What fruits can we get from this tree?

Laudanno, Giovanni

DOI:
10.33612/diss.155031292

Chapter 4

Detecting lineage-specific shifts in diversification: a proper likelihood approach

G. Laudanno, B. Haegeman, D. L. Rabosky, R. S. Etienne
Systematic Biology, 2020

Abstract

The branching patterns of molecular phylogenies are generally assumed to contain information on rates of the underlying speciation and extinction processes. Simple birth-death models with constant, time-varying, or diversity-dependent rates have been invoked to explain these patterns. They have one assumption in common: all lineages have the same set of diversification rates at a given point in time. It seems likely, however, that there is variability in diversification rates across subclades in a phylogenetic tree. This has inspired the construction of models that allow multiple rate regimes across the phylogeny, with instantaneous shifts between these regimes. Several methods exist for calculating the likelihood of a phylogeny under a specified mapping of diversification regimes and for performing inference on the most likely diversification history that gave rise to a particular phylogenetic tree. Here we show that the likelihood computation of these methods is not correct. We provide a new framework to compute the likelihood correctly and show, with simulations of a single shift, that the correct likelihood indeed leads to parameter estimates that are on average in much better agreement with the generating parameters than the incorrect likelihood. Moreover, we show that our corrected likelihood can be extended to multiple rate shifts in time-dependent and diversity-dependent models. We argue that identifying shifts in diversification rates is a non-trivial model selection exercise where one has to choose whether shifts in now-extinct lineages are taken into account or not. Hence, our framework also resolves the recent debate on such unobserved shifts.

4.1 Introduction

The literature abounds with examples of spectacular radiations, where specific clades seem to have an elevated diversification rate against a much slower background rate of diversification (Liem, 1973; Mitter, Farrell, and Wiegmann, 1988; Schluter, 2000; Blount, Borland, and Lenski, 2008; Yoder et al., 2010). This phenomenon may occur due to increased mutation rates (Hua and Bromham, 2017), but may also be caused by ecological opportunity (Simpson, 1944; Simpson, 1955; Wellborn and Langerhans, 2015; Mahler et al., 2010) which may arise in three different ways: (1) antagonist extinction; (2) availability of a new environment, either due to dispersal to a new area or due to environmental change driven; or (3) a key innovation that enables escape from competition for niche space (Heard and Hauser, 1995). The theoretical arena to study macroevolutionary diversification is the framework of stochastic birth-death models, where speciation is modeled as a birth event and extinction as a death event. These birth-death models allow for estimating speciation and extinction rates from phylogenetic trees. Nee, May, and Harvey (1994) provided the mathematical tools to do so for the birth-death model with constant or time-dependent speciation and extinction rates. Their work has been the foundation for biologically more complex models. One example is the diversity-dependent birth-death model where speciation and extinction rates are influenced by the number of species in the same clade (Etienne et al., 2012). Another set of examples are the trait-dependent birth-death models, notably the State-dependent Speciation and Extinction (SSE) models (e.g. BiSSE by Maddison, Midford, and Otto (2007), QuaSSE by FitzJohn (2010), MuSSE by FitzJohn (2012), HiSSE by Beaulieu and O'Meara (2016) and SECSSE by Herrera-Alsina, Els, and Etienne (2019)). These models allow for trait shifts over macroevolutionary time and assign different diversification rate regimes to each trait value.

Methods to detect clades with elevated diversification rates without reference to traits or diversity have been developed and are available in a number of software programs. There are two types of approaches. The first type maps the rate shifts on the tree and then asks whether these rate shifts are statistically supported. Implementations of this type include MEDUSA (Alfaro et al., 2009), BAMM (Rabosky, 2014), the Key Innovation model in DDD (Etienne and Haegeman, 2012), and the split-SSE models in DIVERSITREE (FitzJohn, 2010; FitzJohn, 2012). The second type does not map the shifts explicitly on the phylogeny but assumes a multi-state SSE model with each state having its own speciation and extinction rates, where the shifts in states (and hence in diversification rates) are modelled

dynamically. Implementations of this type include the Lineage-Specific Birth-Death-Shift (LSBDS) models in RevBayes (Höhna et al., 2019), the Multi-State Birth-Death model (MSBD) in BEAST2 (Barido-Sottani, Vaughan, and Stadler, 2020) and ClaDS in RPANDA (Maliet, Hartig, and Morlon, 2019). LSBDS and MSBD assume that lineages change to a different state along a branch, while ClaDS assumes that the state shift occurs during speciation. They are special cases of the SECSSE (Herrera-Alsina, Els, and Etienne, 2019) and MISSE (Caetano, O'Meara, and Beaulieu, 2018) frameworks − which combine features of MuSSE (FitzJohn, 2012), GeoSSE (Goldberg, Lancaster, and Ree, 2011), ClaSSE (Goldberg and Igić, 2012) and HiSSE (Beaulieu and O'Meara, 2016) − applied to many concealed traits (and no examined traits). Here we address implementations of the first type, i.e. with shifts in diversification rates mapped on the phylogeny. These methods rely on the same framework as the SSE models but without state-dependence. We will refer to this framework as the D-E framework with mapped rate shifts, from the names of its core functions (D and E). However, here we show that the D-E framework, while mathematically sound in general (such as in applications to SSE models), cannot be applied to models with mapped shifts in rates. We demonstrate that it can lead to probabilities larger than 1. We propose a new, mathematically correct, analytical likelihood formula based on the functions originally introduced by Nee, May, and Harvey (1994) for the constant-rate birth-death model of diversification without rate shifts. Using simulated data with a single shift, we show that our new likelihood performs better in parameter estimation than the incorrect likelihood. Furthermore, we extend our mathematical reasoning to multiple shifts and time-dependent and diversity-dependent models. Finally, we show that model selection in this framework requires making decisions on whether unobserved shifts (i.e. shifts on extinct lineages) are allowed or not.

4.2 Methods

4.2.1 The D-E framework

The D-E framework uses two variables: $D(t)$, the probability of observing the tipward part of the phylogeny at a given lineage at a given time t in the phylogeny, and $E(t)$, the probability of a lineage alive at time t to have no surviving descendants at the present. To compute the likelihood of the entire phylogeny, one computes $D(t)$ and $E(t)$ by integrating the following set of differential equations,

for every branch from tip to the first rootward node:

$$\dot{D} = -(\lambda + \mu)D + 2\lambda DE \tag{4.2.1}$$

$$\dot{E} = \mu - (\lambda + \mu)E + \lambda E^2 \tag{4.2.2}$$

which has the following solution for initial conditions $D(0) = D_0$ and $E(0) = E_0$,

$$D(t) = D_0 \frac{(\lambda - \alpha)^2 \, \mathrm{e}^{-(\lambda-\mu)t}}{(\lambda - \alpha \, \mathrm{e}^{-(\lambda-\mu)t})^2} \tag{4.2.3}$$

$$E(t) = \frac{\mu - \alpha \, \mathrm{e}^{-(\lambda-\mu)t}}{\lambda - \alpha \, \mathrm{e}^{-(\lambda-\mu)t}} \qquad \text{with} \qquad \alpha = \frac{\mu - \lambda E_0}{1 - E_0} \tag{4.2.4}$$

At a node the two $D(t_{\text{node}})$-values of the two daughter branches are multiplied with one another and the speciation rate to obtain the $D(t_{\text{node}})$ of the parent branch. This value will then serve as the new initial condition to further integrate the system back in time to obtain $D(t)$ at the next rootward node. This is then continued until one reaches the crown or stem of the phylogeny. The $D(t)$ value at the stem or crown is the likelihood of the phylogeny. For $E(t)$ nothing changes at the node, as this extinction probability is independent of observed branching points. We refer to the original papers by Alfaro et al. (2009) and Rabosky (2014), and to Appendix 4.4 for more details. The likelihood computed in this way is correct as long as the same rates are used for all lineages (observed and extinct) at a particular time. Below we show that the likelihood is no longer correct if there are lineage-specific rates.

4.2.2 The D-E framework applied to mapped rate shifts leads to probabilities larger than 1

Rate shifts have been accommodated in the D-E framework in the software packages mentioned in the introduction (e.g. MEDUSA, BAMM, DDD) by using different rates of speciation (λ) and extinction (μ) for the lineage that undergoes a rate shift. There has been some debate on the initial condition for the equation for $E(t)$ at the shift point that will be used for further rootward integration of the equations. Rabosky, Mitchell, and Chang (2017) proposed a "recompute" and a "pass-up" algorithm. The first, "recompute", recomputes the $E(t)$ using the rootward rates, whereas "pass-up" uses as initial condition at the time of the shift the $E(t)$ already computed with the shifted rates for the subclade experiencing the rate shift from the present until the time of the rate shift. The "pass-up" algorithm is incorrect because the extinction rate that is needed in the computation of $D(t)$ is

computed for lineages that will not shift. The "recompute" algorithm is the correct one to compute the extinction rate, but we show here that the D-E framework applied to mapped shifts still suffers from another problem that yields an incorrect likelihood.

We look at a simple example of a phylogeny with only a single extant branch. This ensures that the $D(t)$ we obtain at the root is a real probability and not a probability density due to the multiplication by λ at the nodes (formally the multiplication is with $\lambda\, dt$ for an infinitesimal dt). We assume that at time t_q a clade-wide rate shift in diversification rates occurs: all lineages present at that time undergo this shift. Subsequently, at time t_s another shift occurs involving only one branch, which is the branch that we currently observe. Other branches become extinct before the present.

To study the full process, we divide it into three sub-processes (Figure 4.1), each characterized by a set of speciation and extinction rates (λ_r, μ_r):

- for sub-process M_1: rates λ_{M_1} and μ_{M_1};
- for sub-process M_2: rates λ_{M_2} and μ_{M_2}. These rates do not only govern the diversification dynamics occurring in the interval $[t_q, t_s]$, but also the diversification in the interval $[t_s, t_p]$ for all the lineages that do not undergo the lineage-specific rate-shift at t_s (which is why the "pass-up" algorithm is incorrect);
- for sub-process S: rates λ_S and μ_S.

We have used M_i and S to denote these sub-processes, because the first two processes occur in the main (M) clade that does not undergo a clade-specific shift, but the last one occurs in the a subclade (S) after a shift in a single-lineage.

Adopting the D-E framework and the "recompute" strategy an analytical formula for the likelihood can be derived (Appendix 4.4). In the limit of $t_s \to t_q$, i.e. when the lineage-specific rate shift occurs immediately after the clade-wide range shift, this solution can be written in a compact way using the functions p and u from Nee, May, and Harvey (1994):

$$\mathscr{L}_{\mathrm{DE}} = \frac{p_{M_1}(t_0,t_s)(1-u_{M_1}(t_0,t_s))\, p_S(t_s,t_p)(1-u_S(t_s,t_p))}{(1-u_{M_1}(t_0,t_s)(1-p_{M_2}(t_s,t_p)))^2}, \tag{4.2.5}$$

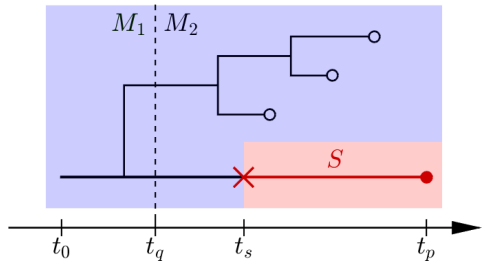

Figure 4.1: Example phylogeny leading to an incorrect likelihood in the D-E framework. The phylogeny consists of a single branch from t_0 to t_p, with a lineage-specific rate shift at t_s (indicated by a $\times$-mark). This rate shift initiates subclade S not included in main clade M. In addition, the rates in the main clade change at t_q from $(\lambda_{M_1}, \mu_{M_1})$ to $(\lambda_{M_2}, \mu_{M_2})$. Open circles indicate species that go extinct before the present and therefore are invisible in the phylogeny. From Eq. 4.2.5 onwards we consider the limit of $t_s \to t_q$.

where

$$p_r(t_1,t_2) = \frac{\lambda_r - \mu_r}{\lambda_r - \mu_r \Lambda_r(t_2 - t_1)} \tag{4.2.6}$$

$$u_r(t_1,t_2) = \frac{\lambda_r(1 - \Lambda_r(t_2 - t_1))}{\lambda_r - \mu_r \Lambda_r(t_2 - t_1)} \qquad \text{with} \qquad \Lambda_r(t) = \mathrm{e}^{-(\lambda_r - \mu_r)t} \tag{4.2.7}$$

with the subscript r referring to the rate regime (M_1, M_2 or S).

Exploring likelihood 4.2.5 – which is a probability – numerically for different values of λ_{M_1} and μ_{M_2} we observe that it exceeds unity in a large part of parameter space (left panel of Figure 4.2), and hence must be incorrect.

We note that we need a clade-wide shift to get probabilities larger than 1. The reason is that to maximize the error in the likelihood, many species need to exist at the time of the shift, which requires a high speciation rate and a low extinction rate. However, in the sampled phylogeny only one lineage remains, so all lineages but one (namely the one that undergoes the shift) then need to go extinct, which requires low speciation rate and high extinction rate. This can only be achieved by having rates change over time, and a clade-wide shift is the simplest way of achieving that. This does not mean that the problem does not occur unless there are clade-wide shifts, but these clade-wide shifts are needed to obtain probabilities larger than 1 which is a clear signal that the likelihood calculation is incorrect. The likelihood calculation remains incorrect when the probabilities are smaller than 1.

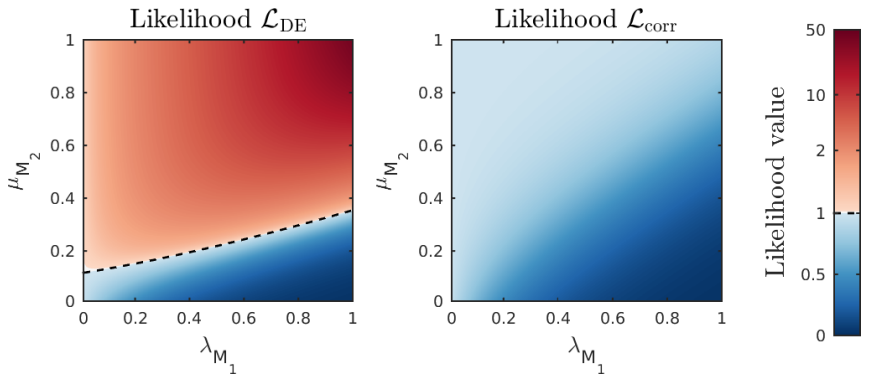

Figure 4.2: Comparison of uncorrected and corrected likelihood for the example of Figure 4.1. We computed the likelihoods as a function of speciation rate λ_{M_1} and extinction rate μ_{M_2} and plotted the results as a heatmap in the $(\lambda_{M_1}, \mu_{M_2})$ plane. The other parameters are kept constant at $\mu_{M_1} = \lambda_{M_2} = \lambda_S = \mu_S = 0$. Left panel: The likelihood $\mathscr{L}_{DE}$ of the D-E framework is larger than one for a large part of parameter space (shades of red) and can reach values that are several orders of magnitude above one. Right panel: The corrected likelihood $\mathscr{L}_{corr}$ is always smaller than one. We have used $t_0 = -10$, $t_q = t_s = -6$, $t_p = 0$ in this numerical example.

4.2.3 Corrected likelihood - Example

The D-E framework only prescribes how to compute the likelihood, but it does not explicitly state the model underlying this likelihood. What has been missing (as pointed out by May and Moore, 2016) in applications of the D-E framework to mapped rate shifts, is a model for these shifts. Here we propose a simple stochastic model for a lineage-specific rate shift, and we use it to yield the right likelihood formula for the example shown in the previous section (Fig. 4.1).

Our model runs from past to present. The simple stochastic model for the rate shift we propose is that at the rate shift time t_s one of the extant species is chosen at random to undergo the rate shift. To construct the likelihood corresponding to this model, we use again functions p (eq. 4.2.6) and u (eq. 4.2.7) from Nee, May, and Harvey, 1994. We denote by n the number of extant species immediately before the rate-shift time. Then, the basic elements of the likelihood are:

- The single species present at the initial time t_0 undergoes a diversification process with rates λ_{M_1} and μ_{M_1}. The probability $P_{M_1}(1,n;t_0,t_s)$ that it has n descendant species immediately before the rate-shift time t_s (i.e. the 1 in the probability corresponds to the single species at t_0 and the n to the number

of descendants at t_s), is

$$P_{M_1}(1,n;t_0,t_s) = p_{M_1}(t_0,t_s)(1-u_{M_1}(t_0,t_s))u_{M_1}(t_0,t_s)^{n-1}. \tag{4.2.8}$$

- The data (the reconstructed tree of Figure 4.1) imposes that the rate shifted species (governed by rates λ_S and μ_S) survives until the present and has only one descendant species. The probability $P_S(1,1;t_s,t_p)$ of this event is

$$P_S(1,1;t_s,t_p) = p_S(t_s,t_p)\big(1-u_S(t_s,t_p)\big). \tag{4.2.9}$$

- The other $n-1$ species extant at time t_s are governed by rates λ_{M_2} and μ_{M_2} and should become extinct in the time interval $[t_s,t_p]$. The probability $P_{M_2}(n-1,0;t_s,t_p)$ of this event is

$$P_{M_2}(n-1,0;t_s,t_p) = (1-p_{M_2}(t_s,t_p))^{n-1}. \tag{4.2.10}$$

Combining these elements we can write the corrected likelihood

$$\begin{aligned}\mathscr{L}_{\text{corr}} &= \sum_{n>0} P_{M_1}(1,n;t_0,t_s)P_S(1,1;t_s,t_p)P_{M_2}(n-1,0;t_s,t_p)\\ &= \sum_{n>0} p_{M_1}(t_0,t_s)(1-u_{M_1}(t_0,t_s))u_{M_1}(t_0,t_s)^{n-1}\\ &\qquad \times (1-p_{M_2}(t_s,t_p))^{n-1}\, p_S(t_s,t_p)(1-u_S(t_s,t_p))\end{aligned} \tag{4.2.11}$$

which can be simplified to

$$\mathscr{L}_{\text{corr}} = \frac{p_{M_1}(t_0,t_s)(1-u_{M_1}(t_0,t_s))\, p_S(t_s,t_p)(1-u_S(t_s,t_p))}{1-u_{M_1}(t_0,t_s)(1-p_{M_2}(t_s,t_p))} \tag{4.2.12}$$

It is instructive to rewrite the likelihood $\mathscr{L}_{\text{DE}}$ in a form similar to eq. 4.2.11,

$$\begin{aligned}\mathscr{L}_{\text{DE}} &= \sum_{n} p_{M_1}(t_0,t_s)(1-u_{M_1}(t_0,t_s))u_{M_1}(t_0,t_s)^{n-1}\\ &\qquad \times n(1-p_{M_2}(t_s,t_p))^{n-1}\, p_S(t_s,t_p)(1-u_S(t_s,t_p))\end{aligned} \tag{4.2.13}$$

showing that the difference with $\mathscr{L}_{\text{corr}}$ resides in the additional factor n in the summand of eq. 4.2.13. This factor n is erroneous, given the explicit rate-shift model we consider here (see Appendix 4.5 for more details), and it also causes the probabilities to exceed unity for some parameter values (left panel of Figure 2).

The corrected likelihood $\mathscr{L}_{\text{corr}}$ solves the problem with probabilities larger than 1. This can be seen from eq. 4.2.12 by noting that $\frac{1-u_{M_1}(t_0,t_s)}{1-u_{M_1}(t_0,t_s)(1-p_{M_2}(t_s,t_p))} \leq 1$. We also checked this numerically (right panel of Figure 2).

We stress that defining a model for how the rate shift occurs is crucial, which explains why SSE-based approaches that model shifts dynamically (Barido-Sottani, Vaughan, and Stadler, 2020; Höhna et al., 2019; Maliet, Hartig, and Morlon, 2019) yield correct likelihoods. However, these models require the introduction of (an arbitrary number of) states and a rate shift parameter and are therefore arguably more complex than the model we propose here.

4.2.4 Corrected likelihood - General case

The corrected likelihood $\mathscr{L}_{\text{corr}}$ can be extended to general trees, as we show in Appendix 4.6. Suppose the rate shift occurs in the j-th lineage of the k_s lineages present at the rate-shift time t_s. Denote by S_j the subclade including all descendants of the shifted lineage j, by M_j the main clade excluding S_j, and by $M^<$ and $M_j^>$ the parts of M_j before and after the rate shift, respectively (Figure 4.3). Furthermore, denote by $I(T)$ the operation of breaking the tree T sensu Nee, May, and Harvey (1994) into separate branches each of which we will index with i. Here T can be either $M^<$, $M_j^>$ or S_j. Strictly, $M_j^>$ needs not be a single tree, but may consist of several trees. For example, in the top-left panel of Figure 4.3, $M_1^>$ consists of three clades which have stem age at t_s: the clades that arise from lineages 2, 3 and 4.

The likelihood that the rate shift occurs in branch j of the main clade, at time t_s, and that it is observed, i.e. that the rate shift is visible in the observed tree,

$$\begin{aligned}\mathscr{L}_{\text{corr}}^{\text{obs},j} = \Big(&\sum_{m_1=0}^{\infty} \cdots \sum_{m_{k_s}=0}^{\infty} \frac{1}{k_s + \sum_{i\in I(M^<)} m_i} \prod_{i\in I(M^<)} (m_i+1)\, p_M(t_i,t_s) \\ &\times (1-u_M(t_i,t_s))\, u_M(t_i,t_s)^{m_i} \left(1-p_M(t_s,t_p)\right)^{m_i}\Big) \\ &\times \Big(\prod_{i\in I(M_j^>)} p_M(t_i,t_p)\,(1-u_M(t_i,t_p))\Big) \\ &\times \Big(\prod_{i\in I(S_j)} p_S(t_i,t_p)\,(1-u_S(t_i,t_p))\Big) \end{aligned} \tag{4.2.14}$$

where k_s denotes the number of observed lineages in the phylogeny at time t_s and the summation index m_i denotes the number of species that come from branch i in $I(M^<)$.

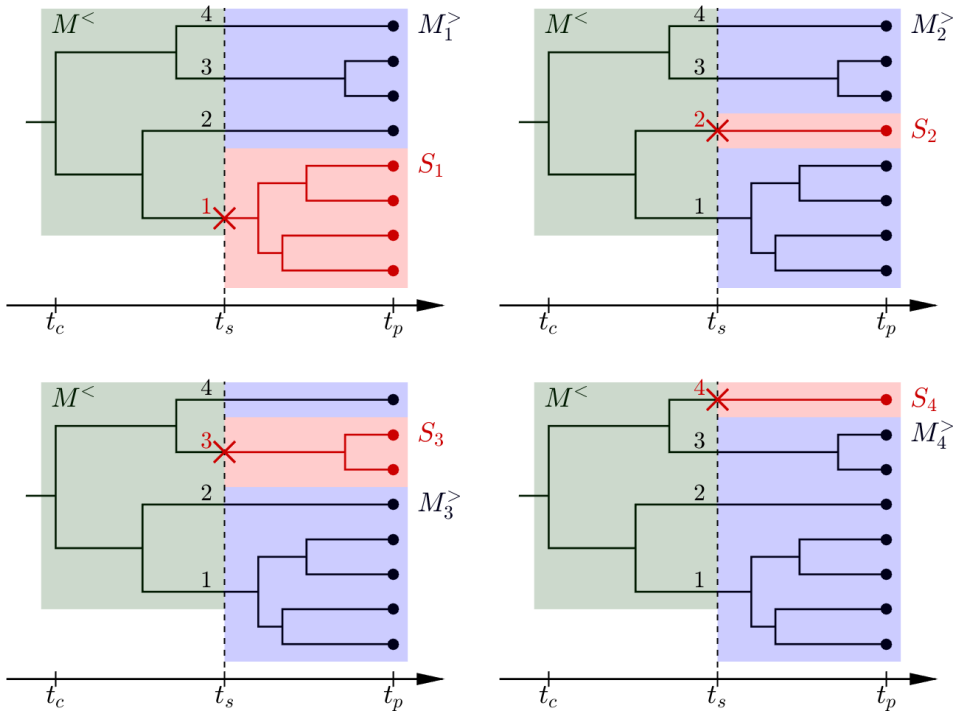

Figure 4.3: Definition of subtrees used in likelihood formula 4.2.14. We consider a phylogeny with $k_s = 4$ lineages at the rate-shift time t_s. Top-left panel: Suppose the rate shift occurs in the first lineage ($j = 1$). The subtree S_1 is the subclade containing all the descendant species of the shifted lineage (in red). The corresponding main clade is M_1, which we decompose in a part $M^<$ before the shift (in green) and a part $M_1^>$ after the shift (in blue). Other panels: If the rate shift occurs in another branch ($j = 2,3,4$), the corresponding main and subclade are different (subtrees S_j and $M_j^>$).

4.2.5 Conditional likelihoods

Likelihoods of diversification models are often conditioned on the existence of the phylogeny, i.e. the survival of the two crown lineages (Nee, May, and Harvey, 1994), but also other conditionings have been discussed in the literature (Stadler, 2012). Here we discuss the standard conditioning on crown lineage survival ($P_{c,0}$) and two additional conditional probabilities. The two additional ones still require that the two crown lineages survive to the present, but we additionally require that the lineage with the rate shift survives to the present with ($P_{c,2}$) or without ($P_{c,1}$) requiring that there is at least one other unshifted surviving lineage of the same crown lineage. To describe these conditional probabilities we need to introduce some notation. We denote by M_L and M_R the two distinct subprocesses arising from left and right crown species, respectively. We assume that M_R undergoes the rate shift at t_s. With S we denote the process arising from the shifted species, as before.

For the standard conditioning on the survival of the two crown lineages we require M_L to survive from the crown to the present, and M_R from the crown to the shift, and either M_R or S to survive from the shift to the present. The conditional probability is given by

$$\begin{aligned} P_{c,0} = 2 \sum_{n_L,n_R>0} & p_M(t_0,t_s)(1-u_M(t_0,t_s))u_M(t_0,t_s)^{n_L-1} \\ & \times p_M(t_0,t_s)(1-u_M(t_0,t_s))u_M(t_0,t_s)^{n_R-1} \\ & \times \frac{n_R}{n_R+n_L}\left(1-(1-p_M(t_s,t_p))^{n_L}\right) \\ & \times \left[p_S(t_s,t_p)+(1-p_S(t_s,t_p))\left(1-(1-p_M(t_s,t_p))^{n_R-1}\right)\right] \end{aligned} \tag{4.2.15}$$

where the factor of two arises from the symmetry of the system: swapping M_L with M_R yields a tree that is indistinguishable from the original one.

For the first new conditioning we require M_L to survive from the crown to the present, M_R to survive from the crown to the shift and S to survive from the shift to the present, but M_R is not required to survive to the present. The conditional

probability $P_{c,1}$ is therefore given by

$$\begin{aligned} P_{c,1} = 2 \sum_{n_L,n_R>0} & p_M(t_0,t_s)(1-u_M(t_0,t_s))u_M(t_0,t_s)^{n_L-1} \\ & \times p_M(t_0,t_s)(1-u_M(t_0,t_s))u_M(t_0,t_s)^{n_R-1} \\ & \times \frac{n_R}{n_R+n_L}\left(1-(1-p_M(t_s,t_p))^{n_L}\right) \\ & \times p_S(t_s,t_p) \end{aligned} \tag{4.2.16}$$

As a second new conditioning we require the survival to the present of both left and right crown species descendants in clade M, as well as the survival of clade S. Its probability, $P_{c,2}$, is given by

$$\begin{aligned} P_{c,2} = 2 \sum_{n_L,n_R>0} & p_M(t_0,t_s)(1-u_M(t_0,t_s))u_M(t_0,t_s)^{n_L-1} \\ & \times p_M(t_0,t_s)(1-u_M(t_0,t_s))u_M(t_0,t_s)^{n_R-1} \\ & \times \frac{n_R}{n_R+n_L}\left(1-(1-p_M(t_s,t_p))^{n_L}\right) \\ & \times \left(1-(1-p_M(t_s,t_p))^{n_R-1}\right) p_S(t_s,t_p) \end{aligned} \tag{4.2.17}$$

The observed tree satisfies the different conditionings. Hence, the conditional likelihood is obtained simply by dividing the likelihood 4.2.14 by either $P_{c,0}$, $P_{c,1}$ or $P_{c,2}$.

4.2.6 Performance of the corrected likelihood in parameter estimation

We tested numerically the performance of the corrected likelihood formula versus the incorrect likelihood resulting from applying the D-E framework to mapped rate shifts for parameter estimation on phylogenies simulated under a constant-rate birth-death model with a single shift for various values of the generating parameters and two conditionings (Table 4.1). We did not use conditioning probability $P_{c,0}$, because to be useful the simulations required the subclade to survive; using $P_{c,0}$ would have introduced biases unrelated to the quality of the estimation method.

We simulated 1000 trees for each parameter setting. We then maximized the likelihood for each tree to infer the best parameter values for λ_M, μ_M, λ_S, and μ_S. We find that the corrected likelihood produces less biased parameter estimates than the likelihood resulting from applying the D-E framework to mapped rate shifts (Figure 4.4).

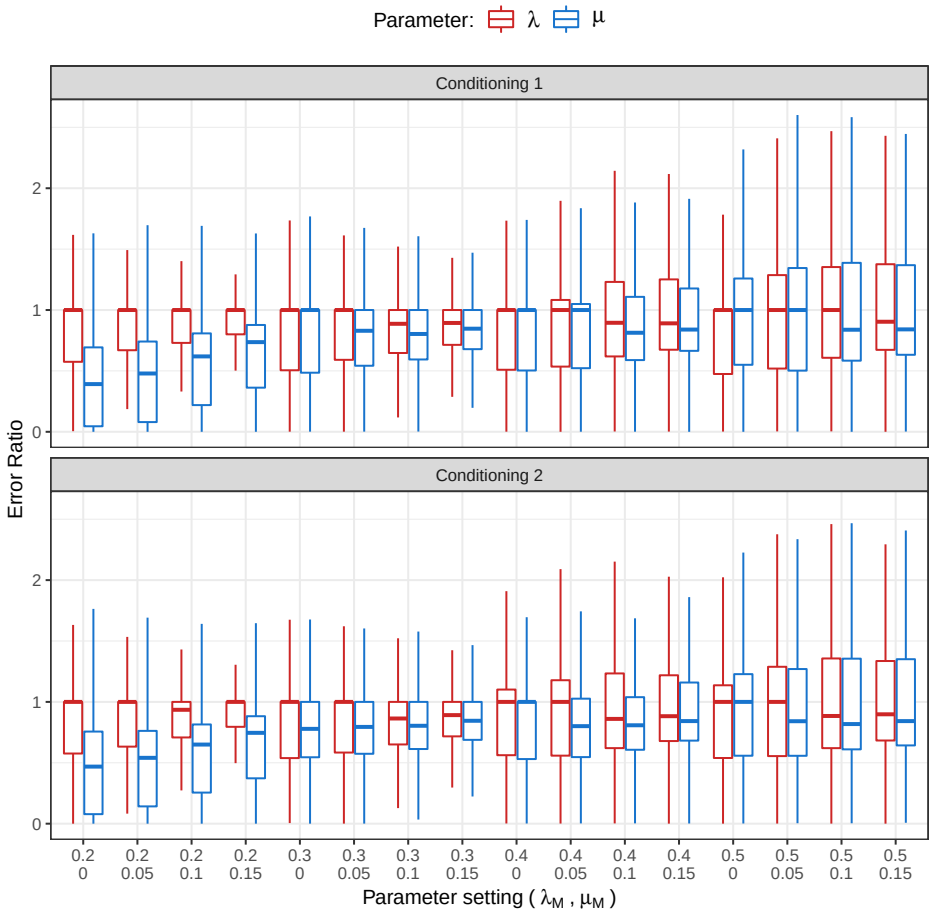

Figure 4.4: Comparison of the parameter estimates of the main clade diversification parameters λ_M and μ_M inferred by the corrected likelihood versus the D-E likelihood. The boxplots show, for various parameter settings displayed on the x-axis, the distributions of the error ratios $\frac{|\lambda_M^{est,corr}-\lambda_M^{true}|}{|\lambda_M^{est,DE}-\lambda_M^{true}|}$ *and* $\frac{|\mu_M^{est,corr}-\mu_M^{true}|}{|\mu_M^{est,DE}-\mu_M^{true}|}$*, where* $\lambda_M^{est,corr}$*,* $\lambda_M^{est,DE}$*,* $\mu_M^{est,corr}$ *and* $\mu_M^{est,DE}$ *were obtained via likelihood maximization and* λ_M^{true} *and* μ_M^{true} *are the known true values used to simulate the trees. Boxplots below 1 imply that the corrected likelihood is closer to the true generating value than the D-E likelihood. Parameter values are given in Table 4.1; the results for shift times* $t_s = -4$ *and* $t_s = -7$ *are shown in the same boxplot. For each unique parameter setting,* 1000 *trees were simulated and hence* 1000 *parameter sets estimated for each likelihood. The two different colors represent the two parameters, while the two panels represent different conditionings (*$P_{c,1}$ *and* $P_{c,2}$*, see eqs. 4.2.16 and 4.2.17).*

Parameter	Values
λ_M	0.2, 0.3, 0.4, 0.5
μ_M	0, 0.05, 0.1, 0.15
t_0	-10
t_s	-4, -7
P_c	$P_{c,1}$, $P_{c,2}$

Table 4.1: Parameter settings used in the simulations. For each parameter setting we simulated 1000 trees. To simulate the subclades parameters $\lambda_S = 0.6$ and $\mu_S = 0.1$ have been used; we did not vary them as they do not influence the inference outcome. We did not use conditioning $P_{c,0}$ because the simulations were conditioned on survival of the shifted subclade.

4.2.7 Extending the corrected likelihood $\mathscr{L}_{\text{corr}}$ to time-dependent and diversity-dependent diversification rates

When the diversification rates depend on time, the basic structure presented in the previous section still holds. However, according to Nee, May, and Harvey, 1994, the core functions 4.2.6 and 4.2.7 have to be replaced with

$$p_r(t_1,t_2) = \left(1 + \int_{t_1}^{t_2} \mu_r(s)\, e^{\rho_r(t_1,s)}\, ds\right)^{-1}, \tag{4.2.18}$$

$$u_r(t_1,t_2) = 1 - p_r(t_1,t_2)\, e^{\rho_r(t_1,t_2)}, \tag{4.2.19}$$

where

$$\rho_r(t_1,t_2) = \int_{t_1}^{t_2} (\mu_r(s) - \lambda_r(s))\, ds. \tag{4.2.20}$$

The corrected likelihood can also be extended to diversity-dependent rates. To do so, we make use of the general framework first introduced in Etienne et al. (2012), which was later used to study the specific case of a single lineage rate shift due to the introduction of a key innovation (Etienne and Haegeman, 2012). The correction comes down to a division by the number of lineages at the shift time (see Appendix 4.8 for details), and has been implemented in version 4.3 of the DDD package (Etienne and Haegeman, 2020).

4.2.8 Extending the corrected likelihood $\mathscr{L}_{\text{corr}}$o multiple shifts

The extension of the corrected likelihood 4.2.14 to the case with multiple shifts is relatively straightforward. Although the formulas become increasingly

cumbersome, the correction is based on the same idea as in the single-shift case. To account correctly for the choice of the species undergoing the rate shift at the rate-shift time, we have to divide by the number of species in which the rate shift can occur. In Appendix C we explicitly work out the correction for the case of two rate shifts occurring in the main clade, leading to likelihood formula 4.6.11.

The conditioning probabilities 4.2.15-4.2.17 can be also extended to the case with multiple shifts. To keep the computations manageable, we consider only the simplest extension. We require that already shifted subclades cannot undergo another shift, i.e., shifts only happen in the main clade. Note that this need not be a strong restriction, because most applications involve large trees with a handful of shifts. In addition, as in conditioning probability $P_{c,2}$, we require that there are surviving species in both crown clades and in all shifted subclades. The corresponding conditional likelihood can be easily evaluated within the diversity-dependent framework of Etienne et al. (2012), and has been implemented in version 4.3 of the DDD package (Etienne and Haegeman, 2020).

4.2.9 Detecting rate shifts in phylogenetic trees

The corrected likelihood 4.2.14 can be used to ask whether a given lineage underwent a rate shift at the designated shift time t_s. To do so, we must compare the likelihood 4.2.14 with a version of 4.2.14 where the rates of the main clade M and the subclade S are the same. That is, we compare a model where the rates actually change at the shift time with a model where the rates remain the same at the shift time (which we will refer to as a dummy shift).

It is important to note that this dummy-shift likelihood is different from Nee, May, and Harvey's likelihood. This can be understood by observing that the former likelihood depends on the predetermined lineage in which the rate shift possibly occurred, while the latter does not distinguish between lineages but treats all lineages equally. To recover Nee, May, and Harvey's likelihood as the reference case in the model comparison, we should ask whether the data contain evidence for a rate shift at a specified time t_s without specifying the rate-shift lineage. To address this question, we have to account for two possibilities: either the rate shift occurred in one of the observed lineages of the phylogeny, or it occurred in a lineage present at time t_s, but that has become extinct after t_s. We refer to these two cases as observed and unobserved rate shifts, respectively. The likelihood of an observed rate shift is simply obtained by summing $\mathscr{L}_{\text{corr}}^{\text{obs},j}$ across all branches

present at time t_s,

$$\mathscr{L}_{\text{corr}}^{\text{obs}} = \sum_{j \in I(M^<)} \mathscr{L}_{\text{corr}}^{\text{obs},j} \tag{4.2.21}$$

The likelihood of an unobserved rate shift is given by

$$\begin{aligned}\mathscr{L}_{\text{corr}}^{\text{unobs}} = \Big(& \sum_{m_1=0}^{\infty} \cdots \sum_{m_{k_s}=0}^{\infty} \frac{\sum_{i \in I(M^<)} m_i}{k_s + \sum_{i \in I(M^<)} m_i} \\ & \times \prod_{i \in I(M^<)} (m_i + 1)\, p_M(t_i, t_s) \\ & \times (1 - u_M(t_i, t_s))\, u_M(t_i, t_s)^{m_i} \left(1 - p_M(t_s, t_p)\right)^{m_i} \Big) \\ & \times \Big(\prod_{i \in I(M_j^>)} p_M(t_i, t_p)(1 - u_M(t_i, t_p)) \Big) \, \frac{1 - p_S(t_s, t_p)}{1 - p_M(t_s, t_p)}\end{aligned} \tag{4.2.22}$$

The likelihoods for observed and unobserved shifts can be added to yield a (generalized) likelihood for a model with a rate shift, on any lineage in the phylogeny that was alive at t_s:

$$\mathscr{L}_{\text{corr}}^{\text{gen}} = \mathscr{L}_{\text{corr}}^{\text{obs}} + \mathscr{L}_{\text{corr}}^{\text{unobs}} \tag{4.2.23}$$

If we take the same rates for main clade M and subclade S (i.e. a dummy shift), we get (see Appendix 4.7)

$$\lim_{\substack{\lambda_S \to \lambda_M \\ \mu_S \to \mu_M}} \mathscr{L}_{\text{corr}}^{\text{gen}} = \mathscr{L}_{\text{Nee}} \tag{4.2.24}$$

where $\mathscr{L}_{\text{Nee}}$ is the likelihood for a phylogeny produced under a constant-rate birth-death model without considering rate shifts, as provided by Nee, May, and Harvey (1994). This shows that the generalized rate-shift likelihood can be compared with the standard likelihood without rate shifts, if we do not specify a priori the lineage in which the rate shift might occur. Note that no such construction is possible when applying the D-E framework to mapped rate shifts, as this systematically overestimates the rate-shift likelihood.

4.3 Discussion

We have shown that several methods developed for detecting shifts in diversification rates at particular points in phylogenies are based on an incorrect likelihood

that could even yield probabilities larger than 1. Our new likelihood formulas – which even apply when diversification rates are time-dependent or diversity-dependent – can be used to correct these methods. This was already been done for the KI model in version 4.1 of the DDD package (Etienne and Haegeman, 2020), which involved only a few lines of code (division by the number of species). For a single shift, DDD allows conditioning in three different ways, corresponding to the diversity-dependent extensions of Eqs. 4.2.15-4.2.17. For multiple shifts, only the last conditioning is feasible, under the additional assumption that shifts can only occur in the main clade. We noticed that the results for two different conditionings were similar for a single shift, so we are confident that this restriction is not very severe. For multiple shifts the DDD package only contains the likelihood calculation, not a function to do likelihood optimization. This could be done by integrating this likelihood calculation with BAMM, MEDUSA, or related multi-shift methods.

Our framework has identified four ways to ask questions on rate shifts at a given shift time. One can ask whether a rate shift occurred in a particular observed lineage in the phylogeny at the shift time, whether it occurred in any observed lineage present at the shift time, whether it occurred in any unobserved lineage present at the shift time, or whether it occurred at all at the given shift time (either in an observed or unobserved lineage). We have provided likelihood formulas for all four cases. This was possible because we provided an explicit model for the choice of the lineage on which the rate shift occurs. Moore et al. (2016) already remarked that such a model was missing in BAMM which they considered problematic for estimating evolutionary rates, although the Moore et al. (2016) critique employed the problematic D-E framework applied to mapped rate shifts. With our formulas we have offered a solution to these problems. In all four cases we have also provided the appropriate null model for model comparison, i.e. a model where there is a shift but rates do not actually change (dummy shift). We have also shown that we recover the well-established likelihood of a constant-rate or time-dependent birth-death model (Nee, May, and Harvey, 1994) in case we allow this dummy shift to occur anywhere in the phylogeny.

While we offer a likelihood that can account for unobserved rate shifts, the use of maximum likelihood to infer shifts on extinct branches is not especially useful in practice. Indeed, this will lead to the estimate of the shifted extinction rate being infinite, as this makes the observed phylogeny most likely. Even when fixing the extinction rate at a given value (or not allowing it to shift), maximizing the likelihood will lead to the shifted speciation rate being estimated to be 0 in unobserved lineages. Fortunately, the generalized likelihood developed to answer

the question whether a shift occurred at all at the given shift time (either in an observed or unobserved lineage) - of which the likelihood for unobserved shifts is an integral part - will not suffer from this problem as any inferred shifted extinction rate applies to both observed and unobserved lineages. To assess whether estimates are biased one should perform an extensive analysis of simulated trees which is beyond the scope of the current paper. An alternative is that for a specific study one conducts simulations with the estimated parameters and checks whether the distribution of parameters estimated from the simulations are in line with the original estimated parameters used to generate the simulated phylogenies. In fact, this parametric bootstrap procedure is useful for any parameter estimation method.

Because the likelihoods implemented in most rate-shift methods suffer from the largely unappreciated issue described in this article, we suggest that researchers interpret results from these methods with caution. The likelihoods obtained with these methods are systematically too large, and hence using these methods to detect rate shifts may be subject to false positives. Simulation studies have generally found that BAMM appears conservative with respect to the inference of rate heterogeneity (Maliet, Hartig, and Morlon, 2019), and rate estimates have been shown to be reasonable across a broad range of parameter space (Title and Rabosky, 2019). However, the fact that the likelihood expression is incorrect implies that BAMM and other approaches may behave unexpectedly in some areas of parameter space or on some datasets. We expect the errors to be largest when there are many species at the time of the shift which do not survive to the present, but the specific conditions under which this situation arises may be hard to identify. Researchers who use these methods should be vigilant in assessing whether results are sensible and should strive to cross-check inferences with alternative methods wherever possible. Our numerical results show how, on sets of simulated trees, the new likelihood 4.2.14 performs consistently better in estimating the rates (i.e. produces less bias) than the likelihoods based on applying the D-E framework to mapped rate shifts. In some simulations the difference was large, in others it seemed small, but it is difficult to say beforehand when estimates will deviate mostly between the corrected likelihood and the likelihood resulting from applying the D-E framework to mapped rate shifts.

As stated in the introduction, alternative approaches to detecting shifts in diversification rates have been developed that do not explicitly map shifts in diversification rates on the phylogeny, but rather look for evidence for multiple rate regimes in general across the phylogeny (Höhna et al., 2019; Barido-Sottani, Vaughan, and Stadler, 2020). These approaches - which rely on the same mathematical equations - do not suffer from the problems we identified here, but they

resort to ancestral state reconstruction to identify what parts of the phylogeny are governed by a rate regime, that is, the shifts in diversification are not fixed, but one obtains a (posterior) probability distribution for the shift positions. Another recently developed approach assumes that each speciation event is accompanied by (usually small) rate shifts in each descendant lineage and allows reconstruction of branch-specific rate estimates (Maliet, Hartig, and Morlon, 2019). However, in each of these frameworks, it is not possible to directly link rate shifts to historical events that happened at specific times: shifts are either assumed to happen with each speciation event (Maliet, Hartig, and Morlon, 2019) or the model determines where the most probable shift locations are (Barido-Sottani, Vaughan, and Stadler, 2020; Höhna et al., 2019). When linking rate shifts to historical events such as glaciation (e.g., Weir et al., 2016) or mountain uplift (e.g., Chaves, Weir, and Smith, 2011) is the ultimate goal of rate shift analyses, our likelihood framework is ideally suited.

Lastly, we emphasize that our approach can also be used for diversity-dependent diversification models, which is not the case for the three recent approaches mentioned above (Barido-Sottani, Vaughan, and Stadler, 2020; Höhna et al., 2019; Maliet, Hartig, and Morlon, 2019).

It has recently been suggested that making inference on diversification scenarios from phylogenies of extant species may be a futile enterprise, because this type of data cannot distinguish between models assuming constant speciation and extinction rates and an infinite set of models with time-dependent speciation and extinction models (Louca and Pennell, 2020), which is a generalization of the results by Nee, May, and Harvey (1994) that a model with constant speciation and extinction rates is equivalent to a model without extinction and time-dependent speciation rate. While this problem of non-identifiability has not yet been mathematically shown to apply also to SSE models (Louca and Pennell, 2020), diversity-dependent models (but see Etienne, Pigot, and Phillimore, 2016) or rate shift models as considered here, the results of Louca and Pennell (2020) have made clear that we can only draw conclusions on the models that we are comparing, and not on the existence of time-dependence, diversity-dependence or rate shifts in general.

In summary, we have described theoretical concerns with the likelihood expression used by a number of diversification rate-shift methodologies, and we have provided a new likelihood formula for rate shifts that is mathematically consistent. We implemented this approach for macroevolutionary scenarios involving a single rate shift and found that the approach performed better than the incorrect D-E framework. We hope that our formulas or algorithms to compute the like-

lihoods will be applied in likelihood-based inference tools such as BAMM and MEDUSA.

4.4 Appendix A: D-E likelihood for example phylogeny of Fig. 4.1

Here we compute the D-E likelihood for the tree of Fig. 4.1. We apply the "recompute" algorithm and use the solutions 4.2.3 and 4.2.4 for the functions D and E.

In the example the phylogeny has only a single extant branch. This implies that the D-E likelihood we obtain at the root is a real probability and not a probability density due to the multiplication by λ at the nodes (formally the multiplication is with $\lambda\, dt$ for an infinitesimal dt). We assume that at time t_q a clade-wide rate shift in diversification rates occurs: all lineages present at that time undergo this shift. Subsequently, at time t_s another shift occurs involving only one branch, which is the branch that we currently observe. Other branches become extinct before the present.

The derivation of the D-E likelihood proceeds in several steps:

- We solve the D-E equations in the interval $[t_s, t_p]$ with rates λ_S and μ_S. For initial conditions $E(t_p) = 0$ and $D(t_p) = 1$ the solution reads

$$\begin{aligned} D_1(t_s) &= D(t_p)\frac{(\lambda_S-\mu_S)^2\Lambda_S(t_s-t_p)}{\big(\lambda_S(1-E(t_p))-(\mu_S-\lambda_S E(t_p))\Lambda_S(t_s-t_p)\big)^2} \\ &= \frac{(\lambda_S-\mu_S)^2\Lambda_S(t_s-t_p)}{\big(\lambda_S-\mu_S\Lambda_S(t_s-t_p)\big)^2} \qquad (4.4.1) \\ E_1(t_s) &= \frac{\mu_S(1-E(t_p))-(\mu_S-\lambda_S E(t_p))\Lambda_S(t_s-t_p)}{\lambda_S(1-E(t_p))-(\mu_S-\lambda_S E(t_p))\Lambda_S(t_s-t_p)} \\ &= \frac{\mu_S-\mu_S\Lambda_S(t_s-t_p)}{\lambda_S-\mu_S\Lambda_S(t_s-t_p)} \qquad (4.4.2) \end{aligned}$$

  where the index 1 refers to the first computation in the interval $[t_s, t_p]$.

- We solve the D-E equations a second time in the interval $[t_s, t_p]$, this time with rates λ_{M_2} and μ_{M_2}. This is required for the "recompute" variant of the D-E framework. The initial conditions are again $E(t_p) = 0$ and $D(t_p) = 1$, so that

$$D_2(t_s) = \frac{(\lambda_{M_2}-\mu_{M_2})^2\Lambda_{M_2}(t_s-t_p)}{\big(\lambda_{M_2}-\mu_{M_2}\Lambda_{M_2}(t_s-t_p)\big)^2} \qquad (4.4.3)$$

$$E_2(t_s) = \frac{\mu_{M_2}-\mu_{M_2}\Lambda_{M_2}(t_s-t_p)}{\lambda_{M_2}-\mu_{M_2}\Lambda_{M_2}(t_s-t_p)} \qquad (4.4.4)$$

where the index 2 refers to the second computation in the interval $[t_s, t_p]$.

- We solve the D-E equations in the interval $[t_q, t_s]$ with rates λ_{M_2} and μ_{M_2}. The initial conditions are $E(t_s) = E_2(t_s)$ (species that originate in $[t_q, t_s]$ and become extinct in $[t_s, t_p]$ are governed by rates λ_{M_2} and μ_{M_2}) and $D(t_s) = D_1(t_s)$ (the observed branch in $[t_s, t_p]$ is governed by rates λ_S and μ_S). We get

$$\begin{aligned} D(t_q) &= D_1(t_s)\frac{(\lambda_{M_2}-\mu_{M_2})^2\Lambda_{M_2}(t_q-t_s)}{(\lambda_{M_2}(1-E_2(t_s))-(\mu_{M_2}-\lambda_{M_2}E_2(t_s))\Lambda_{M_2}(t_q-t_s))^2} \\ &= D_1(t_s)\frac{(\lambda_{M_2}-\mu_{M_2}\Lambda_{M_2}(t_s-t_p))^2\Lambda_{M_2}(t_q-t_s)}{(\lambda_{M_2}-\mu_{M_2}\Lambda_{M_2}(t_q-t_p))^2} \end{aligned} \tag{4.4.5}$$

$$\begin{aligned} E(t_q) &= \frac{\mu_{M_2}(1-E_2(t_s))-(\mu_{M_2}-\lambda_{M_2}E_2(t_s))\Lambda_{M_2}(t_q-t_p)}{\lambda_{M_2}(1-E_2(t_s))-(\mu_{M_2}-\lambda_{M_2}E_2(t_s))\Lambda_{M_2}(t_q-t_p)} \\ &= \frac{\mu_{M_2}-\mu_{M_2}\Lambda_{M_2}(t_q-t_p)}{\lambda_{M_2}-\mu_{M_2}\Lambda_{M_2}(t_q-t_p)} \end{aligned} \tag{4.4.6}$$

- We solve the D-E equations in the interval $[t_0, t_q]$ with rates λ_{M_1} and μ_{M_1}. Using as initial conditions the expressions for $D(t_q)$ and $E(t_q)$, we get

$$D(t_0) = D(t_q)\frac{(\lambda_{M_1}-\mu_{M_1})^2\Lambda_{M_1}(t_0-t_q)}{\left(\lambda_{M_1}(1-E(t_q))-(\mu_{M_1}-\lambda_{M_1}E(t_q))\Lambda_{M_1}(t_0-t_q)\right)^2} \tag{4.4.7}$$

Then, the likelihood as prescribed by the D-E framework (which we are going to show to be incorrect) is $\mathscr{L}_{\mathrm{DE}} = D(t_0)$.

To make formulas clearer, we set $t_q \to t_s$, i.e., the lineage-specific rate shift occurs immediately after the clade-wide range shift. Then, eqs. 4.4.1–4.4.7 can be expressed in terms of the functions p and u introduced by Nee, May, and Harvey, 1994, already stated in eqs. 4.2.6 and 4.2.7, which yields

$$D(t_q) = D_1(t_s) = p_S(t_s,t_p)(1-u_S(t_s,t_p)) \tag{4.4.8}$$

$$E(t_q) = E_2(t_s) = 1 - p_{M_2}(t_q,t_p) \tag{4.4.9}$$

$$\begin{aligned} D(t_0) &= D(t_q)\frac{p_{M_1}(t_0,t_s)(1-u_{M_1}(t_0,t_s))}{\left(1-u_{M_1}(t_0,t_s)(1-p_{M_2}(t_s,t_p))\right)^2} \\ &= \frac{p_S(t_s,t_p)(1-u_S(t_s,t_p))p_{M_1}(t_0,t_s)(1-u_{M_1}(t_0,t_s))}{\left(1-u_{M_1}(t_0,t_s)(1-p_{M_2}(t_s,t_p))\right)^2} \end{aligned} \tag{4.4.10}$$

which establishes eq. 4.2.5.

4.5 Appendix B: Corrected likelihood for phylogeny of Fig. 4.1

In the main text we presented a short derivation of the corrected likelihood for the example phylogeny shown in Fig. 4.1. Here we provide an introduction to the approach of Nee, May, and Harvey (1994) on which the derivation is based, and a comparison of the corrected likelihood with the incorrect likelihood computed within the D-E framework.

4.5.1 A short introduction to Nee et al.

Many properties of the constant-rate birth-death process can be expressed in terms of functions p and u introduced by Nee, May, and Harvey (1994),

$$p(t_1,t_2) = \frac{\lambda-\mu}{\lambda-\mu\,\mathrm{e}^{-(\lambda-\mu)(t_2-t_1)}}$$
$$u(t_1,t_2) = \frac{\lambda\left(1-\mathrm{e}^{-(\lambda-\mu)(t_2-t_1)}\right)}{\lambda-\mu\,\mathrm{e}^{-(\lambda-\mu)(t_2-t_1)}}$$

where λ is the speciation rate and μ the extinction rate. In particular, the probability that the process starting at time t_1 with a single species has n descendant species at a later time t_2 is given by

$$P(1,n;t_1,t_2) = \begin{cases} 1-p(t_1,t_2) & \text{if } n=0 \\ p(t_1,t_2)\left(1-u(t_1,t_2)\right)u(t_1,t_2)^{n-1} & \text{if } n=1,2,\ldots \end{cases} \tag{4.5.1}$$

We see that the number of species is a geometric distribution with parameter $1-u$ with an added zero term (Nee, May, and Harvey, 1994). This formula can be generalized to the probability that n_1 species at time t_1 have n_2 species at time t_2, denoted by $P(n_1,n_2;t_1,t_2)$. The computation exploits the fact that the n_1 species at time t_1 undergo independent dynamics, so that the number of species at time t_2 is the n_1-fold convolution product of $P(1,n_2;t_1,t_2)$. For example, the formula for the case $n_2=0$ reads

$$P(n_1,0;t_1,t_2) = [P(1,0;t_1,t_2)]^{n_1} = \left(1-p(t_1,t_2)\right)^{n_1}. \tag{4.5.2}$$

Nee, May, and Harvey (1994) showed that under the constant-rate birth-death model the likelihood of a phylogeny can be computed as a product over branches. Each of these branches runs from a branching time t_i to the present time t_p. The

corresponding likelihood contribution is equal to the probability that a speciation event occurs at t_i multiplied by the probability that the branch has a single descendant species at the present time t_p. Explicitly,

$$\begin{aligned}&\text{likelihood contribution of branch from } t_i \text{ to } t_p\\ &\quad = \lambda\, \mathrm{d}t_i \times p(t_i,t_p)\left(1-u(t_i,t_p)\right)\end{aligned} \tag{4.5.3}$$

The infinitesimal factor $\mathrm{d}t_i$ is required to impose that the branching time occurs in the infinitesimal interval $[t_i, t_i+\mathrm{d}t_i]$. Because this factor does not affect likelihood maximization, we will leave it out of the likelihood formulas (so that the likelihood is no longer a probability, but a probability density). Taking the product over branches, we obtain the (unconditioned) likelihood of the phylogeny

$$\mathscr{L} = \lambda^s \prod_{i=0}^{s+1} p(t_i,t_p)\left(1-u(t_i,t_p)\right), \tag{4.5.4}$$

where s denotes the number of branching events in the phylogeny and $t_0 = t_1$ denotes the crown age. Because the likelihood is obtained by decomposing the phylogeny into branches, the approach of Nee, May, and Harvey (1994) is sometimes referred to as "breaking the tree". In Appendix 4.6 we present an alternative derivation of eq. 4.5.4, which in contrast to the argument of Nee, May, and Harvey (1994) can be easily generalized to phylogenies with rate shifts.

4.5.2 Comparison of D-E likelihood and corrected likelihood

From the expressions of the likelihoods $\mathscr{L}_{\mathrm{DE}}$ and $\mathscr{L}_{\mathrm{corr}}$, we see that the difference resides in an additional factor n. To interpret this difference, we isolate in both likelihoods the probability that, given that there are n species at the rate-shift time t_s, one of them undergoes the rate shift and has surviving descendant species at the present time t_p, while the other $n-1$ species has no descendant species at t_p. This probability is

$$\text{according to likelihood } \mathscr{L}_{\mathrm{DE}} \quad n\big(1-p_{M_2}(t_s,t_p)\big)^{n-1} p_S(t_s,t_p) \tag{4.5.5}$$

$$\text{according to likelihood } \mathscr{L}_{\mathrm{corr}} \quad \big(1-p_{M_2}(t_s,t_p)\big)^{n-1} p_S(t_s,t_p). \tag{4.5.6}$$

In Figure 4.5 we construct this probability by explicitly considering all possible full trees corresponding to a specific reconstructed tree. Note that in the figure we use simplified notation and set $p_M = p_{M_2}(t_s,t_p)$ and $p_S = p_S(t_s,t_p)$.

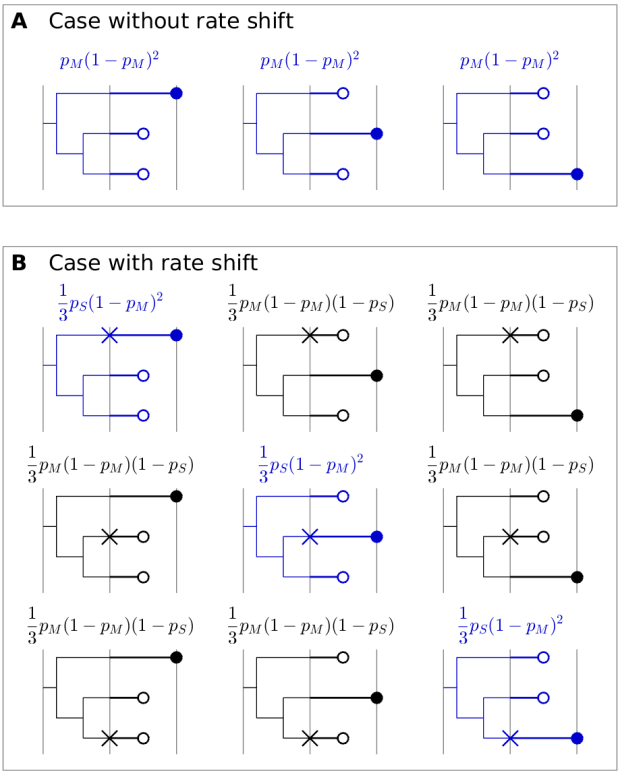

Figure 4.5: Demonstration of correct likelihood formula. (A) Case of phylogeny without rate shift. We consider a reconstructed tree consisting of a single branch running from t_0 to t_p. We assume that there are three extant species at an intermediate time t_s (indicated by the middle vertical line), and consider the dynamics of these three species from t_s to t_p, all governed by rates λ_M and μ_M. There are three possible full trees, each having probability $p_M(1-p_M)^2$, so that the total probability is $3p_M(1-p_M)^2$. (B) Case of phylogeny with rate shift. The reconstructed tree consists of a single branch with an observed rate shift at the intermediate time t_s. We assume again that there are three extant species at t_s, each of them having probability $\frac{1}{3}$ of undergoing the rate shift. The rate-shifted species has rates (λ_S, μ_S); the other species have rates (λ_M, μ_M). We show the nine combinations obtained by first selecting the species undergoing the rate shift, and second selecting the species surviving until the present. Only three of these nine combinations have the correct reconstructed tree (consistent combinations are plotted in blue); the other six combinations correspond to unobserved rate shifts. Each of the three consistent full trees have probability $\frac{1}{3}p_S(1-p_M)^2$, so that the total probability is $p_S(1-p_M)^2$. Symbols have same meaning as in other figures ($\times$-mark: rate shift; filled circle: species surviving until present; open circle: species going extinct before present).

It is instructive to first consider the case without rate shift (Fig. 4.5, panel A). We consider a reconstructed tree consisting of a single branch. If there are n species at the intermediate time t_s, there are n possible full trees, because each of the n species extant at time t_s can be chosen to survive. Each full tree has probability $p_M(1-p_M)^{n-1}$, so that the total probability is $np_M(1-p_M)^{n-1}$. Note that this probability cannot be larger than one; indeed, when adding the probabilities that none or more than one species survives, we get

$$\begin{aligned} np_M(1-p_M)^{n-1} &\leq \sum_{s=0}^{n} \binom{n}{s} p_M^s (1-p_M)^{n-s} \\ &= \big(p_M + (1-p_M)\big)^n = 1. \end{aligned} \tag{4.5.7}$$

Next consider the case with rate shift (Fig. 4.5, panel B). As in the example of Fig. 4.1, we consider a reconstructed tree with a single branch and an observed rate shift. We see that there are n^2 possible full trees, corresponding to two choices, each one having n options. First, we have to choose the species that is undergoing the rate shift. Second, we have to choose the species that is going to survive, which can be either the species that has undergone the rate shift or one of the $n-1$ other species. Importantly, not all of these full trees are consistent with the reconstructed tree. In particular, for $n(n-1)$ of them the rate shift is unobserved. The n full trees for which the rate shift is observed in the corresponding phylogeny each have probability $\frac{1}{n} p_S(1-p_M)^{n-1}$, so that the total probability is $p_S(1-p_M)^{n-1}$. This probability cannot be larger than one, because when adding the probabilities that none or more than one species survives,

$$\begin{aligned} p_S(1-p_M)^{n-1} &\leq \big(p_S + (1-p_S)\big) \sum_{s=0}^{n-1} p_M^s (1-p_M)^{n-1-s} \\ &= \big(p_S + (1-p_S)\big)\big(p_M + (1-p_M)\big)^{n-1} = 1. \end{aligned} \tag{4.5.8}$$

This computation demonstrates that formula 4.5.6, and hence likelihood $\mathscr{L}_{\text{corr}}$, is correct, and that formula 4.5.5, and hence likelihood $\mathscr{L}_{\text{DE}}$, is not. The latter, which can be seen as a naive generalization of the formula without rate shift, does not account correctly for unobserved rate shifts (i.e., rate shifts that occur in a species that is not represented in the phylogeny). Note that formula 4.5.5 is also different from the probability of having either an observed or an unobserved rate shift. The correct probability for this case is

$$p_S(1-p_M)^{n-1} + (1-p_S)(n-1)p_M(1-p_M)^{n-2}. \tag{4.5.9}$$

4.6 Appendix C: Corrected likelihood for general phylogenies

In this appendix we derive the likelihood formula for a general phylogeny with one or several lineage-specific rate shifts. We start the computation, which is related to the "breaking the tree" argument of Nee, May, and Harvey, 1994, by deriving the likelihood again for a general phylogeny without rate shift.

4.6.1 A useful identity

We will repeatedly use the following identity:

$$P(1,n_2;t_0,t_2)n_2 = \sum_{\substack{n_1,n_{2a},n_{2b} \\ n_{2a}+n_{2b}=n_2}} P(1,n_1;t_0,t_1)\, n_1\, P(1,n_{2a};t_1,t_2)\, n_{2a} \\ P(n_1-1,n_{2b};t_1,t_2). \quad (4.6.1)$$

This identity can be understood by introducing a sampling process at time t_2, in which each extant species is sampled with probability ρ. Multiplying the left-hand side of eq. 4.6.1 by $\rho\,(1-\rho)^{n_2-1}$, we see that

$$A = P(1,n_2;t_0,t_2)\, n_2\, \rho\,(1-\rho)^{n_2-1}$$

is the probability that a species at time t_0 has n_2 descendant species at time t_2 and one sampled descendant species. We establish identity 4.6.1 by computing this same probability in a different way. To do so, we consider the n_1 species extant at time t_1. For this group of n_1 species to have one sampled descendant species at time t_1, there should be one of the n_1 species with a single sampled descendant species, and all other species should have no sampled descendant species. Therefore, the probability A can also be computed as

$$\begin{aligned} A &= \sum_{n_1} P(1,n_1;t_0,t_1)\, n_1 \sum_{\substack{n_{2a},n_{2b} \\ n_{2a}+n_{2b}=n_2}} P(1,n_{2a};t_1,t_2)\, n_{2a}\, \rho\,(1-\rho)^{n_{2a}-1} \\ &\qquad P(n_1-1,n_{2b};t_1,t_2)\,(1-\rho)^{n_{2b}} \\ &= \sum_{\substack{n_1,n_{2a},n_{2b} \\ n_{2a}+n_{2b}=n_2}} P(1,n_1;t_0,t_1)\, n_1\, P(1,n_{2a};t_1,t_2)\, n_{2a} \\ &\qquad P(n_1-1,n_{2b};t_1,t_2)\,\rho\,(1-\rho)^{n_2-1}. \end{aligned}$$

Here n_{2a} is the number of species at t_2 (before sampling) descendant from the single species extant at t_1 with a sampled descendant species at t_2; and n_{2b} is the number of species at t_2 (before sampling) descendant from the other n_1-1 species extant at t_1. From the n_{2a} species, one should be sampled; from the n_{2b} species none should be sampled. Dividing the last expression by $\rho\,(1-\rho)^{n_2-1}$, we obtain the right-hand side of eq. 4.6.1.

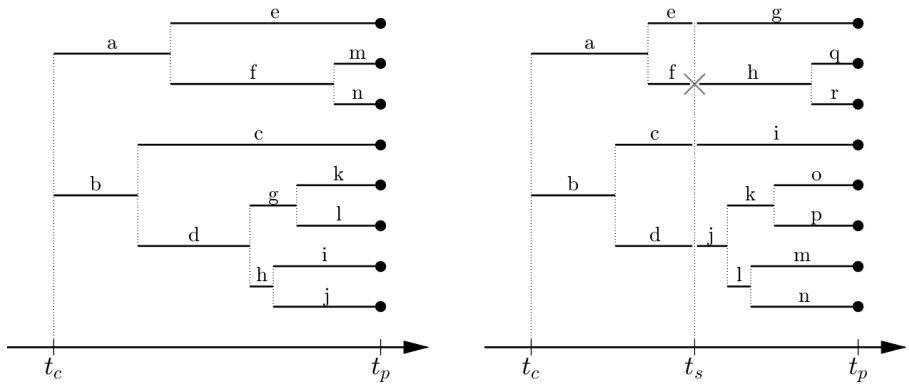

Figure 4.6: Decomposing a phylogeny in simple branches. Left panel: Example tree without rate shift, decomposed in simple branches labeled by letters 'a' to 'n'. Right panel: Same example tree, but this time with rate shift at t_s. Simple branches that contain the rate shift are split, resulting in a larger set of simple branches (labeled by letters 'a' to 'r').

4.6.2 Case without rate shift

We construct the likelihood of a tree under speciation-extinction dynamics without rate shift. This likelihood will be extended below to speciation-extinction dynamics with one or more rate shifts. The argument is related to the "breaking the tree" approach of Nee, May, and Harvey, 1994. As in the proof of eq. 4.6.1 we consider a sampling process at the present time t_p with sampling probability ρ.

We start by decomposing the tree into simple branches, i.e., branches of the reconstructed tree without branching events. See left panel of Figure 4.6, where the simple branches are labeled by letters 'a' to 'n'. We denote the time interval of branch α by $[t_\alpha^b, t_\alpha^e]$. We distinguish internal and boundary simple branches: internal branches are those for which $t_\alpha^e < t_p$, and boundary branches are those for which $t_\alpha^e = t_p$. We denote the set of internal simple branches by B_{int}, and the set of boundary simple branches by B_{ext}. For the example of Fig. 4.6 (left panel) we have

$$B_{int} = \{\mathrm{a},\mathrm{b},\mathrm{d},\mathrm{f},\mathrm{g},\mathrm{h}\} \qquad B_{ext} = \{\mathrm{c},\mathrm{e},\mathrm{i},\mathrm{j},\mathrm{k},\mathrm{l},\mathrm{m},\mathrm{n}\}.$$

For each internal simple branch α, we denote by n_α the number of descendant species at the end of the branch (at time t_α^e). One of these descendant species is

represented in the tree. For the other $n_\alpha - 1$ descendant species we keep track of the number of descendants at t_p; we denote this number by m_α. For each boundary simple branch β, we denote by m_β the number of descendants at t_p in addition to the species represented in the tree (hence, $1 + m_\beta$ descendant species in total).

The numbers n_α and m_α allow us to write down the likelihood

$$\begin{aligned}\mathscr{L} = \lambda^s \sum_{n_\alpha | \alpha \in B_{int}} \sum_{m_\alpha | \alpha \in B_{int}} \Big(\prod_{\alpha \in B_{int}} \\ P(1, n_\alpha; t_\alpha^b, t_\alpha^e)\, n_\alpha\, P(n_\alpha - 1, m_\alpha; t_\alpha^e, t_p)\, (1-\rho)^{m_\alpha} \Big) \\ \times \sum_{m_\beta | \beta \in B_{ext}} \Big(\prod_{\beta \in B_{ext}} P(1, 1+m_\beta; t_\beta^b, t_p)\, (1+m_\beta)\, \rho\, (1-\rho)^{m_\beta} \Big)\end{aligned} \tag{4.6.2}$$

where we used short-hand notation for the multidimensional sums, e.g.,

$$\sum_{n_\alpha | \alpha \in B_{int}} \quad \text{stands for} \quad \sum_{n_{\alpha_1}} \sum_{n_{\alpha_2}} \cdots \sum_{n_{\alpha_s}} \quad \text{assuming} \quad B_{int} = \{\alpha_1, \alpha_2, \ldots, \alpha_s\}$$

Eq. 4.6.2 can be understood as follows. The product on the first line corresponds to internal simple branches. The term for branch α contains the probability of having n_α descendants at t_α^e, a factor n_α related to the selection of the descendant species that is represented in the tree, the probability that the other $n_\alpha - 1$ species at t_α^e have m_α descendants at t_p, and the probability that none of the latter descendants is sampled. The product on the second line corresponds to boundary simple branches. The term for branch β contains the probability of having $1 + m_\beta$ descendants at t_p, a factor $1 + m_\beta$ related to the selection of the sampled descendant species, and the probability of sampling this species and not sampling the other species.

Eq. 4.6.2 can be simplified by combining simple branches. In particular, the terms relating to an internal branch can be incorporated in the terms relating to the boundary branch to which it is connected. For example, referring to Fig. 4.6 (left panel), consider branches 'a' and 'e', which are an internal and a boundary simple branch, respectively. The terms in eq. 4.6.2 associated with branches 'a' and 'e'

are

$$\begin{aligned}
&\sum_{n_\mathrm{a}}\sum_{m_\mathrm{a}} P(1,n_\mathrm{a};t_\mathrm{a}^b,t_\mathrm{a}^e)\, n_\mathrm{a}\, P(n_\mathrm{a}-1,m_\mathrm{a};t_\mathrm{a}^e,t_p)\,(1-\rho)^{m_\mathrm{a}} \\
&\times \sum_{m_\mathrm{e}} P(1,1+m_\mathrm{e};t_\mathrm{e}^b,t_p)\,(1+m_\mathrm{e})\,\rho\,(1-\rho)^{m_\mathrm{e}} \\
&= \sum_{n_\mathrm{a},m_\mathrm{a},m_\mathrm{e}} P(1,n_\mathrm{a};t_\mathrm{a}^b,t_\mathrm{a}^e)\, n_\mathrm{a}\, P(n_\mathrm{a}-1,m_\mathrm{a};t_\mathrm{a}^e,t_p)\,(1-\rho)^{m_\mathrm{a}} \\
&\qquad\qquad P(1,1+m_\mathrm{e};t_\mathrm{e}^b,t_p)\,(1+m_\mathrm{e})\,\rho\,(1-\rho)^{m_\mathrm{e}} \\
&= \sum_{m_\mathrm{ae}} P(1,1+m_\mathrm{ae};t_\mathrm{ae}^b,t_p)\,(1+m_\mathrm{ae})\,\rho\,(1-\rho)^{m_\mathrm{ae}}, \qquad (4.6.3)
\end{aligned}$$

where we have applied eq. 4.6.1 in the last line. We see that the part of the likelihood corresponding to the composed branch 'ae' (composed of simple branches 'a' and 'e', with $t_\mathrm{ae}^b = t_\mathrm{a}^b$ and $t_\mathrm{ae}^e = t_\mathrm{e}^e = t_p$, where $m_{ae} = m_a + m_e$) has the same form as a boundary simple branch β in eq. 4.6.2. Hence, we can absorb an internal simple branch in the boundary simple branch to which it is connected.

By repeatedly absorbing internal simple branches, we obtain boundary branches that are increasingly composed, until there are no internal simple branches left in eq. 4.6.2. Denoting the resulting set of boundary branches by I, we obtain

$$\mathscr{L} = \lambda^s \sum_{m_\alpha|\alpha\in I} \prod_{\alpha\in I} P(1,1+m_\alpha;t_\alpha^b,t_p)\,(1+m_\alpha)\,\rho\,(1-\rho)^{m_\alpha} \qquad (4.6.4)$$

and for the case where all species are sampled (set $\rho = 1$)

$$\mathscr{L} = \lambda^s \prod_{\alpha\in I} P(1,1;t_\alpha^b,t_p). \qquad (4.6.5)$$

This is the "breaking the tree" likelihood of Nee, May, and Harvey (1994), see eq. 4.5.4.

Note that there are several ways of combining simple branches into composed ones. For example, for the tree shown in Fig. 4.6 (left panel), two possible sets of boundary branches are

$$\{\mathrm{ae},\mathrm{fm},\mathrm{n},\mathrm{bc},\mathrm{dgk},\mathrm{l},\mathrm{hi},\mathrm{j}\} \quad \text{and} \quad \{\mathrm{afn},\mathrm{m},\mathrm{e},\mathrm{bdhj},\mathrm{i},\mathrm{gl},\mathrm{k},\mathrm{c}\}.$$

However, these sets lead to the same value of likelihoods 4.6.4 and 4.6.5. In fact, the likelihoods only depend on the branching times, and the latter do not depend on the specific choice of composed branches.

4.6.3 Case of a single rate shift

The "breaking the tree" approach can also be used to construct the likelihood of a tree with a rate shift. For the rate-shift model described in the main text, we have to divide by the total number of species extant at the rate shift time t_s. Hence, we have to keep track of the total number of species at t_s.

First, note that the subclade with the rate shift can be dealt with separately from the main clade. For the subclade we can apply the likelihood formula for a tree without rate shift. The subclade tree starts at the rate-shift time t_s and continues until the present time t_p. Here we derive the likelihood formula for the main clade. For the example phylogeny of Fig. 4.6 (right panel) the main clade corresponds to the entire tree except branches $\{\mathrm{h},\mathrm{q},\mathrm{r}\}$.

We decompose the main clade into simple branches. In the case of a rate shift, all simple branches are split at the rate shift time, see Fig. 4.6 (right panel). We distinguish internal simple branches before the rate shift (with $t_\alpha^e < t_s$; set denoted by $B_{int}^{(M,1)}$), boundary simple branches before the rate shift (with $t_\alpha^e = t_s$; set denoted by $B_{ext}^{(M,1)}$), internal simple branches after the rate shift (with $t_s < t_\alpha^e < t_p$; set denoted by $B_{int}^{(M,2)}$) and boundary simple branches after the rate shift (with $t_\alpha^e = t_p$; set denoted by $B_{ext}^{(M,2)}$). For the example of Fig. 4.6 (right panel),

$$B_{int}^{(M,1)} = \{\mathrm{a},\mathrm{b}\} \quad B_{ext}^{(M,1)} = \{\mathrm{c},\mathrm{d},\mathrm{e},\mathrm{f}\} \quad B_{int}^{(M,2)} = \{\mathrm{j},\mathrm{k},\mathrm{l}\} \quad B_{ext}^{(M,2)} = \{\mathrm{g},\mathrm{i},\mathrm{m},\mathrm{n},\mathrm{o},\mathrm{p}\}$$

As before, we introduce the numbers n_α and m_α for internal simple branch α, and m_β for boundary simple branch β. For an internal branch, n_i stands for the number of descendant species at t_α^e. Only one of the n_α species is represented in the phylogeny; the other $n_\alpha - 1$ species have m_α descendant species at time t_s (for a branch before the rate shift) or at time t_p (for a branch after the rate shift). Similarly, for a boundary branch, m_β stands for the number of descendant species not represented in the phylogeny at time t_s (for a branch before the rate shift) or at time t_p (for a branch after the rate shift).

Using the numbers n_α and m_α we can construct the likelihood for the main

clade

$$\begin{aligned}
\mathscr{L}_M = \lambda_M^s & \sum_{n_\alpha^1 | \alpha \in B_{int}^{(M,1)}} \sum_{m_\alpha^1 | \alpha \in B_{int}^{(M,1)}} \\
& \Big(\prod_{\alpha \in B_{int}^{(M,1)}} P_M(1, n_\alpha^1; t_\alpha^b, t_\alpha^e)\, n_\alpha^1\, P_M(n_\alpha^1 - 1, m_\alpha^1; t_\alpha^e, t_s) \Big) \\
& \times \sum_{m_\beta^1 | \beta \in B_{ext}^{(M,1)}} \Big(\prod_{\beta \in B_{ext}^{(M,1)}} P_M(1, 1 + m_\beta^1; t_\beta^b, t_s)\, (1 + m_\beta^1) \Big) \\
& \times \frac{1}{k_s^1 + \sum_\alpha m_\alpha^1 + \sum_\beta m_\beta^1} \sum_{m_s} P_M(\textstyle\sum_\alpha m_\alpha^1 + \sum_\beta m_\beta^1, m_s; t_s, t_p)(1-\rho)^{m_s} \\
& \times \sum_{n_\alpha^2 | \alpha \in B_{int}^{(M,2)}} \sum_{m_\alpha^2 | \alpha \in B_{int}^{(M,2)}} \\
& \Big(\prod_{\alpha \in B_{int}^{(M,2)}} P_M(1, n_\alpha^2; t_\alpha^b, t_\alpha^e)\, n_\alpha^2\, P_M(n_\alpha^2 - 1, m_\alpha^2; t_\alpha^e, t_p)\, (1-\rho)^{m_\alpha^2} \Big) \\
& \times \sum_{m_\beta^2 | \beta \in B_{ext}^{(M,2)}} \Big(\prod_{\beta \in B_{ext}^{(M,2)}} P_M(1, 1 + m_\beta^2; t_\beta^b, t_p)\, (1 + m_\beta^2)\, \rho\, (1-\rho)^{m_\beta^2} \Big).
\end{aligned} \tag{4.6.6}$$

The first and second line impose the branching times before the rate shift, while keeping track of the total number of species at the rate shift. This number is equal to $k_s^1 + \sum_{\alpha \in B_{int}^{(M,1)}} m_\alpha^1 + \sum_{\beta \in B_{ext}^{(M,1)}} m_\beta^1$, by which we divide in the third line, where $\sum_{\beta \in B_{ext}^{(M,1)}} 1 = k_s^1$ denotes the number of species represented in the phylogeny at the time of the shift (in Fig. 4.6 (right panel) $k_s^1 = 4$). The other factor on the third line imposes that the species that are not represented in the phylogeny, of which there are $\sum_{\alpha \in B_{int}^{(M,1)}} m_\alpha^1 + \sum_{\beta \in B_{ext}^{(M,1)}} m_\beta^1$, do not have sampled species. The fourth and fifth line impose the branching times after the rate shift, and require that the species that are (not) represented in the phylogeny are (not) sampled.

As for the case without rate shift, the likelihood expression can be simplified. We absorb internal into boundary branches, until there are no internal branches left. This leads to a set of boundary branches before the rate shift, and a set of boundary branches after the rate shift, which we denote by $I^{(M,1)}$ and $I^{(M,2)}$,

respectively. For example, for the tree of Fig. 4.6 (right panel), these sets could be

$$I^{(M,1)} = \{\text{ae}, \text{f}, \text{bc}, \text{d}\} \qquad I^{(M,2)} = \{\text{g}, \text{i}, \text{jko}, \text{p}, \text{lm}, \text{n}\}.$$

Other ways of combining branches are possible, leading to different sets $I^{(M,1)}$ and $I^{(M,2)}$, but result in the same likelihood value,

$$\begin{aligned}\mathscr{L}_M = \lambda_M^s & \sum_{m_\gamma^1 | \gamma \in I^{(M,1)}} \Big(\prod_{\gamma \in I^{(M,1)}} P_M(1, 1+m_\gamma^1; t_\gamma^b, t_s)\,(1+m_\gamma^1) \Big) \\ & \times \frac{1}{k_s^1 + \sum_\gamma m_\gamma^1} \sum_{m_s} P_M(\textstyle\sum_\gamma m_\gamma^1, m_s; t_s, t_p)(1-\rho)^{m_s} \\ & \times \sum_{m_\gamma^2 | \gamma \in I^{(M,2)}} \Big(\prod_{\gamma \in I^{(M,2)}} P_M(1, 1+m_\gamma^2; t_\gamma^b, t_p)\,(1+m_\gamma^2)\,\rho\,(1-\rho)^{m_\gamma^2} \Big).\end{aligned} \tag{4.6.7}$$

If all species are sampled, we get (by setting $\rho = 1$)

$$\begin{aligned}\mathscr{L}_M = \lambda_M^s & \sum_{m_\gamma | \gamma \in I^{(M,1)}} \Big(\prod_{\gamma \in I^{(M,1)}} P_M(1, 1+m_\gamma; t_\gamma^b, t_s)\,(1+m_\gamma) \Big) \\ & \times \frac{1}{k_s^1 + \sum_\gamma m_\gamma} P_M(\textstyle\sum_\gamma m_\gamma, 0; t_s, t_p) \\ & \times \prod_{\gamma \in I^{(M,2)}} P_M(1, 1; t_\gamma^b, t_p).\end{aligned} \tag{4.6.8}$$

This is the likelihood formula reported in the main text, see eq. 4.2.14.

Case of multiple rate shifts

Consider two rate shifts at times t_s^1 and t_s^2. We distinguish two cases. First, we assume that the second rate shift occurs in the subclade with the first rate shift (Fig. 4.7, left panel). The corresponding likelihood can be readily constructed from the single rate-shift formula. Indeed, because the main-clade diversification dynamics after t_s^1 are unaffected by the second rate shift, the part of the likelihood dealing with the main clade follows directly from the one for a single rate shift. Similarly, because the subclade diversification dynamics are unaffected by the main clade, also the part of the likelihood dealing with the subclade follows directly from the one for a single rate shift.

Here we work out the second, more complicated case, in which the second rate shift occurs in the main clade (Fig. 4.7, right panel). We denote the sets of simple

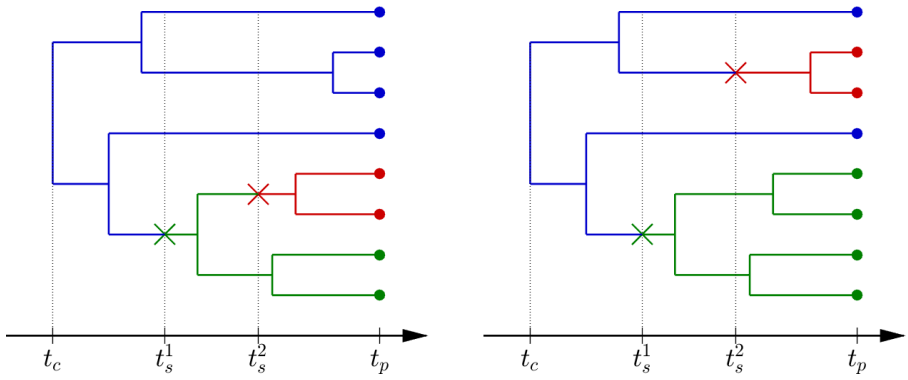

Figure 4.7: Example phylogenies with two rate shifts. Colors indicate the different rate regimes: main clade in blue, subclade initiated by first rate shift in green, subclade initiated by second rate shift in red. Left panel: The second rate shift occurs in the subclade of the first rate shift. The likelihood of this phylogeny is the product of the one-shift likelihood for the main clade, the one-shift likelihood for the first subclade, and the no-shift likelihood for the second subclade. Right panel: The second rate shift occurs in the main clade. The total likelihood is the product of the two-shifts likelihood for the main clade, the no-shift likelihood for the first subclade, and the no-shift likelihood for the second subclade.

branches by $B_{int}^{(M,1)}$ and $B_{ext}^{(M,1)}$ (branches before t_s^1), $B_{int}^{(M,2)}$ and $B_{ext}^{(M,2)}$ (branches between t_s^1 and t_s^2) and $B_{int}^{(M,3)}$ and $B_{ext}^{(M,3)}$ (branches after t_s^2). We again introduce numbers n_α and m_α to keep track of the total number of species at the two rate shifts. Then, the part of the likelihood relating to the main clade (i.e. the blue clade in Fig. 4.7) is

$$
\begin{aligned}
\mathscr{L}_M = \lambda_M^s & \sum_{n_\alpha^1 | \alpha \in B_{int}^{(M,1)}} \sum_{m_\alpha^1 | \alpha \in B_{int}^{(M,1)}} \\
& \Big(\prod_{\alpha \in B_{int}^{(M,1)}} P_M(1, n_\alpha^1; t_\alpha^b, t_\alpha^e)\, n_\alpha^1\, P_M(n_\alpha^1 - 1, m_\alpha^1; t_\alpha^e, t_s^1) \Big) \\
& \times \sum_{m_\beta^1 | \beta \in B_{ext}^{(M,1)}} \Big(\prod_{\beta \in B_{ext}^{(M,1)}} P_M(1, 1 + m_\beta^1; t_\beta^b, t_s^1)\, (1 + m_\beta^1) \Big) \\
& \times \frac{1}{k_s^1 + \sum_\alpha m_\alpha^1 + \sum_\beta m_\beta^1} \sum_{m_s^1} P_M(\textstyle\sum_\alpha m_\alpha^1 + \sum_\beta m_\beta^1, m_s^1; t_s^1, t_s^2) \\
& \times \sum_{n_\alpha^2 | \alpha \in B_{int}^{(M,2)}} \sum_{m_\alpha^2 | \alpha \in B_{int}^{(M,2)}} \\
& \Big(\prod_{\alpha \in B_{int}^{(M,2)}} P_M(1, n_\alpha^2; t_\alpha^b, t_\alpha^e)\, n_\alpha^2\, P_M(n_\alpha^2 - 1, m_\alpha^2; t_\alpha^e, t_s^2) \Big) \\
& \times \sum_{m_\beta^2 | \beta \in B_{ext}^{(M,2)}} \Big(\prod_{\beta \in B_{ext}^{(M,2)}} P_M(1, 1 + m_\beta^2; t_\beta^b, t_s^2)\, (1 + m_\beta^2) \Big) \\
& \times \frac{1}{k_s^2 + m_s^1 + \sum_\alpha m_\alpha^2 + \sum_\beta m_\beta^2} \\
& \times \sum_{m_s^2} P_M(m_s^1 + \textstyle\sum_\alpha m_\alpha^2 + \sum_\beta m_\beta^2, m_s^2; t_s^2, t_p)\, (1 - \rho)^{m_s^2} \\
& \times \sum_{n_\alpha^3 | \alpha \in B_{int}^{(M,3)}} \sum_{m_\alpha^3 | \alpha \in B_{int}^{(M,3)}} \\
& \Big(\prod_{\alpha \in B_{int}^{(M,3)}} P_M(1, n_\alpha^3; t_\alpha^b, t_\alpha^e)\, n_\alpha^3\, P_M(n_\alpha^3 - 1, m_\alpha^3; t_\alpha^e, t_p)\, (1 - \rho)^{m_\alpha^3} \Big) \\
& \times \sum_{m_\beta^3 | \beta \in B_{ext}^{(M,3)}} \Big(\prod_{\beta \in B_{ext}^{(M,3)}} P_M(1, 1 + m_\beta^3; t_\beta^b, t_p)\, (1 + m_\beta^3)\, \rho\, (1 - \rho)^{m_\beta^3} \Big).
\end{aligned}
\tag{4.6.9}
$$

The latter expression can be simplified by incorporating internal branches into

longer boundary branches. Denoting the resulting sets of boundary branches by $I^{(M,1)}$, $I^{(M,2)}$ and $I^{(M,3)}$, we obtain

$$\begin{aligned}\mathscr{L}_M = \lambda_M^s & \sum_{m_i^1 | i\in I^{(M,1)}} \Big(\prod_{i\in I^{(M,1)}} P_M(1,1+m_i^1;t_i^b,t_s^1)\,(1+m_i^1)\Big) \\ & \times \frac{1}{k_s^1+\sum_i m_i^1} \sum_{m_s^1} P_M(\textstyle\sum_i m_i^1, m_s^1; t_s^1, t_s^2) \\ & \times \sum_{m_i^2 | i\in I^{(M,2)}} \Big(\prod_{i\in I^{(M,2)}} P_M(1,1+m_i^2;t_i^b,t_s^2)\,(1+m_i^2)\Big) \\ & \times \frac{1}{k_s^2+m_s^1+\sum_i m_i^2} \sum_{m_s^2} P_M(m_s^1+\textstyle\sum_i m_i^2, m_s^2; t_s^2, t_p)\,(1-\rho)^{m_s^2} \\ & \times \sum_{m_i^3 | i\in I^{(M,3)}} \Big(\prod_{i\in I^{(M,3)}} P_M(1,1+m_i^3;t_i^b,t_p)(1+m_i^3)\,\rho\,(1-\rho)^{m_i^3}\Big).\end{aligned} \tag{4.6.10}$$

If all species are sampled, we get

$$\begin{aligned}\mathscr{L}_M = \lambda_M^s & \sum_{m_i^1 | i\in I^{(M,1)}} \Big(\prod_{i\in I^{(M,1)}} P_M(1,1+m_i^1;t_i^b,t_s^1)\,(1+m_i^1)\Big) \\ & \times \frac{1}{k_s+\sum_i m_i^1} \sum_{m_s^1} P_M(\textstyle\sum_i m_i^1, m_s^1; t_s^1, t_s^2) \\ & \times \sum_{m_i^2 | i\in I^{(M,2)}} \Big(\prod_{i\in I^{(M,2)}} P_M(1,1+m_i^2;t_i^b,t_s^2)\,(1+m_i^2)\Big) \\ & \times \frac{1}{k_s+m_s^1+\sum_i m_i^2} P_M(m_s^1+\textstyle\sum_i m_i^2, 0; t_s^2, t_p) \\ & \times \prod_{i\in I^{(M,3)}} P_M(1,1;t_i^b,t_p).\end{aligned} \tag{4.6.11}$$

4.7 Appendix D: Likelihood for unobserved rate shift

It is intuitively clear that the rate-shift model in which the rate shift has no effect, i.e., when rate-shifted rates (λ_S,μ_S) are equal to non-rate-shifted rates (λ_M,μ_M), should be connected to the model without rate shift. Here we show how the likelihood formula with rate shift should be combined to recover Nee et al.'s (1994) likelihood. More precisely, we prove that

$$\lim_{S\to M}\Big(\mathscr{L}_{\text{corr}}^{\text{obs}}+\mathscr{L}_{\text{corr}}^{\text{unobs}}\Big) = \prod_{i\in I(M^<)\cup I(M_j^>)\cup I(S_j)} P_M(1,1;t_i,t_p) \tag{4.7.1}$$

We start by noting that when the rate shift has no effect (i.e., when setting $S \to M$), the likelihoods $\mathscr{L}_{\text{corr}}^{\text{obs},\,j}$ and $\mathscr{L}_{\text{corr}}^{\text{unobs},\,j}$ can be rewritten as

$$\lim_{S\to M} \mathscr{L}_{\text{corr}}^{\text{obs},\,j} = \sum_{m_1=0}^{\infty} \cdots \sum_{m_{k_s}=0}^{\infty} \frac{1}{k_s + m_s} C(\mathbf{m}) \tag{4.7.2}$$

$$\lim_{S\to M} \mathscr{L}_{\text{corr}}^{\text{unobs},\,j} = \sum_{m_1=0}^{\infty} \cdots \sum_{m_{k_s}=0}^{\infty} \frac{m_j}{k_s + m_s} C(\mathbf{m}) \tag{4.7.3}$$

with common factor $C(\mathbf{m})$ given by

$$\begin{aligned} C(\mathbf{m}) = &\Big(\prod_{i\in I(M^<)} P_M(1, m_i+1; t_i, t_s)\,(m_i+1)\,P_M(m_i, 0; t_s, t_p) \Big) \\ &\times \Big(\prod_{i\in I(M_j^>)} P_M(1,1;t_i,t_p) \Big) \Big(\prod_{i\in I(S_j)} P_M(1,1;t_i,t_p) \Big). \end{aligned} \tag{4.7.4}$$

Hence,

$$\begin{aligned} \lim_{S\to M} \Big(\mathscr{L}_{\text{corr}}^{\text{obs}} + \mathscr{L}_{\text{corr}}^{\text{unobs}} \Big) &= \lim_{S\to M} \sum_{j\in I(M^<)} \mathscr{L}_{\text{corr}}^{\text{obs},\,j} + \sum_{j\in I(M^<)} \mathscr{L}_{\text{corr}}^{\text{unobs},\,j} \\ &= \sum_{m_1=0}^{\infty} \cdots \sum_{m_{k_s}=0}^{\infty} \sum_{j\in I(M^<)} \Big(\frac{1}{k_s+m_s} + \frac{m_j}{k_s+m_s} \Big) C(\mathbf{m}) \\ &= \sum_{m_1=0}^{\infty} \cdots \sum_{m_{k_s}=0}^{\infty} C(\mathbf{m}). \end{aligned} \tag{4.7.5}$$

Substituting the expression for $C(\mathbf{m})$,

$$\begin{aligned} &\lim_{S\to M} \Big(\mathscr{L}_{\text{corr}}^{\text{obs}} + \mathscr{L}_{\text{corr}}^{\text{unobs}} \Big) \\ &\quad = \Big(\sum_{m_1=0}^{\infty} \cdots \sum_{m_{k_s}=0}^{\infty} \prod_{i|t_i\le t_s} P_M(1, m_i+1; t_i, t_s)\,(m_i+1)\,P_M(1,1;t_s,t_p) \\ &\qquad\qquad P_M(m_i, 0; t_s, t_p) \Big) \times \Big(\prod_{i|t_i>t_s} P_M(1,1;t_i,t_p) \Big) \\ &\quad = \Big(\prod_{i|t_i\le t_s} P_M(1,1;t_i,t_p) \Big) \times \Big(\prod_{i|t_i>t_s} P_M(1,1;t_i,t_p) \Big) \\ &\quad = \prod_{i\in I(M^<)\cup I(M_j^>)\cup I(S_j)} P_M(1,1;t_i,t_p) \end{aligned} \tag{4.7.6}$$

where in the second equality we have applied eq. 4.6.1 with $n_2 = 1$ (and hence $n_{2a} = 1$ and $n_{2b} = 0$). This concludes the proof of eq. 4.7.1.

4.8 Appendix E: Rate shifts in diversity-dependent model

The corrected likelihood $\mathscr{L}_{\text{corr}}^{\text{obs},j}$, eq. 4.2.14, can be extended to the case of diversity-dependent diversification. Here we describe how this can be done within the framework of Etienne et al., 2012. In this approach the total number of species, including the unobserved ones, is tracked through time. Therefore, the likelihood correction of eq. 4.2.14, which basically consists in the division by the total number of species present at the rate shift, can be readily implemented.

The framework of Etienne et al., 2012 is built on the quantities $Q_m^k(t)$, the probabilities that the diversification process is consistent with the observed phylogeny from the starting time (typically crown age t_c) to the current time t, with k visible (i.e. represented in the phylogeny) and m invisible (i.e. not represented in the phylogeny, as they will go extinct before the present time t_p or they are unsampled in the data) species at time t. The probabilities $Q_m^k(t)$ are computed forward in time, from t_c to t_p. We introduce the vector $\mathbf{Q}^k(t)$ with components $Q_m^k(t)$ for all m. Then, for a phylogeny with k_p tips and without rate shift,

$$\begin{aligned}\mathbf{Q}^{k_p}(t_p) = A_{k_p}(t_{k_p-1},t_p)\,B_{k_p-1,k_p}\,A_{k_p-1}(t_{k_p-2},t_{k_p-1})\ldots \\ A_4(t_3,t_4)\,B_{3,4}\,A_3(t_2,t_3)\,B_{2,3}\,A_2(t_c,t_2)\,\mathbf{Q}^2(t_c).\end{aligned} \tag{4.8.1}$$

Starting from the initial vector $\mathbf{Q}^2(t_c)$ at crown age, we repeatedly apply matrices $A_k(t_{k-1},t_k)$ between branching times t_{k-1} and t_k and matrices $B_{k,k+1}$ at branching time t_k. Here the branching times are time-ordered (i.e., $t_c < t_2 < t_3 < \ldots < t_{k_p-1} < t_p$), so that at branching time t_k the phylogeny transits from k to $k+1$ branches. The (unconditioned) likelihood of the tree under the diversity-dependent diversification model is then obtained as the $m = 0$ component of $\mathbf{Q}^{k_p}(t_p)$.

The matrices A_k appearing in eq. 4.8.1 is obtained by solving the linear differential equation

$$\frac{d\mathbf{Q}^k(t)}{dt} = T_k\mathbf{Q}^k(t), \tag{4.8.2}$$

so that $A_k(t_{k-1},t_k) = \exp(T_k(t_k - t_{k-1}))$, with the matrices T_k given by

$$T_k = \begin{bmatrix} -k(\mu_k+\lambda_k) & \mu_{k+1} & 0 & \ldots \\ 2k\lambda_k & -(k+1)(\mu_{k+1}+\lambda_{k+1}) & 2\mu_{k+2} & \ldots \\ 0 & (2k+1)\lambda_{k+1} & -(k+2)(\mu_{k+2}+\lambda_{k+2}) & \ldots \\ 0 & 0 & (2k+2)\lambda_{k+2} & \ldots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix}.$$

(4.8.3)

Here the speciation rates λ_n and extinction rates μ_n can depend on the number of extant species $n = k+m$ at the event times. The matrices $B_{k,k+1}$ are given by

$$B_{k,k+1} = \begin{bmatrix} \lambda_k & 0 & 0 & 0 & \dots \\ 0 & \lambda_{k+1} & 0 & 0 & \dots \\ 0 & 0 & \lambda_{k+2} & 0 & \dots \\ 0 & 0 & 0 & \lambda_{k+3} & \dots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix}. \tag{4.8.4}$$

For the case of a lineage-specific rate shift, we apply the same computation from t_c to the rate-shift time t_s. At the rate shift we choose the species that is undergoing the rate shift, which corresponds to dividing by the total number of species (both visible and invisible) extant at t_s. We then continue the computation from t_s to t_p, taking into account that the main clade has lost one branch at the rate shift. Hence, eq. 4.8.1 has to be modified to

$$\begin{aligned} \mathbf{Q}^{k_p}(t_p) = {} & A_{k_p}(t_{k_p}, t_p)\, B_{k_p-1,k_p}\, A_{k_p-1}(t_{k_p-1}, t_{k_p}) \dots \\ & A_{k_s-1}(t_s, t_{k_s})\, C_{k_s,k_s-1}\, A_{k_s}(t_{k_s-1}, t_s) \dots \\ & A_3(t_2, t_3)\, B_{2,3}\, A_2(t_c, t_2)\, \mathbf{Q}^2(t_c) \end{aligned} \tag{4.8.5}$$

with the matrices $C_{k,k-1}$ given by

$$C_{k,k-1} = \begin{bmatrix} \frac{1}{k} & 0 & 0 & 0 & \dots \\ 0 & \frac{1}{k+1} & 0 & 0 & \dots \\ 0 & 0 & \frac{1}{k+2} & 0 & \dots \\ 0 & 0 & 0 & \frac{1}{k+3} & \dots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix}. \tag{4.8.6}$$

Note that at branching time $t_k < t_s$ the main-clade phylogeny goes from k to $k+1$ branches, and at branching time $t_k > t_s$ it goes from $k-1$ to k branches.

The likelihood contribution $\mathscr{L}_M$ of the main clade is obtained as the $m = 0$ component of $\mathbf{Q}^{k_p}(t_p)$ computed with eq. 4.8.5. The likelihood contribution $\mathscr{L}_S$ of the subclade can be computed with eq. 4.8.1 starting at the rate-shift time t_s. By multiplying these two contributions we get the (unconditioned) corrected likelihood for a phylogeny with an observed rate shift in a specific branch j

$$\mathscr{L}_{\text{corr}}^{\text{obs},\,j} = \mathscr{L}_M\, \mathscr{L}_S. \tag{4.8.7}$$

Using techniques of Laudanno, Haegeman, and Etienne, 2020 it can be proven that this formula reduces to the likelihood 4.2.14 in the diversity-independent case.

In the same way as for the likelihood $\mathscr{L}_{\text{corr}}^{\text{obs},j}$ of an observed rate shift, diversity-dependent extensions can be derived for

- the likelihood $\mathscr{L}_{\text{corr}}^{\text{unobs}}$ of an unobserved rate shift. Note however that in the diversity-dependent case a dummy rate shift (i.e., rates before and after rate shift are the same) does have an effect on the diversification dynamics, because the subclade is not subjected to the diversity dependence of the main clade. Hence, there is no diversity-dependent extension of eq. 4.7.1;

- the conditioning probabilities $P_{c,0}$, $P_{c,1}$ and $P_{c,2}$, see eqs. 4.2.15, 4.2.16 and 4.2.17 of the main text; see also Etienne and Haegeman, 2012 where explicit expressions are presented for the diversity-dependent analogue of eq. 4.2.15;

- the likelihood for multiple rate shifts, e.g., eqs. 4.6.10 and 4.6.11 for the case of two rate shifts in the main clade. In fact, while the explicit formulas are cumbersome for multiple rate shifts, the framework of Etienne et al., 2012 with the vector $\mathbf{Q}$ can easily incorporate an arbitrary number of rate shifts and can be used to numerically evaluate the multiple-shifts likelihood. This has been implemented in version 4.3 in the R package DDD (Etienne and Haegeman, 2020).