

Unifying Phylogenetic Birth–Death Models in Epidemiology and Macroevolution

AILENE MACPHERSON^{1,2,3,*}, STILIANOS LOUCA^{4,5}, ANGELA McLAUGHLIN^{6,7}, JEFFREY B. JOY^{6,7,8}
AND MATTHEW W. PENNELL¹

¹Zoology, University of British Columbia, Vancouver V6T 1Z4, Canada; ²Ecology and Evolutionary Biology, University of Toronto, Toronto M5S 3B2, Canada; ³Mathematics, Simon Fraser University, Burnaby V5A 1S6, Canada; ⁴Biology, University of Oregon, Eugene 97403, USA; ⁵Institute of Ecology and Evolution, University of Oregon, Eugene 97403, USA; ⁶British Columbia Centre for Excellence in HIV/AIDS, Vancouver V6Z 1Y6, Canada;

⁷Bioinformatics, University of British Columbia, Vancouver V6T 1Z4, Canada; ⁸Medicine, University of British Columbia, Vancouver V6T 1Z4, Canada

*Correspondence to be sent to: Ecology and Evolutionary Biology, University of Toronto, Toronto M5S 3B2, Canada;
E-mail: ailene_macpherson@sfu.ca.

Received 8 March 2021; reviews returned 9 June 2021; accepted 21 June 2021

Associate Editor: James Albert

Abstract.—Birth–death stochastic processes are the foundations of many phylogenetic models and are widely used to make inferences about epidemiological and macroevolutionary dynamics. There are a large number of birth–death model variants that have been developed; these impose different assumptions about the temporal dynamics of the parameters and about the sampling process. As each of these variants was individually derived, it has been difficult to understand the relationships between them as well as their precise biological and mathematical assumptions. Without a common mathematical foundation, deriving new models is nontrivial. Here, we unify these models into a single framework, prove that many previously developed epidemiological and macroevolutionary models are all special cases of a more general model, and illustrate the connections between these variants. This unification includes both models where the process is the same for all lineages and those in which it varies across types. We also outline a straightforward procedure for deriving likelihood functions for arbitrarily complex birth–death(–sampling) models that will hopefully allow researchers to explore a wider array of scenarios than was previously possible. By rederiving existing single-type birth–death sampling models, we clarify and synthesize the range of explicit and implicit assumptions made by these models. [Birth–death processes; epidemiology; macroevolution; phylogenetics; statistical inference.]

Evolutionary, demographic, and epidemiological processes leave a footprint in the branch length distribution and topology of reconstructed phylogenetic trees. This insight has inspired a huge effort to extract information about these processes by fitting stochastic models. For example, in molecular epidemiology, researchers have leveraged the fact that for many viral pathogens, such as HIV and SARS-CoV-2, accumulate genetic diversity on the timescale of transmission (Drummond et al. 2003; Duffy et al. 2008). This genetic diversity can be used to reconstruct the evolutionary relationships between viral variants sampled from different hosts, which in turn can help elucidate the epidemiological dynamics of a pathogen over time (Grenfell et al. 2004; Volz 2012). Similarly, phylogenetic trees can provide unique insights into the temporal variation in speciation and extinction rates (Morlon 2014).

Phylogenetic branching models can be broadly grouped into two classes. The first, based on Kingman’s coalescent process (Kingman 1982), has been widely used to examine changes in the historical population size of pathogens (Pybus et al. 2000; Strimmer and Pybus 2001; Drummond et al. 2005; Volz et al. 2009). These coalescent methods have also been applied to reconstruct macroevolutionary dynamics (Morlon et al. 2010). Coalescent models are well suited for estimating deterministic population dynamics; however, fitting highly stochastic processes, such as the dynamics of an emerging pathogen, is computationally intensive and in some cases the assumptions of the coalescent may not be appropriate (Stadler et al. 2015; Boskova et al. 2014;

Volz and Frost 2014). The second class of models, which are collectively referred to as birth–death–sampling (BDS) models (Kendall 1948; Maddison et al. 2007; Stadler 2009; Stadler 2010), is well suited for stochastic scenarios, and are thus becoming an increasingly favorable and popular alternative to coalescent models in epidemiology (Stadler et al. 2012) and have long been the foundation of most macroevolutionary studies—both for inferring speciation and extinction dynamics (Raup 1985; Nee et al. 1994) and for estimating divergence times (Gernhard 2008; Heath et al. 2014). As the name implies, the BDS process includes three types of events: birth (pathogen transmission between hosts, or speciation in a macroevolutionary context), death (host death or recovery, or extinction in macroevolution), and sampling (including fossil collection in macroevolution).

In the context of epidemiology, BDS models have the additional property that the model parameters, which can be estimated from viral sequence data, explicitly correspond to parameters in classic structured epidemiological models that are often fit to case surveillance data. If we reparameterize these models, we can describe the dynamics of the basic and effective reproductive ratios (R_0 and R_e , respectively) over time (Stadler et al. 2012; Stadler et al. 2013) (see Box 1). A common research aim is to describe how the frequency of birth, death, and sampling events, and other derived variables such as R_e , change throughout the course of an epidemic. There has been less work in macroevolution linking the parameters of a BDS model to those of an underlying more mechanistic

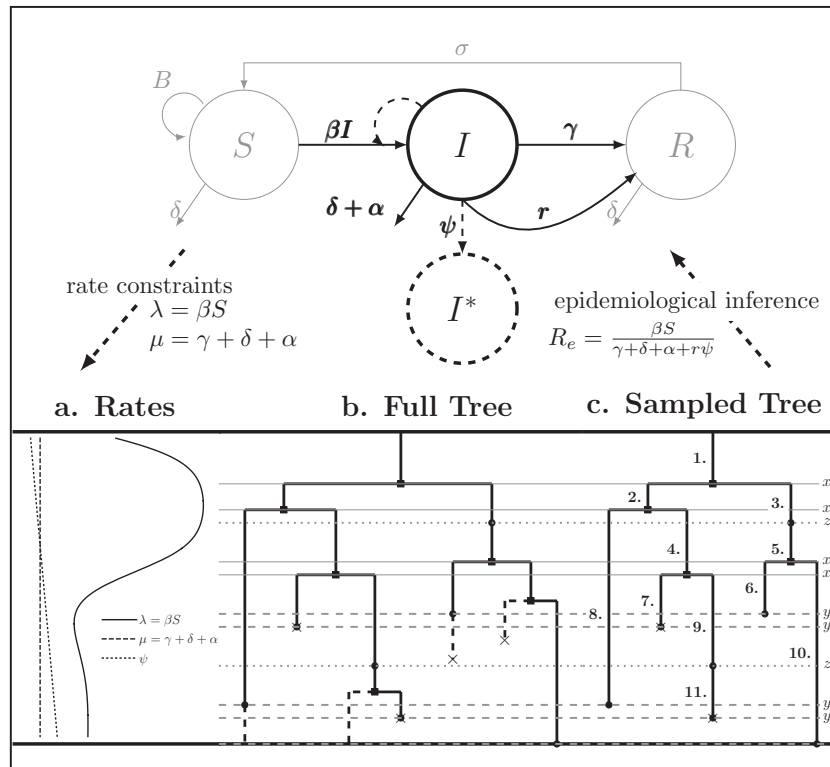


FIGURE B.1. BDS-SIR model connection. Top: The SIR epidemiological model. Black (gray) lines and classes represent rates and variables followed (in)directly by the BDS model. The SIR model can be used to constrain the rates of the BDS model (a). Simulated forward in time, the result of the BDS stochastic processes is a *full tree* (b) giving the complete genealogy of the viral population. Pruning away extinct and unsampled lineages produces the *sampled tree* (c). Arising from a BDS process, this sampled tree can be summarized in two ways. First by the set of edges (labeled 1–11) or as a set of critical times (horizontal lines) including: 1) the time of birth events (solid, x_i) 2) terminal sampling times (dashed, y_i), and 3) ancestral sampling times (dotted, z_k). Given the inferred rates from a reconstructed sampled tree, these rates can be used to estimate characteristic parameters of the SIR model, for example, the basic or effective reproductive number.

BOX 1: THE CONNECTION BETWEEN BDS AND SIR MODELS

The single-type BDS model is intimately related to the SIR compartmental model used in classic theoretical epidemiology. This connection illustrates the explicit and implicit assumptions of the general BDS model and its sub models. Here, we define the SIR epidemiological model, discuss how it can inform and be informed by these diversification models, and examine the shared assumptions of the two frameworks.

The SIR Model:

The SIR model partitions the host population via infection status into susceptible (S), infected (I), and recovered (R) hosts. Infection of susceptible hosts occurs at a per-capita rate βI . Infected hosts may recover (at rate γ), die of virulent cases (at rate α), or be sampled (at rate ψ). The cumulative number of sampled hosts is represented in the SIR model

(Figure B.1 top) by I^* . Upon sampling, infected hosts may be treated and hence effectively recover with probability r . Hosts that have recovered from infection exhibit temporary immunity to future infection which wanes at rate σ . The special case of the SIR model with no immunity (the SIS model) is obtained in the limit as $\sigma \rightarrow \infty$. In addition to these epidemiological processes, the SIR model includes demographic processes, such as host birth (rate B) and death from natural causes (rate δ). While not shown explicitly in the figure, these epidemiological and demographic rates may change over time as a result of host behavioral change, pharmaceutical and nonpharmaceutical interventions, or host/pathogen evolution.

The BDS Model:

The BDS model follows the number of sampled and unsampled viral lineages over time, analogous to the I and I^* classes of the SIR model. A key element of general BDS model is that birth and death rates may vary over time. This time dependence may

be either continuous (Morlon et al. 2011; Rabosky and Lovette 2008b) or discrete (Stadler 2011; Stadler and Bonhoeffer 2013; Gavryushkina et al. 2014; Kühnert et al. 2014). Although arbitrarily time-dependent, the birth, death, and sampling rates in the general BDS model are assumed to be diversity-independent, analogous to the assumption of density-dependent transmission (pseudo mass action) in the SIR model (Keeling and Rohani 2008). Incorporating such diversity dependence into macroevolutionary models has been shown to increase the accuracy of extinction rate estimates and are necessary to accurately capture the saturation of diversity (Etienne et al. 2012). While some forms of diversity-dependence in diversification rates may be incorporated implicitly capturing deterministic diversity dependence as time dependence (Rabosky and Lovette 2008a), stochastic diversity-dependence (Etienne and Rosindell 2012) goes beyond the scope of the BDS models considered here.

The single-type BDS model assumes all viral lineages are exchangeable—this has several implications. First, all viral lineages are epidemiologically identical hence all mutations between them are neutral. Incorporating nonneutral genetic variation requires a multitype approach as in Equation (16). Second, transmission is independent of lineage age. In the macroevolutionary case, such age-dependence has been suggested to reflect niche differentiation in novel species (Hagen et al. 2015) and in the epidemiological case may reflect adaptation towards increased transmissibility following a host species-jumping event. Third, lineage exchangeability is reflected in the absence of an exposed (E) class in the SIR model in which hosts can, for example, transmit infections but not be sampled or vice versa. Finally, the single-type BDS model assumes all lineages are sampled at random and does not include submodels with nonrandom representation of lineages (Stadler et al. 2012).

Model Connections

Given their shared model assumptions, the single-type BDS model can be constrained explicitly to reflect an underlying SIR epidemic by setting the viral birth rate equal to the per-capita transmission rate of the infectious class $\lambda(\tau) = \beta S(\tau)$ and the viral death rate to the infectious recovery or removal rate $\mu(\tau) = \gamma + \delta + \alpha$, whereas the sampling rate $\psi(\tau)$ is identical across models (Figure B1a). While constraining the birth, death, and sampling rates in this manner can be used to parameterize compartmental models (Kühnert et al. 2014) doing

so is an approximation assuming independence between the exact timing of transmission, recovery or removal from the population, and sampling events in the SIR model and birth, death, and sampling events in the diversification model. The resulting tree likelihood in terms of the compartmental model is given by:

$$Pr(\mathcal{T}|\Theta_{\text{SIR}}) = \underbrace{Pr(\mathcal{T}|\Theta_{\text{BDS}})}_{\text{BDS likelihood}} \underbrace{P(\Theta_{\text{BDS}}|\Theta_{\text{SIR}})}_{\text{SIR process}}. \quad (\text{B1})$$

While they are not submodels of the general BDS process given by equation 13, likelihood models have been developed that capture the full nonindependence of viral diversification and epidemiological dynamics for the SIR model specifically (Leventhal et al. 2012) and in compartmental models in general (Vaughan et al. 2019). The connection between the BDS process and SIR epidemiological models can also be used after the diversification rates are inferred to estimate the basic and effective reproductive rates (Stadler et al. 2012; Stadler and Bonhoeffer 2013). Specifically, the effective reproductive rate at time τ before the present day is given by $R_e(\tau) = \frac{\lambda(\tau)}{\mu(\tau) + r(\tau)\psi(\tau)}$. Although the SIR model is a useful epidemiological model for its simplicity, realistically modeling epidemic dynamics requires far more complex compartmental models. As reflected by their shared structure, the application of the single-type BDS model is restricted, however, to the assumptions of the SIR model alone and further methodological advances in multitype modeling are necessary for direct inference for the larger class of epidemiological models.

model (but see Ezard et al. 2016) but this seems like a promising avenue for future development. As we detail below and in the [Supplementary material](https://doi.org/10.5281/zenodo.5028470) available at Dryad at <https://doi.org/10.5281/zenodo.5028470>, there has been an astounding rise in the variety and complexity of BDS model variants. A key assumption in the specification of BDS submodels is whether all lineages alive at some time point are exchangeable (Stadler 2013) (such models are hereafter “single-type” models), meaning they diversify according to the same process, or if rather the diversification process is variable (“multitype” models; e.g., Maddison et al. 2007; FitzJohn 2012; Stadler and Bonhoeffer 2013; Rasmussen and Stadler 2019; Barido-Sottani et al. 2018), with lineages belonging to one of multiple possible states each characterized by a unique process. Each of these diversification processes can then be characterized by different dynamical assumptions. In the epidemiological case, these assumptions specify, for example, the nature of viral transmission and the sampling procedure (Stadler et al. 2013; Kühnert et al. 2014;

Gavryushkina et al. 2014). While typically not explicitly tied to mechanistic evolutionary processes, there are a similar abundance of dynamical assumptions employed in the macroevolutionary context specifying the nature of biodiversity change through time (Nee 2006; Gernhard 2008; Morlon et al. 2011; Stadler 2011; Morlon 2014; Heath et al. 2014; Louca 2020).

This flourishing of methods and models has facilitated critical insights into epidemics (du Plessis and Stadler 2015; Joy et al. 2016) and the origins of contemporary biodiversity (Morlon 2014; Schluter and Pennell 2017). However, this diversity of models has made it difficult to trace the connections between variants and to understand the precise epidemiological, evolutionary, and sampling processes that differ between them. Furthermore, despite their apparent similarities, these models have been derived on a case-by-case basis using different notation and techniques; this creates a substantial barrier for researchers working to develop novel models for new situations. And critically, it is imperative that we understand the general properties of BDS phylogenetic models and the limits of inferences from them (Louca and Pennell 2020a; Louca et al. 2021), and this is difficult to do without considering the full breadth of possible scenarios. Here, we address all of these challenges by unifying the whole class of phylogenetic BDS models. We do so by first deriving a likelihood for general single- and multitype BDS models; in the general case, we do not assume anything about the functional forms (i.e., temporal dynamics) of the various parameters including the sampling rate through time, the possibility of sampling ancestors (or not), or how the process was conditioned. While such general models may be useful for studying the mathematical properties of BDS models as a whole (Lambert and Stadler 2013; Louca and Pennell 2020a; Louca et al. 2021), statistical inference from these models requires researchers to make further constraints on the process. We prove that existing BDS model variants are indeed submodels of the more general case—and thereby clarify the specific assumptions made by different models—and provide a standardized notation and technique for deriving these and other submodels that have not previously been considered in the literature.

The Single-type Birth–Death–Sampling Model

Model specification.—The BDS stochastic process begins with a single lineage at time T before the present day. We note that this may be considerably older than the age of the most recent common ancestor of an observed sample which is given by t_{MRCA} . While we focus primarily on applications to epidemiology, our approach is agnostic to whether the rates are interpreted as describing pathogen transmission or macroevolutionary diversification.

In the model, transmission/speciation results in the birth of a lineage and occurs at rate $\lambda(\tau)$, where τ ($0 \leq \tau \leq T$) is measured in units of time before the present day ($\tau = 0$), such that λ can be time-dependent. We make the common assumption that lineages in the viral phylogeny

coalesce exactly at transmission events, thus ignoring the within-host coalescent processes in the donor (Romero-Severson et al. 2016). Throughout, we will use τ as a general time variable and t_x to denote the time at which a specific event x occurs as measured in units of time before the present day (see Supplementary Table S1). Lineage extinction, resulting from host recovery or death in the epidemiological case or the death of all individuals in a population in the macroevolutionary case, occurs at time-dependent rate $\mu(\tau)$. We allow for two distinct types of sampling: lineages are either sampled according to a Poisson process through time $\psi(\tau)$ or binomially at very short intervals, which we term “concerted sampling attempts” (CSAs), where lineages at some specified time t_l are sampled with probability ρ_l ($\vec{\rho}$ denotes a vector of concerted sampling events at different time points). In macroevolutionary studies based only on extant lineages, there is no Poissonian sampling, but a CSA at the present ($\rho_0 > 0$). In epidemiology, CSAs correspond to large-scale testing efforts (relative to the background rate of testing) in a short amount of time (relative to the rates of viral sequence divergence); for full explanation, see Appendix. We call these attempts rather than events because if ρ is small or the infection is rare in the population, few or no samples may be obtained. CSAs can also be incorporated into the model by including infinitesimally short spikes in the sampling rate ψ (more precisely, appropriately scaled Dirac distributions). Hence, for simplicity, in the main text we focus on the seemingly simpler case of pure Poissonian sampling through time except at the present-day, where we allow for a CSA to facilitate comparisons with macroevolutionary models; the resulting formulas can then be used to derive a likelihood formula for the case where past CSAs are included (see Appendix).

In the epidemiological case, sampling may be concurrent (or not) with host treatment or behavioral changes resulting in the effective extinction of the viral lineage. Hence, we assume that sampling results in the immediate extinction of the lineage with probability $r(\tau)$. As with the CSAs, this arbitrary time dependence allows for the incorporation of Dirac spikes in any of these variables, for example, with mass extinctions (μ) and lagerstätten in the fossil record ($\psi(1-r)$) (Magee and Höhna 2021). Similarly, in the case of past CSAs, we must include the probability, r_l , that sampled hosts are removed from the infectious pool during the CSA at time t_l . Poissonian sampling without the removal of lineages ($r(\tau) < 1$) can be employed in the macroevolutionary case to explicitly model the collection of samples from the fossil record (such as the fossilized-birth–death process; Heath et al. 2014).

For our derivation, we make no assumption about the temporal dynamics of λ , μ , ψ , or r ; each may be constant, or vary according to any arbitrary function of time given that it is biologically valid (nonnegative and between 0 and 1 in the case of r). Specifically, the time-varying functions may be any piecewise-continuous functions of time with at most finite number of discontinuities (see Rate Assumption section in the Appendix). Note that these functions need not be differentiable. We make

the standard assumption that at any given time any given lineage experiences a birth, death or sampling event independently of (and with the same probabilities as) all other lineages. We revisit this assumption in Box 1 where we discuss how the implicit assumptions of the single-type BDS process are well summarized by the diversification model's relationship to the SIR epidemiological model. Our resulting general time-variable BDS process can be fully defined by the parameter set $\Theta_{\text{BDS}} = \{\lambda(\tau), \mu(\tau), \psi(\tau), r(\tau), \bar{\rho}\}$.

In order to make inference about the model parameters, we need to calculate the likelihood, \mathcal{L} , that an observed phylogeny, \mathcal{T} , is the result of a given BDS process. With respect to the BDS process there are two ways to represent the information contained in the phylogeny \mathcal{T} , both of which have been used in the literature, which we call the "edge" and "critical time" representations, respectively. We begin by deriving the likelihood in terms of the edge representation and later demonstrate how to reformulate the likelihood in terms of critical times. In the edge representation, the phylogeny is summarized as a set of edges in the mathematical graph that makes up the phylogeny, numbered 1–11 in Figure B.1c, and the types of events that occurred at each node. We define $g_e(\tau)$ as the probability that the edge e which begins at time s_e and ends at time t_e gives rise to the subsequently observed phylogeny between time τ , ($s_e < \tau < t_e$) and the present day. The likelihood of the model for the observed tree is then, is by definition $g_{\text{stem}}(\mathcal{T})$: the probability density that the stem lineage ($\text{stem} = 1$ in Figure B.1c) gives rise to the observed phylogeny from the origin, T , to the present day. We find that it is more intuitive to derive the likelihood in terms of the edge representation, as we show below; from this it is straightforward to derive the critical times formulation which results in mathematical simplification. Below, we present our five-step technique for the derivation of the tree likelihood.

Step 1. Deriving the Initial Value Problem (IVP) for $g_e(\tau)$.—We derive the IVP for the likelihood density $g_e(\tau)$ using an approach first developed by (Maddison et al., 2007). We begin by deriving the recursion equation for g_e by considering all the possible events that could occur along edge e between time τ and $\tau + \Delta\tau$ assuming that that $\Delta\tau$ is small enough such that at most one event is likely to occur.

$$\begin{aligned}
 g_e(\tau + \Delta\tau) \approx & \underbrace{(1 - \lambda(\tau)\Delta\tau)(1 - \mu(\tau)\Delta\tau)(1 - \psi(\tau)\Delta\tau)}_{\text{nothing happens}} \times g_e(\tau) \\
 & + \underbrace{\lambda(\tau)\Delta\tau(1 - \mu(\tau)\Delta\tau)(1 - \psi(\tau)\Delta\tau)}_{\text{birth event}} \times 2g_e(\tau)E(\tau) \\
 & + \underbrace{\mu(\tau)\Delta\tau(1 - \lambda(\tau)\Delta\tau)(1 - \psi(\tau)\Delta\tau)}_{\text{death event}} \times 0 \\
 & + \underbrace{\psi(\tau)\Delta\tau(1 - \lambda(\tau)\Delta\tau)(1 - \mu(\tau)\Delta\tau)}_{\text{sampling event}} \times 0 + \mathcal{O}(\Delta\tau^2).
 \end{aligned} \tag{1}$$

Here, $E(\tau)$ is the probability that a lineage alive at time τ leaves no sampled descendants at the present day. We will examine this probability in more detail below. Assuming $\Delta\tau$ is small, we can approximate the above recursion equation as the following difference equation.

$$\begin{aligned}
 \Delta g_e(\tau) \approx & -(\lambda(\tau) + \mu(\tau) + \psi(\tau))\Delta\tau g_e(\tau) \\
 & + 2\lambda(\tau)g_e(\tau)E(\tau)\Delta\tau + \mathcal{O}(\Delta\tau^2).
 \end{aligned} \tag{2}$$

By the definition of the derivative, we have:

$$\frac{dg_e(\tau)}{d\tau} = -(\lambda(\tau) + \mu(\tau) + \psi(\tau))g_e(\tau) + 2\lambda(\tau)g_e(\tau)E(\tau). \tag{3}$$

Equation (3) is known as the Kolmogorov backward equation of the BDS process (Feller 1949; Louca and Pennell 2020b). Beginning at time s_e , the initial condition of g_e depends on which event occurred at the beginning of edge e .

$$g_e(s_e) = \begin{cases} \lambda(s_e)g_{e1}(s_e)g_{e2}(s_e) & \text{birth event giving rise} \\ & \text{to edges } e1 \text{ and } e2 \\ (1 - r(s_e))\psi(s_e)g_{e1}(s_e) & \text{ancestral sampling event} \\ \psi(s_e)r(s_e) + \psi(s_e)(1 - r(s_e))E(s_e) & \text{terminal sampling event} \\ \rho_0 & s_e = 0, \text{ extant sample} \end{cases} \tag{4}$$

Together Equations (3) and (4) define the initial value problem for $g_e(\tau)$ as a function of the probability $E(\tau)$.

Because the likelihood density g_e is the solution to a linear differential equation with initial condition at time s_e , we can express its solution as follows:

$$g_e(\tau) = \Psi(s_e, \tau)g_e(s_e), \tag{5}$$

where the auxiliary function, Ψ , is given by:

$$\Psi(s_e, \tau) = \exp\left[\int_{s_e}^{\tau} 2\lambda(x)E(x) - (\lambda(x) + \mu(x) + \psi(x))dx\right]. \tag{6}$$

This function, $\Psi(s, t)$, maps the value of g_e at time s to its value at t , and hence is known as the probability "flow" of the Kolmogorov backward equation (Louca and Pennell 2020b).

Step 2. Deriving the IVP for $E(\tau)$.—We derive the IVP for $E(\tau)$ in a similar manner as above, beginning with a difference equation.

$$\begin{aligned}
 E(\tau + \Delta\tau) = & \underbrace{(1 - \lambda(\tau)\Delta\tau)(1 - \mu(\tau)\Delta\tau)(1 - \psi(\tau)\Delta\tau)}_{\text{nothing happens}} \times E(\tau) \\
 & + \underbrace{\lambda(\tau)\Delta\tau(1 - \mu(\tau)\Delta\tau)(1 - \psi(\tau)\Delta\tau)}_{\text{birth event}} \times E(\tau)^2 \\
 & + \underbrace{\mu(\tau)\Delta\tau(1 - \lambda(\tau)\Delta\tau)(1 - \psi(\tau)\Delta\tau)}_{\text{death event}} \times 1 \\
 & + \underbrace{\psi(\tau)\Delta\tau(1 - \lambda(\tau)\Delta\tau)(1 - \mu(\tau)\Delta\tau)}_{\text{sampling event}} \times 0.
 \end{aligned} \tag{7}$$

By the definition of a derivative, we have:

$$\begin{aligned} \frac{dE(\tau)}{d\tau} &= -(\lambda(\tau) + \mu(\tau) + \psi(\tau))E(\tau) + \lambda(\tau)E(\tau)^2 + \mu(\tau), \\ E(0) &= 1 - \rho_0, \end{aligned} \quad (8)$$

where ρ_0 is the probability a lineage is sampled at the present day. The initial condition at time 0 is therefore the probability that a lineage alive at the present day is not sampled. Given an analytical or numerical general solution to $E(\tau)$, we can find the likelihood by evaluating $g_{stem}(T)$, as follows.

Step 3. Deriving the expression for $g_{stem}(T)$.—Given the linear nature of the differential equation for $g_e(\tau)$ and hence the representation in Equation (5), the likelihood $g_{stem}(\tau)$ is given by the product over all the initial conditions times the product over the probability flow for each edge.

$$\begin{aligned} g_{stem}(T) &= \underbrace{\rho_0^{N_0}}_{\text{extant tips}} \underbrace{\prod_{i=1}^I \lambda(x_i)}_{\text{births}} \underbrace{\prod_{j=1}^n [\psi(y_j)(1-r(y_j))E(y_j) + \psi(y_j)r(y_j)]}_{\text{extinct tips}} \\ &\quad \times \underbrace{\prod_{k=1}^m \psi(z_k)(1-r(z_k))}_{\text{ancestral samples}} \underbrace{\prod_{e \in \mathcal{T}} \Psi(s_e, t_e)}_{\text{edges}}, \end{aligned} \quad (9)$$

where x_i , y_j , and z_k are the times at which individual birth, terminal sampling and ancestral sampling events occur as we elaborated below.

Step 4. Representing $g_{stem}(T)$ in terms of critical times.—Equation (9) can be further simplified by removing the need to enumerate over all the edges of the phylogeny (the last term of Equation (9)) and writing \mathcal{L} in terms of the tree's critical times (horizontal lines in Fig. B.1). The critical times of the tree are made up of three vectors, \vec{x} , \vec{y} , and \vec{z} , as well as the time of origin T . The vector \vec{x} gives the time of each birth event in the phylogeny and has length $I = N_0 + n - 1$, where N_0 is the number of lineages sampled at the present day and n is the number of terminal samples. Unless noted otherwise the elements of vector \vec{x} are listed in decreasing order, such that $x_1 > x_2 > \dots > x_I$ and hence x_1 is the time of the most recent common ancestor $t_{MRC A}$. The vector \vec{y} gives the timing of each terminal sample and hence has length n whereas vector \vec{z} gives the timing of each ancestral sample and has length m . With respect to the BDS likelihood then the sampled tree is summarized by $\mathcal{T} = \{\vec{x}, \vec{y}, \vec{z}, T\}$. We note that the critical times only contain the same information as the edges as a result of the assumptions of the BDS process but are not generally equivalent representations of \mathcal{T} .

As a result of the linear nature of $g_e(\tau)$ it is straightforward to rewrite the likelihood in Equation (9) in terms of the critical-time representation of the

sampled tree. Defining

$$\begin{aligned} \Phi(\tau) &= \Psi(0, \tau) \\ &= \exp \left[\int_0^\tau 2\lambda(x)E(x) - (\lambda(x) + \mu(x) + \psi(x)) dx \right], \end{aligned} \quad (10)$$

the probability flow Ψ can be rewritten as the following ratio:

$$\Psi(s, \tau) = \frac{\Psi(0, \tau)}{\Psi(0, s)} = \frac{\Phi(\tau)}{\Phi(s)}. \quad (11)$$

This relationship allows us to rewrite the likelihood by expressing the product over the edges as two separate products, one over the start of each edge and the other over the end of each edge which in turn allows us to rearrange and cancel terms to obtain an alternative likelihood expression. Edges begin (value of t_e) at either: 1) the tree origin, 2) a birth event resulting to two lineages, or 3) an ancestral sampling event. Edges end (values of s_e) at either: 1) a birth event, 2) an ancestral sampling event, 3) a terminal sampling event, or 4) the present day. Hence we have:

$$\begin{aligned} g_{stem}(T) &= \underbrace{\Phi(T)}_{\text{root}} \times \underbrace{\left(\frac{\rho_0}{\Phi(0)} \right)^{N_0}}_{\text{extant tips}} \times \underbrace{\prod_{i=1}^I \lambda(x_i) \frac{\Phi(x_i)^2}{\Phi(x_i)}}_{\text{births}} \\ &\quad \times \underbrace{\prod_{j=1}^n \frac{\psi(y_j)}{\Phi(y_j)} [(1-r(y_j))E(y_j) + r(y_j)]}_{\text{extinct tips}} \\ &\quad \times \underbrace{\prod_{k=1}^m \frac{\Phi(z_k)}{\Phi(z_k)} \psi(z_k)(1-r(z_k))}_{\text{ancestral samples}}. \end{aligned} \quad (12)$$

Note $\Phi(0)=1$. While Equations (9) and (12) are numerically identical, the critical time expression is more convenient for application as it requires numerically evaluating only a single function $\Phi(\tau)$ as given by Equation (10).

Step 5. Conditioning the likelihood.—While Equation (12) is equal to the basic model likelihood for the phylogeny \mathcal{T} , it is often appropriate to condition the tree likelihood on the tree exhibiting some property, for example the condition there being at least sampled lineage. Imposing a condition on the likelihood is done by multiplying by a factor S . Various conditioning schemes are considered in section A4 and listed in Table S3 with the value of S ranging in complexity from a constant to a general function of the model parameters. The resulting

likelihood expression for the general BDS model is:

$$\begin{aligned} \mathcal{L}(\Theta_{\text{BDS}}, \mathcal{S} | \bar{x}, \bar{y}, \bar{z}, N_0) \\ = S \rho_0^{N_0} \Phi(T) \prod_{i=1}^I \lambda(x_i) \Phi(x_i) \\ \times \prod_{j=1}^n \frac{\Psi(y_j)}{\Phi(y_j)} \left[(1-r(y_j))E(y_j) + r(y_j) \right] \prod_{k=1}^m \Psi(z_k) (1-r(z_k)) \end{aligned} \quad (13)$$

Many Existing Models are Special Cases of this General BDS Model

A large variety of previously published BDS models in epidemiology and macroevolution are special cases of the general model presented here (for a summary of the models we investigated, see [Supplementary Table S2](#)). Indeed, we can obtain the likelihood of these models by adding mathematical constraints (i.e., simplifying assumptions) to the terms in Equation (13). Our work thus not only provides a consistent notation for unifying a multitude of seemingly disparate models, it also provides a concrete and numerically straightforward recipe for computing their likelihood functions. We recognize that there are many valid approaches for deriving tree likelihoods for BDS models with share many similarities with our own (e.g., [Nee et al. 1994](#); [Maddison et al. 2007](#); [Gernhard 2008](#); [Morlon et al. 2011](#); [Lambert and Stadler 2013](#); [Lambert 2018](#); [Laudanno et al. 2020](#); [Louca and Pennell 2020b](#)) and do not claim ours is superior to these; however, we have found our technique to be intuitive and flexible. We have implemented the single-type BDS likelihood in the R package `castor` ([Louca and Doebeli 2018](#)), including routines for maximum-likelihood fitting of BDS models with arbitrary functional forms of the parameters given a phylogeny and routines for simulating phylogenies under the general BDS models (functions `fit_hbds_model_on_grid`, `fit_hbds_model_parametric` and `generate_tree_hbds`).

Figure 1 summarizes the simplifying assumptions that underlie common previously published BDS models; these assumptions generally fall into four categories: 1) assumptions about the functional form of birth, death, and sampling rates over time, 2) assumptions pertaining to the sampling of lineages, 3) the presence of mass-extinction events, and 4) the nature of the tree-conditioning as given by \mathcal{S} . Here, we provide a brief overview of the type of previously invoked constraints which are consistent (or not) with our unified framework; for full details on each specific case, we refer readers to the [Supplementary material](#). While we illustrate these constraints within the single-type context, analogous assumptions can be made within the multitype context examined in the following section.

In regards to rate assumptions, many early BDS models ([Stadler 2009](#); [Stadler 2010](#); [Stadler et al. 2012](#)) assumed that the birth, death, and sampling rates remained constant over time. This is mathematically and computationally convenient since an analytical solution can easily be obtained for $E(\tau)$. In the epidemiological case, holding λ constant, however, implies that the number of susceptible hosts is effectively constant throughout the epidemic and/or that the population does not change its behavior over time; this is an unrealistic assumption given seasonal changes or changes in response to the disease itself. As such, this assumption is only really valid for small time periods or the early stages of an epidemic. This is useful for estimating the basic reproductive number, R_0 , of the SIR model ([Box 1](#)) but not for the effective reproductive number R_e at later time points ([Stadler et al. 2012](#)).

A similarly tractable, but more epidemiologically relevant, model is known as the “birth–death–skyline” variant ([Stadler and Bonhoeffer, 2013](#); [Gavryushkina et al., 2014](#)), in which rates are piecewise-constant functions through time (like the constant rate model, there is also an analytical way to calculate the likelihood of this model; see [Appendix](#)). The BDS skyline model has been implemented under a variety of additional assumptions in the Bayesian phylogenetics software BEAST2 ([Bouckaert et al. 2019](#)). The BDS skyline model has also been extended by ([Kühnert et al., 2014](#)) to infer the parameters of an underlying stochastic SIR model. In this case the diversification model parameters Θ_{BDS} are random variables that emerge from stochastic realizations of the epidemiological model given by Θ_{SIR} , see Equation (B1). Finally, the birth–death skyline model with piecewise constant rates can also be applied in the macroevolutionary case when no sampling occurs through time, $\psi(\tau)=0$ ([Stadler 2011](#)).

In addition to imposing constraints on the temporal variation in the rates, previously derived submodels have considered a variety of different assumptions about the nature of the sampling process. Most notably, in macroevolutionary studies, sampling of molecular data typically occurs only at the present day ([Stadler 2009](#); [Stadler 2011](#); [Morlon et al. 2011](#)) whereas past Poissonian sampling can be introduced to include the sampling of fossil data ([Heath et al. 2014](#)). In epidemiology, concerted sampling at the present day is likely biologically unrealistic ([Stadler et al. 2012](#)), though in some implementations of the models, such a sampling scheme has been imposed. These concerted sampling attempts prior to the present day as well as mass extinction events can be incorporated via the inclusion of Dirac distributions in the sampling and death rates, respectively. Finally, previous models often multiply the likelihood by a factor \mathcal{S} in order to condition on a particular observation (e.g., observing at least one lineage or exactly N_0 lineages), enumerate indistinguishable trees (e.g., accounting for possible orientations or unlabeled trees) ([Gavryushkina et al. 2013](#); [Gavryushkina et al. 2014](#);

| Model | Rates | Sampling | Mass Extinction | Conditioning |
|----------------------------------|------------------------|--|---------------------------|---|
| Stadler 2009 S1.1 | constant | present-day sampling | | constant + N_0 samples |
| Stadler 2010 S1.2 | constant | | fossils $r \leq 1$ | N_0 samples or t_{MRCA} or ≥ 1 extant sample |
| Morlon et al., 2011 S1.3 | | present-day sampling | | ≥ 1 sample |
| Stadler 2011 S1.4 | piecewise constant | present-day sampling | mass extinction | ≥ 1 extant sample or t_{MRCA} |
| Stadler et al. 2012 S1.5 | constant | no present-day sampling | | |
| Stadler et al. 2013 S1.6 | piecewise constant | CSAs | | ≥ 1 sample |
| Gavryushkina et al. 2014 S1.6 | piecewise constant | CSAs + fossils $r \leq 1$ | | constant + ≥ 1 sample |
| Heath et al. 2014 S1.7 | constant | present-day sampling + fossils $r \leq 1$ | | constant + t_{MRCA} + ≥ 1 sample |
| Kühnert et al. 2014 S1.7 | stochastic SIR | | | |

FIGURE 1. Submodel assumptions. Rate, sampling, mass extinction, and conditioning assumptions of existing submodels of the general time-variable BDS process. The key points are that i) each of the previously developed models we considered can be obtained by adding specific combinations of constraints to the various parameters of the general BDS model; and ii) that there are many plausible, and potentially biologically informative combinations of constraints that have not been considered by researchers in epidemiology or macroevolution.

Stadler 2009), or to reflect known uncertainties. The “fossilized-birth–death” likelihood derived by Heath et al. (2014) for example, includes a factor that reflects the uncertainty in the attachment and placement of fossils on the macroevolutionary tree. This

fossilized-birth–death process has been used to estimate divergence times and to model lineage diversification (Gavryushkina et al. 2017; Landis et al. 2021). Variants of the fossilized-birth–death process, for example including mass extinction events, are feasible and can

be derived using our approach. We also note that models similar to the time-variable fossilized-birth–death process have been developed for cases when phylogenetic data is not available (i.e., when only including fossil occurrence data; see [Silvestro et al. 2014](#); [Lehtonen et al. 2017](#)); we have not investigated how these models relate to our generalized BDS model but we speculate that it would be possible to also bring these models into a common framework with those that we have discussed. [Supplementary material](#) demonstrates how these submodels can be re-derived by either imposing the necessary constraints on the general likelihood formula given in Equation (13) or, alternatively, by starting from the combinations of assumptions and using the five-step procedure outlined above.

The Multitype Birth–Death–Sampling Model

A common extension of the single-type diversification models explored above is to consider cases where the diversification rates (λ, μ, ψ) and probabilities (r, ρ) vary among lineages as a function of a categorical “lineage type”. This lineage type can be defined in terms of specific ([Maddison et al. 2007](#); [Rasmussen and Stadler 2019](#)) or unspecified traits ([Beaulieu and O’Meara 2016](#)) or trait combinations ([FitzJohn 2012](#)) (for reviews of these models see [Morlon 2014](#); [Ng and Smith 2014](#)). Representing these lineage types as colors at nodes and along branches of the tree, we first extend the single-type model above by deriving the likelihood of a fully colored tree with topology \mathcal{T} where the states along all edges of the phylogeny are known as given by \mathcal{C} . The resulting likelihood is an extension of the likelihood first developed by [Barido-Sottani et al. \(2018\)](#), where the diversification rates and probabilities are allowed to vary arbitrarily through time. To illustrate that our derivation is indeed quite general, we follow the model developed (independently) by [Magnuson-Ford and Otto \(2012\)](#) and [Goldberg and Igić \(2012\)](#), where the state of lineages can change either anagenetically, with a lineage of type a mutating to a type b at rate $\gamma_{a,b}(\tau)$ or cladogenetically, with a lineage of type a giving rise to a daughter lineage of type b at rate $\lambda_{a,b}(\tau)$. Lineages go extinct at a state-dependent rate $\mu_a(\tau)$ and are sampled at rate $\psi_a(\tau)$. As in the single-type model, upon sampling lineages are removed from the population with probability $r_a(\tau)$ whereas all lineages alive at the present day are sampled with a probability $\rho_a(\tau)$. As discussed in depth by [Goldberg and Igić \(2012\)](#), the other discrete variations of state-dependent diversification models ([FitzJohn et al. 2009](#); [Goldberg et al. 2011](#); [FitzJohn 2012](#)) fall out as special cases of this model. (See [Ng and Smith 2014](#); [Caetano et al. 2018](#); [Louca and Pennell 2020b](#) for further discussion of the connection between multitype models.)

We use the five-step technique specified above for the single-type case to derive the probability of observing a given colored tree under a general multitype model (see

[Supplementary material II](#)). We first derive the initial value problem for the probability $g_{e,a}(\tau)$ that an edge e of type a in the tree at time τ gives rise to the subsequently observed phylogeny. The edge e here refers not to an edge in the topological tree, but to a segment of the tree all of one state between birth, sampling, or mutation events.

$$\begin{aligned} \frac{dg_{e,a}(\tau)}{d\tau} = & - \left(\sum_b \lambda_{a,b}(\tau) + \mu_a(\tau) + \psi_a(\tau) + \sum_b \gamma_{a,b}(\tau) \right) g_{e,a}(\tau) \\ & + \sum_b \iota_{a,b} \lambda_{a,b}(\tau) g_{e,a}(\tau) E_b(\tau) \\ g_{e,a}(s_e) = & \begin{cases} \lambda_{a,b}(s_e) g_{e_1,a}(s_e) g_{e_2,b}(s_e) & \text{birth event } a \rightarrow a+b \\ (1-r_a(s_e)) \psi_a(s_e) g_{e_1,a}(s_e) & \text{ancestral sampling event} \\ r_a(s_e) \psi_a(s_e) + (1-r_a(s_e)) \psi_a(s_e) E_a(s_e) & \text{terminal sampling event} \\ (\gamma_{a,b}(s_e) + \lambda_{a,b} E_a) g_{e_1,b}(s_e) & \text{mutation/hidden birth} \\ \rho_a & \text{event } a \rightarrow b \\ & \text{sampled at present day.} \end{cases} \end{aligned} \quad (14)$$

Equation (16) distinguishes between multiple types of birth events as pictured in [Supplementary Figure S1](#). Birth events may be symmetric, with both daughter lineages inheriting the parental type. The exchangeability of the resulting daughter lineages is reflected in the indicator variable $\iota_{a,b}$ which takes on value of 2 if $a=b$ and 1 otherwise. In contrast asymmetric birth events the resulting daughter lineages differ in type due to cladogenetic change. Importantly, the differential equation for $g_{e,a}$ is linear and hence has a known general solution $g_{e,a} = g_{e,a}(s_e) \Psi(s_e, \tau)$. As in the single-type model $\Psi(s_e, \tau)$ is the probability flow ([Louca and Pennell 2020b](#)) mapping the probability $g_{e,a}$ from the initial state at time s_e to the probability at time τ .

An analogous initial value problem can be derived for the probability $E_a(\tau)$, that a lineage of type a alive at time τ leaves no observed descendants in the sampled tree.

$$\begin{aligned} \frac{dE_a(\tau)}{d\tau} = & - \left(\sum_b \lambda_{a,b}(\tau) + \mu_a(\tau) + \psi_a(\tau) + \sum_b \gamma_{a,b}(\tau) \right) E_a(\tau) \\ & + \sum_b \lambda_{a,b}(\tau) E_a(\tau) E_b(\tau) + \mu_a(\tau) + \sum_b \gamma_{a,b}(\tau) E_b(\tau) \\ E_a(0) = & 1 - \rho_a \end{aligned} \quad (15)$$

This is a nonlinear differential equation and must be solved numerically. Given the solution of $g_{e,a}$ and E_a the likelihood for the fully colored tree is characterized by a series of critical times: first, $\vec{x}_{a,b}$ the times at which a lineage of type a gives birth to a lineage of type b , \vec{y}_a the ages of tip samples of type a , \vec{z}_a the ages of ancestral samples of type a , and $\vec{w}_{a,b}$ the times at which lineages are observed to transition events from type a to type b .

The resulting likelihood is given by:

$$\begin{aligned} \mathcal{L}(\Theta_{\text{MBDS}}|\mathcal{T}, \mathcal{C}) &= \mathcal{S} \times \Phi_{\mathcal{C}^*}(T) \times \left[\prod_{a=1}^A \rho_a^{N_a} \right] \times \left[\prod_{a=1}^A \prod_{b=1}^A \prod_{i=1}^{I_{a,b}} \lambda_{a,b}(x_{a,b,i}) \Phi_b(x_{a,b,i}) \right] \\ &\times \left[\prod_{a=1}^A \prod_{j=1}^{J_a} \left[\Psi_a(y_{a,j})(1-r_a(y_{a,j}))E_a(y_{a,j}) + \Psi_a(y_{a,j})r_a(y_{a,j}) \right] \frac{1}{\Phi_a(y_{a,j})} \right] \quad (16) \\ &\times \left[\prod_{a=1}^A \prod_{k=1}^{K_a} \Psi_a(z_{a,k})(1-r_a(z_{a,k})) \right] \\ &\times \left[\prod_{a=1}^A \prod_{b \neq a}^A \prod_{l=1}^{L_{a,b}} [\gamma_{a,b}(w_{a,b,l}) + \lambda_{a,b}(w_{a,b,l})E_a(w_{a,b,l})] \frac{\Phi_b(w_{a,b,l})}{\Phi_a(w_{a,b,l})} \right] \end{aligned}$$

Here, \mathcal{S} is an arbitrary form of conditioning as in Equation (13) and $\Phi_a(\tau) = \Psi_a(\tau, 0)$, a complete list of notation is given in Table S4.

Equation (16) gives the likelihood of a fully colored tree, the tree topology plus the state along each branch and at each node in the tree. This likelihood is a generalization of that presented by (Barido-Sottani et al., 2018, 2020). Maximizing Equation (16) while incrementally adding and removing changes in state along the branches of the tree can be used to identify clades with distinct diversification parameters. This method can be used, for example, to identify transmission clusters within a disease outbreak (Barido-Sottani et al. 2018). This likelihood is distinct from but related to post-traversal likelihood methods developed to infer state-dependent diversification rates given the known state of sampled lineages (e.g., Maddison et al. 2007; Magnuson-Ford and Otto 2012; Stadler and Bonhoeffer 2013). Specifically, these methods give the likelihood $\mathcal{L}(\Theta_{\text{MBDS}}|\mathcal{T}, \mathcal{C}_\bullet)$ where $\mathcal{C}_\bullet = \{C_\rho, C_y, C_z\}$ is the state of present-day, C_ρ , past C_y , and ancestral, C_z , sampled lineages. The relationship between the numerically obtained post-traversal likelihood and the closed-form fully colored likelihood (Equation (16)) is given by:

$$\mathcal{L}(\Theta_{\text{MBDS}}|\mathcal{T}, \mathcal{C}_\bullet) = \frac{\mathcal{L}(\Theta_{\text{MBDS}}|\mathcal{T}, \mathcal{C}^*)}{Pr(\mathcal{C}^*|\mathcal{T}, \mathcal{C}_\bullet, \Theta_{\text{MBDS}})}. \quad (17)$$

Here, \mathcal{C}^* is one specific coloring of the tree \mathcal{T} (e.g., a maximum parsimony ancestral state reconstruction) that is consistent with the observed states. We include Equation (17) as it clarifies the relationship between these two different approaches that have been used to calculate multitype likelihoods in phylogenetics. Whether or not this is useful for inference is an open question as $Pr(\mathcal{C}^*|\mathcal{T}, \mathcal{C}_\bullet, \Theta_{\text{MBDS}})$ is challenging to compute (the details of which are beyond the scope of the present paper).

Concluding Remarks

In this article, we have unified a broad class of BDS models that have been widely used both in epidemiology and macroevolution. And in doing so, we have also presented a standardized notation and

approach that can be used both for deriving the various submodels that have previously been studied as well as novel combinations of assumptions about the model parameters. The unification of these models clarifies the connections between BDS variants, facilitates the development of new variants tailored to specific scenarios, and provides a structure for understanding how results depend on model assumptions (Kirkpatrick et al. 2002; Lafferty et al. 2015; Louca and Pennell 2020a). And importantly, given the recent discovery of widespread nonidentifiability in birth–death processes fit to extant-only (Louca and Pennell 2020a) and serially sampled (Louca et al. 2021) phylogenetic data, there is a critical need to explore a much broader range of BDS models than were previously considered and the mathematical generalization presented here will be able to enable this.

SUPPLEMENTARY MATERIAL

Data available from the Dryad Digital Repository: <https://doi.org/10.5281/zenodo.5028470>.

ACKNOWLEDGMENTS

We would like to thank Sally Otto for her thoughtful comments on this work.

FUNDING

This work was supported by a *Grant for Catalyzing Research Clusters* awarded to the UBC Biodiversity Research Centre, NSF DEB Grant [#2028986 award to S.L. and M.W.P.]; a NSERC Discovery Grant to M.W.P.; and in part by the EEB department Postdoctoral Fellowship from the University of Toronto to A.M.; in part by a Canadian Institutes of Health Research (CIHR) doctoral award [#6557] to A.M.c.L.; a Genome Canada Bioinformatics and Computational Biology grant [287PHY], CIHR coronavirus rapid response program grant [440371] to J.B.J. and is grateful to the British Columbia Centre for Excellence in HIV / AIDS for additional funding support.

APPENDIX: ADDING ASSUMPTIONS TO THE GENERAL MODEL

In this appendix, we demonstrate how one can obtain the likelihood of submodels with different sets of assumptions by applying constraints to the general likelihood. There are four classes of assumptions that are commonly applied in epidemiological and macroevolutionary studies. First, researchers can make assumptions about the functional form of the birth, death, and sampling rates. Here, we address two such unique assumptions: i) birth, death, and sampling rates are constant (see the Rate Assumptions section, Sections S1.1, S1.2, and S1.5 of the Supplementary material;

and ii) birth, death, and sampling rates are piecewise-constant functions of time (see Piecewise-constant Rates section, [Section S1.6](#) of the [Supplementary material](#)). The cases where birth, death, and sampling rates are defined by a stochastic or deterministic SIR model are mathematically analogous to the cases of the piecewise-constant and general time-variable models respectively. All additional constraints imposed will depend on the exact compartmental model used and hence we will not discuss them in detail in this section. The second major class of assumptions pertains to sampling. There are four such sampling assumptions: i) sampling happens only at the present day as in a birth–death model (see No Sampling at the Present Section, [Sections S1.1](#), [S1.3](#), and [S1.4](#) of the [Supplementary material](#)) or as implemented in the “Birth Death Skyline Contemporary” prior in the BDSKY package in BEAST2; ii) the absence of concerted present-day sampling (see Birth–Death Models section, [Section S1.5](#) of the [Supplementary material](#)); iii) the inclusion of ancestral samples with sampled descendants ([Sections S1.6](#) and [S1.7](#) of the [Supplementary material](#); and iv) concerted sampling attempts (CSA) during which all lineages are sampled with a given probability (see Concerted Sampling Attempts section, [Section S1.6](#) of the [Supplementary material](#)). The third assumption class considers the presence of mass extinction events (see Mass Extinction section, [Section S1.5](#) of the [Supplementary material](#)). The fourth and final major class of assumptions deal with the conditioning of the likelihood. The various conditioning schemes are explored in below and summarized in [Supplementary Table S3](#) available on Dryad.

RATE ASSUMPTIONS

Constant Rates

- *Model assumptions:* Constant diversification rates: $\lambda(t) = \lambda$, $\mu(t) = \mu$, $\psi(t) = \psi$, and constant removal probability $r(t) = r$.

- *The IVP for $g_e(\tau)$:*

$$\frac{dg_e(\tau)}{d\tau} = -(\lambda + \mu + \psi)g_e(\tau) + 2\lambda g_e(\tau)E(\tau)$$

$$g_e(s_e) = \begin{cases} \lambda g_{e1}(s_e)g_{e2}(s_e) & \text{birth event giving rise to edges} \\ & \text{e1 and e2} \\ \psi(1-r)g_{e1}(s_e) & \text{ancestral sampling event} \\ \psi r + \psi(1-r)E(s_e) & \text{terminal sampling event} \\ \rho_0 & s_e = 0, \text{ edge sampled at} \\ & \text{present day} \end{cases}$$

- *The IVP for $E(\tau)$:*

$$\frac{dE(\tau)}{d\tau} = -(\lambda + \mu + \psi)E(\tau) + \lambda E(\tau)^2 + \mu \quad E(0) = 1 - \rho_0.$$

In this case the IVP for $E(\tau)$ is a Bernoulli differential equation and has a known analytical solution. As given by Equation 1 in ([Stadler, 2010](#))

this solution is given by:

$$E(\tau) = \frac{\lambda + \mu + \psi}{2\lambda} + \frac{c_1}{2\lambda} \frac{e^{-c_1\tau}(1-c_2) - (1+c_2)}{e^{-c_1\tau}(1-c_2) + (1+c_2)}$$

$$c_1 = \left| \sqrt{(\lambda - \mu - \psi)^2 + 4\lambda\psi} \right| \quad (\text{A1})$$

$$c_2 = -\frac{\lambda - \mu - 2\lambda\rho_0 - \psi}{c_1}.$$

- *The probability flow:*

$$\Phi(\tau) = \exp \left[\int_0^\tau 2\lambda E(x) - (\lambda + \mu + \psi) dx \right].$$

- *The likelihood:*

$$\mathcal{L}_C = S \rho_0^{N_0} \Phi(T) \lambda^L \psi^{n+m} (1-r)^m \prod_{i=1}^L \Phi(x_i) \prod_{j=1}^n \frac{1}{\Phi(y_j)} \left[(1-r)E(y_j) + r \right] \quad (\text{A2})$$

Piecewise-Constant Rates

- *Model assumptions:* Divide time into $L+1$ intervals defined by *transition times* $0 = t_0 < t_1 < t_2 < \dots < t_L < t_{L+1} = T$. Define rates and removal probabilities constant within a given interval.

$$\lambda(\tau) = \lambda_l \quad t_l < \tau \leq t_{l+1}$$

$$\mu(\tau) = \mu_l \quad t_l < \tau \leq t_{l+1}$$

$$\psi(\tau) = \psi_l \quad t_l < \tau \leq t_{l+1}$$

$$r(\tau) = r_l \quad t_l < \tau < t_{l+1}$$

- *The IVP and solution for $g_e(\tau)$:* Given the definitions of $\lambda(\tau)$, $\mu(\tau)$, $\psi(\tau)$, and $r(\tau)$ within each time interval the IVP for $g_e(\tau)$ is identical to that given in Equations (3) and (4). If $g_{l,e}(\tau)$ is the probability density within time interval l than $g_{l,e}(t_l) = g_{l-1,e}(t_l)$.

- *The IVP and solution for $E(\tau)$:* As with $g_e(\tau)$, the IVP for $E(\tau)$ is given by Equation (8). With the piecewise-constant rate assumptions, however, the general solution for $E(\tau)$ between $t_l < \tau \leq t_{l+1}$ is known (similar to Equation (A1)). Defining $E_l(\tau) = E(\tau)$ where $t_l < \tau \leq t_{l+1}$ and $E_l(t_l) = E_{l-1}(t_l)$ we have:

$$E_l(\tau) = \frac{\lambda_l + \mu_l + \psi_l}{2\lambda_l} + \frac{c_1}{2\lambda_l} \frac{e^{-c_1\tau}(1-c_2) - (1+c_2)}{e^{-c_1\tau}(1-c_2) + (1+c_2)}$$

$$c_1 = \left| \sqrt{(\lambda_l - \mu_l - \psi_l)^2 + 4\lambda_l\psi_l} \right|$$

$$c_2 = -\frac{\lambda_l - \mu_l - 2\lambda_l(1 - E_l(t_l)) - \psi_l}{c_1},$$

where $E_l(t_l) = E_{l-1}(t_l)$ for $l > 0$ and $E_0(t_0) = 1 - \rho_0$.

- *The probability flow:* We define a probability *subflow* within each time interval. Specifically, in the l th time interval.

$$\Phi_l(\tau) = \exp \left[\int_{t_l}^{\tau} 2\lambda_l E_l(x) - (\lambda_l + \mu_l + \psi_l) dx \right].$$

The complete flow can be expressed as a function of the subflows in the following manner:

$$\begin{aligned} \Phi(\tau) &= \Phi_{L_\tau}(\tau) \prod_{l=1}^{L_\tau} \Phi_{l-1}(t_l) \\ \Phi(t_k) &= \underbrace{\Phi_k(t_k)}_1 \prod_{l=1}^k \Phi_{l-1}(t_l) = \prod_{l=1}^k \Phi_{l-1}(t_l), \end{aligned} \quad (A3)$$

where L_t is the index of the time t_l at or after time t , that is, the largest index such that $t_l \leq \tau$.

- *The likelihood:* Given these piecewise definitions we substitute them into the general BDS likelihood (13).

$$\begin{aligned} \mathcal{L}_{PC} &= \mathcal{S} \underbrace{\rho_0^{N_0}}_{\text{extant tips}} \underbrace{\Phi_L(T) \prod_{l=1}^L \Phi_{l-1}(t_l)}_{\text{root}} \\ &\times \underbrace{\prod_{i=1}^I \left[\lambda_{L_{x_i}} \Phi_{L_{x_i}}(x_i) \prod_{l=1}^{L_{x_i}} \Phi_{l-1}(t_l) \right]}_{\text{births}} \\ &\times \underbrace{\prod_{j=1}^n \frac{\psi_{L_{y_j}} \left((1-r_{L_{y_j}}) E_{L_{y_j}}(y_j) + r_{L_{y_j}} \right)}{\Phi_{L_{y_j}} \prod_{l=1}^{L_{y_j}} \Phi_{l-1}(t_l)}}_{\text{extinct tips}} \\ &\times \underbrace{\prod_{k=1}^m \psi_{L_{z_k}} (1-r_{L_{z_k}})}_{\text{ancestral samples}}, \end{aligned}$$

where we use *PC* to denote the piecewise-constant assumption.

We can simplify several of these products. Let α_l be the number of birth events $\geq t_l$ and σ_l the number of sampling events $\geq t_l$.

$$\begin{aligned} \prod_{i=1}^I \prod_{l=1}^{L_{x_i}} \Phi_{l-1}(t_l) &= \prod_{l=1}^L [\Phi_{l-1}(t_l)]^{\alpha_l} \\ \prod_{j=1}^n \prod_{l=1}^{L_{y_j}} \frac{1}{\Phi_{l-1}(t_l)} &= \prod_{l=1}^L [\Phi_{l-1}(t_l)]^{-\sigma_l}. \end{aligned} \quad (A4)$$

Let n_l be the number of observed lineages alive at time t_l . Because the number of observed lineages increases with each birth and decreases with each sampled tip, counting the root we have $n_l = \alpha_l - \sigma_l + 1$. Substituting the expressions for the into the likelihood and using the definition of n_l we have:

$$\begin{aligned} \mathcal{L}_{PC} &= \mathcal{S} \rho_0^{N_0} \Phi_L(T) \prod_{i=1}^I \lambda_{L_{x_i}} \Phi_{L_{x_i}}(x_i) \\ &\times \prod_{j=1}^n \frac{\psi_{L_{y_j}}}{\Phi_{L_{y_j}}} \left((1-r_{L_{y_j}}) E_{L_{y_j}}(y_j) + r_{L_{y_j}} \right) \prod_{k=1}^m \psi_{L_{z_k}} (1-r_{L_{z_k}}) \\ &\times \prod_{l=1}^L \Phi_{l-1}(t_l)^{n_l}. \end{aligned} \quad (A5)$$

SAMPLING ASSUMPTIONS

Birth–Death Models

- *Model assumptions:* The birth–death model assumes that $\psi(\tau) = 0$. Note that the probability of sampling a lineage given it is alive at the present day remains as ρ_0 (incomplete sampling).

- *IVP for $g_e(\tau)$:*

$$\begin{aligned} \frac{dg_e(\tau)}{d\tau} &= -(\lambda(\tau) + \mu(\tau))g_e(\tau) + 2\lambda(\tau)g_e(\tau)E(\tau) \\ g_e(s_e) &= \begin{cases} \lambda(s_e)g_{e1}(s_e)g_{e2}(s_e) & \text{birth event giving rise to} \\ & \text{edges } e1 \text{ and } e2 \\ \rho_0 & s_e = 0, \text{ edge sampled at} \\ & \text{present day} \end{cases} \end{aligned}$$

- *IVP for $E(\tau)$:*

$$\begin{aligned} \frac{dE(\tau)}{d\tau} &= -(\lambda(\tau) + \mu(\tau))E(\tau) \\ &\quad + \lambda(\tau)E(\tau)^2 + \mu(\tau) \quad E(0) = 1 - \rho_0. \end{aligned}$$

Note in this case $E(\tau)$ equals $\hat{E}(\tau)$, the probability a lineage leaves no sampled extant descendants. As demonstrated by (Morlon et al., 2011) there exists a general solution to this initial value problem, see Section ?? for more details. This general solution is given by:

$$E(\tau) = 1 - \frac{\rho_0 \exp \left[\int_0^\tau (\lambda(u) - \mu(u)) du \right]}{1 + \int_0^\tau \rho_0 \exp \left[\int_0^x (\lambda(u) - \mu(u)) du \right] dx}.$$

- *The probability flow:* From (Morlon et al., 2011), the probability flow can be written as the following:

$$\Phi(\tau) = \exp \left[\int_0^\tau (\lambda(\sigma) - \mu(\sigma)) d\sigma \right]$$

$$\left[1 + \frac{\int_0^\tau \rho_0 \lambda(u) \exp\left[\int_0^u (\lambda(\sigma) - \mu(\sigma)) d\sigma\right] du}{1 + \rho_0} \right]^{-2}$$

- *The likelihood:*

$$\mathcal{L}_{BD} = S \rho_0^{N_0} \Phi(T) \prod_{i=1}^I \lambda(x_i) \Phi(x_i) \quad (A6)$$

No Sampling at the Present

Here, we consider the case when $\rho_0 = 0$. The likelihood follows exactly as in the general model case. The resulting likelihood expression is given by:

$$\begin{aligned} \mathcal{L}_{\rho_0=0} = & S \Phi(T) \prod_{i=1}^I \lambda(x_i) \Phi(x_i) \\ & \times \prod_{j=1}^n \frac{\Psi(y_j)}{\Phi(y_j)} \left[(1-r(y_j))E(y_j) + r(y_j) \right] \prod_{k=1}^m \Psi(z_k) (1-r(z_k)). \end{aligned} \quad (A7)$$

Note that in this case $I = n - 1$.

Concerted Sampling Attempts

- *Model assumptions:* Here, we introduce L concerted sampling attempts (CSA) at known points in time, $t_l \in \{1, 2, \dots, L\}$. Like the CSA at the present day, and in contrast to the background Poissonian sampling rate, during the CSA at time t_l every lineages is sampled with a fixed probability ρ_l . In the derivation of the likelihood below, we must distinguish between three different sampling event types. First, *past Poissonian sampling events* are those that do not occur during CSAs. Second, *past concerted sampling events* are those that occur during a CSA at time $t_l \in \{1, 2, \dots, L\}$. Finally, *present concerted sampling events* are those that occur at the present day $\tau = 0$. Past concerted sampling attempts can be included in the general model above by adding L Dirac distributions to the Poisson sampling rate function. Namely,

$$\psi(\tau) = \bar{\psi}(\tau) + \sum_{l=1}^L [w_l * \delta(\tau - t_l)], \quad (A8)$$

where $\bar{\psi}(\tau)$ is the background Poissonian sampling rate and $w_l = -\ln(1 - \rho_l)$. The definition of w_l comes from solving the CDF of the exponentially distribution for the "effective sampling rate" such that the probability of a lineage being sampled is ρ_l .

- *IVP for $g_e(\tau)$:*

$$\frac{dg_e(\tau)}{d\tau} = -(\lambda(\tau) + \mu(\tau) + \psi(\tau))g_e(\tau) + 2\lambda(\tau)g_e(\tau)E(\tau)$$

$$g_e(s_e) = \begin{cases} \lambda(s_e)g_{e1}(s_e)g_{e2}(s_e) & \text{rise birth event giving to edges e1 and e2} \\ (1-r(s_e))\bar{\psi}(s_e)g_{e1}(s_e) & \text{Poisson ancestral sampling event} \\ \bar{\psi}(s_e)r(s_e) + \bar{\psi}(s_e)(1-r(s_e))E(s_e) & \text{Poisson terminal sampling event} \\ (1-r(t_l))\rho_l g_{e1}(t_l) & \text{ancestral sample at } t_l \\ \rho_l r(t_l) + \rho_l(1-r(t_l))E(t_l) & \text{terminal sample at } t_l \\ \rho_0 & s_e = 0, \text{ edge sampled at present day} \end{cases}$$

The solution to $g_e(\tau)$ is given by Equations (5) and (6).

- *IVP for $E(\tau)$:* As with $g_e(\tau)$, the IVP for $E(\tau)$ is identical to that given for the general model in Equation (8). Except in rare cases the IVP must be solved numerically hence requiring numerical integration over Dirac distributions which can prove to be problematic.

Note however, that when examining the integrals over the CSAs, a priori, it is a matter of convention whether the Dirac distribution should be considered as "integrated over" when located at the upper integration bound $\int_a^b \delta(s-b) ds = 1$ or at the lower integration bound $\int_a^b \delta(s-a) ds = 1$. Whichever convention we chose, we must rigorously obey it so that the ratio $\Phi(t)/\Phi(s)$ correctly evaluates to $\Psi(s, t)$ whenever $s \leq t$. Using the former convention, we can rewrite the probability $E(t_l)$ at each concerted sampling time t_l as:

$$E(t_l) = E(t_l^-)e^{w_l} = E(t_l^-)(1 - \rho_l),$$

where t_l^- denotes the limit as time approaches t_l from below. Hence, the probability $E(\tau)$ at any time τ can be evaluated numerically by considering the dynamics between successive CSAs and at each CSA separately.

- *The probability flow:* The probability flow is given by:

$$\Phi(\tau) = \exp\left[\int_0^\tau 2\lambda(x)E(x) - (\lambda(x) + \mu(x) + \bar{\psi}(x) + \sum_{l=1}^L w_l \delta(x - t_l)) dx\right].$$

As with $E(\tau)$ integration over the dirac distributions can be problematic and hence we rewrite this expression separating out these terms. Let L_τ be the oldest CSA occurring at or after time τ , that is, the largest index for which

$t_l \leq \tau$.

$$\begin{aligned} \Phi(\tau) &= \exp \left[\int_0^\tau 2\lambda(x)E(x) \right. \\ &\quad \left. - \left(\lambda(x) + \mu(x) + \bar{\psi}(x) + \sum_{l=1}^{L_\tau} w_l \delta(x - t_l) \right) dx \right] \\ &= \exp \left[\int_0^\tau 2\lambda(x)E(x) - (\lambda(x) + \mu(x) + \bar{\psi}(x)) dx \right] \prod_{l=1}^{L_\tau} e^{-w_l} \\ &= \exp \left[\int_0^\tau 2\lambda(x)E(x) - (\lambda(x) + \mu(x) + \bar{\psi}(x)) dx \right] \prod_{l=1}^{L_\tau} (1 - \rho_l). \end{aligned} \quad (\text{A9})$$

We define:

$$\bar{\Phi}(\tau) = \exp \left[\int_0^\tau 2\lambda(x)E(x) - (\lambda(x) + \mu(x) + \bar{\psi}(x)) dx \right], \quad (\text{A10})$$

which means that we can rewrite Equation (A9) as:

$$\Phi(\tau) = \bar{\Phi}(\tau) \prod_{l=1}^{L_\tau} (1 - \rho_l). \quad (\text{A11})$$

- *The likelihood:* The edge representation of g_{stem} is given by:

$$\begin{aligned} g_{\text{stem}}(T) &= \rho_0^{N_0} \prod_{i=1}^L \lambda(x_i) \prod_{j=1}^n \psi(y_j) \\ &\quad \left[(1 - r(y_j))E(y_j) + r(y_j) \right] \prod_{k=1}^m \psi(z_k)(1 - r(z_k)) \\ &\quad \times \prod_{l=1}^L \rho_l [(1 - r_l)E(t_l) + r_l]^{N_l} \\ &\quad \times \prod_{l=1}^L [\rho_l(1 - r_l)]^{M_l} \prod_{\text{edges}} \Psi(s_e, t_e). \end{aligned}$$

The critical time representation of g_{stem} is given by:

$$\begin{aligned} g_{\text{stem}}(T) &= \underbrace{\rho_0^{N_0}}_{\text{extant tips}} \underbrace{\bar{\Phi}(T) \prod_{l=1}^L (1 - \rho_l)}_{\text{root}} \underbrace{\prod_{i=1}^I \lambda(x_i) \bar{\Phi}(x_i)}_{\text{births}} \left[\prod_{l=1}^{L_{x_i}} (1 - \rho_l) \right] \\ &\quad \times \underbrace{\prod_{j=1}^n \frac{\psi(y_j)}{\bar{\Phi}(y_j)} [(1 - r(y_j))E(y_j) + r(y_j)]}_{\text{Pois. extinct tips}} \left[\prod_{l=1}^{L_{y_j}} \frac{1}{1 - \rho_l} \right] \\ &\quad \underbrace{\prod_{k=1}^m \psi(z_k)(1 - r(z_k))}_{\text{Pois. ances. samples}} \end{aligned}$$

$$\begin{aligned} &\times \underbrace{\prod_{l=1}^L \left(\frac{\rho_l}{\bar{\Phi}(t_l)} [(1 - r_l)E(t_l) + r_l] \right)^{N_l} \left[\prod_{j=1}^l \frac{1}{(1 - \rho_j)^{N_l}} \right]}_{\text{CSA extinct tips}} \\ &\quad \underbrace{\prod_{l=1}^L [\rho_l(1 - r_l)]^{M_l}}_{\text{CSA ances. samples}}, \end{aligned}$$

where N_l is the number of tip samples (samples without descendants) obtained during the l th CSA and M_l is the number of ancestral samples (sequences with descendants). By changing how we enumerate birth, death, and sampling events we can greatly simplify this likelihood. First, let α_l be the number of branching events at or before the l th CSA. In other words, α_l is the number of branching events if the tree were trimmed at the l th CSA. Then:

$$\prod_i \left[\prod_{l=1}^{L_{x_i}} (1 - \rho_l) \right] = \prod_{l=1}^L (1 - \rho_l)^{\alpha_l}. \quad (\text{A12})$$

Second, let σ_l be the number of past Poissonian sampling events before time t_l . Then:

$$\prod_j^n \left[\prod_{l=1}^{L_{y_j}} \frac{1}{(1 - \rho_l)} \right] = \prod_{l=1}^L \frac{1}{(1 - \rho_l)^{\sigma_l}}. \quad (\text{A13})$$

Finally, let β_l be the number of past lineages sampled during a CSA at or before the CSA at time t_l . Hence, $\beta_l = N_l + N_{l+1} + \dots + N_L$. Then:

$$\prod_{l=1}^L \left[\prod_{j=1}^l \frac{1}{(1 - \rho_j)^{N_l}} \right] = \prod_{l=1}^L \frac{1}{(1 - \rho_l)^{\beta_l}}. \quad (\text{A14})$$

The likelihood hence simplifies to:

$$\begin{aligned} g_{\text{stem}}(T) &= \rho_0^{N_0} \bar{\Phi}(T) \prod_{l=1}^L (1 - \rho_l)^{\alpha_l - \beta_l - \sigma_l + 1} \prod_{i=1}^I \lambda(x_i) \bar{\Phi}(x_i) \\ &\quad \times \prod_{j=1}^n \frac{\psi(y_j)}{\bar{\Phi}(y_j)} [(1 - r(y_j))E(y_j) + r(y_j)] \\ &\quad \prod_{k=1}^m \psi(z_k)(1 - r(z_k)) \\ &\quad \times \prod_{l=1}^L \left(\frac{\rho_l}{\bar{\Phi}(t_l)} [(1 - r_l)E(t_l) + r_l] \right)^{N_l} \prod_{l=1}^L [\rho_l(1 - r_l)]^{M_l}. \end{aligned}$$

Let n_l be the number of lineages that cross t_l , that is, the number of lineages alive at time t_l with sampled descendants at some younger age. Note that by this definition $n_0 = 0$. Then $b_l + \sigma_l + n_l$ is the number of tips in the tree had it been trimmed at age t_l whereas α_l is the number of branching events.

Therefore we must have $\alpha_l = b_l + \sigma_l + n_l - 1$. This allows us to simplify the conditioned likelihood given below.

$$\begin{aligned} \mathcal{L}_{\text{CSA}} = & \mathcal{S} \rho_0^{N_0} \Phi(T) \prod_{l=1}^L (1 - \rho_l)^{n_l} \prod_{i=1}^I \lambda(x_i) \Phi(x_i) \\ & \times \prod_{j=1}^n \frac{\Psi(y_j)}{\bar{\Phi}(y_j)} \left[(1 - r(y_j)) E(y_j) + r(y_j) \right] \prod_{k=1}^m \Psi(z_k) (1 - r(z_k)) \\ & \times \prod_{l=1}^L \left(\frac{\rho_l}{\bar{\Phi}(t_l)} [(1 - r_l) E(t_l) + r_l] \right)^{N_l} \prod_{l=1}^L [\rho_l (1 - r_l)]^{M_l} \end{aligned} \quad (\text{A15})$$

MASS EXTINCTION

- *Model assumptions:* In addition to the Poisson birth death and sampling events considered in the general model, there are L mass extinctions events occurring at times $t_1 > t_2 > \dots > t_L$. During the l th mass extinction event each lineage goes extinct with probability v_l . As with concerted sampling such mass extinction events can be introduced into the model by adding a set of dirac-delta functions to the Poisson death rate, $\bar{\mu}(\tau)$.

$$\bar{\mu}(\tau) = \bar{\mu}(\tau) + \sum_{l=1}^L m_l \delta(\tau - t_l), \quad (\text{A16})$$

where $m_l = -\ln(1 - v_l)$.

- *IVP for $g_e(\tau)$:* The initial value problem for $g_e(\tau)$ is identical to that given in equation by Equations (3) and (4) except that μ is now includes the mass extinction events.
- *IVP for $E(\tau)$:* The IVP for $E(\tau)$ to that given by Equation (8) except where the extinction rate is given by Equation (A16). The solution to $E(\tau)$ is obtained by numerical integration. Given the dirac-delta functions this numerical integration can be carried out in a piecewise manner integrating separately between and over each mass extinction event. Defining $E(t_l^-)$ as the solution up to but not including the mass extinction event at time t_l , we have:

$$E(t_l) = (1 - v_l) E(t_l^-) + v_l.$$

The first term reflects the probability that a lineage that does not go extinct during the l th mass extinction event leaves no observable offspring (with probability $E(t_l^-)$) whereas the second term reflects the fact that all lineages that go extinct during the l th mass extinction leave no observed descendants with probability 1.

- *The probability flow:* The solution to the IVP is once again given by $g_e(\tau) = g_e(s_e) \Psi(s_e, \tau) =$

$g_e(s_e) \frac{\Phi(\tau)}{s_e}$ where:

$$\begin{aligned} \Phi(\tau) = & \exp \left[\int_0^\tau 2\lambda(x) E(x) \right. \\ & \left. - \left(\lambda(x) + \bar{\mu}(x) + \sum_{l=1}^L m_l \delta(x - t_l) + \psi(x) \right) dx \right]. \end{aligned}$$

As with the CSAs, let L_τ be the last index l such that $t_l < \tau$. We can separate out the mass extinction terms in the following way.

$$\begin{aligned} \Phi(\tau) = & \exp \left[\int_0^\tau 2\lambda(x) E(x) - (\lambda(x) + \bar{\mu}(x) + \psi(x)) dx \right] \prod_{l=1}^{L_\tau} e^{-m_l} \\ = & \exp \left[\int_0^\tau 2\lambda(x) E(x) - (\lambda(x) + \bar{\mu}(x) + \psi(x)) dx \right] \prod_{l=1}^{L_\tau} (1 - v_l) \\ = & \bar{\Phi}(\tau) \prod_{l=1}^{L_\tau} (1 - v_l), \end{aligned}$$

where $\bar{\Phi}(\tau)$ is defined as in Equation (A10).

- *The likelihood:* Given these initial value problems the likelihood follows as in the general model.

$$\begin{aligned} \mathcal{L}_{\text{ME}} = & \mathcal{S} \rho_0^{N_0} \Phi(T) \prod_{l=1}^L (1 - v_l) \prod_{i=1}^I \left[\lambda(x_i) \Phi(x_i) \prod_{l=1}^{L_{x_i}} (1 - v_l) \right] \\ & \times \prod_{j=1}^n \left[\frac{\Psi(y_j) [(1 - r(y_j)) E(y_j) + r(y_j)]}{\bar{\Phi}(y_j) \prod_{l=1}^{L_{y_j}} (1 - v_l)} \right] \prod_{k=1}^m \Psi(z_k) (1 - r(z_k)). \end{aligned}$$

As with the CSAs we can use relations analogous to Equations A12–A14 to rewrite the likelihood:

$$\begin{aligned} \mathcal{L}_{\text{ME}} = & \mathcal{S} \rho_0^{N_0} \Phi(T) \prod_{l=1}^L (1 - v_l)^{n_l} \\ & \times \prod_{i=1}^I \lambda(x_i) \Phi(x_i) \prod_{j=1}^n \frac{\Psi(y_j)}{\bar{\Phi}(y_j)} \\ & \left[(1 - r(y_j)) E(y_j) + r(y_j) \right] \prod_{k=1}^m \Psi(z_k) (1 - r(z_k)), \end{aligned} \quad (\text{A17})$$

where n_l is defined as before as the number of lineages present at time t_l .

ALTERNATIVE CONDITIONING

Supplementary Table S3 lists a number of possible conditionings, \mathcal{S} that can be applied to the tree likelihood. First, is the trivial case of no conditioning $\mathcal{S}_0 = 1$ which gives the probability of the observed tree including the stem edge between time T and t_{MRCA} . To obtain the model likelihood excluding the stem edge, that is, conditioning of the t_{MRCA} , can be obtained by

setting $\mathcal{S} = \mathcal{S}_1 = \frac{\Phi(x_1)}{\Phi(T)}$. Recall that the elements of \vec{x} are ordered such that $x_1 = t_{\text{MRCA}}$ is the first (oldest) birth event.

Acknowledging that one would not reconstruct a phylogeny without any sampled lineages, we can condition the likelihood on observing at least one sampled lineage (either at or before the present day) given the time of origin, $\mathcal{S}_2 = \frac{1}{1-E(T)}$. Or as with \mathcal{S}_1 , conditioning on at least one sampled lineage given the t_{MRCA} . In order to have at least one sampled lineage *and* a most recent common ancestor, however, each daughter lineage of the common ancestor must have at least one descendent. Hence, we have $\mathcal{S}_3 = \frac{\Phi(x_1)}{\Phi(T)} \frac{1}{(1-E(x_1))^2}$.

The general birth–death–sampling model assumes that all lineages alive at the present day are sampled with probability ρ_0 . As with concerted sampling attempts (CSAs) prior to the present day, this present day CSA may include the sampling of multiple lineages as well as possibly resulting in no sampled lineages. As with \mathcal{S}_2 and \mathcal{S}_3 we can condition the tree likelihood on observing at least one extant lineage at the present day. To do so, we define $\hat{E}(\tau) = E(\tau | \psi = 0)$, the probability that a lineage alive at time τ has no extant descendants. Conditioning on the time of origin we have: $\mathcal{S}_4 = \frac{1}{1-\hat{E}(T)}$. Conditioning on the time of the most recent common ancestor we have:

$$\mathcal{S}_5 = \frac{\Phi(x_1)}{\Phi(T)} \frac{1}{2(1-\hat{E}(x_1))(1-E(x_1))},$$

where now at least one of the two daughter lineages of the common ancestor has a present day sample. In many cases \mathcal{S}_5 is modified, however, to condition on *both* daughter lineages having an extant sampled descendent: $\mathcal{S}'_5 = \frac{\Phi(x_1)}{\Phi(T)} \frac{1}{(1-\hat{E}(x_1))^2}$.

As an alternative to conditioning on at least one extant sampled descent, tree likelihoods can be conditioned on having *exactly* N_0 sampled (extant) descendants. Let $\hat{E}_{N_0}(\tau)$ be the probability a lineage alive at time τ has exactly N_0 descendants. Although a general expression for $\hat{E}_{N_0}(\tau)$ is unknown, in the case of constant birth, death, and sampling rates (the case in which this form of conditioning has been applied), the expression for $\hat{E}_{N_0}(\tau)$ is given by (Gernhard, 2008), (Kendall, 1948) and Theorem 3.3 by (Stadler, 2010):

$$\hat{E}_{N_0}(\tau) = \rho_0 \hat{\Phi}(\tau) \left(\frac{\lambda}{\mu} \hat{E}(\tau) \right)^{N_0-1}$$

$$\hat{E}_{N_0}(\tau) = \rho_0 \hat{\Phi}(\tau) \left(\frac{\rho_0 \lambda (1 - e^{-(\lambda-\mu)t})}{\lambda \rho_0 + (\lambda(1-\rho_0) - \mu) e^{-(\lambda-\mu)t}} \right)^{N_0-1},$$

where, like \hat{E} , $\hat{\Phi}$ is given by Equation (10) evaluated with where $\psi = 0$. Given the time of origin we can condition on observing exactly N_0 lineages by setting $\mathcal{S} = \mathcal{S}_6 = \frac{1}{\hat{E}_{N_0}(T)}$. When t_{MRCA} is given instead, then the number

of descendants of the two daughter lineages must add up to N_0 while both daughter lineages must still have at least one descendant (see (Stadler, 2010) Corollary 3.9).

$$\mathcal{S} = \mathcal{S}_7 = \frac{\Phi(x_1)}{\Phi(t_{\text{or}})} \left(\sum_{i=1}^{N_0-1} \hat{E}_i(x_1) \hat{E}_{N_0-i}(x_1) \right)^{-1}$$

$$= \frac{\Phi(x_1)}{\Phi(t_{\text{or}})} \left[(N_0-1)(\rho_0 \hat{\Phi}(x_1))^2 \left(\frac{\rho_0 \lambda (1 - e^{-(\lambda-\mu)t})}{\lambda \rho_0 + (\lambda(1-\rho_0) - \mu) e^{-(\lambda-\mu)t}} \right)^{N_0-2} \right]^{-1}.$$

While early BDS models often employed such conditioning (Stadler 2009; Stadler 2010), this form of conditioning has not been employed in many later models perhaps because the biological justification for such conditioning is vague.

The final form of conditioning used in the literature, which we will represent simply as \mathcal{S}_8 , is the multiplication of the BDS likelihood by a constant to account for the enumeration over the possible indistinguishable representations of a given tree. The value of this constant depends on whether the tree considered is “labeled” and “oriented” (Gavryushkina et al. 2013) and whether, as in the derivation here, the vector of birth events, \vec{x} , is (un)ordered. Inclusion of such a constant should have no effect on the maximum likelihood inference of the model parameters given a specified tree. In cases where the constant is a function of the critical times (Heath et al. 2014), it can influence the inference when the parameters and the tree are jointly estimated.

REFERENCES

- Barido-Sottani J., Vaughan T.G., Stadler T. 2018. Detection of HIV transmission clusters from phylogenetic trees using a multi-state birth–death model. *J. R. Soc. Interface* 15:20180512.
- Barido-Sottani J., Vaughan T.G., Stadler T. 2020. A multitype birth–death model for Bayesian inference of lineage-specific birth and death rates. *Syst. Biol.* 69:973–986.
- 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.
- Boskova V., Bonhoeffer S., Stadler T. 2014. Inference of epidemiological dynamics based on simulated phylogenies using birth–death and coalescent models. *PLoS Comput. Biol.* 10:1–18.
- Bouckaert R., Vaughan T.G., Barido-Sottani J., Duchêne S., Fourment M., Gavryushkina A., Heled J., Jones G., Kühnert D., Maio N.D., Matschiner M., Mendes F.K., Müller N.F., Ogilvie H.A., du Plessis, Poppinga A., Rambaut A., Rasmussen D., Siveroni I., Suchard M.A., Wu C.-H., Xie D., Zhang C., Stadler T., Drummond A.J. 2019. BEAST 2.5: an advanced software platform for Bayesian evolutionary analysis. *PLoS Comput. Biol.* 15:e1006650.
- Caetano D.S., O’Meara B.C., Beaulieu J.M. 2018. Hidden state models improve state-dependent diversification approaches, including biogeographical models. *Evolution* 72:2308–2324.
- Drummond A.J., Pybus O.G., Rambaut A., Forsberg R., Rodrigo A.G. 2003. Measurably evolving populations. *Trends Ecol. Evol.* 18:481–488.
- Drummond A.J., Rambaut A., Shapiro B., Pybus O.G. 2005. Bayesian coalescent inference of past population dynamics from molecular sequences. *Mol. Biol. Evol.* 22:1185–1192.
- du Plessis L., Stadler T. 2015. Getting to the root of epidemic spread with phylodynamic analysis of genomic data. *Trends Microbiol.* 23:383–386.

- Duffy S., Shackelton L.A., Holmes E.C. 2008. Rates of evolutionary change in viruses: patterns and determinants. *Nat. Rev. Genetics* 9:267–276.
- Etienne R.S., Haegeman B., Stadler T., Aze T., Pearson P.N., Purvis A., Phillimore A.B. 2012. Diversity-dependence brings molecular phylogenies closer to agreement with the fossil record. *Proc. R. Soc. B* 279:1300–1309.
- Etienne R.S., Rosindell J. 2012. Prolonging the past counteracts the pull of the present: protracted speciation can explain observed slowdowns in diversification. *Syst. Biol.* 61:204–204.
- Ezard T.H.G., Quental T.B., Benton M.J. 2016. The challenges to inferring the regulators of biodiversity in deep time. *Philos. Trans. R. Soc. B* 371:20150216.
- Feller W. 1949. *Proceedings of the [First] Berkeley Symposium on Mathematical Statistics and Probability*. University of California Press, Berkeley California.
- FitzJohn R.G. 2012. Diversitree: comparative phylogenetic analyses of diversification in R. *Methods Ecol. Evol.* 3:1084–1092.
- FitzJohn R.G., Maddison W.P., Otto S.P. 2009. Estimating trait-dependent speciation and extinction rates from incompletely resolved phylogenies. *Syst. Biol.* 58:595–611.
- Gavryushkina A., Heath T.A., Ksepka D.T., Stadler T., Welch D., Drummond A.J. 2017. Bayesian total-evidence dating reveals the recent crown radiation of penguins. *Syst. Biol.* 66:57–73.
- Gavryushkina A., Welch D., Drummond A.J. 2013. Recursive algorithms for phylogenetic tree counting. *Algorithms Mol. Biol.* 8:26.
- Gavryushkina A., Welch D., Stadler T., Drummond A.J. 2014. Bayesian inference of sampled ancestor trees for epidemiology and fossil calibration. *PLoS Comput. Biol.* 10:1–15.
- Gernhard T. 2008. The conditioned reconstructed process. *J. Theor. Biol.* 253:769–778.
- Goldberg E.E., Igić B. 2012. Tempo and mode in plant breeding system evolution. *Evolution* 66:3701–3709.
- Goldberg E.E., Lancaster L.T., Ree R.H. 2011. Phylogenetic inference of reciprocal effects between geographic range evolution and diversification. *Syst. Biol.* 60:451–465.
- Grenfell B.T., Pybus O.G., Gog J.R., Wood L.N., Daly J.M., Mumford J.A., Holmes E.C. 2004. Unifying the epidemiological and evolutionary dynamics of pathogens. *Science* 303:327–332.
- Hagen O., Hartmann K., Steel M., Stadler T. 2015. Age-dependent speciation can explain the shape of empirical phylogenies. *Syst. Biol.* 64:432–440.
- Heath T.A., Huelsenbeck J.P., Stadler T. 2014. The fossilized birth–death process for coherent calibration of divergence-time estimates. *Proc. Natl. Acad. Sci. USA* 111:E2957–E2966.
- Joy J.B., McCloskey R.M., Nguyen T., Liang R.H., Khudyakov Y., Olmstead A., Krajden M., Ward J.W., Harrigan P.R., Montaner J.S.G., Poon A.F.Y. 2016. The spread of hepatitis C virus genotype 1a in North America: a retrospective phylogenetic study. *Lancet Infectious Dis.* 16:698–702.
- Keeling M.J., Rohani P. 2008. *Modeling infectious diseases: in humans and animals*. Princeton University Press.
- Kendall D.G. 1948. On the generalized “birth-and-death” process. *Ann. Math. Stat.* 19:1–15.
- Kingman J.F.C. 1982. On the genealogy of large populations. *J. Appl. Prob.* 19:27–43.
- Kirkpatrick M., Johnson T., Barton N. 2002. General models of multilocus evolution. *Genetics* 161:1727–1750.
- Kühnert D., Stadler T., Vaughan T.G., Drummond A.J. 2014. Simultaneous reconstruction of evolutionary history and epidemiological dynamics from viral sequences with the birth–death SIR model. *J. R. Soc. Interface* 11:20131106.
- Lafferty K.D., DeLeo G., Briggs C.J., Dobson A.P., Gross T., Kuris A.M. 2015. A general consumer–resource population model. *Science* 349:854–857.
- Lambert A. 2018. The coalescent of a sample from a binary branching process. *Theor. Popul. Biol.* 122:30–35.
- Lambert A., Stadler T. 2013. Birth–death models and coalescent point processes: the shape and probability of reconstructed phylogenies. *Theor. Popul. Biol.* 90:113–128.
- Landis M.J., Eaton D.A.R., Clement W.L., Park B., Spriggs E.L., Sweeney P.W., Edwards E.J., Donoghue M.J. 2021. Joint phylogenetic estimation of geographic movements and biome shifts during the global diversification of viburnum. *Syst. Biol.* 70:67–85.
- Laudanno G., Haegeman B., Rabosky D.L., Etienne R.S. 2020. Detecting lineage-specific shifts in diversification: a proper likelihood approach. *Syst. Biol.* 70:389–407.
- Lehtonen S., Silvestro D., Karger D.N., Scotese C., Tuomisto H., Kessler M., Peña C., Wahlberg N., Antonelli A. 2017. Environmentally driven extinction and opportunistic origination explain fern diversification patterns. *Sci. Rep.* 7:4831.
- Leventhal G.E., Kouyos R., Stadler T., von Wyl V., Yerly S., Böni J., Cellerai C., Klimkait T., Günthard H.F., Bonhoeffer S. 2012. Inferring epidemic contact structure from phylogenetic trees. *PLoS Comput. Biol.* 8:e1002413.
- Louca S. 2020. Simulating trees with millions of species. *Bioinformatics* 26:2907–2908.
- Louca S., Doebeli M. 2018. Efficient comparative phylogenetics on large trees. *Bioinformatics* 34:1053–1055.
- Louca S., McLaughlin A., MacPherson A., Joy J.B., Pennell M.W. 2021. Fundamental identifiability limits in molecular epidemiology. *Mol. Biol. Evol.*
- Louca S., Pennell M.W. 2020a. Extant timetrees are consistent with a myriad of diversification histories. *Nature* 580:1–4.
- Louca S., Pennell M.W. 2020b. A general and efficient algorithm for the likelihood of diversification and discrete-trait evolutionary models. *Syst. Biol.* 69:545–556.
- 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.
- Magee A.F., Höhna S. 2021. Impact of K-Pg mass extinction event on crocodylomorpha inferred from phylogeny of extinct and extant taxa. *bioRxiv* 2021.01.14.426715. Available from: <https://www.biorxiv.org/content/10.1101/2021.01.14.426715v1>.
- Magnuson-Ford K., Otto S.P. 2012. Linking the investigations of character evolution and species diversification. *Am. Nat.* 180:225–245.
- Morlon H. 2014. Phylogenetic approaches for studying diversification. *Ecol. Lett.* 17:508–525.
- Morlon H., Parsons T.L., Plotkin J.B. 2011. Reconciling molecular phylogenies with the fossil record. *Proc. Natl. Acad. Sci. USA* 108:16327–16332.
- Morlon H., Potts M.D., Plotkin J. B. 2010. Inferring the dynamics of diversification: a coalescent approach. *PLoS Biol.* 8:e1000493.
- Nee S. 2006. Birth-death models in macroevolution. *Annu. Rev. Ecol. Syst.* 37:1–17.
- Nee S., May R.M., Harvey P.H. 1994. The reconstructed evolutionary process. *Philos. Trans. R. Soc. B* 344:7.
- Ng J., Smith S.D. 2014. How traits shape trees: new approaches for detecting character state-dependent lineage diversification. *J. Evol. Biol.* 27:2035–2045.
- Pybus O.G., Rambaut A., Harvey P.H. 2000. An integrated framework for the inference of viral population history from reconstructed genealogies. *Genetics* 155:1429–1437.
- Rabosky D.L., Lovette I.J. 2008a. Density-dependent diversification in North American wood warblers. *Proc. R. Soc. B* 275:2363–2371.
- Rabosky D.L., Lovette I.J. 2008b. Explosive evolutionary radiation: decreasing speciation or increasing extinction through time? *Evolution* 62:1866–1875.
- Rasmussen D.A., Stadler T. 2019. Coupling adaptive molecular evolution to phylodynamics using fitness-dependent birth-death models. *eLife* 8:1–24.
- Raup D.M. 1985. Mathematical models of cladogenesis. *Paleobiology* 11:42–52.
- Romero-Severson E.O., Bulla I., Leitner T. 2016. Phylogenetically resolving epidemiologic linkage. *Proc. Natl. Acad. Sci. USA* 113:2690–2695.
- Schluter D., Pennell M.W. 2017. Speciation gradients and the distribution of biodiversity. *Nature* 546:48–55.
- Silvestro D., Schnitzler J., Liow L.H., Antonelli A., Salamin N. 2014. Bayesian estimation of speciation and extinction from incomplete fossil occurrence data. *Syst. Biol.* 63:349–367.
- Stadler T. 2009. On incomplete sampling under birth-death models and connections to the sampling-based coalescent. *J. Theor. Biol.* 261:58–66.

- Stadler T. 2010. Sampling-through-time in birth-death trees. *J. Theor. Biol.* 267:396–404.
- Stadler T. 2011. Mammalian phylogeny reveals recent diversification rate shifts. *Proc. Natl. Acad. Sci. USA* 108:6187–6192.
- Stadler T. 2013. Recovering speciation and extinction dynamics based on phylogenies. *J. Evol. Biol.* 26:1203–1219.
- Stadler T., Bonhoeffer S. 2013. Uncovering epidemiological dynamics in heterogeneous host populations using phylogenetic methods. *Philos. Trans. R. Soc. B* 368:20120198.
- Stadler T., Kouyos R.D., von Wyl V., Yearly S., Böni J. 2012. Estimating the basic reproductive number from viral sequence data. *Mol. Biol. Evol.* 29:347–357.
- Stadler T., Kühnert D., Bonhoeffer S., Drummond A.J. 2013. Birth-death skyline plot reveals temporal changes of epidemic spread in HIV and hepatitis C virus (HCV). *Proc. Natl. Acad. Sci. USA* 110:228–233.
- Stadler T., Vaughan T.G., Gavryushkin A., Guindon S., Kühnert D., Leventhal G.E., Drummond A.J. 2015. How well can the exponential-growth coalescent approximate constant-rate birth-death population dynamics? *Proc. Biol. Sci.* 282:20150420.
- Strimmer K., Pybus O.G. 2001. Exploring the demographic history of DNA sequences using the generalized skyline plot. *Mol. Biol. Evol.* 18:2298–2305.
- Vaughan T.G., Leventhal G.E., Rasmussen D.A., Drummond A.J., Welch D., Stadler T. 2019. Estimating epidemic incidence and prevalence from genomic data. *Mol. Biol. Evol.* 36:1804–1816.
- Volz E.M. 2012. Complex population dynamics and the coalescent under neutrality. *Genetics* 190:187–201.
- Volz E.M., Frost S.D.W. 2014. Sampling through time and phylodynamic inference with coalescent and birth–death models. *J. R. Soc. Interface* 11:20140945.
- Volz E.M., Kosakovsky Pond S.L., Ward M.J., Leigh Brown A.J., Frost S.D.W. 2009. Phylodynamics of infectious disease epidemics. *Genetics* 183:1421–1430.