ability to reproduce in case of rarity of mates. Therefore, transition from outcrossing to selfing occurs and is thought to be one of the most frequent evolutionary changes in angiosperms. However, once the outcrossing ability of a species is lost, it is difficult to regain due to, for instance, the accumulation of loss-of-function mutations on self-incompatibility genes (Igić et al., 2008). The main advantage of outcrossing is avoiding inbreeding depression, which is presumably purged by a species that reproduces through selfing for some time. In the long run, selfing species suffer from negative genetic consequences:

selfing reduces effective population sizes and recombination, which globally lessens the efficacy of selection. Selfing species are thus, more prone to the accumulation of deleterious mutations and less able to adapt to changing environments, which eventually drives them towards extinction (Wright et al., 2013). In agreement with this prediction and using the BiSSE model (Maddison et al., 2007), the net diversification rates are found to be higher in outcrossing than in selfing Solanaceae species (Goldberg et al., 2010). In parallel, the negative genetic effects of selfing, which can explain higher extinction rates, were tested by comparing the molecular evolutionary rates between selfing and outcrossing lineages. The efficacy of selection can be assessed through the ratio of nonsynonymous to synonymous substitutions,  $dN/dS$ ; a higher ratio corresponding to less efficient selection. Phylogenetic analyses have been carried out, but usually they either detect only weak effects, or fail at detecting any effect of mating system on  $dN/dS$  at all (Glémin and Muyle, 2014). A likely explanation for such results could be the recent origin of selfing, corresponding to a small fraction of the branch leading to a selfing species. Anyhow, the diversification processes underlying the observed trees, have not been incorporated so far, in such  $dN/dS$  analyses which implicitly assume a model where trait shifts occur at speciation times.

In this work, we examine, simultaneously, trait evolution, species diversification and molecular evolutionary rates. We first present a probabilistic modeling framework for a random, binary trait, species tree. The characterization of the tree properties then allows us to study a trait-dependent substitution process running along the branches of the tree. In particular, the branch lengths of the two trait components of the tree, are the key elements which we use to precisely determine the trait-dependent evolutionary rates. Similar models have been used in statistical inference, but, to our knowledge, their detailed mathematical properties have not been studied yet.

The edges of our tree model are grouped into two categories based on a generic trait. Over time, the trait has influenced and shaped the ancestral family tree of the extant species and led eventually to the observable mixture of traits associated with

the present tree branches. The model specifies expected values of various random functionals of the tree in terms of basic Markov chain parameters. Not only the expected number of species of each trait, but also the expected total branch lengths in the ancestral tree, with existing ancestors at the time of observation, are provided. While the tree model is constructed in forward time, it is the traits and mutations in the ancestral tree seen backwards from the extant set of species, that determine the current set of states. In the model, this shift of view corresponds to studying the reduced branching tree, obtained from the original species tree by the removal of extinct species. The model also allows us to study the accumulated number of mutations in relation to the observed distribution of traits. The rate of fixation, as the species accumulate mutations, depends on the trait value, due, for example, to a varying degree of selection associated with the traits.

Our analysis helps to understand the interplay between the evolution of a trait associated to rates of speciation and extinction on one hand, and molecular rates on the other. This theoretical framework should be useful for further developments of statistical methods combining trait-dependent diversification processes and trait-dependent molecular evolutionary rates.

## 2. Modeling the species tree

We consider evolution of binary traits on a species tree, starting with a single species of known trait at the root of the tree. Each species in the tree, throughout its lifetime, carries trait value 0 or 1. The ancestral species carries trait-0. The species family evolves as a branching tree with births of new species and extinction of existing species. The tree runs in continuous time over an interval  $[0, t]$ , where  $t$  is the time span from the time of origin of the first species at  $t = 0$ , up to the time of observation, i.e., today. The speciation process in the model is the simplest possible where new species arise instantly, either as a branch point or a transition from an earlier species. A speciation event could simply be replacement of a single species with two new ones of the same type ( $0 \rightarrow 0 + 0$  and  $1 \rightarrow 1 + 1$ ), or replacement of a single trait-0 species with a combination of

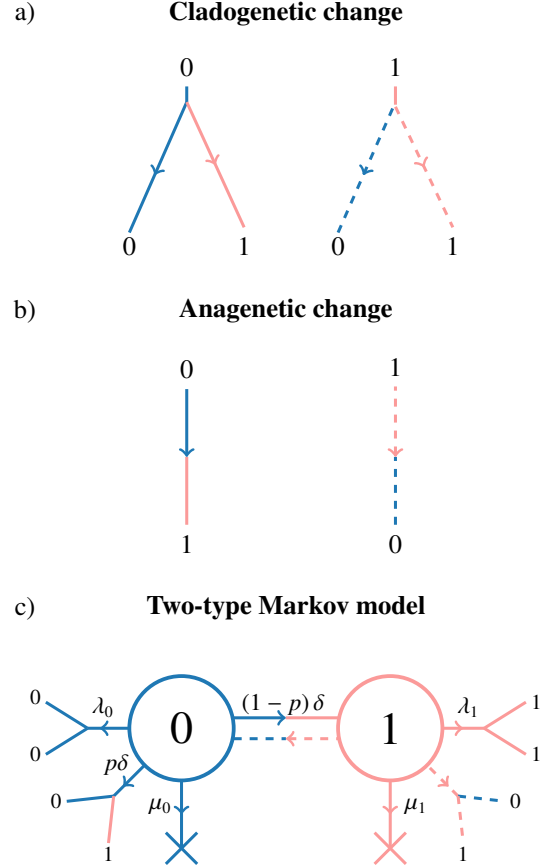


Figure 1: Diagrammatic representation of cladogenetic state change, anagenetic state change, and the Markov branching process. The 0 and 1 labels represent trait marks. In a), cladogenetic change  $0 \rightarrow 0 + 1$  is given by solid blue-red lines, and  $1 \rightarrow 0 + 1$  by dashed blue-red lines. In b), anagenetic change  $0 \rightarrow 1$  is given by a solid blue-red line, and  $1 \rightarrow 0$  by a dashed red-blue line. Panel c) depicts the two-type Markov model, as well as various rates of the branching process. Transitions from trait-1 to trait-0, represented by dashed lines in all three panels, are not considered further in this paper.

the two traits ( $0 \rightarrow 0 + 1$ ). The latter is called ‘cladogenetic’ state change and is represented by solid lines in Figure 1a. The state change could also be ‘anagenetic’ in nature, in which a single trait-0 species is replaced by a single trait-1 species along a branch ( $0 \rightarrow 1$ ), as represented by the solid line in Figure 1b. Moreover, it is assumed that cladogenetic and anagenetic state changes from trait-1 to trait-0 (dashed lines in Fig. 1a and Fig. 1b, respectively) do not occur.

The number of species and their traits change according to an integer-valued, two-type, continuous time Markov branching process, where we use the terms ‘type’ and ‘trait’ inter-

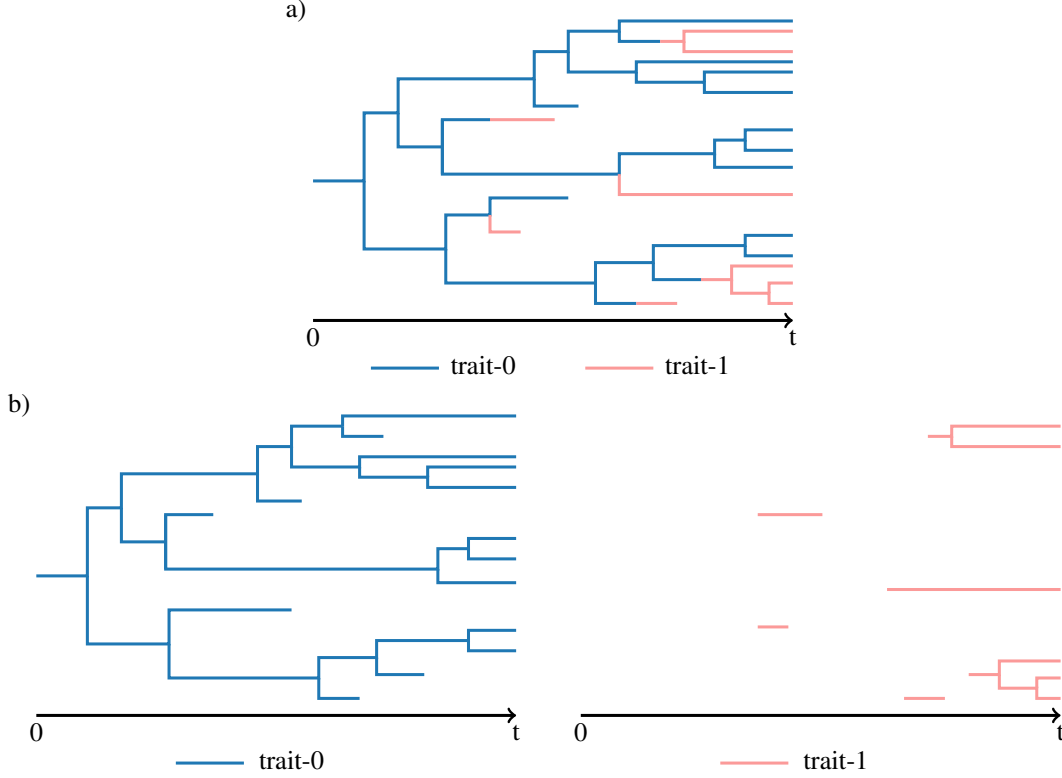


Figure 2: Diagram of a species tree  $\mathcal{T}$ . Panel a): rooted tree  $\mathcal{T} = \mathcal{T}^0 \cup \mathcal{T}^1$ , with trait-0 species colored blue and trait-1 species colored red. Panel b): rooted type-0 tree  $\mathcal{T}^0$ , and edges of type-1 tree  $\mathcal{T}^1$ , plotted separately in blue and red, respectively.

changeably (Athreya and Ney, 1972; Taylor and Karlin, 1984).

The Markov intensities for birth events are  $\lambda_0$  for a change  $0 \rightarrow 0 + 0$ ,  $\lambda_1$  for  $1 \rightarrow 1 + 1$ ,  $p\delta$  for  $0 \rightarrow 0 + 1$  and  $(1 - p)\delta$  for  $0 \rightarrow 1$ . Here, the rates  $\lambda_0$ ,  $\lambda_1$  and  $\delta$  are non-negative, whereas  $p$  ( $0 \leq p \leq 1$ ) is the probability of cladogenetic change of states from type-0 to type-1 species. Each species is exposed to extinction with intensities  $\mu_0$  for type-0 and  $\mu_1$  for type-1. This two-type Markov branching process is represented in Fig. 1c. The resulting species family is described by a rooted tree  $\mathcal{T}$ , with vertices given by the birth or extinction of species. The edges are marked by 0 or 1 recording the trait of a species and the edge lengths represent the species lifetime. The species tree is composed of disjoint parts

$$\mathcal{T} = \mathcal{T}^0 \cup \mathcal{T}^1,$$

where  $\mathcal{T}^0$  is a rooted type-0 tree and  $\mathcal{T}^1$  consists of all edges of type-1. Figure 2a shows such a tree  $\mathcal{T}$ , restricted to the interval  $[0, t]$ , with species of trait-0 plotted in blue and species of trait-1 plotted in red. Figure 2b shows  $\mathcal{T}^0$  and  $\mathcal{T}^1$  separately, again

cut off at time  $t$ .

Putting  $K_t$  = the number of type-0 species and  $L_t$  = the number of type-1 species,  $K_t + L_t$  is the total number of species at time  $t$ , and

$$X_t = (K_t, L_t), \quad t \geq 0, \quad (1)$$

is a two-type, continuous time Markov branching process, with branching rates

$$(k, \ell) \mapsto \begin{cases} (k + 1, \ell) & \lambda_0 k \\ (k - 1, \ell + 1) & (1 - p)\delta k \\ (k - 1, \ell) & \mu_0 k \\ (k, \ell + 1) & p\delta k + \lambda_1 \ell \\ (k, \ell - 1) & \mu_1 \ell. \end{cases} \quad (2)$$

Here,  $X_0 = (1, 0)$ , that is,  $K_0 = 1$  and  $L_0 = 0$ .  $K_t$ , the number of species in the sub-tree  $\mathcal{T}^0$ , is an ordinary one-type, continuous time branching process with parameters  $(\lambda, \mu)$ , where

$\lambda = \lambda_0$  is the rate of binary splitting of trait-0 species

Table 1: Description of parameters.

Parameter	Description
$K_t$	number of type-0 species
$L_t$	number of type-1 species
$\lambda_0 = \lambda$	speciation rate of type-0 species
$\lambda_1$	speciation rate of type-1 species
$\mu_0$	extinction rate of type-0 species
$\mu_1$	extinction rate of type-1 species
$p\delta$	rate of cladogenetic change to type-1 species
$(1-p)\delta$	rate of anagenetic change to type-1 species
$\delta$	total rate of state change to type-1 species
$\mu$	$\mu_0 + (1-p)\delta$
$r_0$	$\lambda_0 - \mu_0$ , diversification rate of type-0 species
$r_1 = \gamma_1$	$\lambda_1 - \mu_1$ , diversification rate of type-1 species
$\gamma_0$	$\lambda - \mu = r_0 - (1-p)\delta$
$\mathcal{T}$	complete species tree
$\mathcal{T}^0$	type-0 component in $\mathcal{T}$
$\mathcal{T}^1$	type-1 component in $\mathcal{T}$
$\mathcal{T}_t$	reduced species tree at time $t$
$(\mathcal{T}_t)^0$	type-0 component in $\mathcal{T}_t$
$(\mathcal{T}_t)^1$	type-1 component in $\mathcal{T}_t$
$(\mathcal{T}^0)_t$	type-0 species tree, reduced at time $t$
$(\mathcal{T}^1)_t$	type-1 species tree, reduced at time $t$
$A_t$	total branch length of $(\mathcal{T}^0)_t$
$A'_t$	total branch length of $(\mathcal{T}_t)^0$
$B_t$	total branch length of $(\mathcal{T}^1)_t = (\mathcal{T}_t)^1$
$T_{\text{tot}}$	branch length estimate of $\mathcal{T}_t$
$T_{\text{tot}}^1$	branch length estimate of $(\mathcal{T}_t)^1$
$\tilde{K}_{s,t}$	number of type-0 species in $(\mathcal{T}^0)_t$ at time $s$
$\tilde{K}'_{s,t}$	number of type-0 species in $(\mathcal{T}_t)^0$ at time $s$
$\tilde{L}_{s,t}$	number of type-1 species in $(\mathcal{T}_t)^1$ at time $s$
$\omega_0$	$dN/dS$ in type-0 species
$\omega_1$	$dN/dS$ in type-1 species
$\widehat{\omega}_0$	$dN/dS$ estimate in type-0 species
$\widehat{\omega}_1$	$dN/dS$ estimate in type-1 species
$\widehat{\omega}$	$dN/dS$ estimate over the whole species tree

and

$\mu = \mu_0 + (1-p)\delta$  is the rate of removal of trait-0 species.

The exact distribution of  $X_t$  is known for the pure anagenetic case ( $p = 0$ ) in terms of generating functions (Antal and Krapivsky, 2011). In the model, the net diversification rates for the two traits are  $r_0$  and  $r_1$  given as

$$r_0 = \lambda_0 - \mu_0 \quad \text{and} \quad r_1 = \lambda_1 - \mu_1.$$

The eigenvalues of the mean offspring matrix are given by

$$\gamma_0 = \lambda - \mu = r_0 - (1-p)\delta \quad \text{and} \quad \gamma_1 = \lambda_1 - \mu_1 = r_1.$$

A list of all important parameters used in this paper, is given in Table 1. Appendix A summarizes the mathematical properties of the two-type branching process  $X_t$ . The more general model of two-sided transitions, where type-1 species may generate species of type-0 at birth, is not discussed further in this work.

### 3. Analyzing the reduced species tree

The reduced species tree is what remains if we fix a time point and remove all branches of the species tree  $\mathcal{T}$ , that are extinct at that time. More formally, we denote by  $\mathcal{T}_t$ ,  $t \geq 0$ , the sequence of the reduced species trees defined for each fixed  $t$ .  $\mathcal{T}_t$  is pruned of any extinct species and hence composed of only those branches which exist at time  $t$  (Fig. 3a). In case all species are extinct at  $t$ , the reduced tree is empty. The species traits in the original tree provide a record of types for the reduced tree, and hence  $\mathcal{T}_t$  splits up into disjoint trees

$$\mathcal{T}_t = (\mathcal{T}_t)^0 \cup (\mathcal{T}_t)^1,$$

where  $(\mathcal{T}_t)^0$  is a single connected tree and  $(\mathcal{T}_t)^1$  may consist of several components (Fig. 3b). Moreover, we let  $(\mathcal{T}^0)_t$  and  $(\mathcal{T}^1)_t$  be the reduced trait-0 and trait-1 trees, respectively, obtained by first splitting the full tree  $\mathcal{T}$  into disjoint parts, and then reducing it at  $t$  (Fig. 3c). We observe, however, that

$$(\mathcal{T}_t)^0 \supseteq (\mathcal{T}^0)_t \quad \text{and} \quad (\mathcal{T}_t)^1 = (\mathcal{T}^1)_t,$$

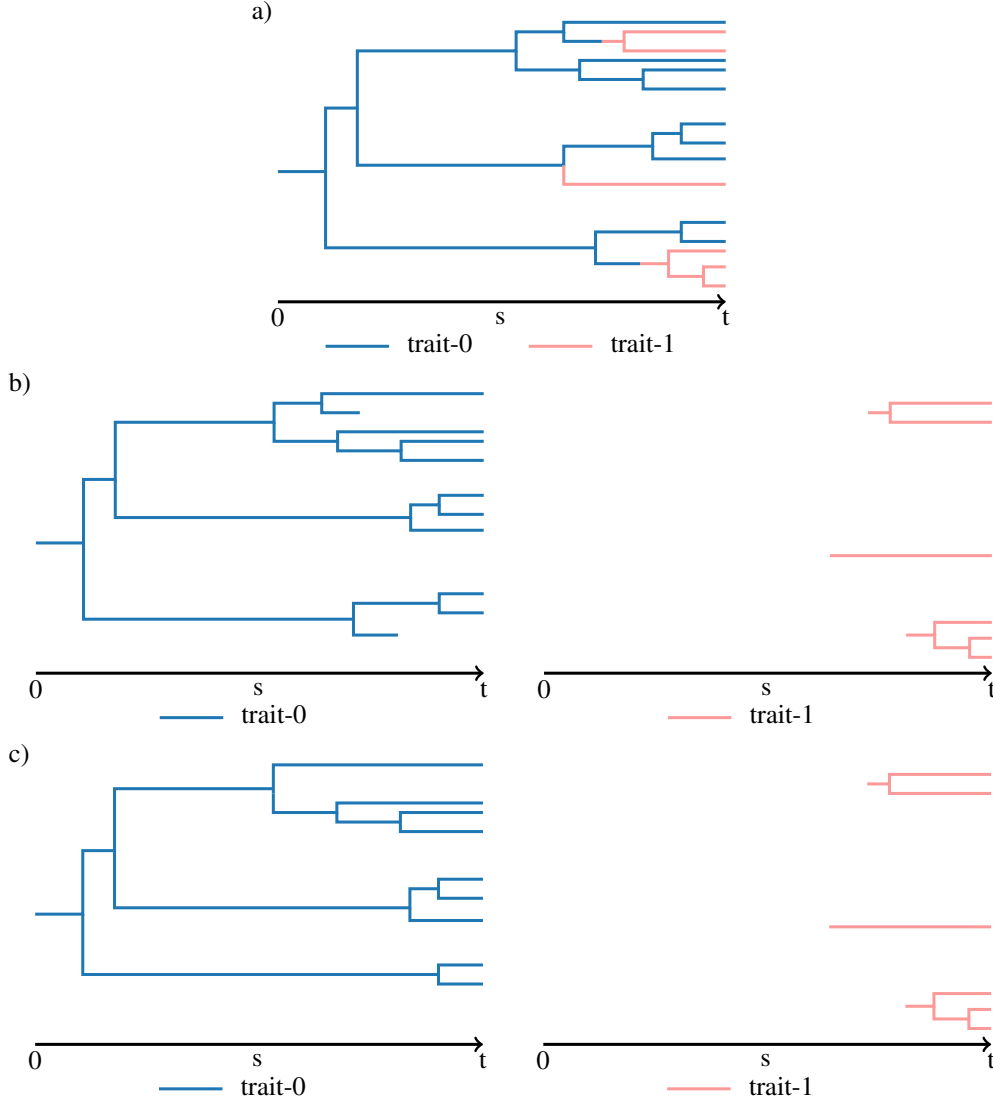


Figure 3: Diagrammatic representation of reduced species trees. Panel a) shows the reduced tree  $\mathcal{T}_t$ , given the original tree  $\mathcal{T}$  from Figure 2a. Trait-0 species are colored blue and trait-1 species are colored red. In panel b), the trees are obtained by splitting  $\mathcal{T}_t$  into disjoint parts, that is,  $(\mathcal{T}_t)^0$  and  $(\mathcal{T}_t)^1$ , which are plotted separately, in blue and red, respectively. Panel c) shows the rooted trait-0 tree  $(\mathcal{T}^0)_t$  in blue, and the trait-1 edges  $(\mathcal{T}^1)_t$  in red, which are obtained if the original tree  $\mathcal{T}$  (from Fig. 2a) is first split into type-0 and type-1 disjoint parts (as in Fig. 2b) and then reduced at  $t$ . In all three panels,  $t \geq 0$  and  $0 \leq s \leq t$ .

as shown in Figure 3b and Figure 3c. Indeed, for  $0 \leq s \leq t$ , letting  $\tilde{K}_{s,t}$  be the size at time  $s$  of the tree  $(\mathcal{T}^0)_t$  obtained by first splitting and then reducing the full tree, and  $\tilde{K}'_{s,t}$  be the size at  $s$  of the corresponding tree  $(\mathcal{T}_t)^0$  obtained by first reducing and then splitting, one has

$$\tilde{K}_{s,t} = \text{number of 0-species at } s \text{ with at least one descendant 0-species at } t,$$

$$\tilde{K}'_{s,t} = \text{number of 0-species at } s \text{ with at least one descendant 0- or 1-species at } t,$$

so that  $\tilde{K}_{s,t} \leq \tilde{K}'_{s,t}$ , where the assumption of one-sided transitions makes the inequality true. For trait-1 species, the order of splitting and reducing the tree has no effect and we write  $\tilde{L}_{s,t}$  for the total number of 1-species at time  $s$  in the reduced tree  $(\mathcal{T}_t)^1 = (\mathcal{T}^1)_t$ , consisting of the collection of reduced trees generated by any existing 1-species at  $t$ . This collection of branches could be empty, consist of a single reduced tree of siblings, or be composed of a cluster of disconnected trees with at least one descendant, each at time  $t$ . Clearly,  $\tilde{K}_{t,t} = K_t$  and  $\tilde{L}_{t,t} = L_t$ . When  $|X_t| = K_t + L_t = 0$ , the reduced tree is void and

$\widetilde{K}_{s,t} = \widetilde{L}_{s,t} = 0$ , where  $0 \leq s \leq t$ .

We now let

$$A_t = \int_0^t \widetilde{K}_{s,t} ds \quad (3)$$

be the total branch length of  $(\mathcal{T}^0)_t$ , and

$$A'_t = \int_0^t \widetilde{K}'_{s,t} ds \quad (4)$$

the total branch length of  $(\mathcal{T}_t)^0$ , so that  $A'_t \geq A_t$  almost surely. Similarly, the total branch length of  $(\mathcal{T}^1)_t = (\mathcal{T}_t)^1$  is given by

$$B_t = \int_0^t \widetilde{L}_{s,t} ds. \quad (5)$$

The purpose of the following sections is to find the expected values of  $A_t$  and  $B_t$ , and to relate these quantities to model parameters. Different approaches towards analyzing the reduced trait-0 and trait-1 species tree are reviewed and extended. The terminology of reduced trees used here, is well established in the theory of branching processes. Other options are reconstructed tree or reconstructed evolutionary process, being aware that statistically oriented phylogeneticists might use other kinds of reconstructed trees.

### 3.1. The reduced trait-0 species tree

In this section, we examine certain mathematical properties of the reduced trait-0 tree  $(\mathcal{T}^0)_t$ . In particular, we determine  $E(A_t|K_t > 0)$ , that is, the expected total branch length of  $(\mathcal{T}^0)_t$  conditional on non-extinction. To do this, we recall the conditional expectations of the branching process  $K_t$ , and the reduced branching process  $\widetilde{K}_{s,t}$  (see e.g., Kendall (1948) and Nee et al. (1994)). With  $\lambda = \lambda_0$  and  $\mu = \mu_0 + (1-p)\delta$  being the speciation and extinction rates, respectively, of the trait-0 species tree, and restricting to  $\gamma_0 = \lambda - \mu \neq 0$ , put

$$p_0(t) = \frac{\mu(1 - e^{-\gamma_0 t})}{\lambda - \mu e^{-\gamma_0 t}},$$

where  $p_0(t)$  denotes the probability of the tree going extinct before time  $t$ . We also let

$$u_t = \frac{\lambda p_0(t)}{\mu} = \frac{\lambda(1 - e^{-\gamma_0 t})}{\lambda - \mu e^{-\gamma_0 t}},$$

and

$$v_{s,t} = \frac{1 - p_0(t)}{1 - p_0(s)} u_s = \frac{\lambda(1 - e^{-\gamma_0 s})}{\lambda - \mu e^{-\gamma_0 t}}.$$

Then

$$E(K_t|K_t > 0) = \frac{1}{1 - u_t} = \frac{\lambda e^{\gamma_0 t} - \mu}{\gamma_0}, \quad (6)$$

and

$$E(K_t) = e^{\gamma_0 t}.$$

More generally, for  $s \leq t$ ,

$$\begin{aligned} E(K_s|K_t > 0) &= \frac{1}{1 - u_s} + \frac{u_s p_0(t-s)}{1 - u_s p_0(t-s)} \\ &= \frac{\lambda e^{\gamma_0 s} - \mu}{\gamma_0} + \frac{\lambda \mu (e^{\gamma_0 s} - 1)(e^{\gamma_0(t-s)} - 1)}{\gamma_0(\lambda e^{\gamma_0 t} - \mu)}, \end{aligned} \quad (7)$$

$$E(\widetilde{K}_{s,t}|K_t > 0) = \frac{1}{1 - v_{s,t}} = \frac{\lambda e^{\gamma_0 t} - \mu}{\lambda e^{\gamma_0(t-s)} - \mu}, \quad (8)$$

and

$$E(\widetilde{K}_{s,t}) = \frac{1 - p_0(t)}{1 - v_{s,t}} = \frac{\gamma_0 e^{\gamma_0 t}}{\lambda e^{\gamma_0(t-s)} - \mu}. \quad (9)$$

Thus, using (3) and (8)

$$\begin{aligned} E(A_t|K_t > 0) &= \int_0^t E(\widetilde{K}_{s,t}|K_t > 0) ds \\ &= \int_0^t \frac{\lambda e^{\gamma_0 t} - \mu}{\lambda e^{\gamma_0(t-s)} - \mu} ds. \end{aligned} \quad (10)$$

Evaluating the above integral, we obtain

$$E(A_t|K_t > 0) = \frac{\mu - \lambda e^{\gamma_0 t}}{\mu \gamma_0} \log \left( \frac{\lambda - \mu}{\lambda - \mu e^{-\gamma_0 t}} \right).$$

The relation in (10) simplifies for the critical case  $\gamma_0 = \lambda - \mu = 0$ , for example

$$E(A_t|K_t > 0) = \int_0^t \frac{1 + \lambda t}{1 + \lambda(t-s)} ds = \frac{1 + \lambda t}{\lambda} \log(1 + \lambda t).$$

Moreover, we can also describe properties of the trait-0 species tree conditioned on observing a fixed number of species at time  $t$  ( $K_t = n$ ), where the single-type linear branching process  $(K_t)_{t \geq 0}$ , is now restricted to the critical or supercritical case,  $\gamma_0 = \lambda - \mu \geq 0$ . Given  $n$  trait-0 species at time  $t$ , the  $n-1$  bifurcation times,  $S_1, \dots, S_{n-1}$ , are the time intervals from the tips of the tree at  $t$  backwards until two species merge. The bifurcation times of the complete species tree are identical to the bifurcation times of the reduced tree, given  $K_t = n$ . It turns out that



the joint distribution of the bifurcation times is the same as that of  $n - 1$  i.i.d. observations sampled from a particular family of distribution functions  $F_t(s)$ ,  $0 \leq s \leq t$ , depending on  $\lambda$  and  $\mu$  (Thompson, 1975; Gernhard, 2008). For the supercritical case  $\lambda > \mu$ ,

$$F_t(s) = \frac{\lambda - \mu e^{-\gamma_0 t}}{1 - e^{-\gamma_0 t}} \frac{1 - e^{-\gamma_0 s}}{\lambda - \mu e^{-\gamma_0 s}}, \quad 0 \leq s \leq t. \quad (11)$$

Writing  $S_{(k)}$ ,  $1 \leq k \leq n - 1$ , for the ordered bifurcation times of the reduced type-0 tree and adding  $S_{(0)} = 0$  and  $S_{(n)} = t$ , so that  $0 = S_{(0)} \leq S_{(1)} \leq \dots \leq S_{(n-1)} \leq S_{(n)} = t$ , it follows that the total branch length  $A_t$ , in (3), has the representation

$$\begin{aligned} A_t &= n(S_{(1)} - S_{(0)}) + (n - 1)(S_{(2)} - S_{(1)}) + \dots \\ &\quad + 1 \cdot (S_{(n)} - S_{(n-1)}) \\ &= S_{(1)} + S_{(2)} + \dots + S_{(n-1)} + t \\ &= S_1 + S_2 + \dots + S_{n-1} + t. \end{aligned}$$

Hence,

$$E(A_t | K_t = n) = (n - 1)E[S] + t,$$

where

$$E[S] = \int_0^t (1 - F_t(s)) ds,$$

and  $F_t(s)$  is obtained from (11).

### 3.2. The reduced trait-1 species tree

The trait-1 species tree is composed of a collection of branches, which are injected at random times and locations on top of the initially existing trait-0 tree, to obtain the complete species tree. The total intensity at which species of trait-1 enter the tree at any time  $s \geq 0$ , is  $\delta K_s$ , hence proportional to the current number of trait-0 species  $K_s$ , in the system. A new trait-1 species is the result of cladogenetic splitting with probability  $p$  and of anagenetic transition with probability  $1 - p$ . Each new 1-species potentially initiates a sub-tree, which preserves its trait during the subsequent path to extinction or supercritical growth. Let  $L_t^s$ ,  $s \leq t$ , denote the branching process with initial time  $s$ , with  $L_s^s = 1$ , which counts the number of type-1's at  $t$  originating from a new type-1 at  $s$ . Then, the total number of trait-1 species at  $t$  is a random sum

$$L_t = \sum_{i: s_i \leq t} L_t^{s_i}, \quad t \geq 0, \quad (12)$$

where  $L_t^{s_i}$ ,  $i \geq 1$ , are independent copies of the type-1 branching process. Under cladogenetic splitting, the process  $K_t$  is independent of the number of 1-species and (12) is a Poisson sum representation of  $L_t$ . The dynamics of an anagenetic transition is more involved, as  $K_t$  decreases by one whenever a type-1 species enters the tree, and (12) is a self-regulating process rather than a Poisson process. In both cases, however, the expected number of 1-species at  $t$  is

$$E_0(L_t) = \delta \int_0^t E_0(K_s) E_1(L_t^s) ds, \quad t \geq 0,$$

where  $E_0$  is the expectation starting from one species of trait-0 and  $E_1$  is the expectation given an initial species of trait-1. Similarly,

$$E_0(L_t | K_t > 0) = \delta \int_0^t E_0(K_s | K_t > 0) E_1(L_t^s) ds, \quad (13)$$

where

$$E_1(L_t^s) = e^{\gamma_1(t-s)}, \quad (14)$$

and, using (7),

$$E_0(K_s | K_t > 0) = \frac{\lambda e^{\gamma_0 s} - \mu}{\gamma_0} + \frac{\lambda \mu (e^{\gamma_0 s} - 1)(e^{\gamma_0(t-s)} - 1)}{\gamma_0(\lambda e^{\gamma_0 t} - \mu)}.$$

Using (5), and keeping the condition of at least one 0-species at  $t$ , the expected branch length of trait-1 species equals

$$E_0(B_t | K_t > 0) = \int_0^t E_0(\tilde{L}_{s,t} | K_t > 0) ds, \quad (15)$$

where  $\tilde{L}_{s,t}$  is a summation of contributing reduced branching processes  $\tilde{L}_{s,t}^u$ ,  $u \leq s \leq t$ , with  $L_{u,t}^u = 1$ , which originate from some point of the non-reduced, trait-0 tree at time  $u$ . Hence,

$$E_0(\tilde{L}_{s,t} | K_t > 0) = \delta \int_0^s E_0(K_u | K_t > 0) E_1(\tilde{L}_{s,t}^u) du.$$

Using the above equation in (15) yields

$$E_0(B_t | K_t > 0) = \delta \int_0^t \int_0^s E_0(K_u | K_t > 0) E_1(\tilde{L}_{s,t}^u) du ds, \quad (16)$$

where  $E_0(K_u | K_t > 0)$  is given in (7) and  $E_1(\tilde{L}_{s,t}^u)$  is obtained as in (9), by replacing  $\lambda$ ,  $\mu$ , and  $\gamma_0$  by  $\lambda_1$ ,  $\mu_1$ , and  $\gamma_1$ , respectively, that is,

$$E_1(\widetilde{L}_{s,t}^\mu) = \frac{1 - p_0(t-u)}{1 - v_{s-u,t-u}} = \frac{\gamma_1 e^{\gamma_1(t-u)}}{\lambda_1 e^{\gamma_1(t-s)} - \mu_1}.$$

From (16), it can be seen that the expected trait-1 branch length depends not only on the expected number of trait-0 species conditioned on non-extinction, but also on the expected number of trait-1 species already present in the tree. This equation serves as an important tool for evaluating the trait-1 branch lengths, and will be essential later on in studying  $dN/dS$  over the species tree.

### 3.3. The number of trait-1 species clusters

Let  $C_t$  be the number of separate clusters of trait-1 species at time  $t$ , that is, the number of sub-trees in  $\mathcal{T}^1$  at  $t$ . Clearly,  $1 \leq C_t \leq L_t$ . The expected number of clusters is obtained by modifying (13) as

$$E_0(C_t|K_t > 0) = \delta \int_0^t E_0(K_s|K_t > 0) P_1(L_t^s > 0) ds, \quad (17)$$

where  $E_0(K_s|K_t > 0)$  is given in (7) and

$$P_1(L_t^s > 0) = \frac{\gamma_1}{\lambda_1 - \mu_1 e^{-\gamma_1(t-s)}}. \quad (18)$$

Now, by (14) and (18), it can be seen that for  $0 \leq \lambda_1 \leq \mu_1$ ,

$$1 \leq \frac{E_1(L_t^s)}{P_1(L_t^s > 0)} \leq \frac{\mu_1 - \lambda_1 e^{-(\mu_1 - \lambda_1)t}}{\mu_1 - \lambda_1}.$$

Hence, the ratio of the expected number of trait-1 species to the expected number of trait-1 clusters satisfies

$$1 \leq \frac{E(L_t|K_t > 0)}{E(C_t|K_t > 0)} \leq \frac{\mu_1 - \lambda_1 e^{-(\mu_1 - \lambda_1)t}}{\mu_1 - \lambda_1}, \quad (19)$$

for each value of  $p$  and  $\delta$ . It is straightforward to verify that

$$1 = \frac{E(L_t|K_t > 0)}{E(C_t|K_t > 0)} \quad \text{if and only if} \quad \lambda_1 = 0. \quad (20)$$

A conclusion of (20) is that if we observe a phylogenetic tree where each species of trait-1 at  $t$  forms its own singleton cluster with no other trait-1 species as a closest neighbor (hence,  $C_t = L_t$ ), it is natural to make the parameter estimation  $\lambda_1 = 0$ . In Appendix B, we have derived further estimates for the case when  $\lambda_1 = 0$ , under certain special conditions on the species tree.

### 3.4. An illustration using arbitrary parameters

To illustrate the species tree model, we first notice that the trait-0 tree, corresponding to the blue colored sub-tree  $\mathcal{T}^0$  in Figure 2b, depends only on the parameters  $\lambda$  and  $\mu$ . In particular, the expected number of such species at  $t$ , given at least one existing species, is given by (6) as

$$E(K_t|K_t > 0) = \frac{\lambda e^{\gamma_0 t} - \mu}{\gamma_0},$$

and the expected branch length of the reduced tree, which is the blue-colored subtree  $(\mathcal{T}^0)_t$  in Figure 3c, is obtained in (10) as

$$E(A_t|K_t > 0) = \int_0^t \frac{\lambda e^{\gamma_0 t} - \mu}{\lambda e^{\gamma_0(t-s)} - \mu} ds, \quad \gamma_0 = \lambda - \mu.$$

In these expressions, however, the extinction parameter  $\mu = \mu_0 + (1-p)\delta$  consists of trait-0 extinction at rate  $\mu_0$ , and trait-0 to trait-1 anagenetic transition at rate  $(1-p)\delta$ . As we now move on and derive expected values for the trait-1 species tree, the resulting expressions are stated as functions of  $\lambda, \mu, \delta, \lambda_1$  and  $\mu_1$ , so that the dependence of  $\mu_0, p$  and  $\delta$  arises in  $\mu$ . Starting with the expected number of trait-1 species at  $t$ , given at least one existing trait-0 species, corresponding to the number of tips in the red colored sub-tree  $\mathcal{T}^1$  in Figure 2b, we have by (13),

$$E(L_t|K_t > 0) = \delta \int_0^t \left( \frac{\lambda e^{\gamma_0 s} - \mu}{\gamma_0} + \frac{\lambda \mu (e^{\gamma_0 s} - 1)(e^{\gamma_0(t-s)} - 1)}{\gamma_0 (\lambda e^{\gamma_0 t} - \mu)} \right) e^{\gamma_1(t-s)} ds,$$

$\gamma_1 = \lambda_1 - \mu_1$ . The expected trait-1 branch length, which is consistent with the branch length of the red colored sub-tree  $(\mathcal{T}^1)_t$  in Figure 3c, is obtained in (16) as

$$E(B_t|K_t > 0) = \delta \int_0^t \int_0^s \left( \frac{\lambda e^{\gamma_0 u} - \mu}{\gamma_0} + \frac{\lambda \mu (e^{\gamma_0 u} - 1)(e^{\gamma_0(t-u)} - 1)}{\gamma_0 (\lambda e^{\gamma_0 t} - \mu)} \right) \times \frac{\gamma_1 e^{\gamma_1(t-u)}}{\lambda_1 e^{\gamma_1(t-s)} - \mu_1} du ds.$$

Finally, the expected number of clusters of trait-1 species, corresponding to the red-colored subtree  $(\mathcal{T}^1)_t$  in Figure 3c, is derived in (17) as

$$E(C_t|K_t > 0) = \delta \int_0^t \left( \frac{\lambda e^{\gamma_0 s} - \mu}{\gamma_0} + \frac{\lambda \mu (e^{\gamma_0 s} - 1)(e^{\gamma_0(t-s)} - 1)}{\gamma_0 (\lambda e^{\gamma_0 t} - \mu)} \right) \times \frac{\gamma_1}{\lambda_1 - \mu_1 e^{-\gamma_1(t-s)}} ds.$$

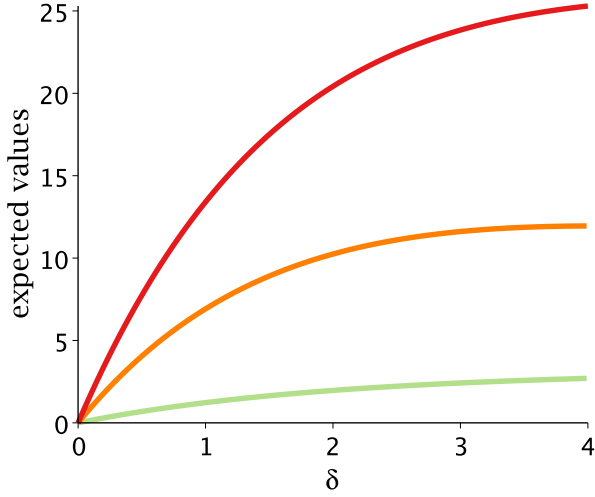


Figure 4: Plots of expected values versus  $\delta$ ;  $E(L_t|K_t > 0)$  versus  $\delta$  in red,  $E(C_t|K_t > 0)$  versus  $\delta$  in orange, and  $E(B_t|K_t > 0)$  versus  $\delta$  in green, when  $t = 1$ ,  $p = 0.5$ ,  $\lambda = \lambda_0 = 8$ ,  $\mu_0 = 4$ ,  $\lambda_1 = 10$ ,  $\mu_1 = 12$  and  $\mu = \mu_0 + (1 - p)\delta = 4 + 0.5\delta$ .

Let us take a fixed value of the probability  $p$  of cladogenetic state change. If we assume that point estimates of the parameters  $\lambda_0$  and  $\mu_0$  are known, this means that the trait-0 speciation rate  $\lambda = \lambda_0$  is known, whereas the trait-0 extinction rate  $\mu = \mu_0 + (1 - p)\delta$  is given as a function of  $\delta$ . From this view point, for a set of arbitrary parameters, Figure 4 shows plots of the three trait-1 functionals listed above, illustrating the dependence on the trait transition intensity  $\delta$ .

Next, we consider the number of trait-1 species per trait-1 cluster and use the shorthand notation  $L/C$  to denote the ratio  $E(L_t|K_t > 0)/E(C_t|K_t > 0)$ . Using the same set of parameters as in Figure 4, we now plot  $L/C$  as a function of  $p$  and  $\delta$ . Plots of  $L/C$  as a function of  $\delta$ , for fixed values of  $p$ , are shown in Figure 5a. Similarly, plots of  $L/C$  versus  $p$ , for selected values of  $\delta$ , are given in Figure 5b. As shown earlier in (19), re-written below for convenience

$$1 \leq \frac{L}{C} \leq \frac{\mu_1 - \lambda_1 e^{-(\mu_1 - \lambda_1)t}}{\mu_1 - \lambda_1},$$

it can be seen that for any value of  $p$  and  $\delta$ , the expected number of trait-1 species per cluster lies within a given bound. The upper bound is conservative, and for a wide range of arbitrary parameter values,  $L/C$  remains closer to the lower bound, meaning that it is unlikely to observe large clades of trait-1 species.

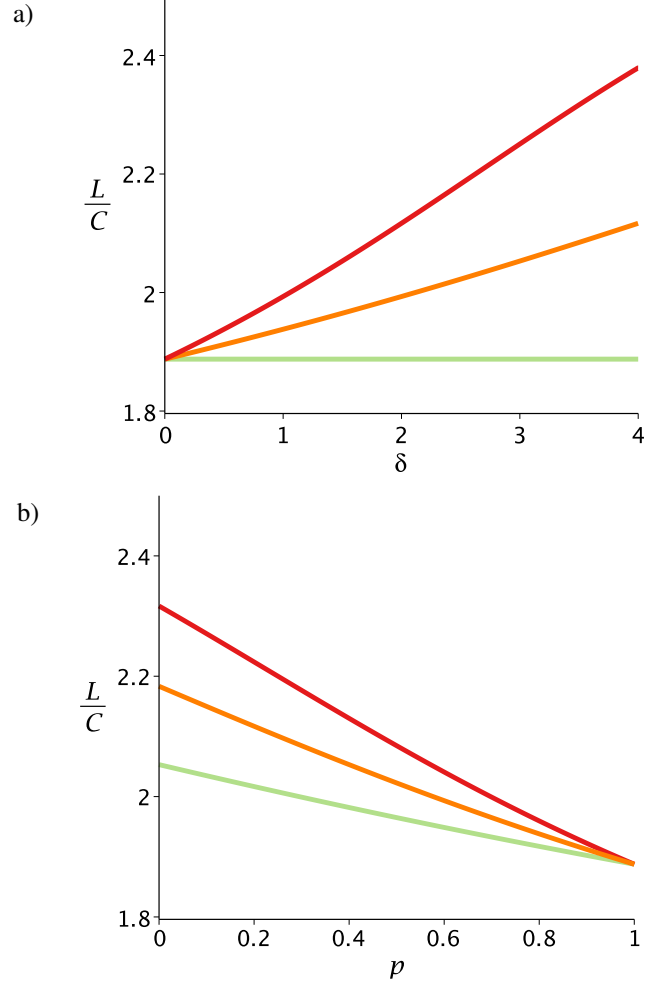


Figure 5: Plots of  $L/C = E(L_t|K_t > 0)/E(C_t|K_t > 0)$  versus  $\delta$  and  $p$ . Plot of  $L/C$  versus  $\delta$ , for  $p = 0$  in red,  $p = 0.5$  in orange and  $p = 1$  in green, is shown in a), while plot of  $L/C$  versus  $p$ , for  $\delta = 1.5$  in green,  $\delta = 2.5$  in orange and  $\delta = 3.5$  in red, is given in b). In both a) and b),  $t = 1$ ,  $\lambda = \lambda_0 = 8$ ,  $\mu_0 = 4$ ,  $\lambda_1 = 10$ ,  $\mu_1 = 12$  and  $\mu = 4 + (1 - p)\delta$ .

#### 4. Mutations on the species tree

Mutation events occur randomly, according to a fixed Poisson molecular clock of evolution, and the resulting Poisson intensity  $\theta > 0$  of mutations per time unit, is the same for all species. The total mutation intensity splits into two contributions,  $\theta = \theta_{\text{syn}} + \theta_{\text{non}}$ , where  $\theta_{\text{syn}}$  represents synonymous mutations, while  $\theta_{\text{non}}$  represents nonsynonymous mutations. The precise fractions  $\theta_{\text{syn}}$  and  $\theta_{\text{non}}$  can be obtained from a detailed mutation model (Mugal et al., 2014), or estimated from data. The actual marks of mutation, such as nucleotide substitutions, codon substitutions, etc., will be called ‘fixed mutations’ for

short. The rate of fixed mutations depends on the trait. Indeed, letting  $\omega_0$  and  $\omega_1$  be the trait-dependent scaled fixation rates, the fixed mutations accumulate as a Poisson process running along all branches of the species tree. Synonymous substitutions build up neutrally over the whole tree at scaled rate  $\theta_{\text{syn}}$ , while the substitution rate of nonsynonymous mutations is reduced to  $\theta_{\text{non}}\omega_0$  for trait-0 species and to  $\theta_{\text{non}}\omega_1$  for trait-1 species where

$$\omega_0 \leq \omega_1 < 1. \quad (21)$$

The ordering in (21) indicates that both traits are under negative selection, with the efficacy of selection in removing deleterious mutations higher in trait-0 species.

#### 4.1. Application to $dN/dS$

The  $dN/dS$ -ratio measures the normalized ratio of nonsynonymous to synonymous substitutions on the species tree. If we observe a single trait-0 species known to exist over a fixed time duration  $t$ , the expected number of nonsynonymous and synonymous substitutions is  $\theta_{\text{non}}\omega_0 t$  and  $\theta_{\text{syn}}t$ , respectively, and we understand the normalized ratio to be simply  $\omega_0$ . Similarly, for a species known to carry trait-1 over a fixed time interval, the corresponding ratio is  $\omega_1$ .

In our context, it is natural to associate  $dN/dS$  with the average number of substitutions that occurred anywhere in the species family tree and are observable today. Substitutions observable today must have occurred on the reduced tree. Of course, a meaningful  $dN/dS$  concept is naturally conditioned on survival of some species today. To capture the accumulation of fixed mutations on the species tree in more detail, we run independent Poisson processes along branches of the tree. First, a collection of Poisson points  $\mathcal{S}$  with intensity  $\theta_{\text{syn}}$ , representing synonymous substitutions, is placed on top of the entire tree. Next, nonsynonymous substitutions are generated by a Poisson measure  $\mathcal{N}^{(0)}$  with intensity  $\omega_0\theta_{\text{non}}$  along all branches carrying trait-0, and by a measure  $\mathcal{N}^{(1)}$  with intensity  $\omega_1\theta_{\text{non}}$  along the branches representing trait-1. Starting with the trait-0 tree reduced at  $t$ ,  $(\mathcal{T}^0)_t$ , the random variable  $\mathcal{S}_t^0 = \mathcal{S}((\mathcal{T}^0)_t)$  counts the

number of synonymous substitutions and  $\mathcal{N}_t^0 = \mathcal{N}^{(0)}((\mathcal{T}^0)_t)$  the number of nonsynonymous substitutions in trait-0 species existing at  $t$ . Thus,  $(\mathcal{S}_t^0)_{t \geq 0}$  and  $(\mathcal{N}_t^0)_{t \geq 0}$  are Poisson processes modulated by the stochastic intensities  $\theta_{\text{syn}}A_t$  and  $\theta_{\text{non}}\omega_0A_t$ , where  $A_t$ , introduced in (3), is the total branch length of  $(\mathcal{T}^0)_t$ . An empirical investigation would proceed from here by estimating  $\theta_{\text{non}}$  and  $\theta_{\text{syn}}$  and comparing observed values of  $\mathcal{N}_t^0$  and  $\mathcal{S}_t^0$  to obtain a  $dN/dS$ -ratio for  $\omega_0$  as

$$\widehat{\omega}_0 = \frac{\mathcal{N}_t^0/\theta_{\text{non}}}{\mathcal{S}_t^0/\theta_{\text{syn}}}. \quad (22)$$

Next, when we apply  $\mathcal{S}$  and  $\mathcal{N}^{(i)}$  to the entire reduced species tree at  $t$ ,  $\mathcal{T}_t = (\mathcal{T}_t)^0 \cup (\mathcal{T}_t)^1$ , we obtain Poisson processes

$\mathcal{S}_t = \#$  accumulated synonymous substitutions

in all existing species at  $t$ ,

$\mathcal{N}_t^{(i)} = \#$  accumulated nonsynonymous substitutions

in trait- $i$  up to time  $t$ ,  $i = 0, 1$ ,  $t \geq 0$ .

Here,  $\mathcal{S}_t$  is modulated by the stochastic intensity  $\theta_{\text{syn}}(A'_t + B_t)$ , where  $A'_t$  and  $B_t$ , introduced in (4) and (5), respectively, are the total lengths of trait-0 and trait-1 branches in the reduced tree  $\mathcal{T}_t$ . Indeed, conditionally, given  $t$  and given  $A'_t$  and  $B_t$ ,  $\mathcal{S}_t$  has a Poisson distribution with mean  $\theta_{\text{syn}}(A'_t + B_t)$ . Similarly,  $\mathcal{N}_t^{(i)}$  are Poisson processes modulated by the random intensities  $\theta_{\text{non}}\omega_0 A'_t$  for  $i = 0$  and  $\theta_{\text{non}}\omega_1 B_t$  for  $i = 1$ .

Putting these tools together, we consider the  $dN/dS$ -ratio

$$\begin{aligned} dN/dS &\sim \frac{(E(\mathcal{N}_t^{(0)}|K_t > 0) + E(\mathcal{N}_t^{(1)}|K_t > 0))/\theta_{\text{non}}}{E(\mathcal{S}_t|K_t > 0)/\theta_{\text{syn}}} \\ &= \frac{\omega_0 E(A'_t|K_t > 0) + \omega_1 E(B_t|K_t > 0)}{E(A'_t + B_t|K_t > 0)}. \end{aligned}$$

The conditioning scheme of assuming at least one trait-0 species at  $t$ ,  $K_t > 0$ , is, to some degree, arbitrary. Alternatives, such as assuming  $K_t = k$  or imposing a condition involving both  $K_t$  and  $L_t$ , are equally natural. Our choice is computationally more convenient and hence we define

$$dN/dS|_t = \frac{\omega_0 E(A'_t|K_t > 0) + \omega_1 E(B_t|K_t > 0)}{E(A'_t|K_t > 0) + E(B_t|K_t > 0)}. \quad (23)$$

Because of the appearance of  $A'_t$ , it is not immediately clear as to how we can compute this expression, even when the model

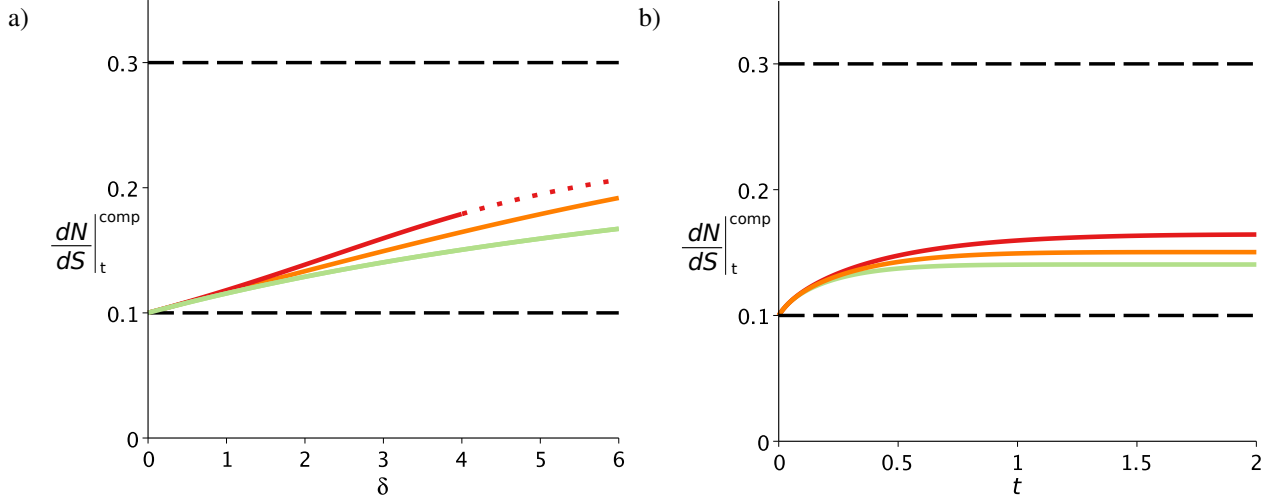


Figure 6: Plots of  $dN/dS|_t^{\text{comp}}$  versus  $\delta$  in a) and  $dN/dS|_t^{\text{comp}}$  versus  $t$  in b), for selected values of  $p$ . In a),  $t = 1$ ,  $\omega_0 = 0.1$ ,  $\omega_1 = 0.3$ ,  $\lambda_0 = 8$ ,  $\mu_0 = 4$ ,  $\lambda_1 = 10$ ,  $\mu_1 = 12$ , while  $\mu = \mu_0 + (1 - p)\delta$  varies with  $p$  and  $\delta$ . The dotted red curve gives those values of  $dN/dS|_1^{\text{comp}}$  which are obtained for a sub-critical process ( $\gamma_0 = \lambda - \mu < 0$ ). In b),  $\delta = 3$ ,  $\omega_0 = 0.1$ ,  $\omega_1 = 0.3$ ,  $\lambda_0 = 8$ ,  $\mu_0 = 4$ ,  $\lambda_1 = 10$ ,  $\mu_1 = 12$ , while  $\mu = \mu_0 + (1 - p)\delta$  depends on the value of  $p$ . In both panels, the red curve is obtained for  $p = 0$ , orange for  $p = 0.5$ , green for  $p = 1$ , and the dashed black lines represent the value of  $\omega_0$  and  $\omega_1$ .

parameters are known. A possible approach is replacing  $A'_t$  by the branch length  $A_t \leq A'_t$  of the trait-0 reduced tree  $(\mathcal{T}^0)_t$ , in both numerator and denominator of (23), and obtain the approximation

$$dN/dS|_t^{\text{comp}} = \frac{\omega_0 E(A_t|K_t > 0) + \omega_1 E(B_t|K_t > 0)}{E(A_t|K_t > 0) + E(B_t|K_t > 0)},$$

where  $E(A_t|K_t > 0)$  and  $E(B_t|K_t > 0)$  are functions of the model parameters derived in (10) and (16), respectively. For  $t = 1$ , fixed mutation rates  $\omega_0$  and  $\omega_1$ , and arbitrary values of parameters  $\lambda_0$ ,  $\mu_0$ ,  $\lambda_1$  and  $\mu_1$ , Figure 6a illustrates the shape of  $dN/dS|_t^{\text{comp}}$  versus  $\delta$ , for selected values of  $p$ . On the other hand, if  $\delta$  is fixed,  $dN/dS|_t^{\text{comp}}$  as a function of divergence time  $t$ , increases from  $\omega_0$  at  $t = 0$  to a limiting value  $dN/dS|_\infty^{\text{comp}}$ , as shown in Figure 6b. By analyzing (10) and (16) in more detail it can be seen that

$$dN/dS|_\infty^{\text{comp}} = \frac{D}{1 + D} \omega_0 + \frac{1}{1 + D} \omega_1, \quad (24)$$

where

$$D = \frac{\gamma_0(\gamma_0 - \gamma_1)}{\delta\gamma_1} \frac{\int_0^\infty \frac{1}{\lambda e^{\gamma_0 s} - \mu} ds}{\int_0^\infty \frac{1}{\lambda_1 e^{\gamma_0 s} - \mu_1 e^{(\gamma_0 - \gamma_1)s}} ds}.$$

While Figure 6 gives an idea of the dynamics of  $dN/dS$  for known parameter values, we now describe a similar, but slightly

different approximation approach that can be obtained from a given estimated tree. Assuming that the model parameters have already been estimated from the tree, we would like to avoid the replacement of  $A'$  with  $A$  in (23). In addition, we require estimates of  $\omega_0$  and  $\omega_1$ . An estimate of  $\omega_0$  on a species tree can be found fairly easily by making use of existing methods, but it is difficult to obtain a value for  $\omega_1$ , since we do not know exactly when transition from trait-0 to trait-1 takes place along a branch. In order to achieve this, suppose that we have an estimate of the total branch length  $T_{\text{tot}}$ , of the whole species tree, which is naturally associated with the denominator of (23), i.e.,

$$T_{\text{tot}} = E(A'_t|K_t > 0) + E(B_t|K_t > 0).$$

The analytical expression (16), giving the expected total branch length of trait-1 species tree, that is,

$$E(B_t|K_t > 0) = \delta \int_0^t \int_0^s \left( \frac{\lambda e^{\gamma_0 u} - \mu}{\gamma_0} + \frac{\lambda \mu (e^{\gamma_0 u} - 1)(e^{\gamma_0(t-u)} - 1)}{\gamma_0(\lambda e^{\gamma_0 t} - \mu)} \right) \times \frac{\gamma_1 e^{\gamma_1(t-u)}}{\lambda_1 e^{\gamma_1(t-s)} - \mu_1} du ds,$$

yields an estimate  $T_{\text{tot}}^1 = E(B_t|K_t > 0)$ , and hence we obtain from (23),

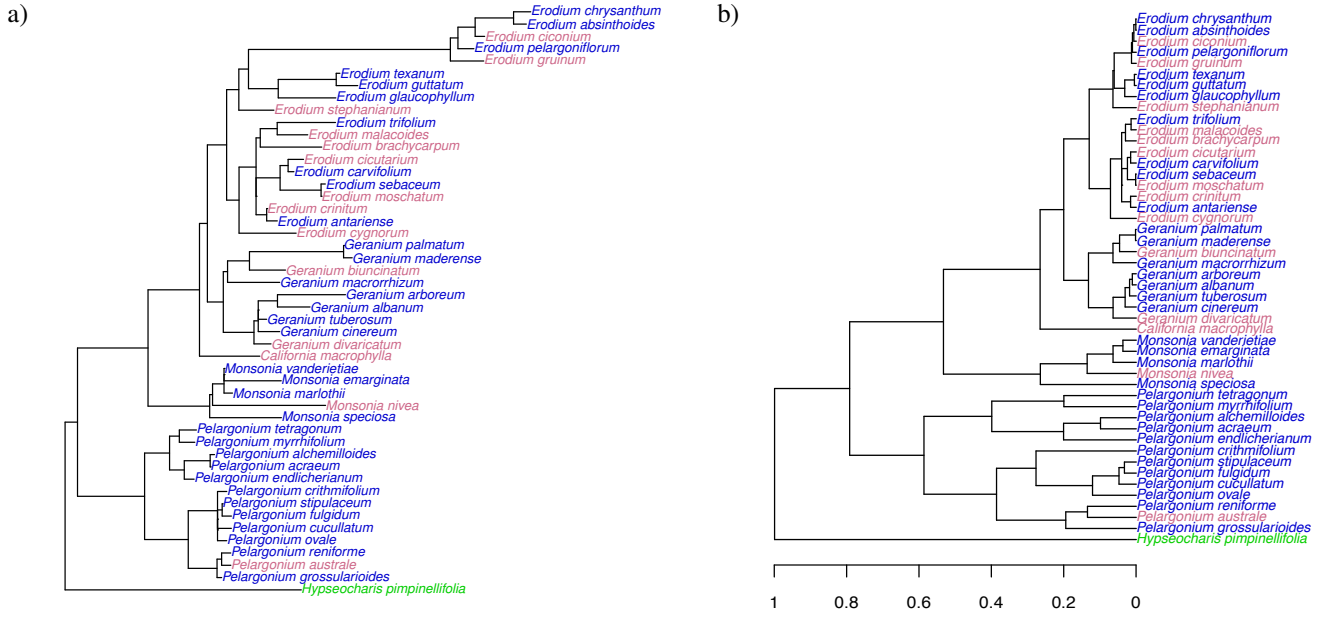


Figure 7: Phylogenetic trees. Panel a) gives the tree consisting of 47 species in the Geraniaceae family and 1 outgroup species, while in panel b), ultrametric version of the tree in a) is shown. In both a) and b), outcrossing species are colored in blue, selfing species in red, and the outgroup species in green.

$$dN/dS|_t \sim \left(1 - \frac{T_{\text{tot}}^1}{T_{\text{tot}}}\right) \omega_0 + \frac{T_{\text{tot}}^1}{T_{\text{tot}}} \omega_1.$$

Also available from the tree, as indicated in (22), is an estimate  $\widehat{\omega}_0$  of  $\omega_0$ . By applying the technique in (22) to the entire species tree, one can similarly obtain an estimate  $\widehat{\omega}$  of  $dN/dS|_t$ . Then, the above equation becomes

$$\widehat{\omega} = \left(1 - \frac{T_{\text{tot}}^1}{T_{\text{tot}}}\right) \widehat{\omega}_0 + \frac{T_{\text{tot}}^1}{T_{\text{tot}}} \omega_1,$$

and therefore by rearranging the above equation, we derive an estimate of  $\omega_1$  as

$$\widehat{\omega}_1 = \widehat{\omega}_0 + (\widehat{\omega} - \widehat{\omega}_0) \frac{T_{\text{tot}}}{T_{\text{tot}}^1}. \quad (25)$$

This relation shows that incorporating information from the underlying diversification process can help estimating  $\omega_1$ , without having to replace  $A'$  with  $A$  in (23). To implement (25) in practice, we require values for  $T_{\text{tot}}$  and  $T_{\text{tot}}^1$ , the estimates  $\widehat{\omega}_0$  and  $\widehat{\omega}$ , and diversification parameters of the tree. In the next section, we demonstrate the use of (25) by applying it to a real data set. While the estimates of  $\omega_0$  and  $\omega$ , as well as tree parameters, are obtained from already established techniques, the accuracy of the estimates of  $T_{\text{tot}}$  and  $T_{\text{tot}}^1$  are tested through simulations, and the results are summarized in Appendix C.

#### 4.2. An example: outcrossing and selfing plant species

In this section, we analyze a set of outcrossing and selfing plant species belonging to the Geraniaceae family. In doing so, we also illustrate how the diversification process can affect the  $dN/dS$  estimates. The outcrossing species, or outcrossers, are the ancestral species and represent trait-0, while the selfing species, or selfers, represent trait-1. For a total of  $m = 48$  species, a sequence alignment of chloroplastic genes, consisting of 1425 codons, is obtained from Glémin and Muyle (2014). Out of the 48 species,  $k = 33$  are outcrossers,  $\ell = 14$  are selfers, and one species serves as the outgroup. For all the species, the phylogeny analysis software PhyML (Guindon and Gascuel, 2003), with nucleotide substitution model K80, is used to reconstruct the phylogenetic tree and obtain corresponding branch lengths. This tree is given in Figure 7a. The one-ratio model in the software PAML (Yang, 1997), with 3xF4 option for codon frequencies, is used to obtain  $\widehat{\omega} = 0.095$ , an estimate of the global  $dN/dS$  ratio over the tree (excluding the outgroup). The outcrosser species sub-tree, is then obtained by removing the outgroup and all selfing branches from the initial tree in Figure 7a. This sub-tree and the corresponding outcross-

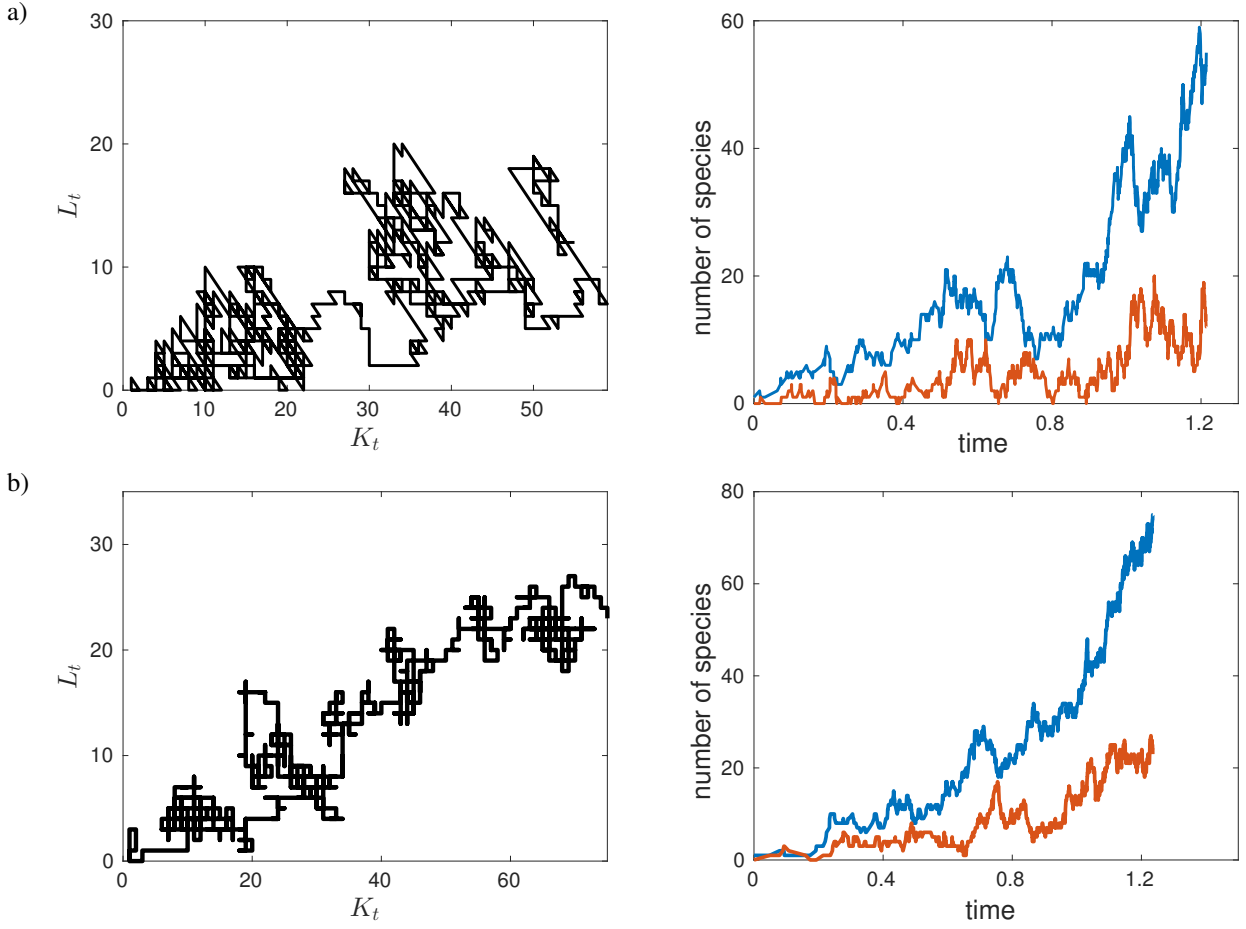


Figure 8: Simulations of the two-type branching process  $X_t$ . In a), simulations were carried out with BiSSE model parameters ( $\lambda_1 = 0$ ,  $\lambda_0 = 17.601$ ,  $p = 0$ ,  $\delta = 17.322$ ,  $\mu_0 = 0$ ,  $\mu_1 = 53.716$ ), while in b), ClaSSE model parameters ( $\lambda_1 = 0$ ,  $\lambda_0 = 11.761$ ,  $p = 1$ ,  $\delta = 8.581$ ,  $\mu_0 = 10.406$ ,  $\mu_1 = 19.494$ ) were used. Plots in the left panel represent the trace of  $(K_t, L_t)$  in the  $(k, \ell)$  plane, while plots in the right panel show the paths over time of  $K_t$  and  $L_t$  in blue and red, respectively.

ing sequence data yields the  $dN/dS$ -value, essentially an estimate of  $\omega_0$ , as  $\widehat{\omega}_0 = 0.075$ . As before, we use the one-ratio model in PAML to obtain the estimate for  $\omega_0$ . The observed ordering  $\widehat{\omega}_0 < \widehat{\omega}$  is consistent with our basic hypothesis in (21) that  $\omega_0 \leq \omega_1$ . The tree in Figure 7a is then made ultrametric using the package ‘ape’ (Paradis et al., 2004) in the software R (R Core Team, 2016), and normalized to have  $t = 1$ . This ultrametric tree is given in Figure 7b. The total branch length of the ultrametric tree, excluding the outgroup, is  $T_{\text{tot}} = 6.926$ .

To see which values of  $\omega_1$  would be consistent with the observations, we proceed by estimating parameters on the tree by maximum likelihood techniques. First of all, we consider the case when  $p = 0$  by fitting a BiSSE model (Maddison et al., 2007) to the ultrametric tree in Figure 7b, using the package ‘di-

versitree’ (FitzJohn, 2012) in R. The parameters of the BiSSE model are constrained to prohibit transitions from selfing to outcrossing. This gives  $\lambda = \lambda_0 = 17.601$ ,  $\lambda_1 = 0$ ,  $\mu_0 = 0$ ,  $\mu_1 = 53.716$ ,  $\delta = 17.322$  and  $\mu = \mu_0 + (1 - p)\delta = 17.322$ . The corresponding two-type branching process is supercritical with extinction probability  $\mu/\lambda = 0.984$ . The left and right panels in Figure 8a show a fairly typical realization of the branching process, conditional on non-extinction. Using the parameter values in (16) along with  $t = 1$ , we obtain an estimate of the total selfer branch length as  $T_{\text{tot}}^1 = E(B_1 | K_1 > 0) = 0.126$ . Since the values for  $T_{\text{tot}}$  and  $T_{\text{tot}}^1$  are now known, and the estimates  $\widehat{\omega} = 0.095$  and  $\widehat{\omega}_0 = 0.075$  have been obtained earlier as well, we can derive an estimate of  $\omega_1$  using (25), that is

$$\widehat{\omega}_1 = \widehat{\omega}_0 + (\widehat{\omega} - \widehat{\omega}_0) \frac{T_{\text{tot}}}{T_{\text{tot}}^1} = 1.174.$$

The general view regarding the role of selection in removing deleterious mutations suggests that  $\omega_1$  should be less than 1, as assumed in (21). Obtaining  $\omega_1 > 1$  in the above equation violates (21), and is indicative of a pure anagenetic ( $p = 0$ ) model being unsuitable to describe this particular family of species. Also, the difference between the maximum likelihood estimates of  $\lambda$  and  $\mu$  is very small ( $\gamma_0 = \lambda - \mu = 0.279$ ), suggesting that since the outcrossing parameters are bordering on the critical case, it will be very unlikely to observe a tree in this case.

Now, if we go to the other extreme and assume as much selfing in the tree as possible, i.e., a pure cladogenetic model, we can obtain an estimate of  $\omega_1$ . Using the classical approach of a two-ratio branch model in PAML, we get  $\omega_0 = 0.079$  (which is consistent with our estimate  $\widehat{\omega}_0 = 0.075$ ) and  $\omega_1 = 0.167$ . Since the values of  $\omega_0$  and  $\omega_1$  are quite close to one another, it seems that the strength of selection is almost the same for both traits. Moreover, in this scenario, we can neither determine the branching tree rates, nor the corresponding value of the parameter  $p$ . Here, we would also like to stress on the fact that assuming a pure cladogenetic case ( $p = 1$ ) for the full species tree does not imply that the reduced tree is also cladogenetic (as assumed by PAML) in our model, as illustrated in Figure 9.

Coming back to our example tree in Figure 7b, we now try to obtain intermediate values of the parameters by fitting a ClaSSE (Goldberg and Igić, 2012) model to the ultrametric tree, constrained in manner so that  $p = 1$  and only one-sided transitions from outcrossing to selfing are allowed. The ClaSSE model, incorporated in the package ‘diversitree’ (FitzJohn, 2012) in R, gives the maximum likelihood estimates  $\lambda_0 = 11.761$ ,  $\lambda_1 = 0$ ,  $\mu = \mu_0 = 10.406$ ,  $\mu_1 = 19.494$  and  $\delta = 8.581$ . In this case, the corresponding two-type branching process is supercritical ( $\gamma_0 = \lambda - \mu = 1.355 > 0$ ) with extinction probability  $\mu/\lambda = 0.885$ . The left and right panels in Figure 8b show a typical realization of this branching process. Using the parameter values in (16) with  $t = 1$ , we obtain  $T_{\text{tot}}^1 = 0.510$  and from (25), a suitable estimate of  $\omega_1$  as

$$\widehat{\omega}_1 = \widehat{\omega}_0 + (\widehat{\omega} - \widehat{\omega}_0) \frac{T_{\text{tot}}^1}{T_{\text{tot}}^1} = 0.347.$$

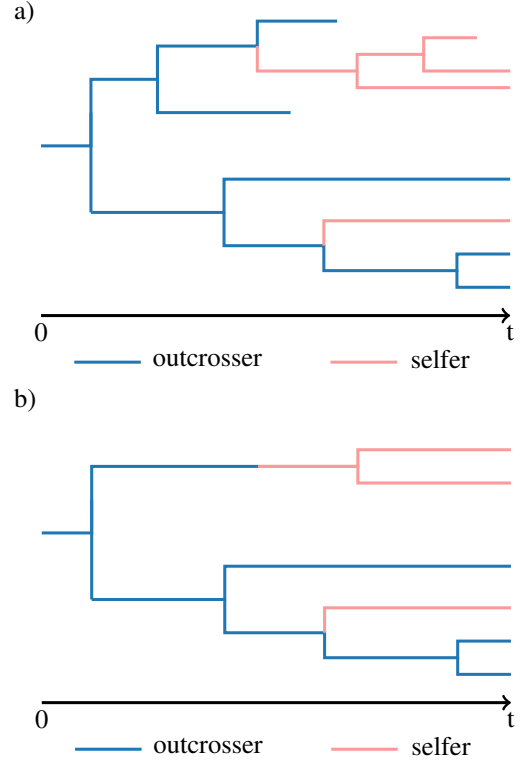


Figure 9: Illustration of a tree with cladogenetic state changes. In panel a), the tree constructed in forward time with parameter value  $p = 1$  is shown, while in panel b), the same tree in a), reduced at time  $t$ , is given. It can be seen that even though  $p = 1$ , the reduced tree (constructed backwards in time) is not cladogenetic. In both panels, the blue and red species represent outcrossers and selfers, respectively.

Furthermore, using the ClaSSE model parameters along with the estimated values of  $\widehat{\omega}_0$  and  $\widehat{\omega}$  in (24), we obtain  $\widehat{\omega}_1 = 0.305$ , which is very close to the estimated value of  $\widehat{\omega}_1$  in the above equation. As a concluding observation, we see that if the estimates of different parameters including  $\omega_0$  and  $\omega_1$  are known, then after a sufficient amount of time has passed, the value of  $dN/dS$ , as well as the fraction of outcrossing and selfing species in the species tree, reaches a steady state, which can be determined using (24).

## 5. Discussion

In this paper, we linked trait-dependent diversification process with trait-dependent substitution process. We focused on binary-trait models based on the ‘evolutionary dead-end’ hypothesis, with unidirectional shift from a trait-0 to a trait-1,



where trait-0 is assumed to be the dominating type. To do so, we first described several properties of diversification models that have not been obtained before. In particular, we introduced a novel way to decompose and analyze the reduced trait-0 and trait-1 species trees. We then showed how trait-dependent diversification processes affect the inference and interpretation of relationships between traits and molecular evolutionary rates.

We derived several expressions describing the tree characteristics, such as the expected sizes of the reduced trees as well as their expected branch lengths. An interesting result was obtained, in particular, regarding the expected number of trait-1 species per cluster; in (19), we showed that this value always remains within a bounded region. Assigning outcrossing and selfing species as trait-0 and trait-1, respectively, this formally confirms the general observation that large clades of selfing species are rare. Larger clusters of selfing species can only be obtained when the selfing diversification rate is positive ( $\lambda_1 > \mu_1$ ), which corresponds to conditions where selfing is no more an evolutionary dead-end. Observation of large clusters of selfing species thus suggests that dead-end conditions may not apply uniformly on a given tree. For example, a clade of 21 self-compatible species was observed in the Solanaceae dataset used in Goldberg et al. (2010). Such a pattern would be compatible with a scenario where selfing lineages senesce in diversification rates, with increasing extinction rates through time (Ho and Agrawal, 2017).

A special case occurs when selfing species do not form any clusters, but appear only as singletons. Under this condition, testing the difference of  $dN/dS$  between outcrossing and selfing branches can be difficult. For longer branches leading to the selfing species, we could expect to have more power to distinguish between  $\omega_0$  and  $\omega_1$ , simply because more substitutions occur and  $\omega_0$ ,  $\omega_1$  values are better estimated. However, the proportion of the branch with  $\omega_1$  decreases as the branch length increases, making  $\omega_0$  and  $\omega_1$  values closer to one another. For shorter branches, the mutation substitution process cannot be considered as instantaneous, and the polymorphic transitory phase – currently not included in our modeling frame-

work – must be taken into account to avoid detection of spurious changes in  $dN/dS$  (Mugal et al., 2014). Thus, not detecting any effect of selfing on  $dN/dS$  does not necessarily mean that the effect is weak. The same rationale and conclusions may be applied to other binary traits, such as hermaphroditism versus dioecy (Käfer et al., 2013), sexuality versus clonality (Henry et al., 2012) and solitary life versus sociality (Romiguier et al., 2014).

Overall, our results show that trait-dependent diversification processes can have a strong impact on the relationship between traits and molecular evolution. Testing the effect of dead-end traits on molecular evolutionary rates can be especially problematic in the presence of small clusters, as discussed above. In addition, if large clusters corresponds to parts of the tree where dead-end conditions are not yet met, weak or no effect of traits on molecular evolutionary rates should also be expected. The possible bias depends on the diversification process parameters, for example, under a pure cladogenetic model, branch-specific  $dN/dS$  estimates correspond to  $\omega_0$  and  $\omega_1$ . Fitting a diversification model prior to molecular evolutionary analysis should help interpreting results. A further step would be to develop statistical methods allowing to jointly infer the effect of traits on both the diversification process and the molecular evolutionary rates. Since, under the dead-end hypothesis, accumulation of deleterious non synonymous mutations directly affects the long-term survival of species (or is, at least, a signature of reduced selection efficacy that can drive species towards extinction), the rate of extinction of a species could be considered directly proportional to the rate of fixation of mutations in the corresponding trait. Therefore, a possible extension of this work would be to integrate the species diversification process with molecular evolution by combining the substitution and extinction rates. This could be achieved by means of, say for instance, a linear function, such as using a control parameter  $c > 0$  and assuming that  $\mu_0 = c\omega_0$  and  $\mu_1 = c\omega_1$ . This would be a straightforward way to connect the two processes, but more realistic parameterizations could also be considered, as long as they bring new insights without complicating the treatment or changing the rationale.

Allowing for two-sided transitions (reversion from selfing to outcrossing, in our example) would also be a natural extension to this work, but would lead to more complex scenarios, since both trait-0 and trait-1 trees could be disjoint. We hope that the modeling framework presented in our paper, will be a useful starting point for further development in this field of research.

### Supplementary material

The codon sequence file of the Geraniaceae data set, used in our example in Section 4.2, and the corresponding estimated phylogeny, depicted in Figure 7a, can be found online at <http://dx.doi.org/10.17632/d2fpg6dkrv.1>.

### Appendix A. Mathematical properties

This section elaborates the mathematical properties of the two-type, continuous time Markov branching process  $X_t = (K_t, L_t)$ , given in (1). Here, we analyze the process following the same approach and notation as in Athreya and Ney (1972).

The branching rates of  $X_t$ , given in (2), are rewritten below for convenience

$$(k, \ell) \mapsto \begin{cases} (k+1, \ell) & \lambda_0 k \\ (k-1, \ell+1) & (1-p)\delta k \\ (k-1, \ell) & \mu_0 k \\ (k, \ell+1) & p\delta k + \lambda_1 \ell \\ (k, \ell-1) & \mu_1 \ell. \end{cases}$$

The life length of type- $i$ ,  $i = 0, 1$ , is exponentially distributed with parameter  $\mathbf{a} = (a_0, a_1)$ , such that

$$a_0 = \lambda_0 + \mu_0 + \delta \quad \text{and} \quad a_1 = \lambda_1 + \mu_1.$$

The offspring distribution of the two types is given by  $\mathbf{p}(\mathbf{j}) = (p^{(0)}(\mathbf{j}), p^{(1)}(\mathbf{j}))$ , where

$$\begin{aligned} p^{(0)}(2, 0) &= \frac{\lambda_0}{\lambda_0 + \delta + \mu_0}, & p^{(0)}(0, 1) &= \frac{(1-p)\delta}{\lambda_0 + \delta + \mu_0}, \\ p^{(0)}(0, 0) &= \frac{\mu_0}{\lambda_0 + \delta + \mu_0}, & p^{(0)}(1, 1) &= \frac{p\delta}{\lambda_0 + \delta + \mu_0}, \\ p^{(1)}(0, 2) &= \frac{\lambda_1}{\lambda_1 + \mu_1}, & p^{(1)}(0, 0) &= \frac{\mu_1}{\lambda_1 + \mu_1}, \end{aligned}$$

and

$$\sum_j p^{(i)}(\mathbf{j}) = 1.$$

The generating function is of the form  $\mathbf{f}(\mathbf{s}) = (\mathbf{f}^{(0)}(\mathbf{s}), \mathbf{f}^{(1)}(\mathbf{s}))$ , where  $\mathbf{f}^{(i)}(\mathbf{s}) = \sum_j p^{(i)}(\mathbf{j}) \mathbf{s}^j$ , that is

$$f^{(0)}(s_0, s_1) = \frac{\lambda_0 s_0^2 + p\delta s_0 s_1 + (1-p)\delta s_1 + \mu_0}{\lambda_0 + \delta + \mu_0},$$

and

$$f^{(1)}(s_0, s_1) = \frac{\lambda_1 s_1^2 + \mu_1}{\lambda_1 + \mu_1}.$$

The infinitesimal generating function is given by  $\mathbf{u}^i(\mathbf{s}) = a_i[\mathbf{f}^{(i)}(\mathbf{s}) - s_i]$ , that is,

$$\begin{aligned} u^{(0)}(s_0, s_1) &= \lambda_0 s_0^2 + p\delta s_0 s_1 + (1-p)\delta s_1 \\ &\quad - (\lambda_0 + \delta + \mu_0)s_0 + \mu_0, \end{aligned}$$

and

$$u^{(1)}(s_0, s_1) = \lambda_1 s_1^2 - (\lambda_1 + \mu_1)s_1 + \mu_1.$$

The mean offspring matrix  $A$  is defined as

$$A = (a_{ij}), \quad \text{where} \quad a_{ij} = a_i \left[ \frac{\partial f^{(i)}(\mathbf{s})}{\partial s_j} \Big|_{\mathbf{s}=\mathbf{1}} - \delta_{ij} \right],$$

and

$$\delta(i, j) = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{otherwise.} \end{cases}$$

Hence,

$$A = \begin{pmatrix} \lambda_0 - (1-p)\delta - \mu_0 & \delta \\ 0 & \lambda_1 - \mu_1 \end{pmatrix}.$$

The eigenvalues of  $A$ ,

$$\gamma_0 = \lambda_0 - (1-p)\delta - \mu_0 \quad \text{and} \quad \gamma_1 = \lambda_1 - \mu_1,$$

classify the process  $X_t$  as

$$\gamma_+ = \max(\gamma_0, \gamma_1) \quad \begin{cases} < 0 & X_t \text{ is subcritical} \\ = 0 & X_t \text{ is critical} \\ > 0 & X_t \text{ is supercritical.} \end{cases}$$

Let the extinction probabilities of  $X_t = (K_t, L_t)$  be

$$q_0 = P[\text{extinction} | X_0 = (1, 0)], \quad q_1 = P[\text{extinction} | X_0 = (0, 1)].$$

Then,  $q_0$  and  $q_1$  are minimal solutions of the system

$$q_0 = f^{(0)}(q_0, q_1), \quad q_1 = f^{(1)}(q_0, q_1).$$

Since type-1 species are unable to produce type-0 offspring,  $q_1 = 1$ . Hence, the only minimal solution of the above system, that provides a nonnegative chance of species survival, is

$$q_0 = \frac{\mu_0 + (1-p)\delta}{\lambda_0} = \frac{\mu}{\lambda}, \quad q_1 = 1.$$

To find the mean values, assume that

$$m_{00}(t) = E(K_t | K_0 = 1), \quad m_{11}(t) = E(L_t | L_0 = 1),$$

$$m_{01}(t) = E(L_t | K_0 = 1), \quad m_{10}(t) = E(K_t | L_0 = 1) = 0,$$

and

$$M(t) = \begin{pmatrix} m_{00}(t) & m_{01}(t) \\ 0 & m_{11}(t) \end{pmatrix}.$$

Then, we have that  $M(s+t) = M(s)M(t)$  and  $M(t) \rightarrow I$  as  $t \rightarrow 0$ . Hence,  $M(t) = e^{At}$ , and

$$m_{00}(t) = e^{(r_0 - (1-p)\delta)t}, \quad m_{11}(t) = e^{r_1 t},$$

$$m_{01}(t) = \frac{\delta}{r_0 - (1-p)\delta - r_1} [e^{(r_0 - (1-p)\delta)t} - e^{r_1 t}].$$

Here,  $M(t)$  is *not* positively regular. However,

$$\frac{E(K_t)}{E(K_t) + E(L_t)} = \frac{m_{00}(t)}{m_{00}(t) + m_{01}(t)} \rightarrow 1 - \frac{\delta}{r_0 - r_1 + p\delta},$$

and

$$\frac{E(L_t)}{E(K_t)} = \frac{m_{01}(t)}{m_{00}(t)} \rightarrow \frac{\delta}{r_0 - (1-p)\delta - r_1},$$

as  $t \rightarrow \infty$ .

## Appendix B. Estimates for the case when $\lambda_1 = 0$

We examine a particular case, in which we assume that complete species sampling has taken place, and that each observed trait-1 species pairs up with a species of trait-0, at the most recent branching bifurcation point. Hence, the trait-1 species occur only as singletons, and we estimate  $\lambda_1 = 0$ . Let us consider such a species pair at time  $t$ , which traces back to a joint ancestor at time  $t - s$  (Fig. B.1). Since  $\lambda_1 = 0$ , the joint ancestor is necessarily a species of trait-0. The total divergence time

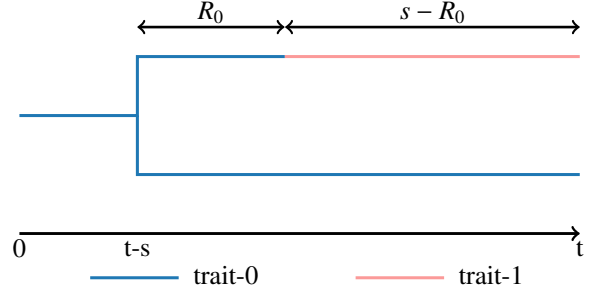


Figure B.1: An observed pair of trait-0 (blue) and trait-1 (red) species at time  $t$ , with a common trait-0 ancestor at time  $t - s$ .

of the pair is  $2s$ . One branch of length  $s$  is trait-0 throughout, while the other branch divides into  $s = R_0 + (s - R_0)$ , where  $R_0$  is the time spent as trait-0. In particular, if the splitting event at  $t - s$  produces one species of each trait, then  $R_0 = 0$ .

We now attempt to find an estimate of the expected value of  $R_0$ . Suppose that we observe a total of  $\ell$  species of trait-1 at  $t$ , each having a trait-0 species as their nearest neighbor species backwards in the tree. Let the divergence times of each of the pairs be  $t - s_i$ ,  $i = 1, \dots, \ell$ . Let  $R_0(s_i)$  be the corresponding times represented by trait-0 species since divergence. Then

$$A'_t = A_t + \sum_{i=1}^{\ell} R_0(s_i),$$

$$E(A'_t | K_t > 0) \approx E(A_t | K_t > 0) + \sum_{i=1}^{\ell} r_0(s_i),$$

where the  $r_0(s_i)$ 's denote the expected time spent as trait-0 on a branch of length  $s$ , and are computed as follows.

Let  $R$  denote an exponential random variable with rate  $(1-p)\delta$ , and let  $R_0$  have a mixed distribution so that  $R_0 = 0$  with probability  $p\delta/(\lambda + p\delta)$  and  $R_0$  is given by  $R$  otherwise. Also, let  $V_0$  and  $V_1$  be exponential extinction times of rate  $\mu_0$  and  $\mu_1$ , respectively. Here,  $R_0$  represents the time as trait-0 in the branch ending up as trait-1, given that the species survive to  $t$ . Thus, using notation of the type  $E(X|A) = E(X, A)/P(A)$ ,

$$\begin{aligned} r_0(s) &= E(R_0 | R_0 < s, V_0 > R_0, V_1 > s - R_0) \\ &= \frac{E(R_0, R_0 < s, V_0 > R_0, V_1 > s - R_0)}{P(R_0 < s, V_0 > R_0, V_1 > s - R_0)}. \end{aligned}$$

The expectation simplifies as

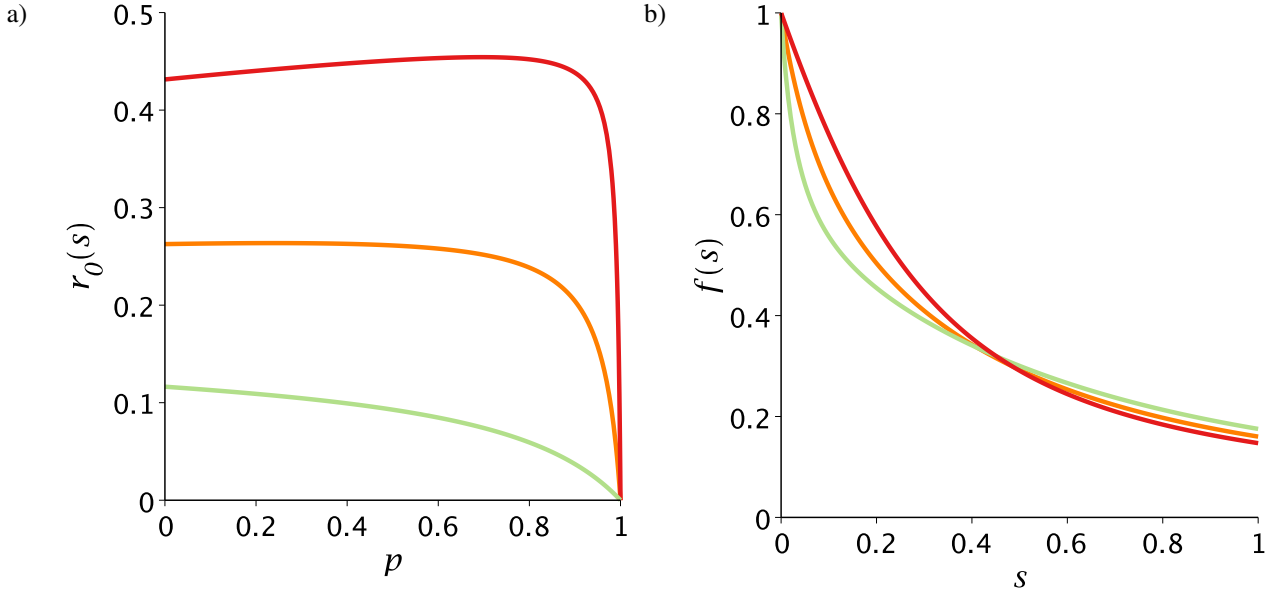


Figure B.2: The left panel a) gives plots of  $r_0(s)$  versus  $p$  for  $s = 0.2$  in green,  $s = 0.4$  in orange, and  $s = 0.6$  in red. The right panel b) gives plot of  $f(s)$  versus  $s$  for  $p = 0.2$  in green,  $p = 0.4$  in orange, and  $p = 0.6$  in red. In both plots,  $\delta = 3$ ,  $\lambda_0 = 8$ ,  $\mu_0 = 4$ ,  $\mu_1 = 12$ ,  $\lambda_1 = 0$  and  $\mu = \mu_0 + (1 - p)\delta$ .

$$\begin{aligned}
& E(R_0, R_0 < s, V_0 > R_0, V_1 > s - R_0) \\
&= \frac{\lambda}{\lambda + p\delta} E(R, R < s, V_0 > R, V_1 > s - R) \\
&= \frac{\lambda}{\lambda + p\delta} \int_0^s r e^{-\mu_0 r} e^{-\mu_1(s-r)} (1-p)\delta e^{-(1-p)\delta r} dr.
\end{aligned}$$

Furthermore,

$$\begin{aligned}
& P(R_0 < s, V_0 > R_0, V_1 > s - R_0) \\
&= \frac{p\delta}{\lambda + p\delta} P(V_1 > s) + \frac{\lambda}{\lambda + p\delta} P(R < s, V_0 > R, V_1 > s - R) \\
&= \frac{p\delta e^{-\mu_1 s}}{\lambda + p\delta} + \frac{\lambda}{\lambda + p\delta} \int_0^s e^{-\mu_0 r - \mu_1(s-r)} (1-p)\delta e^{-(1-p)\delta r} dr,
\end{aligned}$$

and hence, recalling that  $\mu = \mu_0 + (1 - p)\delta$ ,

$$r_0(s) = \frac{\lambda(1-p) \int_0^s r e^{(\mu_1 - \mu)r} dr}{p + \lambda(1-p) \int_0^s e^{(\mu_1 - \mu)r} dr},$$

or, evaluating the integrals,

$$r_0(s) = \frac{\lambda(1-p) \left( (\mu_1 - \mu)s - 1 + e^{-(\mu_1 - \mu)s} \right)}{p(\mu_1 - \mu)^2 e^{-(\mu_1 - \mu)s} + \lambda(1-p)(\mu_1 - \mu)(1 - e^{-(\mu_1 - \mu)s})}.$$

Figure B.2a shows how  $r_0(s)$  varies with  $p$  for different values of  $s$ , when  $\lambda_1 = 0$ . In particular, it can be seen that when changes are not purely cladogenetic ( $p < 1$ ), a substantial fraction of the branch evolves as trait-0.

Define  $f(s)$  to be the fraction of type-1 in a branch carrying both trait-0 and trait-1, as shown in Figure B.1, that is,

$$f(s) := \frac{s - r_0(s)}{s}.$$

Figure B.2b gives an illustration of  $f(s)$  versus  $s$  for different values of  $p$ , with  $\lambda_1 = 0$ . It can be seen that  $f(s)$  decreases with  $s$  whenever  $\mu_1 > \mu = \mu_0 + (1 - p)\delta$ . This means that proportionally, the longer the branch leading to type-1, the more recent the transition event would be.

### Appendix C. Simulation results

To test the accuracy of the relevant parameters namely,  $T_{\text{tot}}^1$  and  $T_{\text{tot}}$ , used in our example in Section 4.2, we simulated 100 phylogenetic trees with  $\lambda_0 = 11.761$ ,  $\lambda_1 = 0$ ,  $\mu = \mu_0 = 10.406$ ,  $\mu_1 = 19.494$ ,  $\delta = 8.581$  and  $p = 1$ , as initial values. (Note that these are the same values as those estimated from the ClaSSE model in Section 4.2.) Maximum likelihood estimation method of ClaSSE, constrained to allow only one-sided transitions, was then applied to each of the 100 trees to obtain different parameter values. Out of the 100 set of parameter values, we disregarded 16 since they did not conform to our model ( $\gamma_0 = \lambda - \mu$  was negative for these trees). From the remaining 84 sets of

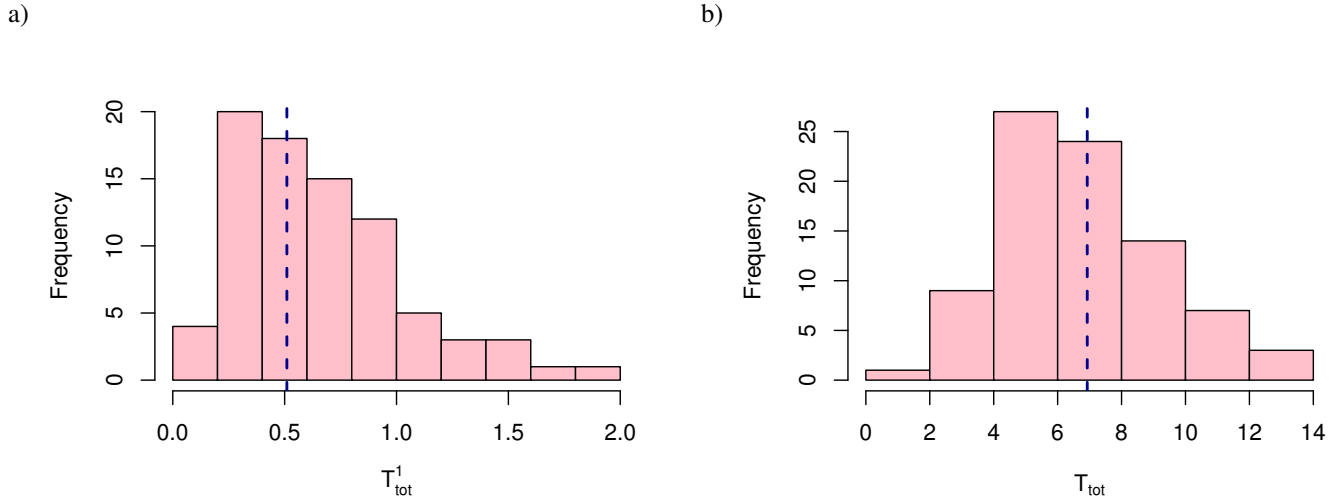


Figure C.1: Histograms showing probability distributions for  $T_{\text{tot}}^1$  in a) and  $T_{\text{tot}}$  in b), for 84 simulated trees. In the left and right panels, the observed values  $T_{\text{tot}}^1 = 0.510$  and  $T_{\text{tot}} = 6.926$ , respectively, are indicated by a dashed blue vertical line.

parameter values, we calculated the expected selfer tree branch lengths  $T_{\text{tot}}^1$  using (16), and from the corresponding 84 simulated trees, we also recorded the observed total branch length  $T_{\text{tot}}$  of each tree. The results are shown in the form of histograms given in Figure C.1a and C.1b. From these figures, we conclude that the estimates  $T_{\text{tot}}^1 = 0.510$  and  $T_{\text{tot}} = 6.926$ , used in our example in Section 4.2, are typically consistent with the model.

## Acknowledgements

This work was supported by the Marie Curie Intra-European Fellowship (grant number IEF-623486, project SELFADAPT to S.G. and M.L.). D.T. is supported by The Centre for Interdisciplinary Mathematics in Uppsala University.

## References

Antal, T., Krapivsky, P.L., 2011. Exact solution of a two type branching process: models of tumor progression. *J. Stat. Mech.* P08018.

Athreya, K.B., Ney, P.E., 1972. *Branching processes*. Springer, New York.

Beaulieu, J.M., O'Meara, B.C., 2016. Detecting hidden diversification shifts in models of trait-dependent speciation and extinction. *Syst. Biol.* 65, 583–601.

Figuet, E., Nabholz, B., Bonneau, M., Carrio, E.M., Brzyska, K.N., Ellegren, H., Galtier, N., 2016. Life history traits, protein evolution, and the nearly neutral theory in amniotes. *Mol. Biol. Evol.* 33, 1517–1527.

Fisher, R.A., 1941. Average excess and average effect of a gene substitution. *Ann. Hum. Genet.* 11, 53–63.

FitzJohn, R.G., 2012. Diversitree: comparative phylogenetic analyses of diversification in R. *Methods Ecol. Evol.* 3, 1084–1092.

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

Glémin, S., Muyle, A., 2014. Mating systems and selection efficacy: a test using chloroplastic sequence data in angiosperms. *J. Evol. Biol.* 27, 1386–1399.

Goldberg, E.E., Igić, B., 2008. On phylogenetic tests of irreversible evolution. *Evolution*. 62, 2727–2741.

Goldberg, E.E., Kohn, J.R., Lande, R., Robertson, K.A., Smith, S.A., Igić, B., 2010. Species selection maintains self-incompatibility. *Science*. 330, 493–495.

Goldberg, E.E., Igić, B., 2012. Tempo and mode in plant breeding system evolution. *Evolution*. 66, 3701–3709.

Guindon, S., Gascuel, O., 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst. Biol.* 52, 696–704.

Henry, L., Schwander, T., Crespi, B.J., 2012. Deleterious mutation accumulation in asexual *Timema* stick insects. *Mol. Biol. Evol.* 29, 401–408.

Ho, E.K.H., Agrawal, A.F., 2017. Aging asexual lineages and the evolutionary maintenance of sex. *Evolution*. 71, 1865–1875.

Igić, B., Lande, R., Kohn, J.R., 2008. Loss of self-incompatibility and its evolutionary consequences. *Int. J. Plant Sci.* 169, 93–104.

Igić, B., Busch, J.W., 2013. Is self fertilization an evolutionary dead end? *New Phytol.* 198, 386–397.

Käfer, J., Talianová, M., Bigot, T., Michu, E., Guéguen, L., Widmer, A., Žluvová, J., Glémin, S., Marais, G.A., 2013. Patterns of molecular evolution in dioecious and non-dioecious *Silene*. *J. Evol. Biol.* 26, 335–346.

Kendall, D.G., 1948. On the generalized birth-and-death process. *Ann. Math.*

Stat. 19, 1–15.

- Levy Karin, E., Ashkenazy, H., Wicke, S., Pupko, T., Mayrose, I., 2017. TraitRateProp: a web server for the detection of trait-dependent evolutionary rate shifts in sequence sites. *Nucleic Acids Res.* 45, W260–W264.
- Levy Karin, E., Wicke, S., Pupko, T., Mayrose, I., 2017. An integrated model of phenotypic trait changes and site-specific sequence evolution. *Syst. Biol.* 66, 917–933.
- Maddison, W.P., Midford, P.E., Otto, S.P., 2007. Estimating a binary character's effect on speciation and extinction. *Syst. Biol.* 56, 701–710.
- Mayrose, I., Otto, S.P., 2011. A likelihood method for detecting trait-dependent shifts in the rate of molecular evolution. *Mol. Biol. Evol.* 28, 759–770.
- Mayrose, I., Zhan, S.H., Rothfels, C.J., Magnuson-Ford, K., Barker, M.S., Rieseberg, L.H., Otto, S.P., 2011. Recently formed polyploid plants diversify at lower rates. *Science*. 333, 1257–1257.
- Mugal, C.F., Wolf, J.B., Kaj, I., 2014. Why time matters: codon evolution and the temporal dynamics of  $dN/dS$ . *Mol. Biol. Evol.* 31, 212–231.
- Nee, S., May, R.M., Harvey, P.H., 1994. The reconstructed evolutionary process. *Philos. Trans.: Biol. Sci.* 344, 305–311.
- Nikolaev, S.I., Montoya-Burgos, J.I., Popadin, K., Parand, L., Margulies, E.H., Antonarakis, S.E., 2007. Life-history traits drive the evolutionary rates of mammalian coding and noncoding genomic elements. *Proc. Natl. Acad. Sci. U. S. A.* 104, 20443–20448.
- Paradis, E., Claude, J., Strimmer, K., 2004. APE: Analyses of Phylogenetics and Evolution in R language. *Bioinformatics*. 20, 289–290.
- R Core Team, 2016. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.
- Romiguier, J., Lourenco, J., Gayral, P., Faivre, N., Weinert, L.A., Ravel, S., Ballenghien, M., Cahais, V., Bernard, A., Loire, E., Keller, L., Galtier, N., 2014. Population genomics of eusocial insects: the costs of a vertebrate-like effective population size. *J. Evol. Biol.* 27, 593–603.
- Smith, S.A., Donoghue, M.J., 2008. Rates of molecular evolution are linked to life history in flowering plants. *Science*. 322, 86–89.
- Stebbins, G.L., 1957. Self fertilization and population variability in the higher plants. *Am. Nat.* 91, 337–354.
- Taylor, H.M., Karlin, S., 1984. An introduction to stochastic modeling. Academic Press, New York.
- Thompson, E.A., 1975. Human Evolutionary Trees. Cambridge University Press, Cambridge.
- Wright, S.I., Kalisz, S., Slotte, T., 2013. Evolutionary consequences of self-fertilization in plants. *Proc. R. Soc. B* 280, 20130133.
- Yang, Z., 1997. PAML: a program package for phylogenetic analysis by maximum likelihood. *Comput. Appl. Biosci.* 13, 555–556.

## Figure legends

Figure 1: Diagrammatic representation of cladogenetic state change, anagenetic state change, and the Markov branching process. The 0 and 1 labels represent trait marks. In a), cladogenetic change  $0 \rightarrow 0 + 1$  is given by solid blue-red lines, and  $1 \rightarrow 0 + 1$  by dashed blue-red lines. In b), anagenetic change  $0 \rightarrow 1$  is given by a solid blue-red line, and  $1 \rightarrow 0$  by a dashed red-blue line. Panel c) depicts the two-type Markov model, as well as various rates of the branching process. Transitions from trait-1 to trait-0, represented by dashed lines in all three panels, are not considered further in this paper.

Figure 2: Diagram of a species tree  $\mathcal{T}$ . Panel a): rooted tree  $\mathcal{T} = \mathcal{T}^0 \cup \mathcal{T}^1$ , with trait-0 species colored blue and trait-1 species colored red. Panel b): rooted type-0 tree  $\mathcal{T}^0$ , and edges of type-1 tree  $\mathcal{T}^1$ , plotted separately in blue and red, respectively.

Figure 3: Diagrammatic representation of reduced species trees. Panel a) shows the reduced tree  $\mathcal{T}_t$ , given the original tree  $\mathcal{T}$  from Figure 2a. Trait-0 species are colored blue and trait-1 species are colored red. In panel b), the trees are obtained by splitting  $\mathcal{T}_t$  into disjoint parts, that is,  $(\mathcal{T}_t)^0$  and  $(\mathcal{T}_t)^1$ , which are plotted separately, in blue and red, respectively. Panel c) shows the rooted trait-0 tree  $(\mathcal{T}^0)_t$  in blue, and the trait-1 edges  $(\mathcal{T}^1)_t$  in red, which are obtained if the original tree  $\mathcal{T}$  (from Fig. 2a) is first split into type-0 and type-1 disjoint parts (as in Fig. 2b) and then reduced at  $t$ . In all three panels,  $t \geq 0$  and  $0 \leq s \leq t$ .

Figure 4: Plots of expected values versus  $\delta$ ;  $E(L_t|K_t > 0)$  versus  $\delta$  in red,  $E(C_t|K_t > 0)$  versus  $\delta$  in orange, and  $E(B_t|K_t > 0)$  versus  $\delta$  in green, when  $t = 1$ ,  $p = 0.5$ ,  $\lambda = \lambda_0 = 8$ ,  $\mu_0 = 4$ ,  $\lambda_1 = 10$ ,  $\mu_1 = 12$  and  $\mu = \mu_0 + (1 - p)\delta = 4 + 0.5\delta$ .

Figure 5: Plots of  $L/C = E(L_t|K_t > 0)/E(C_t|K_t > 0)$  versus  $\delta$  and  $p$ . Plot of  $L/C$  versus  $\delta$ , for  $p = 0$  in red,  $p = 0.5$  in orange and  $p = 1$  in green, is shown in a), while plot of  $L/C$  versus  $p$ , for  $\delta = 1.5$  in green,  $\delta = 2.5$  in orange and  $\delta = 3.5$  in red, is given in b). In both a) and b),  $t = 1$ ,  $\lambda = \lambda_0 = 8$ ,  $\mu_0 = 4$ ,  $\lambda_1 = 10$ ,  $\mu_1 = 12$  and  $\mu = 4 + (1 - p)\delta$ .

Figure 6: Plots of  $dN/dS|_t^{\text{comp}}$  versus  $\delta$  in a) and  $dN/dS|_t^{\text{comp}}$  versus  $t$  in b), for selected values of  $p$ . In a),  $t = 1$ ,  $\omega_0 = 0.1$ ,  $\omega_1 = 0.3$ ,  $\lambda_0 = 8$ ,  $\mu_0 = 4$ ,  $\lambda_1 = 10$ ,  $\mu_1 = 12$ , while  $\mu = \mu_0 + (1 - p)\delta$  varies with  $p$  and  $\delta$ . The dotted red curve gives those values of  $dN/dS|_1^{\text{comp}}$  which are obtained for a sub-critical process ( $\gamma_0 = \lambda - \mu < 0$ ). In b),  $\delta = 3$ ,  $\omega_0 = 0.1$ ,  $\omega_1 = 0.3$ ,  $\lambda_0 = 8$ ,  $\mu_0 = 4$ ,  $\lambda_1 = 10$ ,  $\mu_1 = 12$ , while  $\mu = \mu_0 + (1 - p)\delta$  depends on the value of  $p$ . In both panels, the red curve is obtained for  $p = 0$ , orange for  $p = 0.5$ , green for  $p = 1$ , and the dashed black lines represent the value of  $\omega_0$  and  $\omega_1$ .

Figure 7: Phylogenetic trees. Panel a) gives the tree consisting of 47 species in the Geraniaceae family and 1 outgroup species, while in panel b), ultrametric version of the tree in a) is shown. In both a) and b), outcrossing species are colored in blue, selfing species in red, and the outgroup species in green.

Figure 8: Simulations of the two-type branching process  $X_t$ . In a), simulations were carried out with BiSSE model parameters ( $\lambda_1 = 0$ ,  $\lambda_0 = 17.601$ ,  $p = 0$ ,  $\delta = 17.322$ ,  $\mu_0 = 0$ ,  $\mu_1 = 53.716$ ), while in b), ClaSSE model parameters ( $\lambda_1 = 0$ ,  $\lambda_0 = 11.761$ ,  $p = 1$ ,  $\delta = 8.581$ ,  $\mu_0 = 10.406$ ,  $\mu_1 = 19.494$ ) were used. Plots in the left panel represent the trace of  $(K_t, L_t)$  in the  $(k, \ell)$  plane, while plots in the right panel show the paths over time of  $K_t$  and  $L_t$  in blue and red, respectively.

Figure 9: Illustration of a tree with cladogenetic state changes. In panel a), the tree constructed in forward time with parameter value  $p = 1$  is shown, while in panel b), the same tree in a), reduced at time  $t$ , is given. It can be seen that even though  $p = 1$ , the

reduced tree (constructed backwards in time) is not cladogenetic. In both panels, the blue and red species represent outcrossers and selfers, respectively.

Figure B.1: An observed pair of trait-0 (blue) and trait-1 (red) species at time  $t$ , with a common trait-0 ancestor at time  $t - s$ .

Figure B.2: The left panel a) gives plots of  $r_0(s)$  versus  $p$  for  $s = 0.2$  in green,  $s = 0.4$  in orange, and  $s = 0.6$  in red. The right panel b) gives plot of  $f(s)$  versus  $s$  for  $p = 0.2$  in green,  $p = 0.4$  in orange, and  $p = 0.6$  in red. In both plots,  $\delta = 3$ ,  $\lambda_0 = 8$ ,  $\mu_0 = 4$ ,  $\mu_1 = 12$ ,  $\lambda_1 = 0$  and  $\mu = \mu_0 + (1 - p)\delta$ .

Figure C.1: Histograms showing probability distributions for  $T_{\text{tot}}^1$  in a) and  $T_{\text{tot}}$  in b), for 84 simulated trees. In the left and right panels, the observed values  $T_{\text{tot}}^1 = 0.510$  and  $T_{\text{tot}} = 6.926$ , respectively, are indicated by a dashed blue vertical line.

### Table legends

Table 1: Description of parameters.