Unifying Phylogenetic Birth-Death Models in Epidemiology and Macroevolution

Ailene MacPherson1,2,*, Stilianos Louca3,4, Angela McLaughlin5,6, Jeffrey B. Joy5,6,7 & Matthew W. Pennell1, *

1Department of Zoology and Biodiversity Research Centre, University of British Columbia, Vancouver, Canada
2Department of Ecology and Evolutionary Biology, University of Toronto, Toronto, Canada
3Department of Biology, University of Oregon, USA
4Institute of Ecology and Evolution, University of Oregon, USA
5British Columbia Centre for Excellence in HIV/AIDS, Vancouver, Canada
6Bioinformatics, University of British Columbia, Vancouver, Canada
7Department of Medicine, University of British Columbia, Vancouver, Canada

*Correspondence should be addressed to: pennell@zoology.ubc.ca and ailene.macpherson@utoronto.ca

Running Head: Unifying birth-death models

Abstract

Birth-death stochastic processes are the foundation 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 non-trivial. 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 re-deriving existing single-type birth-death sampling models we clarify and synthesize the range of explicit and implicit assumptions made by these models.

Keywords: epidemiology; macroevolution; phylogenetics; birth-death processes; 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 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, 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 re-parameterize these models, we can describe the dynamics of the basic and effective reproductive ratios ($R_0$ and $R_e$, respectively) over time...
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 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, there has been an astounding rise in the variety and complexity of BDS model variants. A key assumption in the specification of BDS sub-models 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 (“multi-type” 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.
do so by first deriving a likelihood for general single- and multi-type 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 sub-models 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 sub-models 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 $\tau (0 \leq \tau \leq T)$ is measured in units of time before the present day ($\tau = 0$), such that $\lambda$ 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 $\tau$ as a general time variable and $t_\times$ to denote the time at which a specific event $\times$ occurs as measured in units of time before the present day (see 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 $\rho_l$ ($\vec{\rho}$ 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
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 \( \rho \) 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 \( \psi \) (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 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 behavioural 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 (\( \mu \)) 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 \( \lambda, \mu, \psi, \) or \( r \); each may be constant, or vary according to any arbitrary function of time given that it is biologically valid (non-negative 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 A1.2). 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 \( \Theta_{BDS} = \{ \lambda(\tau), \mu(\tau), \psi(\tau), r(\tau), \rho \} \).
In order to make inference about the model parameters, we need to calculate the likelihood, \( L \), that an observed phylogeny, \( T \), 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 \( T \), 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 B1c, 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 \( \tau \), \((s_e < \tau < t_e)\) and the present day. The likelihood of the model for the observed tree is then, is by definition \( g_{\text{stem}}(T) \): the probability density that the stem lineage (stem = 1 in Figure B1c) 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 \( \tau \) and \( \tau + \Delta \tau \) assuming that \( \Delta \tau \) is small enough such that at most one event is likely to occur.

\[
g_e(\tau + \Delta \tau) \approx (1 - \lambda(\tau)\Delta \tau)(1 - \mu(\tau)\Delta \tau)(1 - \psi(\tau)\Delta \tau) \times g_e(\tau) \]

\[
\quad + \lambda(\tau)\Delta \tau(1 - \mu(\tau)\Delta \tau)(1 - \psi(\tau)\Delta \tau) \times 2g_e(\tau)E(\tau) \]

\[
\quad + \mu(\tau)\Delta \tau(1 - \lambda(\tau)\Delta \tau)(1 - \psi(\tau)\Delta \tau) \times 0 \]

\[
\quad + \psi(\tau)\Delta \tau(1 - \lambda(\tau)\Delta \tau)(1 - \mu(\tau)\Delta \tau) \times 0 + O(\Delta \tau^2).
\]

Here, \( E(\tau) \) is the probability that a lineage alive at time \( \tau \) 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.

\[ \Delta g_e(\tau) \approx -(\lambda(\tau) + \mu(\tau) + \psi(\tau))\Delta \tau g_e(\tau) + 2\lambda(\tau)g_e(\tau)E(\tau)\Delta \tau + O(\Delta \tau^2). \]  

(2)

By the definition of the derivative we have:

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

(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 \( e \).

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

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:

\[ g_e(\tau) = \Psi(s_e, \tau)g_e(s_e), \]  

(5)

where the auxiliary function, \( \Psi \), is given by:

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

(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.

\[
E(\tau + \Delta \tau) - E(\tau) = \left( 1 - \lambda(\tau) \Delta \tau \right) \left( 1 - \mu(\tau) \Delta \tau \right) \left( 1 - \psi(\tau) \Delta \tau \right) \times E(\tau)
\]

nothing happens

\[
+ \lambda(\tau) \Delta \tau (1 - \mu(\tau) \Delta \tau) (1 - \psi(\tau) \Delta \tau) \times E(\tau)^2
\]

birth event

\[
+ \mu(\tau) \Delta \tau (1 - \lambda(\tau) \Delta \tau) (1 - \psi(\tau) \Delta \tau) \times 1
\]

death event

\[
+ \psi(\tau) \Delta \tau (1 - \lambda(\tau) \Delta \tau) (1 - \mu(\tau) \Delta \tau) \times 0.
\]

sampling event

(7)

By the definition of a derivative we have:

\[
\frac{dE(\tau)}{d\tau} = - (\lambda(\tau) + \mu(\tau) + \psi(\tau)) E(\tau) + \lambda(\tau) E(\tau)^2 + \mu(\tau),
\]

(8)

\[
E(0) = 1 - \rho_0,
\]

where \( \rho_0 \) 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.

\[
g_{stem}(T) = \rho_0^N \prod_{i=1}^I \lambda(x_i) \prod_{j=1}^n \left[ \psi(y_j)(1 - r(y_j)) E(y_j) + \psi(y_j) r(y_j) \right] \prod_{k=1}^m \psi(z_k)(1 - r(z_k)) \prod_{e \in \mathcal{T}} \Psi(s_e, t_e).
\]

(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 Figure B1). 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 $x_1 > x_2 > \ldots x_I$ and hence $x_1$ is the time of the most recent common ancestor
$t_{MRCA}$. 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 $g_c(\tau)$ it is straightforward to rewrite the likelihood in Equation (9) in
terms of the critical-time representation of the sampled tree. Defining

$$
\Phi(\tau) = \Psi(0, \tau) = \exp \left[ \int_0^\tau 2\lambda(x)E(x) - (\lambda(x) + \mu(x) + \psi(x)) dx \right],
$$

(10)

the probability flow $\Psi$ can be rewritten as the following ratio:

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

(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:

\[ g_{stem}(T) = \Phi(T) \times \left( \frac{\rho_0}{\Phi(\tau)} \right)^{N_0} \times \prod_{i=1}^{I} \lambda(x_i) \frac{\Phi(x_i)^2}{\Phi(x_i)} \times \prod_{j=1}^{n} \frac{\psi(y_j)}{\Phi(y_j)} \left[ (1 - r(y_j)) E(y_j) + r(y_j) \right] \times \prod_{k=1}^{m} \frac{\Phi(z_k)}{\Phi(z_k)} \psi(z_k) (1 - r(z_k)) \]

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 \( T \), 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:

\[ L(T_{BDS}, S | \bar{x}, \bar{y}, \bar{z}, N_0) = S \rho_0^{N_0} \Phi(T) \prod_{i=1}^{I} \lambda(x_i) \Phi(x_i) \times \prod_{j=1}^{n} \frac{\psi(y_j)}{\Phi(y_j)} \left[ (1 - r(y_j)) E(y_j) + r(y_j) \right] \times \prod_{k=1}^{m} \frac{\Phi(z_k)}{\Phi(z_k)} \psi(z_k) (1 - r(z_k)) \]

**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 Table S2; proofs in Supplemental Material). 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 $S$. 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 multi-type context examined in the following section. In regards to rate assumptions, many early BDS models (Stadler, 2009, 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 $\lambda$ 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. (Kühnert et al., 2014) to infer the the parameters
of an underlying stochastic SIR model. In this case the diversification model parameters $\Theta_{BDS}$ are random variables that emerge from stochastic realizations of the epidemiological model given by $\Theta_{SIR}$, 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 sub-models 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, 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 $S$ 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, 2014; 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. The Supplementary Material demonstrates how these sub-models 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 multi-type birth-death-sampling model

A common extension of the single-type diversification models explored above is to consider cases where the diversification rates ($\lambda, \mu, \psi$) and probabilities ($r, \rho$) 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 colours at nodes and along branches of the tree, we first extend the single-type model above by deriving the likelihood of a fully coloured tree with topology $\mathcal{G}$ 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, and Louca and Pennell, 2020b for further discussion of the connection between multi-type models.)

We use the five-step technique specified above for the single-type case to derive the probability of observing a given coloured tree under a general multi-type 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 $\tau$ gives rise to the subsequently observed phylogeny. The edge $e$ here refers not to an edge in the topological tree, but to a
supplemental information. The segment of the tree all of one state between birth, sampling, or mutation events.

\[
\frac{d g_{e,a}(\tau)}{d\tau} = - \left( \sum_b \lambda_{a,b}(\tau) + \mu_a(\tau) + \psi_a(\tau) + \sum_b \gamma_{a,b}(\tau) \right) g_{e,a}(\tau) + \sum_b \epsilon_{a,b} \lambda_{a,b}(\tau) g_{e,a}(\tau) E_b(\tau)
\]

\[
ge_{e,a}(s_e) = \begin{cases} 
\lambda_{a,b}(s_e) g_{e_{1,a}}(s_e) g_{e_{2,b}}(s_e) & \text{birth event } a \to a + b \\
(1 - r_a(s_e)) \psi_a(s_e) g_{e_{1,a}}(s_e) & \text{ancestral sampling event} \\
r_a(s_e) \psi_a(s_e) + (1 - r_a(s_e)) \psi_a(s_e) E_a(s_e) & \text{terminal sampling event} \\
(\gamma_{a,b}(s_e) + \lambda_{a,b} E_a) g_{e_{1,b}}(s_e) & \text{mutation/hidden birth event } a \to b \\
\rho_a & \text{sampled at present day}
\end{cases}
\]

Equation (16) distinguishes between multiple types of birth events as pictured in 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 \( \epsilon_{a,b} \) 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 caldogenetic 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 \( \tau \).

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

\[
\frac{d E_a(\tau)}{d\tau} = - \left( \sum_b \lambda_{a,b}(\tau) + \mu_a(\tau) + \psi_a(\tau) + \sum_b \gamma_{a,b}(\tau) \right) E_a(\tau) + \sum_b \lambda_{a,b}(\tau) E_a(\tau) E_b(\tau) + \mu_a(\tau) + \sum_b \gamma_{a,b}(\tau) E_b(\tau)
\]

\[
E_a(0) = 1 - \rho_a
\]

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

\[
\mathcal{L}(\Theta_{\text{MBDS}}|\mathcal{T}, \mathcal{C}) = S \times \Phi_{c^*}(\mathcal{T}) \times \prod_{a=1}^{A} \rho_{a}^{N_{a}} \times \prod_{a=1}^{A} \prod_{b=1}^{A} \prod_{i=1}^{l_{a,b}} \lambda_{a,b}(x_{a,b,i}) \Phi_{b}(x_{a,b,i}) \\
\times \left[ \prod_{a=1}^{A} \prod_{j=1}^{J_{a}} \left[ \psi_{a}(y_{a,j})(1 - r_{a}(y_{a,j}))E_{a}(y_{a,j}) + \psi_{a}(y_{a,j})r_{a}(y_{a,j}) \right] \frac{1}{\Phi_{a}(y_{a,j})} \right] \\
\times \left[ \prod_{a=1}^{A} \prod_{k=1}^{K_{a}} \psi_{a}(z_{a,k})(1 - r_{a}(z_{a,k})) \right] \\
\times \left[ \prod_{a=1}^{A} \prod_{b \neq a} \prod_{i=1}^{l_{a,b}} \left[ \lambda_{a,b}(w_{a,b,i}) + \lambda_{a,b}(w_{a,b,i})E_{a}(w_{a,b,i}) \right] \frac{\Phi_{b}(w_{a,b,l})}{\Phi_{a}(w_{a,b,l})} \right]^{16}
\]

Here \( S \) 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 coloured 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 \( \mathcal{L}(\Theta_{\text{MBDS}}|\mathcal{T}, \mathcal{C}^*) \) where \( \mathcal{C}^* = \{ \mathcal{C}_{\rho}, \mathcal{C}_{y}, \mathcal{C}_{z} \} \) is the state of present-day, \( \mathcal{C}_{\rho} \), past \( \mathcal{C}_{y} \), and ancestral, \( \mathcal{C}_{z} \), sampled lineages. The relationship between the numerically obtained post-traversal likelihood and the closed-form fully coloured likelihood (Equation (16)) is given by:

\[
\mathcal{L}(\Theta_{\text{MBDS}}|\mathcal{T}, \mathcal{C}^*) = \frac{\mathcal{L}(\Theta_{\text{MBDS}}|\mathcal{T}, \mathcal{C}^*)}{Pr(\mathcal{C}^*|\mathcal{T}, \mathcal{C}^*, \Theta_{\text{MBDS}})}.
\]

(17)

Here \( \mathcal{C}^* \) is one specific colouring 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 multi-type likelihoods in phylogenetics. Whether or not this is useful for inference is an open question as \( Pr(\mathcal{C}^*|\mathcal{T}, \mathcal{C}^*, \Theta_{\text{MBDS}}) \) is challenging to compute (the
Concluding remarks

In this paper 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 sub-models 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 non-identifiability 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 enable this.
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 $\beta I$. Infected hosts may recover (at rate $\gamma$), die of virulent cases (at rate $\alpha$), or be sampled (at rate $\psi$). The cumulative number of sampled hosts is represented in the SIR model (Figure B1 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 $\sigma$. The special case of the SIR model with no immunity (the SIS model) is obtained in the limit as $\sigma \to \infty$. In addition to these epidemiological processes, the SIR model includes demographic processes, such as host birth (rate $B$) and death from natural causes (rate $\delta$). While not shown explicitly in the figure, these epidemiological and demographic rates may change over time as a result of host behavioural change, pharmaceutical and non-pharmaceutical 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 non-neutral genetic variation requires a multi-type 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 sub-models with non-random 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 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:

\[
Pr(\mathcal{T}|\Theta_{SIR}) = Pr(\mathcal{T}|\Theta_{BDS}) Pr(\Theta_{BDS}|\Theta_{SIR}).
\]

(B1)
While they are not sub-models of the general BDS process, likelihood models have been developed that capture the full non-independence 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 $\tau$ before the present day is given by

$$ R_e(\tau) = \frac{\lambda(\tau)}{\mu(\tau) + \psi(\tau) + \psi(\tau)} \cdot $$

Although the SIR model is a useful epidemiological model for its simplicity, realistically modelling 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 multi-type modelling are necessary for direct inference for the larger class of epidemiological models.

Acknowledgments

We would like to thank Sally Otto for her thoughtful comments on this work. This work was supported by a Grant for Catalyzing Research Clusters awarded to the UBC Biodiversity Research Centre, NSF DEB Grant #2028986 award to SL and MWP. MWP was supported by a NSERC Discovery Grant and AM was supported in part by the EEB department Postdoctoral Fellowship from the University of Toronto. AMcL was supported in part by a Canadian Institutes of Health Research (CIHR) doctoral award (#6557). JBJ is supported by a Genome Canada Bioinformatics and Computational Biology grant (287PHY), CIHR coronavirus rapid response program grant (440371) and is grateful to the British Columbia Centre for Excellence in HIV/AIDS for additional funding support.
References

Barido-Sottani, J., T. G. Vaughan, and T. Stadler. 2018. Detection of HIV transmission clusters from phylogenetic trees using a multi-state birth–death model. Journal of The Royal Society Interface 15:20180512.

Barido-Sottani, J., T. G. Vaughan, and T. Stadler. 2020. A Multitype Birth–Death Model for Bayesian Inference of Lineage-Specific Birth and Death Rates. Systematic Biology.

Beaulieu, J. M. and B. C. O’Meara. 2016. Detecting Hidden Diversification Shifts in Models of Trait-Dependent Speciation and Extinction. Systematic Biology 65:583–601.

Boskova, V., S. Bonhoeffer, and T. Stadler. 2014. Inference of Epidemiological Dynamics Based on Simulated Phylogenies Using Birth-Death and Coalescent Models. PLoS Computational Biology 10.

Bouckaert, R., T. G. Vaughan, J. Barido-Sottani, S. Duchêne, M. Fourment, A. Gavryushkina, J. Heled, G. Jones, D. Kühnert, N. D. Maio, M. Matschiner, F. K. Mendes, N. F. Müller, H. A. Ogilvie, L. du Plessis, A. Popinga, A. Rambaut, D. Rasmussen, I. Siveroni, M. A. Suchard, C.-H. Wu, D. Xie, C. Zhang, T. Stadler, and A. J. Drummond. 2019. BEAST 2.5: An advanced software platform for Bayesian evolutionary analysis. PLOS Computational Biology 15:e1006650.

Caetano, D. S., B. C. O’Meara, and J. M. Beaulieu. 2018. Hidden state models improve state-dependent diversification approaches, including biogeographical models. Evolution 72:2308–2324.

Drummond, A. J., O. G. Pybus, A. Rambaut, R. Forsberg, and A. G. Rodrigo. 2003. Measurably evolving populations. Trends in Ecology & Evolution 18:481–488.

Drummond, A. J., A. Rambaut, B. Shapiro, and O. G. Pybus. 2005. Bayesian Coalescent Inference of Past Population Dynamics from Molecular Sequences. Molecular Biology and Evolution 22:1185–1192.

du Plessis, L. and T. Stadler. 2015. Getting to the root of epidemic spread with phylodynamic analysis of genomic data. Trends in Microbiology 23:383–386.

Duffy, S., L. A. Shackelton, and E. C. Holmes. 2008. Rates of evolutionary change in viruses: Patterns and determinants. Nature Reviews Genetics 9:267–276.
Etienne, R. S., B. Haegeman, T. Stadler, T. Aze, P. N. Pearson, A. Purvis, and A. B. Phillimore. 2012. Diversity-dependence brings molecular phylogenies closer to agreement with the fossil record. Proceedings of the Royal Society B: Biological Sciences 279:1300–1309.

Etienne, R. S. and J. Rosindell. 2012. Prolonging the Past Counteracts the Pull of the Present: Protracted Speciation Can Explain Observed Slowdowns in Diversification. Systematic Biology 61:204–204.

Ezard, T. H. G., T. B. Quental, and M. J. Benton. 2016. The challenges to inferring the regulators of biodiversity in deep time. Philosophical Transactions of the Royal Society B: Biological Sciences 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 in Ecology and Evolution 3:1084–1092.

FitzJohn, R. G., W. P. Maddison, and S. P. Otto. 2009. Estimating Trait-Dependent Speciation and Extinction Rates from Incompletely Resolved Phylogenies. Systematic Biology 58:595–611.

Gavryushkina, A., T. A. Heath, D. T. Ksepka, T. Stadler, D. Welch, and A. J. Drummond. 2017. Bayesian Total-Evidence Dating Reveals the Recent Crown Radiation of Penguins. Systematic Biology 66:57–73.

Gavryushkina, A., D. Welch, and A. J. Drummond. 2013. Recursive algorithms for phylogenetic tree counting. Algorithms for Molecular Biology 8:26.

Gavryushkina, A., D. Welch, T. Stadler, and A. J. Drummond. 2014. Bayesian Inference of Sampled Ancestor Trees for Epidemiology and Fossil Calibration. PLoS Computational Biology 10.

Gernhard, T. 2008. The conditioned reconstructed process. Journal of Theoretical Biology 253:769–778.

Goldberg, E. E. and B. Igić. 2012. Tempo and Mode in Plant Breeding System Evolution. Evolution 66:3701–3709.

Goldberg, E. E., L. T. Lancaster, and R. H. Ree. 2011. Phylogenetic Inference of Reciprocal Effects between Geographic Range Evolution and Diversification. Systematic Biology 60:451–465.
Grenfell, B. T., O. G. Pybus, J. R. Gog, J. L. N. Wood, J. M. Daly, J. A. Mumford, and E. C. Holmes. 2004. Unifying the epidemiological and evolutionary dynamics of pathogens. Science 303:327–332.

Hagen, O., K. Hartmann, M. Steel, and T. Stadler. 2015. Age-Dependent Speciation Can Explain the Shape of Empirical Phylogenies. Systematic Biology 64:432–440.

Heath, T. A., J. P. Huelsenbeck, and T. Stadler. 2014. The fossilized birth–death process for coherent calibration of divergence-time estimates. Proceedings of the National Academy of Sciences 111:E2957–E2966.

Joy, J. B., R. M. McCloskey, T. Nguyen, R. H. Liang, Y. Khudyakov, A. Olmstead, M. Krajden, J. W. Ward, P. R. Harrigan, J. S. G. Montaner, and A. F. Y. Poon. 2016. The spread of hepatitis C virus genotype 1a in North America: A retrospective phylogenetic study. The Lancet. Infectious Diseases 16:698–702.

Keeling, M. J. and P. Rohani. 2008. Modeling Infectious Diseases: In Humans and Animals. Princeton University Press.

Kendall, D. G. 1948. On the Generalized "Birth-and-Death" Process. Annals of Mathematical Statistics 19:1–15.

Kingman, J. F. C. 1982. On the Genealogy of Large Populations. Journal of Applied Probability 19:27–43.

Kirkpatrick, M., T. Johnson, and N. Barton. 2002. General models of multilocus evolution. Genetics 161:1727–1750.

Kühnert, D., T. Stadler, T. G. Vaughan, and A. J. Drummond. 2014. Simultaneous reconstruction of evolutionary history and epidemiological dynamics from viral sequences with the birth–death SIR model. Journal of The Royal Society Interface 11:20131106.

Lafferty, K. D., G. DeLeo, C. J. Briggs, A. P. Dobson, T. Gross, and A. M. Kuris. 2015. A general consumer-resource population model. Science 349:854–857.

Lambert, A. 2018. The coalescent of a sample from a binary branching process. Theoretical Population Biology 122:30–35.

Lambert, A. and T. Stadler. 2013. Birth–death models and coalescent point processes: The shape and probability of reconstructed phylogenies. Theoretical Population Biology 90:113–128.
Landis, M. J., D. A. R. Eaton, W. L. Clement, B. Park, E. L. Spriggs, P. W. Sweeney, E. J. Edwards, and M. J. Donoghue. 2021. Joint Phylogenetic Estimation of Geographic Movements and Biome Shifts during the Global Diversification of Viburnum. Systematic Biology 70:67–85.

Laudanno, G., B. Haegeman, D. L. Rabosky, and R. S. Etienne. 2020. Detecting Lineage-Specific Shifts in Diversification: A Proper Likelihood Approach. Systematic Biology.

Lehtonen, S., D. Silvestro, D. N. Karger, C. Scotese, H. Tuomisto, M. Kessler, C. Peña, N. Wahlberg, and A. Antonelli. 2017. Environmentally driven extinction and opportunistic origination explain fern diversification patterns. Scientific Reports 7:4831.

Leventhal, G. E., R. Kouyos, T. Stadler, V. von Wyl, S. Yerly, J. Böni, C. Celleraï, T. Klimkait, H. F. Günthard, and S. Bonhoeffer. 2012. Inferring Epidemic Contact Structure from Phylogenetic Trees. PLOS Computational Biology 8:e1002413.

Louca, S. 2020. Simulating trees with millions of species. Bioinformatics.

Louca, S. and M. Doebeli. 2018. Efficient comparative phylogenetics on large trees. Bioinformatics 34:1053–1055.

Louca, S., A. McLaughlin, A. MacPherson, J. B. Joy, and M. W. Pennell. 2021. Fundamental identifiability limits in molecular epidemiology. Molecular Biology and Evolution.

Louca, S. and M. W. Pennell. 2020a. Extant timetrees are consistent with a myriad of diversification histories. Nature Pages 1–4.

Louca, S. and M. W. Pennell. 2020b. A General and Efficient Algorithm for the Likelihood of Diversification and Discrete-Trait Evolutionary Models. Systematic Biology 69:545–556.

Maddison, W. P., P. E. Midford, and S. P. Otto. 2007. Estimating a Binary Character’s Effect on Speciation and Extinction. Systematic Biology 56:701–710.

Magee, A. F. and S. Höhna. 2021. Impact of K-Pg Mass Extinction Event on Crocodylomorpha Inferred from Phylogeny of Extinct and Extant Taxa. bioRxiv Page 2021.01.14.426715.

Magnuson-Ford, K. and S. P. Otto. 2012. Linking the investigations of character evolution and species diversification. The American Naturalist 180:225–245.
Morlon, H. 2014. Phylogenetic approaches for studying diversification. Ecology Letters 17:508–525.

Morlon, H., T. L. Parsons, and J. B. Plotkin. 2011. Reconciling molecular phylogenies with the fossil record. Proceedings of the National Academy of Sciences of the United States of America 108:16327–16332.

Morlon, H., M. D. Potts, and J. B. Plotkin. 2010. Inferring the Dynamics of Diversification: A Coalescent Approach. PLOS Biology 8:e1000493.

Nee, S. 2006. Birth-Death Models in Macroevolution. Annual Review of Ecology, Evolution, and Systematics 37:1–17.

Nee, S., R. M. May, and P. H. Harvey. 1994. The reconstructed evolutionary process. Phil. Trans. R. Soc. B Page 7.

Ng, J. and S. D. Smith. 2014. How traits shape trees: New approaches for detecting character state-dependent lineage diversification. Journal of Evolutionary Biology 27:2035–2045.

Pybus, O. G., A. Rambaut, and P. H. Harvey. 2000. An Integrated Framework for the Inference of Viral Population History From Reconstructed Genealogies. Genetics 155:1429–1437.

Rabosky, D. L. and I. J. Lovette. 2008a. Density-dependent diversification in North American wood warblers. Proceedings of the Royal Society B: Biological Sciences 275:2363–2371.

Rabosky, D. L. and I. J. Lovette. 2008b. Explosive evolutionry radiation: Decreasing speciation or increasing extinction through time? Evolution 62:1866–1875.

Rasmussen, D. A. and T. Stadler. 2019. Coupling adaptive molecular evolution to phylodynamics using fitness-dependent birth-death models. eLife 8.

Raup, D. M. 1985. Mathematical models of cladogenesis. Paleobiology 11:42–52.

Romero-Severson, E. O., I. Bulla, and T. Leitner. 2016. Phylogenetically resolving epidemiologic linkage. Proceedings of the National Academy of Sciences 113:2690–2695.

Schluter, D. and M. W. Pennell. 2017. Speciation gradients and the distribution of biodiversity. Nature 546:48–55.
Silvestro, D., J. Schnitzler, L. H. Liow, A. Antonelli, and N. Salamin. 2014. Bayesian Estimation of Speciation and Extinction from Incomplete Fossil Occurrence Data. Systematic Biology 63:349–367.

Stadler, T. 2009. On incomplete sampling under birth-death models and connections to the sampling-based coalescent. Journal of Theoretical Biology 261:58–66.

Stadler, T. 2010. Sampling-through-time in birth-death trees. Journal of Theoretical Biology 267:396–404.

Stadler, T. 2011. Mammalian phylogeny reveals recent diversification rate shifts. Proceedings of the National Academy of Sciences of the United States of America 108:6187–6192.

Stadler, T. 2013. Recovering speciation and extinction dynamics based on phylogenies. Journal of Evolutionary Biology 26:1203–1219.

Stadler, T. and S. Bonhoeffer. 2013. Uncovering epidemiological dynamics in heterogeneous host populations using phylogenetic methods. Phil. Trans. R. Soc. B 368:20120198.

Stadler, T., R. D. Kouyos, V. von Wyl, S. Yearly, and J. Böni. 2012. Estimating the Basic Reproductive Number from Viral Sequence Data. Mol. Biol. Evol. .

Stadler, T., D. Kühnert, S. Bonhoeffer, and A. J. Drummond. 2013. Birth-death skyline plot reveals temporal changes of epidemic spread in HIV and hepatitis C virus (HCV). PNAS 110:228–233.

Stadler, T., T. G. Vaughan, A. Gavryushkin, S. Guindon, D. Kühnert, G. E. Leventhal, and A. J. Drummond. 2015. How well can the exponential-growth coalescent approximate constant-rate birth-death population dynamics? Proceedings. Biological Sciences 282:20150420.

Strimmer, K. and O. G. Pybus. 2001. Exploring the Demographic History of DNA Sequences Using the Generalized Skyline Plot. Molecular Biology and Evolution 18:2298–2305.

Vaughan, T. G., G. E. Leventhal, D. A. Rasmussen, A. J. Drummond, D. Welch, and T. Stadler. 2019. Estimating Epidemic Incidence and Prevalence from Genomic Data. Molecular Biology and Evolution 36:1804–1816.

Volz, E. M. 2012. Complex Population Dynamics and the Coalescent Under Neutrality. Genetics 190:187–201.
Volz, E. M. and S. D. W. Frost. 2014. Sampling through time and phylodynamic inference with coalescent and birth–death models. Journal of The Royal Society Interface 11:20140945.

Volz, E. M., S. L. Kosakovsky Pond, M. J. Ward, A. J. Leigh Brown, and S. D. W. Frost. 2009. Phylodynamics of Infectious Disease Epidemics. Genetics 183:1421–1430.
Figure 1: **Sub-model assumptions.** Rate, sampling, mass extinction, and conditioning assumptions of existing sub-models 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.

Figure B1: **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 (panel a). Simulated forward in time, the result of the BDS stochastic processes is a full tree (panel b) giving the complete genealogy of the viral population. Pruning away extinct and unsampled lineages produces the sampled tree (panel 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, $y_j$), 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.
| Model                  | Rates   | Sampling          | Mass Extinction | Conditioning          |
|------------------------|---------|-------------------|-----------------|-----------------------|
| Stadler 2009 S1.1      | constant| present-day       | constant        | N₀ samples            |
| Stadler 2010 S1.2      | constant| fossils            |                |                       |
| Morlon et al., 2011 S1.3 | present-day | fossils            | N₀ samples      |                       |
| Stadler 2011 S1.4      | piecewise constant | present-day       |                |                       |
| Stadler et al. 2012 S1.5 | constant | no present-day    |                |                       |
| Stadler et al. 2013 S1.6 | piecewise constant | CSAs              |                |                       |
| Gavryushkina et al. 2014 S1.6 | piecewise constant | CSAs              |                |                       |
| Heath et al. 2014 S1.7 | constant | present-day       | fossils         |                       |
| Kühnert et al. 2014 S1.7 | stochastic SIR | fossils          | constant        | ≥ 1 sample            |
rate constraints
\[ \lambda = \beta S \]
\[ \mu = \gamma + \delta + \alpha \]
epidemiological inference
\[ R_e = \frac{\beta S}{\gamma + \delta + \alpha + \psi} \]

a. Rates
b. Full Tree
c. Sampled Tree