EXACT PHYLODYNAMIC LIKELIHOOD
VIA STRUCTURED MARKOV GENEALOGY PROCESSES

AARON A. KING, QIANYING LIN, AND EDWARD L. IONIDES

ABSTRACT. We show that each member of a broad class of Markovian population models induces a unique stochastic process on the space of genealogies. We construct this genealogy process and derive exact expressions for the likelihood of an observed genealogy in terms of a filter equation, the structure of which is completely determined by the population model. We show that existing phylodynamic methods based on either the coalescent or the linear birth-death processes are special cases. We derive some properties of filter equations and describe a class of algorithms that can be used to numerically solve them. Our results open the door to statistically efficient likelihood-based phylodynamic inference for a much wider class of models than is currently possible.

1. Introduction

When the genome of an infectious agent accumulates mutations on timescales similar to those of transmission and infection progression, the resulting pattern of differences among genomes contains information on the history of the pathogen's passage through individual hosts and the host population. As Grenfell et al. (2004) observed, one can extract this information to gain insight into the structure and dynamics of the host-pathogen system. In particular, one can formalize mathematical models of transmission, estimate their parameters, and compare their ability to explain data, following standard statistical paradigms. This collection of tasks is known as *phylodynamic* inference; Alizon (2024) provides a recent review.

A common approach to phylodynamic inference involves building a mathematical linkage between the tree-like *genealogy* or *phylogeny* that expresses the relationships of shared ancestry among sampled genomes and a model of the dynamics of the transmission system. Various linkages are possible, but, to attain maximal statistical efficiency (i.e., lose the least information), it is desirable to be able to compute the likelihood function for models of interest. The likelihood function is the probability density of a given set of data, conditional on a given model, viewed as a function of the parameters of that model. While various kinds of data may be available, we focus here on the case where the data are a set of genomic sequences. Notationally, if the data, S, is a set of genome sequences, Φ a genealogical tree relating these sequences, E a model of sequence evolution, and D a dynamic transmission model, then the full likelihood is

$$\mathcal{L}(D, E) = f(S|D, E) = \int f(S|\Phi, E) \, f(\Phi|D) \, \mathrm{d}\Phi,$$

where the integral is taken over all possible genealogies and we somewhat loosely use the symbol f for the various distinct probability densities, the nature of each of which is clear from its arguments. In this expression, $f(S|\Phi, E)$ is typically the Felsenstein (2004) phylogenetic likelihood. The function $f(\Phi|D)$, which links the phylogeny to the dynamic model, is the *phylodynamic likelihood*. In the Bayesian context, the phylodynamic likelihood is sometimes referred to as a *tree prior* (Möller et al., 2018; Volz & Siveroni, 2018). Its computation has remained out of reach, except in several special cases. This paper presents theory that facilitates its computation for a very broad range of dynamic models.

With few exceptions, existing approaches to the phylodynamic likelihood have been based on one of two mathematical idealizations. The first is the Kingman (1982a) coalescent, by which the likelihood of a given genealogy is computed using a reverse-time argument. This computation provides the exact likelihood for a genealogy resulting from a particular, constant population-size, dynamic model (the Moran model, e.g., Moran, 1958; Kingman, 1982b; Möhle, 2000; King et al., 2020, 2022). Extensions of this approach develop approximate likelihoods for

the case when the population size varies as a function of time (Griffiths & Tavaré, 1994; Drummond et al., 2005) or according to an SIR process (Volz et al., 2009; Rasmussen et al., 2011), as long as the population size is large and the sample-fraction remains negligible. The second idealization is the linear birth-death process, for which the likelihood is available in closed form (Stadler, 2010). Linearity in this context amounts to the assumption that distinct lineages do not interact: it is the resulting self-similarity of genealogies that renders the likelihood analytically tractable. Extensions of this approach develop approximations via linearization of nonlinear processes or restriction to scenarios in which population growth is nearly linear (e.g., MacPherson et al., 2021). Although the tractability of these approaches is appealing, concern naturally arises as to validity of the approximations in specific cases, the biases introduced by them, and the amount of information in data left uncaptured by these approximate methods. For this reason, there is interest in improved phylodynamic inference techniques.

What would an ideal phylodynamic inference method look like? First, it would allow us to ask the questions we wish to ask. Since the set of scientifically interesting hypotheses is not contained within the set of statistically convenient models, an ideal phylodynamic inference methodology would put no restrictions on the form of the models that it can accommodate. In particular, since nonlinearity, nonstationarity, noise, and measurement error are prominent and ubiquitous in epidemiology, such a method would accommodate nonlinear, time-inhomogeneous, stochastic transmission models. Moreover, because many of the most scientifically and practically important uncertainties concern heterogeneities in transmission rates and the susceptibility, behavior, age, and location of hosts, it would accommodate host populations structured by these and other factors. Second, it would be both statistically efficient and robust to model misspecification. In particular, it would achieve maximal statistical efficiency by being based on the exact phylodynamic likelihood, which would also facilitate objective comparison among parameterizations and models. At the same time, model predictions would be robust to small deviations from the model assumptions. Finally, an ideal inference method would be computationally efficient. If numerical computations were absolutely required for its use, these would scale well with model complexity and data volume.

Of course, there is no hope for the realization of such an ideal method. In particular, in selecting methodology, one must navigate trade-offs between statistical efficiency and robustness, and between computational expense and fidelity of model to question. With respect to these trade-offs, far more attention has been paid to date on relatively inexpensive approaches that are restricted to models that can only approximate the motivating questions to a greater or lesser extent. The present paper offers a complementary perspective: we develop mathematical expressions, and corresponding numerical algorithms, for the phylodynamic likelihood of almost arbitrarily complex models. Our focus on the exact likelihood offers maximum information capture, even as the computations it involves are somewhat delicate. Numerical examples are postponed to Appendix C, and computational scalability is discussed in §5.

While the theory presented here greatly expands the class of models for which exact likelihood is computable, it is limited to models with *discrete structure*. Thus, while structuring factors such as age, stage of infection progression, and spatial location are most naturally expressed in terms of continuous variables, to apply the results presented here, one would have to discretize these factors. We suspect that this limitation may often be relatively painless in practice, since discretely structured models have repeatedly proved their value in epidemiology. In particular, compartmental models offer great flexibility and have often been used as approximations when continuous structure leads to uncomfortably high model dimension.

To connect a model at the level of a population with genealogies based on samples taken from individual hosts, it is necessary to make assumptions about the individuals in the population. The simplest such assumption is that the individuals that are identical with respect to the population dynamics are fully statistically identical. That is, that they are *exchangeable*. In a compartmental model, this is tantamount to the assumption that the residence times of the individuals within each compartment are identically distributed, though not independent. Although exchangeability is indeed an additional assumption, it is so natural that it is frequently unrecognized as such, and one often reads statements to the effect that exchangeability of individuals is a consequence of the Markovian assumption. Nonetheless, since it adds minimal additional structure, it is the natural assumption, and the one we will make in this paper.

From the mathematical point of view, the essential difficulty in determining expressions for the likelihood of a genealogy generated by a given population process is the fact that a genealogy's structure is determined by events occurring at widely separated times. Existing approaches to phylodynamic inference employ one of a

few tricks to get around this non-locality. In particular, both coalescent- and linear birth-death process-based approaches construct arguments in reverse time that depend on restrictions on the form of the population process. The main mathematical contribution of the present paper is to show how one can reformulate the phylodynamic likelihood in such a way that all needed computations are time-localized without placing restrictive assumptions on the population process. In particular, models for which forward simulation is possible but reversal of transition densities is intractable can be treated in this way.

In the following, we take as our starting point a population model in the form of a discretely structured continuous-time Markov process. We show how such a process uniquely induces several stochastic processes in the space of genealogies. We go on to show that exact likelihoods for these genealogies can be computed by solving a filter equation. In particular, for each given population-level model there is a specific filter equation. We then show that our approach generalizes both the coalescent- and linear birth-death-process-based approaches. Indeed, the filter equations for the models on which these approaches are based can be solved to obtain the familiar likelihood formulae. In appendices, we develop the properties of filter equations generally, provide a sequential Monte Carlo algorithm that can be used when numerical integration is needed for their solution, and work out the filter equations and their solutions in two examples.

Code sufficient for the reproduction of all the figures in this paper are freely available for download at https://github.com/kingaa/structured-genealogy-process-paper. An archival version of these will be stored on Zotero upon publication of a peer-reviewed version of this paper. The open-source **R** package **phylopomp** (https://github.com/kingaa/phylopomp) implements the simulation and likelihood-computation algorithms employed here.

2. Mathematical preliminaries

2.1. Notation. Throughout the paper, we will adopt the convention that a bold-face symbol (e.g., $\mathbf{X}$), denotes a random element. We will be concerned with a variety of stochastic processes, in both discrete and continuous time. In both cases, we will use a subscript to indicate the time parameter: e.g., $\mathbf{X}_t$ or $\mathbf{G}_k$, where t takes values in the non-negative reals $\mathbb{R}_+$ and k in the non-negative integers $\mathbb{Z}_+$. In the case of continuous-time processes, we will find ourselves needing to refer to their left- or right-limits. Accordingly, if $\mathbf{\Phi}_t$ is any random process, let

$$\widetilde{\mathbf{\Phi}}_t \coloneqq \lim_{s \uparrow t} \mathbf{\Phi}_s \qquad \text{and} \qquad \underset{\approx}{\mathbf{\Phi}}_t \coloneqq \lim_{s \downarrow t} \mathbf{\Phi}_s.$$

We will typically assume that the random processes treated here are right-continuous with left limits (i.e., càdlàg, so that $\underset{\approx}{\mathbf{\Phi}}_t = \mathbf{\Phi}_t$). Note that, in this case, $\widetilde{\mathbf{\Phi}}_t$ is left-continuous with right limits (càglàd).

If $\mathbf{\Phi}_t$, $t \in \mathbb{R}_+$ is a pure jump process, knowledge of its sample path is equivalent to knowledge of the number, $\mathbf{K}_t$, of jumps it has taken as of time t, the jump times $\hat{\mathbf{T}}_k$, and the embedded chain $\hat{\mathbf{\Phi}}_k \coloneqq \mathbf{\Phi}_{\hat{\mathbf{T}}_k}$, $k = 0, \dots, \mathbf{K}_t$. In particular, if we adopt the convention that $\hat{\mathbf{T}}_0 = 0$ and $\hat{\mathbf{T}}_{\mathbf{K}_t+1} = t$, then $\mathbf{\Phi}_t = \hat{\mathbf{\Phi}}_k$ for $t \in \left[\hat{\mathbf{T}}_k, \hat{\mathbf{T}}_{k+1}\right)$, $k = 0, \dots, \mathbf{K}_t$.

2.2. Population process. We are motivated by the desire for exact phylodynamic inference methods for as wide a class of epidemiological models as possible. In particular, we would like to be able to formulate and parameterize an arbitrary compartmental model and to quantify its ability to explain data using likelihood. Fig. 1 depicts a few of the simplest such models in order to give a sense of the kinds of complexities that can arise. Of course, with the ability to entertain models with countably many compartments, much greater complexity is possible. In particular, one can model not only complex infection progression, but also strain structure, behavioral structure, age structure, and spatial structure using compartmental models. As is well known, one can discretize continuous structure-variables and employ the linear chain trick to accommodate non-exponential residence times. While the utility of these approximations will vary, a very wide range of model assumptions lie within the scope of the theory presented here.

We will refer to the stochastic process defined by the given model as the *population process*. We will assume that the population process is a time-inhomogeneous Markov jump process, $\mathbf{X}_t$, $t \in \mathbb{R}_+$, taking values in some space $\mathbb{X}$. In earlier work (King et al., 2022), we limited ourselves to the case $\mathbb{X} = \mathbb{Z}^d$, but here we assume only that $\mathbb{X}$ is a complete metric measure space with a countable dense subset. The population process is completely specified
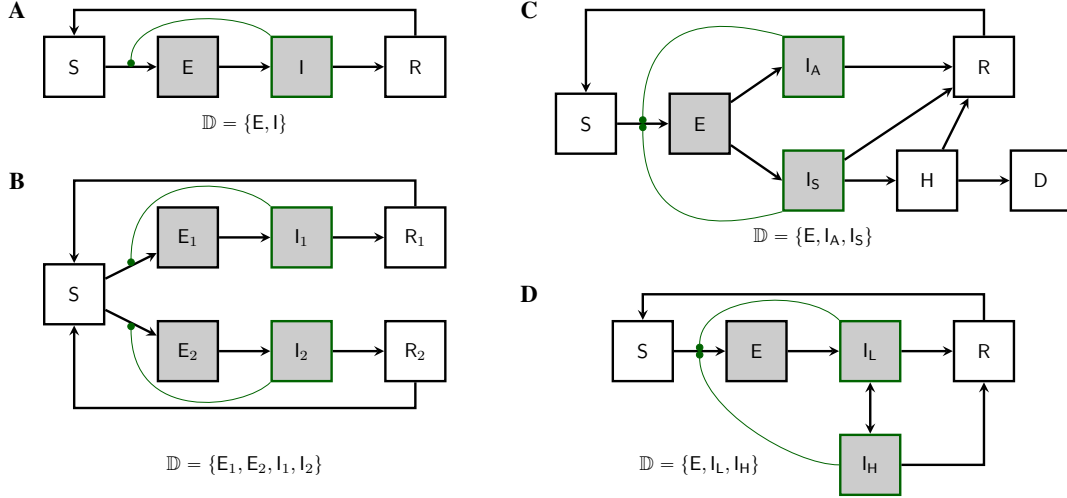
FIGURE 1. Examples of discretely-structured population models. Demes are shaded. Compartments containing infectious hosts are outlined in green. Curved green lines connect transmission rates with the compartments whose occupancies control their modulation; each such connection gives rise to a nonlinearity in the model. **(A)** An SEIRS model. Susceptible individuals (S), once infected, enter a transient incubation phase (E) before they become infectious (I). Upon recovery (R), individuals experience immunity from reinfection. If this immunity wanes, they re-enter the susceptible compartment. Pathogen lineages are to be found in hosts within the E and I compartments only. Accordingly, there are two demes: $\mathbb{D} = \{E, I\}$. If there is exactly one lineage per host, then the occupancy, $n(\mathbf{X}_t) = (n_E(\mathbf{X}_t), n_I(\mathbf{X}_t))$, is the integer 2-vector giving the numbers of hosts in the respective compartments. See §2.6 for definition and discussion of demes and deme occupancy. **(B)** In this four-deme model, two distinct pathogen strains compete for susceptibles. **(C)** A three-deme model according to which, after an incubation period, hosts may develop asymptomatic infection (I_A). If they do not recover, symptomatically infected hosts (I_S) can progress to hospitalization (H) and death (D). **(D)** A three-deme model with heterogeneity in transmission behavior. Contagious individuals move randomly between low-transmission (I_L) and high-transmission (I_H) behaviors.

by its initial-state density, p_0, and its transition rates α. In particular, we suppose that

$$\text{Prob}\left[\mathbf{X}_0 \in \mathcal{E}\right] = \int_{\mathcal{E}} p_0(x)\,\mathrm{d}x \tag{1}$$

for all measurable sets $\mathcal{E} \subseteq \mathbb{X}$. For any $t \in \mathbb{R}_+$, $x, x' \in \mathbb{X}$, we think of the quantity $\alpha(t, x, x')$ as the instantaneous hazard of a jump from x to x'. More precisely, the transition rates have the following properties:

$$\alpha(t, x, x') \geqslant 0, \qquad \int_{\mathbb{X}} \alpha(t, x, x')\,\mathrm{d}x' < \infty,$$

for all $t \in \mathbb{R}_+$ and $x, x' \in \mathbb{X}$ and that, as a function of time, α is càdlàg and continuous almost everywhere. Henceforth, we understand that integrals are taken over all of $\mathbb{X}$ unless otherwise specified. Let $\mathbf{K}_t$ be the number

of jumps that $\mathbf{X}$ has taken by time t. We assume that $\mathbf{K}_t$ is a simple counting process so that

$$\mathsf{Prob}\left[\mathbf{K}_{t+\Delta} = n + 1 \mid \mathbf{K}_t = n\right] = \Delta \int \alpha(t, x, x')\, \mathrm{d}x' + o(\Delta),$$

$$\mathsf{Prob}\left[\mathbf{K}_{t+\Delta} > n + 1 \mid \mathbf{K}_t = n\right] = o(\Delta),$$

$$\mathsf{Prob}\left[\mathbf{X}_{t+\Delta} \in \mathcal{E} \backslash \{x\} \mid \mathbf{X}_t = x\right] = \Delta \int_{\mathcal{E} \backslash \{x\}} \alpha(t, x, x')\, \mathrm{d}x' + o(\Delta),$$

$$\mathsf{Prob}\left[\mathbf{X}_{t+\Delta} = x \mid \mathbf{X}_t = x\right] = 1 - \Delta \int \alpha(t, x, x')\, \mathrm{d}x' + o(\Delta).$$

We further assume that $\mathbf{X}_t$ is *non-explosive*, i.e., that $\mathsf{Prob}\left[\mathbf{K}_t < \infty\right] = 1$ for all $t < \infty$.

2.3. Kolmogorov equations. The above may be compactly summarized by stating that if $v(t, x)$ satisfies the Kolmogorov forward equation (KFE),

$$(2) \qquad \frac{\partial v}{\partial t}(t, x) = \int v(t, x')\, \alpha(t, x', x)\, \mathrm{d}x' - \int v(t, x)\, \alpha(t, x, x')\, \mathrm{d}x', \qquad v(0, x) = p_0(x),$$

for $t \in [0, T]$, then $\int_{\mathcal{E}} v(t, x)\, \mathrm{d}x = \mathsf{Prob}\left[\mathbf{X}_t \in \mathcal{E}\right]$ for every $t \in [0, T]$ and measurable $\mathcal{E} \subseteq \mathbb{X}$. Eq. 2 is sometimes called the *master equation* for $\mathbf{X}_t$. The adjoint form of the KFE is the Kolmogorov backward equation,

$$(3) \qquad -\frac{\partial F}{\partial s}(s, x) = \int \alpha(t, x, x')\left[F(s, x') - F(s, x)\right]\mathrm{d}x', \qquad F(T, x) = f(x).$$

If F satisfies Eq. 3 for $s \in [0, T]$ and $x \in \mathbb{X}$, then $F(s, x) = \mathbb{E}\left[f(\mathbf{X}_T) \mid \mathbf{X}_s = x\right]$ for all $t \in [0, T]$ and $x \in \mathbb{X}$.

2.4. Inclusion of jumps at deterministic times. For modeling purposes, it is sometimes desirable to insist that certain events occur at specified times. For example, if samples are collected at specific times in such a way that the timing itself conveys no information about the process, one might wish to condition on the sampling time. We can expand the class of population models to allow for this as follows. Suppose that $S = \{s_1, s_2, \ldots,\} \subset \mathbb{R}_+$ is a sequence of times. Let us postulate that, at each of these times, an event occurs at which $\mathbf{X}_t$ jumps according to a given probability kernel π. In particular, for any state $x \in \mathbb{X}$ and measurable $\mathcal{E} \subset \mathbb{X}$, $\pi(s_i, x, \mathcal{E})$ is the probability that the jump at time s_i is to $\mathcal{E}$, conditional on $\widetilde{\mathbf{X}}_{s_i} = x$. With this notation, the KFE for the process becomes

$$(4) \qquad \frac{\partial v}{\partial t}(t, x) = \int v(t, x')\, \alpha(t, x', x)\, \mathrm{d}x' - \int v(t, x)\, \alpha(t, x, x')\, \mathrm{d}x', \qquad t \notin S,$$

$$(5) \qquad v(t, x)\, \mathrm{d}x = \int \widetilde{v}(t, x')\, \pi(t, x', \mathrm{d}x)\, \mathrm{d}x', \qquad t \in S.$$

Note that Eq. 4 is identical to Eq. 2; we call this the *regular part* of the KFE. We refer to Eq. 5 as the *singular part* of the KFE. In this notation, $\pi(t, x', \mathrm{d}x)/\mathrm{d}x$ is the density (i.e., Radon-Nikodym derivative) of π with respect to the base measure on $\mathbb{X}$.

As a matter of notation, one can represent Eqs. 4 and 5 as a single equation in the form of Eq. 2. In particular, if in Eq. 2 we make the substitution

$$\alpha(t, x, x')\, \mathrm{d}x' \mapsto \alpha(t, x, x')\, \mathrm{d}x' + \sum_{s \in S} \delta(t, s)\, \pi(t, x, \mathrm{d}x'),$$

we obtain an equation which we can view as shorthand for Eqs. 4 and 5. Here, $\delta(t, s)$ is a one-sided Dirac delta function.

2.5. Jump marks. It will be useful to divide the jumps of the population process $\mathbf{X}_t$ into distinct categories which differ with respect to the changes they induce in a genealogy. For this purpose, we let $\mathbb{U}$ be a countable set of jump *marks* such that

$$\alpha(t, x, x') = \sum_{u \in \mathbb{U}} \alpha_u(t, x, x').$$

Fig. 2 shows an example for which $\mathbb{U}$ has five elements. In the following, sums over u are to be taken over the whole of $\mathbb{U}$ unless otherwise indicated.

Let us define the *jump mark process*, $\mathbf{U}_t$, to be the mark of the latest jump as of time t. As usual, we take the sample paths of $\mathbf{U}_t$ to be càdlàg. Observe that $\mathbf{X}_t$ and $(\mathbf{X}_t, \mathbf{U}_t)$ are Markov processes, but that $\mathbf{U}_t$ is not.
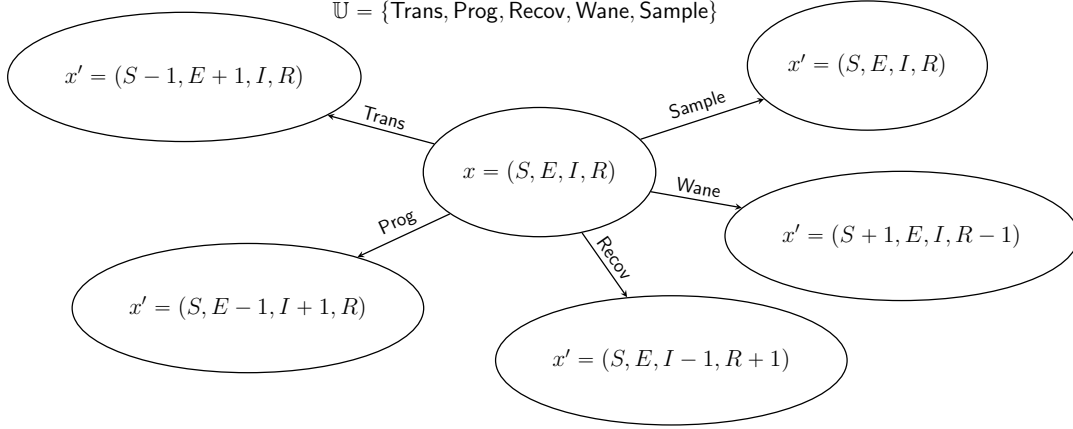
FIGURE 2. Markov state transition diagram for the SEIRS model depicted in Fig. 1A. The state, x, is characterized by four numbers, S, E, I, and R. From a given state x, there are five possible kinds of jumps $x \mapsto x'$. Accordingly, the set, $\mathbb{U}$, of jump marks has five elements. In the terminology of §3.2, each of these is of a different type: Trans (transmission) is of birth type, Prog (progression) is of migration type, Recov (recovery) is of death type, Sample (sampling) is of sample type, and Wane (loss or waning of immunity) is of neutral type. Note that, in this formulation, when a sampling event occurs, the state does not change.

2.6. Demes and deme occupancy. Our first goal in this paper is to show how a given population process induces a stochastic process on the space of genealogies. At each time, this genealogy will represent the relationships of shared ancestry among a population of lineages extant at that time. To accommodate the structure of the population, this population of lineages will itself be subdivided into discrete categories. In particular, we suppose that there are a countable set of subpopulations, within each of which individual lineages are exchangeable. We call these subpopulations *demes* and let $\mathbb{D}$ be an index set for them. We note that other authors have used different terminology for the same concept, including "colony", "type", "state", and "population" (Takahata, 1988; Notohara, 1990; Boskova et al., 2014; Volz & Siveroni, 2018; Barido-Sottani et al., 2020; Seidel et al., 2024; Vaughan & Stadler, 2025). Fig. 1 illustrates this concept in the context of several specific models.

We define the *deme occupancy* function $n : \mathbb{D} \times \mathbb{X} \to \mathbb{Z}_+$ so that for $i \in \mathbb{D}$, $x \in \mathbb{X}$, $n_i(x)$ is the number of lineages in deme i when the population is in state x.

2.7. Examples. The class of population models to which the theory presented here applies is very broad. In particular, it encompasses the entire class of compartmental models with time-dependent flow rates. Here, to give a sense of this breadth, we briefly describe a few models of interest.

SIRS model. King et al. (2022) developed formulae for the exact likelihood of a genealogy induced by an SIRS model. The theory developed in this paper applies, but since there is only one deme in this model, it is a simple case. In Appendix C, we illustrate the theory by working it out in detail for this model.

SEIRS model. Modifying the SIRS model by incorporating an incubation period results in the SEIRS model, which has two demes (Fig. 1A): $\mathbb{D} = \{\mathsf{E}, \mathsf{I}\}$. We can take the state space to be $\mathbb{Z}_+^4$: the state $x = (S, E, I, R)$ is defined by the numbers of hosts in each of the four compartments. The deme occupancy function in this case is $n(x) = (E, I)$. In Appendix C, we work out the theory for this model in detail.

Two-strain competition model. A simple model for the competition of two pathogen strains is depicted in Fig. 1B. In this model, the state vector consists of seven numbers: $x = (S, E_1, E_2, I_1, I_2, R_1, R_2)$. There are four demes ($\mathbb{D} = \{\mathsf{E}_1, \mathsf{E}_2, \mathsf{I}_1, \mathsf{I}_2\}$) and the occupancy function is $n(x) = (E_1, E_2, I_1, I_2)$.

Complex infection progression. Fig. 1C depicts a model in which infections are heterogeneous with respect to the severity of the disease they engender. There are three demes ($\mathbb{D} = \{\mathsf{E}, \mathsf{I}_\mathsf{A}, \mathsf{I}_\mathsf{S}\}$). Asymptomatic infections (I_A) are unobservable, while symptomatic infections (I_S) may develop into hospitalized cases (H) and deaths (D).

Superspreading model. Compartmental models can be used to capture heterogeneities of various kinds. Fig. 1D depicts a model with behavioral heterogeneity. There are three demes ($\mathbb{D} = \{\mathsf{E}, \mathsf{I_L}, \mathsf{I_H}\}$). Hosts in the low-transmitting $\mathsf{I_L}$ compartment have fewer contacts than those in the high-transmitting $\mathsf{I_H}$ compartment, and individuals move between these compartments as they engage in bouts of risky or cautious behavior.

Kingman coalescent and Moran model. The Kingman (1982a) coalescent is one of the two foundations upon which most existing phylodynamic approaches have been constructed. It is the ancestral process for the Moran model, in which a fixed population of n lineages experiences events at times distributed according to a rate-μ Poisson process. At each such event, an individual lineage selected uniformly at random dies and is replaced by the offspring of a second randomly selected lineage. In §4.3.1, we show that the Kingman coalescent likelihood is a special case of the theory presented in this paper.

Linear birth-death model. The linear birth-death process, the second basis for widely-used phylodynamic methods, is also a special case of the theory presented here. For this process, we have $\mathbb{X} = \mathbb{Z}_+$ and there is a single deme. $\mathbf{X}_t$ represents the size of a population and $n(x) = x$. In §4.3.2, we derive exact expressions for the likelihood under this model.

2.8. History process. Consider the Markov process $(\mathbf{X}_t, \mathbf{U}_t)$. We define its *history process*, $\mathbf{H}_t$, to be the restriction of the random function $s \mapsto (\mathbf{X}_s, \mathbf{U}_s)$ to the interval $[0, t]$. Note that $\mathbf{H}_t$ is itself trivially a Markov process, since it contains its own history. Alternatively, one can identify $\mathbf{H}_t$ with the sequence $\left(\left(\hat{\mathbf{T}}_k, \hat{\mathbf{X}}_k, \hat{\mathbf{U}}_k \right) \right)_{k=0}^{\mathbf{K}_t}$. In particular, conditional on $\mathbf{H}_t$, both $\mathbf{X}_t$ and $\mathbf{U}_t$ are deterministic, as are $\mathbf{K}_t$, the embedded chains, $\hat{\mathbf{X}}_k$, $\hat{\mathbf{U}}_k$, and the point process of event times $\hat{\mathbf{T}}_k$. The probability measure on the space of histories can be expressed in terms of these:

$$(6) \quad \mathsf{Prob}\left[\mathrm{d}\mathbf{H}_t \right] = p_0(\hat{\mathbf{X}}_0) \, \mathrm{d}\hat{\mathbf{X}}_0 \prod_{k=1}^{K_t} \alpha_{\hat{\mathbf{U}}_k}\left(\hat{\mathbf{T}}_k, \hat{\mathbf{X}}_{k-1}, \hat{\mathbf{X}}_k \right) \, \mathrm{d}\hat{\mathbf{X}}_k \, \mathrm{d}\hat{\mathbf{T}}_k \, \exp\left(-\sum_{k=0}^{K_t} \int_{\hat{\mathbf{T}}_k}^{\hat{\mathbf{T}}_{k+1}} \sum_u \int \alpha_u(t', \hat{\mathbf{X}}_k, x') \, \mathrm{d}x' \, \mathrm{d}t' \right),$$

where again, by convention, $\hat{\mathbf{T}}_0 = 0$ and $\hat{\mathbf{T}}_{K_t+1} = t$.

If H is a potential value of a history process, $\mathbf{H}_t$, we define $\mathsf{t}(\mathrm{H})$ to be the right endpoint of its domain and use the notation $\mathsf{ev}(\mathrm{H}) \coloneqq \left\{ \hat{\mathbf{T}}_1, \ldots, \hat{\mathbf{T}}_{K_t} \right\} \subset [0, \mathsf{t}(\mathrm{H})]$ to denote the set of its jump times.

3. Genealogy processes

3.1. Genealogies. A *genealogy*, G, encapsulates the relationships of shared ancestry among a set of lineages that are extant at some time $\mathsf{t}(G) \in \mathbb{R}_+$ and perhaps a set of samples collected at earlier times (Fig. 3). A genealogy has a tree- or forest-like structure, with four distinct kinds of nodes: (i) *tip nodes*, which represent labeled extant lineages; (ii) *internal nodes*, which represent events at which lineages diverged and/or moved from one deme to another; (iii) *sample nodes*, which represent labeled samples; and (iv) *root nodes*, at the base of each tree. Each node a is associated with a specific time, $\mathsf{t}(a)$. In particular, if a is a tip node in G, then $\mathsf{t}(a) = \mathsf{t}(G)$; if a is a sample node, then $\mathsf{t}(a) \leqslant \mathsf{t}(G)$ is the time at which the sample was taken. Moreover, if node a is ancestral to node a', then $\mathsf{t}(a) \leqslant \mathsf{t}(a')$ and $\mathsf{t}(a') - \mathsf{t}(a)$ is the distance between a and a' along the genealogy. Without loss of generality we assume that $\mathsf{t}(a) = 0$ for all root nodes a. We let $\mathsf{ev}(G)$ denote the set of all internal and sample node-times of the genealogy G; we refer to these as *genealogical event times*.

Importantly, a genealogy informs us not only about the shared ancestry of any pair of lineages, but also about where in the set of demes any given lineage was at all times. Accordingly, we can visualize a genealogy as a graph, the nodes and edges of which are painted with a distinct color for each deme (Fig. 3). Note that a genealogy will in general have *branch-point nodes*, i.e., internal nodes with more than one descendant, but may also have internal nodes with only one descendant. We refer to such nodes as *inline nodes*. These occur whenever the color changes along a branch, but can also occur without a color-change.

Formally, we define a genealogy, G, to be a triple, (T, Z, Y), where $T = \mathsf{t}(G) \in \mathbb{R}_+$ is the *genealogy time*, Z specifies the genealogy's *tree structure*, and Y gives the *coloring*. In particular, let $\mathbb{L}$ be a countable set of labels—referring to samples and/or extant lineages—and let $\mathsf{partit}(\mathbb{L})$ be the set of all collections of finite, mutually-disjoint subsets of $\mathbb{L}$. That is, an element $z \in \mathsf{partit}(\mathbb{L})$ is a partition of the finite set $\bigcup z \subseteq \mathbb{L}$. Partition
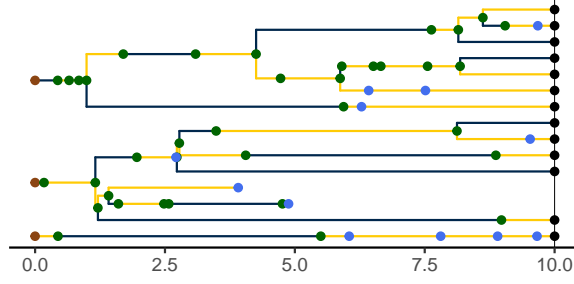
FIGURE 3. A genealogy, G, specifies the relationships of shared ancestry (via its tree-structure) and deme occupancy histories (via the coloring of its branches) of a set of lineages extant at some time $\mathsf{t}(G)$, as well as some samples gathered at earlier times. Here, $\mathsf{t}(G) = 10$ and there are two demes, $\mathbb{D} = \{\mathsf{blue}, \mathsf{yellow}\}$. Tip nodes, denoting extant lineages, are shown as black dots; sample nodes are shown as blue dots; internal nodes are indicated in green. Note that internal nodes occur not only at branch-points, but also inline (i.e., along branches). Wherever a lineage moves from one deme (color) to another, an internal node occurs; the converse does not necessarily hold.

fineness defines a partial order on $\mathsf{partit}(\mathbb{L})$. Specifically, for $z, z' \in \mathsf{partit}(\mathbb{L})$, we say $z \preccurlyeq z'$ if and only if for every $b' \in z'$ there is $b \in z$ such that $b \supseteq b'$. The tree structure of G is a càdlàg map $Z : [0, T] \to \mathsf{partit}(\mathbb{L})$ that is monotone in the sense that $t_1 \leqslant t_2$ implies $Z_{t_1} \preccurlyeq Z_{t_2}$. An element $b \in Z_t$ is a set of labels; it represents the branch of the tree that bears the corresponding lineages. We use the notation $\mathsf{ev}(Z)$ to denote the set of times at which Z is discontinuous. Note that $\mathsf{ev}(Z)$ includes the times of all tip, sample, and branch-point nodes, but excludes root and non-sample inline nodes. Therefore, $\mathsf{ev}(Z) \subseteq \mathsf{ev}(G)$.

The third element of G specifies the coloring of branches and locations of tip, sample, and internal nodes (including inline nodes). Mathematically, if $G = (T, Z, Y)$, then Y is a càdlàg function that maps each point on the genealogy to a deme and a non-negative integer. In particular, if $t \in [0, T]$ and a is the label of any tip or sample node, $Y_t(a) = (Y_t^{\mathsf{d}}(a), Y_t^{\mathsf{m}}(a)) \in \mathbb{D} \times \mathbb{Z}_+$, where $Y_t^{\mathsf{d}}(a)$ is the deme in which the lineage of a is located at time t and $Y_t^{\mathsf{m}}(a)$ is the number of internal or sample nodes encountered along the lineage of a in going from time 0 to time t. In particular, $Y_t^{\mathsf{m}}(a)$ is a simple counting process, with $Y_0^{\mathsf{m}}(a) = 0$ for all a. Since $a, a' \in b \in Z_t$ implies $Y_t(a) = Y_t(a')$, one can equally well think of Y_t as a map $Z_t \to \mathbb{D} \times \mathbb{Z}_+$. Given a tree Z, we let $\mathsf{Y}(Z)$ denote the set of colorings Y that are compatible with Z. We moreover define $\mathsf{Y}_t(Z) := \{Y_t \mid Y \in \mathsf{Y}(Z)\}$. Formally speaking, $\mathsf{Y}(Z)$ is a fiber bundle over Z, each $\mathsf{Y}_t(Z)$ being a fiber.

It will sometimes be convenient to make use of notation whereby the tree structure of genealogy G is G^{Z} and the coloring is G^{Y}, so that $G = (\mathsf{t}(G), G^{\mathsf{Z}}, G^{\mathsf{Y}})$.

3.2. Event types. To see how events in the population process leave a trace in the genealogy, we assume that, at each jump in the population process, a corresponding change occurs in the genealogy, according to whether lineages branch, die, move between demes, or are sampled. For this purpose, there are five distinct *pure types* of events:

(a) *Birth-type events* result in the branching of one or more new lineages from an existing lineage. Examples of birth-type events include transmission events, speciations, and actual births. Importantly, we assume that all new lineages arising from a birth event share the same parent and that at most one birth event occurs at a time, almost surely. While it is possible to relax this assumption, the resulting notational complexities are nontrivial, so we postpone consideration of this more general case.

(b) *Death-type events* result in the extinction of one or more lineages. Examples include recovery from infection, death of a host, and species extinctions. We allow for the possibility that multiple lineages, potentially in multiple demes, die simultaneously.

(c) *Migration-type events* result in the movement of a lineage from one deme to another. Spatial movements, changes in host age or behavior, and progression of an infection can all be represented as migration-type

events. In a migration-type event, one or more lineages within exactly one deme may move simultaneously, and they may move to different demes. Again, while it possible to relax the assumption that exactly one deme is the source of all migrating lineages within any one event, we postpone consideration of this case for the present.

(d) *Sample-type events* result in the collection of a sample from a lineage. We allow for the possibility that multiple samples are collected simultaneously, though we require that, in this case, each extant lineage is sampled at most once and that all sampled lineages be in the same deme. It is possible to allow for simultaneous sampling of multiple demes but, again, we postpone consideration of this case to ease the exposition.

(e) *Neutral-type events* result in no change to any of the lineages.

Fig. 2 depicts an example with jumps of all five pure types. It is not necessary that an event be of a pure type; *compound events* partake of more than one type. For example, a sample/death-type event, in which a lineage is simultaneously sampled and removed, has been employed (Leventhal et al., 2014), as have birth/death events in which one lineage reproduces at the same moment that another dies (e.g., the Moran (1958) process). In combining the pure types, the main restriction is that, at any compound event, all parent lineages must be drawn from the same deme. Beyond this, the theory presented here places few restrictions on the complexity of the events that can occur by combining events of the various pure types.

3.3. Genealogy process. We now show how a given population process induces a stochastic process, $\mathbf{G}_t$, on the space of genealogies. In the case of unstructured population processes (i.e., those having a single deme), King et al. (2022) gave a related construction that is equivalent to the one presented here.

At each jump in the population process, a change is made to the genealogy, according to the mark, u, of the jump (Fig. 4). In particular:

(a) If u is of birth-type (Fig. 4A), it results in the creation of one new internal node, call it b. A tip node, a, of the appropriate deme is chosen with uniform probability from among those present and b is inserted so that its ancestor is that of a, while a takes b as its ancestor. One new tip node, of the appropriate deme, is created for each of the children, all of which take b as their immediate ancestor.

(b) If u is of death-type (Fig. 4B), one or more tip nodes of the appropriate demes are selected with uniform probability from among those present. These are deleted. Next, non-sample internal nodes without children are recursively removed. Sample nodes are never removed.

(c) At a migration-type event (Fig. 4C), the appropriate number of migrating lineages are selected at random with uniform probability, from among those present in the appropriate deme. For each selected lineage, one new branch node is inserted between the selected tip node and its ancestor. The color of the descendant branch changes accordingly.

(d) At a sample-type event (Fig. 4D), the appropriate number of sampled lineages are selected at random from among the tip nodes, with uniform probability according to deme. One new sample node is introduced for each selected lineage: each is inserted between a selected tip node and its ancestor.

(e) At a neutral-type event (Fig. 4E), no change is made to the genealogy.

(f) Finally, events of compound type (e.g., Fig. 4F–H) are accommodated by combining the foregoing rules.

In each of these events, the new node or nodes that are introduced have node-times equal to the time of the jump.

3.3.1. Emergent lineages and production. At each event, the lineages which descend from a node inserted at that event are said to *emerge* from the event. Thus, after a birth-type event, the emerging lineages include all the new offspring as well as the parent. Likewise, at pure migration- or sample-type events, each migrating or sampled lineage emerges from the event. At pure death-type events, no lineages emerge. In general, at an event of mark u, there are r_i^u emergent lineages in deme i. We require that r_i^u be a constant, for each u and i. Thus there is a function $r : \mathbb{U} \times \mathbb{D} \to \mathbb{Z}_+$, such that r_i^u lineages of deme i emerge from each event of mark u. Since, in applications, one is free to expand the set of jump-marks $\mathbb{U}$ as needed, this is not a restriction on the models that the theory can accommodate. We say $r^u := (r_i^u)_{i \in \mathbb{D}}$ is the *production* of an event of mark u. Note that the lineages that die as a result of an event do not count in the production but that a parent lineage that survives the event does count.
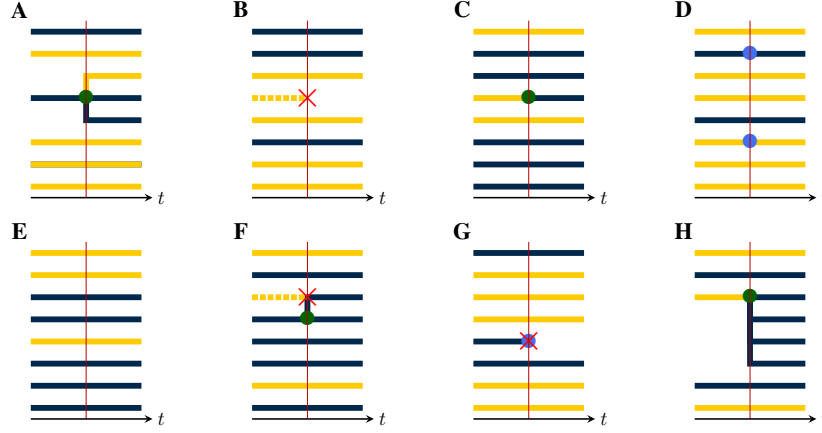
FIGURE 4. Event types differ by their effects on the genealogy. This can be seen by examining the local structure of the genealogy in the neighborhood of a jump. **(A)** A birth-type jump results in the branching of one or more child lineages from the parent. There can be only one parent, though the demes of the child lineages may differ from that of their parent. Here, a parent of the blue deme sires one child lineage in each of the blue and yellow demes. The *production* of an event is an integer vector, with one entry for each deme. The production of this event is therefore $r = (r_{\text{blue}}, r_{\text{yellow}}) = (2, 1)$. The *deme occupancy* of an event is the number of lineages in each deme just to the right of the event. The deme occupancy at this event is therefore $n = (n_{\text{blue}}, n_{\text{yellow}}) = (3, 5)$. **(B)** A death-type event causes the extinction of a lineage. Since internal nodes without children are recursively removed, the affected branch is dropped. The production of this event is $r = (0, 0)$ and the deme occupancy is $n = (3, 4)$. **(C)** A migration-type event results in the movement of one or more lineages from one deme to another. Here, one lineage moves from the yellow to the blue deme. The production of this event is $r = (1, 0)$, i.e., the production is 1 for the blue deme and 0 for the yellow. The deme occupancy is $n = (6, 2)$. **(D)** In a sample-type event, one or more sample nodes (blue circles) are inserted. Here, there are two samples, one in each of the blue and yellow demes. Accordingly, $r = (1, 1)$ and $n = (2, 6)$. **(E)** A neutral-type event has no effect on the genealogy and zero production in all demes: $r = (0, 0)$, $n = (5, 3)$. **(F)** The theory presented here allows for compound events. As an example, here a birth/death-type event occurs, wherein one yellow lineage is extinguished and a blue lineage simultaneously sires a blue child. For this event, we have $r = (2, 0)$ and $n = (6, 2)$. **(G)** Here, a compound sample/death-type event with $r = (0, 0)$ and $n = (2, 5)$ occurs. A blue lineage is sampled and simultaneously extinguished. Note that recursive removal does not occur, since sample nodes are never removed. **(H)** A compound birth/migration-type event with $r = (4, 0)$ and $n = (6, 2)$.

3.3.2. Conditional independence and exchangeability. Application of these rules at each jump of $\mathbf{X}_t$ constructs a chain of genealogies $\hat{\mathbf{G}}_k$. In particular, at each jump-time $\hat{\mathbf{T}}_k$, the genealogy $\hat{\mathbf{G}}_{k-1}$ is modified according to the jump-mark $\hat{\mathbf{U}}_k$ to yield $\hat{\mathbf{G}}_k$. We view $\hat{\mathbf{G}}_k$, $k = 0, 1, 2, \ldots$, as the embedded chain of the continuous-time genealogy process $\mathbf{G}_t$. It is very important to note that, conditional on $(\hat{\mathbf{X}}_k, \hat{\mathbf{U}}_k)$, the number of parents and number of offspring in each deme is determined and the random choice of which lineages die, migrate, are sampled, or sire offspring is independent of these choices at any other times and independent of $(\hat{\mathbf{X}}_j, \hat{\mathbf{U}}_j)$ for all $j \neq k$. Moreover, by assumption, the lineages within each deme are exchangeable: any lineage within a deme is as likely as any other lineage in that deme to be selected as a parent or for death, sampling, or migration. Finally, note that $\mathbf{G}_t$ does not have the Markov property, though $(\mathbf{X}_t, \mathbf{U}_t, \mathbf{G}_t)$ and $(\mathbf{X}_t, \mathbf{G}_t)$ do. In passing, we observe that, if instead of dropping tip nodes at death events we were to retain them as we do samples, the resulting genealogy—which we might call the *complete genealogy*—would have the Markov property.
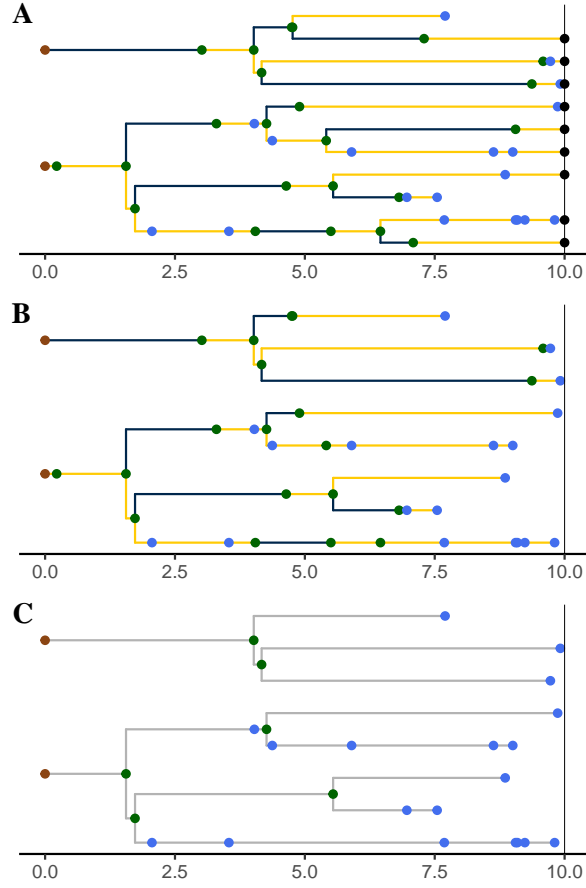
FIGURE 5. Unpruned, pruned, and obscured genealogies from a single realization of the genealogy process induced by the SEIRS model depicted in Figs. 1 and 2. **(A)** A realization of the unpruned genealogy process $\mathbf{G}_t$ is shown at $t = 10$. Tip nodes, corresponding to lineages alive at time $t = 10$ are indicated with black points. Blue points represent samples; green points, internal nodes. Branches are colored according to the deme in which the corresponding lineage resided at that point in time: blue denotes E and yellow, I. **(B)** The genealogy is *pruned* by deleting all tip nodes and then recursively pruning away childless internal nodes. Sample nodes are never removed. **(C)** A pruned genealogy is *obscured* by effacing all deme information from lineage histories: the colors are erased, as are all inline nodes. See the text (§§3.1, 3.4.1, and 3.4.4) for more detail.

3.4. Pruned and obscured genealogies. The process just described yields a genealogy that relates all extant members of the population, and all samples. Moreover, it details each lineage's complete history of movement through the various demes. However, the data we ultimately wish to analyze will be based only on samples. Nor, in general, will the histories of deme occupancy be observable. A generative model must account for this loss of information. We therefore now describe how genealogies are *pruned* to yield sample-only genealogies and then *obscured* via the erasure of color from their branches (Fig. 5).

3.4.1. Pruned genealogy. Given a genealogy G, one obtains the *pruned genealogy*, $P = \mathrm{prune}(G)$ by first dropping every tip node and then recursively dropping every childless internal node (Fig. 5A–B). In a pruned genealogy only internal and sample nodes remain, and sample nodes are found at all of the leaves and possibly some of the interior nodes of the genealogy. Observe that a pruned genealogy is a colored genealogy: it retains information about where among the demes each of its lineages was through time (Fig. 5B). Note also that a pruned genealogy P is characterized by its time, $\mathsf{t}(P)$ and the functions P^{Y} and P^{Z} just as an unpruned genealogy is.
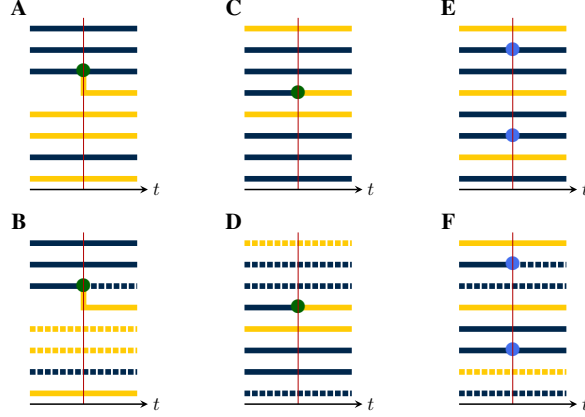
FIGURE 6. **Lineage count and saturation.** Each panel shows the neighborhood of a single event in the unpruned genealogy (top row) and the corresponding pruned genealogy (bottom row). Pruning consists of the removal of all branches that are not ancestral to some sample. In the bottom row of panels, pruned branches are indicated using broken lines. **(A)** A birth-type event with production $r = (r_{\mathsf{blue}}, r_{\mathsf{yellow}}) = (1, 1)$ occurs. **(B)** Suppose that pruning results in the removal of the dashed lineages. Then the lineage count at this event-time is $\ell = (\ell_{\mathsf{blue}}, \ell_{\mathsf{yellow}}) = (2, 2)$. The saturation is $s = (0, 1)$ since only a single, yellow lineage emerges from the event. **(C)** A migration-type event with production $r = (0, 1)$ occurs. **(D)** After pruning, $\ell = (2, 2)$ and $s = (0, 1)$. **(E)** A sample-type event occurs in which two blue lineages are sampled (production $r = (2, 0)$). **(F)** After pruning, $\ell = (2, 2)$ and $s = (1, 0)$. Observe that in panels B and D, the local structures of the pruned genealogies are identical, though they arise from events of different type.

Crucially, since it contains within itself all of its past history, the pruned genealogy process $\mathbf{P}_t = \mathsf{prune}(\mathbf{G}_t)$ is Markov, even though the unpruned genealogy process, $\mathbf{G}_t$, is not.

3.4.2. Lineage count and saturation. In the following, we will find that we need to count the deme-specific numbers of lineages present in a given pruned genealogy at a given time. Accordingly, suppose $P = (T, Z, Y)$ is a pruned genealogy and suppose $t \in [0, T]$. Let ℓ_i denote the number of lineages in deme i at time t and $\ell \coloneqq (\ell_i)_{i \in \mathbb{D}} \in \mathbb{Z}_+^{\mathbb{D}}$. Clearly, ℓ depends only Y_t. Therefore, we can define ℓ as a function such that, whenever $P = (T, Z, Y)$ is a pruned genealogy, $\ell(Y_t)$ is the vector of deme-specific lineage counts at time t. We refer to ℓ as the *lineage-count* function (cf. Fig. 6).

We will also have occasion to refer to the deme-specific number of lineages emerging from a given event. In particular, given a node time t in a pruned genealogy $P = (T, Y, Z)$, the number s_i of lineages of deme i emerging from all nodes with time t is well defined and we can write $s \coloneqq (s_i)_{i \in \mathbb{D}}$. Like the lineage-count function, s depends only on the local structure of P. However, s depends not only on Y_t, but also on $\widetilde{Y}_t$. Thus, we can define the *saturation* function such that, whenever $P = (T, Y, Z)$ is a pruned genealogy, $s(\widetilde{Y}_t, Y_t)$ is the integer vector of deme-specific numbers of emerging lineages at time t. Fig. 6 illustrates.

3.4.3. Compatibility. Suppose P is a pruned genealogy, with $\mathsf{t}(P) = T$ and $t \in [0, T]$. The local structure of P at t is, in general, compatible with only a subset of the possible jumps $\mathbb{U}$. For example, if there is a branch node or a sample node at t, then it is compatible only with birth-type or sample-type jumps, respectively. Similarly, if a lineage moves from deme i to deme i' at t, then u must be either of $i \to i'$ migration type or of a birth type with parent in i and $r_{i'}^u > 0$. To succinctly accommodate all possibilities, let us introduce the indicator function Q such that $Q = 1$ if the local genealogy structure—which is captured by the values of P^{Y} just before and at t—is compatible with an event of type u and $Q = 0$ otherwise. That is, $Q_u(y, y') = 1$ if and only if there is a feasible genealogy, G $= (T, Z, Y)$, and history, H, and a $t \in [0, T]$ such that, given $\mathbf{G}_T = $ G and $\mathbf{H}_T = $ H, we have $U_t = u$, $\widetilde{Y}_t = y$, and $Y_t = y'$. We refer to Q as the *compatibility indicator*.

3.4.4. Obscured genealogy. The *obscured genealogy* is obtained by discarding all information about demes and events not visible from the topology of the tree alone (Fig. 5B–C). In particular, if $P = (T, Z, Y)$ is a pruned genealogy, we write $\mathsf{obs}(P) = (T, Z)$ to denote the obscured genealogy.

3.5. Binomial ratio. The statement of the theorems to come is made easier with the following definition. For $n, r, \ell, s \in \mathbb{Z}_+^{\mathbb{D}}$, we define the *binomial ratio*

$$
\begin{pmatrix} n & \ell \\ r & s \end{pmatrix} := \begin{cases} \dfrac{\displaystyle\prod_{i \in \mathbb{D}} \binom{n_i - \ell_i}{r_i - s_i}}{\displaystyle\prod_{i \in \mathbb{D}} \binom{n_i}{r_i}}, & \text{if } \forall i \ n_i \geqslant \{\ell_i, r_i\} \geqslant s_i \geqslant 0, \\[6pt] 0, & \text{otherwise.} \end{cases}
$$

where $\binom{a}{b} = \frac{a!}{b!\,(a-b)!}$ is the binomial coefficient. Observe that $\begin{pmatrix} n & \ell \\ r & s \end{pmatrix} \in [0, 1]$. Moreover, in consequence of the Chu-Vandermonde identity, we have

$$
\sum_{s \in \mathbb{Z}_+^{\mathbb{D}}} \begin{pmatrix} n & \ell \\ r & s \end{pmatrix} \binom{\ell}{s} = 1,
$$

whenever $n_i \geqslant \{\ell_i, r_i\} \geqslant 0$ for all i.

4. Results

4.1. Likelihood for pruned genealogies. Our first result will be an expression for the likelihood of a given pruned genealogy given the history of the population process.

Theorem 1. *Suppose* $P = (T, Z, Y)$ *is a given pruned genealogy. Define*

$$
(7) \qquad \phi_u(x, y, y') := \begin{pmatrix} n(x) & \ell(y') \\ r^u & s(y, y') \end{pmatrix} Q_u(y, y'),
$$

where n is the deme occupancy (§2.6), r^u is the production (§3.3.1), ℓ and s are the lineage-count and saturation functions, respectively (§3.4.2), Q is the compatibility indicator (§3.4.3), and the binomial ratio is as defined in §3.5. Then

$$
\mathsf{Prob}\left[\mathbf{P}_T = P \mid \mathbf{H}_T = H\right] = \mathbb{1}\{\mathsf{ev}(H) \supseteq \mathsf{ev}(P)\} \prod_{t \in \mathsf{ev}(H)} \phi_{U_t}(X_t, \widetilde{Y}_t, Y_t).
$$

The proof is given in Appendix A.

Next, we show how the likelihood of a pruned genealogy, unconditional on the history, can be computed. For this, we use the filter equation technology developed in Appendix B. In particular, the following theorem follows immediately from Theorem 1 and Lemma B2.

Theorem 2. *Suppose that* $P = (T, Z, Y)$ *is a given pruned genealogy. Suppose that $w = w(t, x)$ is càdlàg in t and satisfies the initial condition $w(0, x) = p_0(x)$ and the filter equation*

$$
(8) \qquad \begin{aligned} \frac{\partial w}{\partial t}(t, x) &= \sum_u \int w(t, x')\, \alpha_u(t, x', x)\, \phi_u(x, \widetilde{Y}_t, Y_t)\, \mathrm{d}x' - \sum_u \int w(t, x)\, \alpha_u(t, x, x')\, \mathrm{d}x', \quad t \notin \mathsf{ev}(P), \\ w(t, x) &= \sum_u \int \widetilde{w}(t, x')\, \alpha_u(t, x', x)\, \phi_u(x, \widetilde{Y}_t, Y_t)\, \mathrm{d}x', \qquad\qquad\qquad\qquad\quad t \in \mathsf{ev}(P), \end{aligned}
$$

where ϕ is defined in Eq. 7. Then the likelihood of P is

$$
\mathcal{L} = \int w(T, x)\, \mathrm{d}x.
$$

Alternatively, one can express the likelihood of a pruned genealogy using the adjoint form of the filter equation (see §B.3).

Corollary 3. *If* $P = (T, Z, Y)$ *is a given pruned genealogy and* $F = F(s, x)$ *is càglàd in* s *and satisfies the final condition* $\underset{\sim}{F}(T, x) = 1$ *and the adjoint filter equation*

$$
\begin{aligned}
-\frac{\partial F}{\partial s}(s, x) &= \sum_u \int \widetilde{\alpha}_u(s, x, x') \left[\phi_u(x', \widetilde{Y}_s, Y_s) F(s, x') - F(s, x) \right] \mathrm{d}x', \quad s \notin \mathsf{ev}(P), \\
F(s, x) &= \sum_u \int \alpha_u(s, x, x') \, \phi_u(x', \widetilde{Y}_s, Y_s) \, \underset{\sim}{F}(s, x') \mathrm{d}x', \qquad\qquad s \in \mathsf{ev}(P),
\end{aligned}
$$

(9)

where $0 \leqslant s \leqslant T$ *and* ϕ *is defined in Eq. 7, then the likelihood of* P *is*

$$
\mathcal{L} = \int F(0, x) \, p_0(x) \, \mathrm{d}x.
$$

Remark 1. Note that the likelihood given by the foregoing expressions is the likelihood of a given genealogy *in which the samples are ordered* and that, while this order should be compatible with time-ordering, it may not be completely determined by it. In particular, if n samples are taken at one time, there are $n!$ possible orderings of the samples, each of which has an identical likelihood. In consequence, the likelihood given above will differ by a factor of $n!$ from the likelihood computed ignoring ordering of samples. From the standpoint of parameter estimation, this is immaterial, since factors that depend only on the data drop out of the score function. This consideration is relevant, however, in the context of model comparison.

4.2. Likelihood for obscured genealogies.
Our next result concerns the likelihood of a given obscured genealogy conditional on the history.

Theorem 4. *Suppose that* (T, Z) *is a given obscured genealogy. Let* q *and* π *be probability kernels, such that for all* $x \in \mathbb{X}$ *and* $y \in Y_0(Z)$,

$$
q(x, y) \geqslant 0, \qquad \sum_{y \in Y_0(Z)} q(x, y) = 1,
$$

and, for all $u \in \mathbb{U}$, $t \in \mathbb{R}_+$, $x, x' \in \mathbb{X}$, $y, y' \in Y_t(Z)$,

$$
\pi_u(t, x, x', y, y') \geqslant 0, \qquad \sum_{y' \in Y_t(Z)} \pi_u(t, x, x', y, y') = 1.
$$

Suppose moreover that $\pi_u(t, x, x', y, y') > 0$ *whenever* $\alpha_u(t, x, x') \, Q_u(y, y') > 0$ *and that* $q(x, y) > 0$ *whenever* $\mathsf{Prob}\left[\mathbf{P}_0^Y = y \mid \mathbf{X}_0 = x \right] > 0$. *Then there is a stochastic jump process* $\mathbf{y}_t$ *with sample paths in* $Y(Z)$ *such that* $(\mathbf{X}_t, \mathbf{U}_t, \mathbf{y}_t)$ *is Markov and*

$$
\mathsf{Prob}\left[\mathbf{P}_T^Z = Z \mid \mathbf{H}_T = H \right] = \mathbb{1}\{\mathsf{ev}(H) \supseteq \mathsf{ev}(Z)\} \, \mathbb{E}\left[\frac{1}{q(X_0, \mathbf{y}_0)} \prod_{t \in \mathsf{ev}(H)} \frac{\phi_{U_t}(X_t, \widetilde{\mathbf{y}}_t, \mathbf{y}_t)}{\pi_{U_t}(t, \widetilde{X}_t, X_t, \widetilde{\mathbf{y}}_t, \mathbf{y}_t)} \right],
$$

where ϕ *is defined in Eq. 7 and the expectation is taken over the sample paths of* $\mathbf{y}_t$.

Proof. First, observe that, since obs is a deterministic operator,

(10)
$$
\mathsf{Prob}\left[\mathbf{P}_T^Z = Z \mid \mathbf{H}_T = H \right] = \mathbb{E}\left[\mathbb{1}\{\mathbf{P}_T^Z = Z\} \mid \mathbf{H}_T = H \right].
$$

Our strategy will be to evaluate Eq. 10 using a change of measure: we will propose pruned genealogies compatible with Z as sample paths from a stochastic process driven by $\mathbf{X}_t$ and evaluate the the expectation in Eq. 10 by a suitably weighted expectation over these paths. Conditional on $\mathbf{H}_T = H$, the initial distribution q and probability kernel π generate a Markov chain, $\hat{\mathbf{y}}_k$ such that

$$
\mathsf{Prob}\left[\hat{\mathbf{y}}_0 \mid \mathbf{H}_T = H \right] = q(X_0, \hat{\mathbf{y}}_0), \qquad \mathsf{Prob}\left[\hat{\mathbf{y}}_k \mid \hat{\mathbf{y}}_{k-1}, \mathbf{H}_T = H \right] = \pi_{\hat{U}_k}(\hat{T}_k, \hat{X}_{k-1}, \hat{X}_k, \hat{\mathbf{y}}_{k-1}, \hat{\mathbf{y}}_k).
$$

The required process $\mathbf{y}_t$ is the unique càdlàg process with event times $\hat{T}_k$ and $\hat{\mathbf{y}}_k$ as its embedded chain. This construction of $\mathbf{y}_t$ obviously guarantees that $\mathsf{ev}(H) \supseteq \mathsf{ev}(\mathbf{y}) \supseteq \mathsf{ev}(Z)$ and that $(\mathbf{X}_t, \mathbf{U}_t, \mathbf{y}_t)$ is Markov.

Now, for $y \in Y(Z)$, let us define $C(y) = (T, Z, y)$. Then, by construction, $\mathsf{obs}(C(y)) = (T, Z)$ and, conversely, for every pruned genealogy P satisfying $\mathsf{t}(P) = T$ and $P^Z = Z$, $C(P^Y) = P$. Moreover, the conditions on the

kernels q and π guarantee that, if $\mathsf{Prob}\left[\mathbf{P}_{\mathrm{T}} = \mathrm{P} \mid \mathbf{H}_{\mathrm{T}} = \mathrm{H}\right] > 0$ and $\mathrm{P}^{\mathsf{Z}} = \mathrm{Z}$, then $\mathsf{Prob}\left[\mathbf{y} = \mathrm{P}^{\mathsf{Y}} \mid \mathbf{H}_{\mathrm{T}} = \mathrm{H}\right] > 0$. We therefore have that

$$\mathsf{Prob}\left[\mathbf{P}_{\mathrm{T}}^{\mathsf{Z}} = \mathrm{Z} \;\middle|\; \mathbf{H}_{\mathrm{T}} = \mathrm{H}\right] = \mathbb{E}\left[\frac{\mathsf{Prob}\left[\mathbf{P}_{\mathrm{T}} = C(\mathbf{y}) \mid \mathbf{H}_{\mathrm{T}} = \mathrm{H}\right]}{\pi(\mathbf{y}|\mathrm{H})}\right],$$

the expectation being taken with respect to the random process $\mathbf{y}$. Here, by definition,

$$\pi(\mathbf{y}|\mathrm{H}) = q(\mathrm{X}_0, \mathbf{y}_0) \prod_{t \in \mathsf{ev}(\mathrm{H})} \pi_{\mathrm{U}_t}(t, \widetilde{\mathrm{X}}_t, \mathrm{X}_t, \widetilde{\mathbf{y}}_t, \mathbf{y}_t).$$

The result then follows from Theorem 1. $\hfill\square$

Note that, since $\mathsf{Y}_t(\mathrm{Z})$ is finite, it is permissible, for example, to choose q and π to be uniform.

The final result shows how to compute the likelihood of an obscured genealogy. It is an immediate consequence of Theorem 4 and Lemma B2.

Theorem 5. *Let $V = (\mathrm{T}, \mathrm{Z})$ be a given obscured genealogy. Then there are probability kernels q and π as in Theorem 4 such that if*

$$\beta_u(t, x, x', y, y') = \alpha_u(t, x, x')\, \pi_u(t, x, x', y, y'), \qquad \Psi_u(t, x, x', y, y') = \frac{\phi_u(x', y, y')}{\pi_u(t, x, x', y, y')},$$

and if $w = w(t, x, y)$ satisfies the initial condition $w(0, x, y) = p_0(x)\, \mathbb{1}\{q(x, y) > 0\}$ and the filter equation

$$\frac{\partial w}{\partial t} = \sum_{uy'} \int w(t, x', y')\, \beta_u(t, x', x, y', y)\, \Psi_u(t, x', x, y', y)\, \mathrm{d}x' - \sum_{uy'} \int w(t, x, y)\, \beta_u(t, x, x', y, y')\, \mathrm{d}x', \quad t \notin \mathsf{ev}(\mathrm{Z}),$$

$$w(t, x, y) = \sum_{uy'} \int \widetilde{w}(t, x', y')\, \beta_u(t, x', x, y', y)\, \Psi_u(t, x', x, y', y)\, \mathrm{d}x', \qquad\qquad t \in \mathsf{ev}(\mathrm{Z}),$$

then the likelihood of V is

$$\mathcal{L} = \sum_{y \in \mathsf{Y}_T(\mathrm{Z})} \int w(\mathrm{T}, x, y)\, \mathrm{d}x.$$

Lemma B3 shows how this can be computed via sequential Monte Carlo. As before, there is also an adjoint form of Theorem 5, to wit:

Corollary 6. *Let $V = (\mathrm{T}, \mathrm{Z})$ be a given obscured genealogy and let β and Ψ be defined as in Theorem 5. Suppose F is càglàd and satisfies the final condition $\underset{\sim}{F}(T, x, y) = 1$ for all $x \in \mathbb{X}$, $y \in \mathsf{Y}_T(\mathrm{Z})$ and also the adjoint filter equation*

$$-\frac{\partial F}{\partial s} = \sum_{uy'} \int \widetilde{\beta}_u(s, x, x', y, y') \left[\widetilde{\Psi}_u(s, x, x', y, y')\, F(s, x', y') - F(s, x, y)\right] \mathrm{d}x', \quad s \notin \mathsf{ev}(\mathrm{Z}),$$

$$F(s, x, y) = \sum_{uy'} \int \beta_u(s, x, x', y, y')\, \Psi_u(s, x, x', y, y')\, \underset{\sim}{F}(s, x', y')\, \mathrm{d}x', \qquad\qquad s \in \mathsf{ev}(\mathrm{Z}),$$

for $0 \leqslant s \leqslant T$. Then the likelihood of V is

$$\mathcal{L} = \sum_{y \in \mathsf{Y}_0(\mathrm{Z})} \int F(0, x, y)\, p_0(x)\, \mathrm{d}x.$$

4.3. Special cases.

4.3.1. Moran model and the Kingman coalescent. In the Moran model (Moran, 1958; King et al., 2020), events occur according to a rate-μ Poisson process. At each event, a compound birth-death jump occurs so that the population size, n, remains constant (cf. Fig. 4F). If we let $\mathbf{X}_t$ be the number of events that have occurred by time t, then $\mathbf{X}_t$ is a simple counting process, which we can use to define the state of the population process. Its KFE is then

$$\frac{\partial v}{\partial t} = \mu\, v(t, x-1) - \mu\, v(t,x), \qquad v(0,x) = \begin{cases} 1, & x = 0, \\ 0, & x > 0. \end{cases}$$

Since there is only a single deme, and since nothing depends on the state, in writing the corresponding filter equation (Theorem 5), we can take w to be independent of both x and y.

In the classical case (Kingman, 1982a), m samples are taken simultaneously at a single time, T. Then, if B is the set of branch-times and $\ell(t)$ is the number of lineages in the genealogy at time t, upon summing over x and y, the filter equation reads

$$(11) \qquad w(0) = 1, \qquad \frac{\partial w}{\partial t} = \mu\, w(t) \left(1 - \frac{\binom{\ell(t)}{2}}{\binom{n}{2}} \right) - \mu\, w(t), \quad t \notin B, \qquad w(t) = \frac{\mu}{\binom{n}{2}}\, \widetilde{w}(t), \quad t \in B.$$

Integrating Eqs. 11 and taking logarithms yields

$$(12) \qquad \log \mathcal{L} = |B| \log \frac{\mu}{\binom{n}{2}} - \frac{\mu}{\binom{n}{2}} \sum_{i=1}^{m} \binom{i}{2} s_i,$$

where the $s_i := \int \mathbb{1}\{\ell(t) = i\}\, \mathrm{d}t$ are the durations of the *coalescent intervals*, i.e., intervals between successive branch-points. We recognize Eq. 12 as the expression for the Kingman (1982a) coalescent likelihood (e.g., Wakeley, 2009).

More generally, if in addition samples are taken according to a rate-ψ Poisson process such that the set of sample-times in the genealogy is $S = S_0 \cup S_1$, where S_0, S_1 are the sets of times of terminal and inline samples, respectively, then the filter equation reads

$$(13) \qquad w(0) = 1, \qquad \frac{\partial w}{\partial t} = -\mu\, \frac{\binom{\ell(t)}{2}}{\binom{n}{2}}\, w(t), \quad t \notin S \cup B, \qquad \frac{w(t)}{\widetilde{w}(t)} = \begin{cases} \frac{\mu}{\binom{n}{2}}, & t \in B, \\ \psi \left(1 - \frac{\ell(t)}{n} \right), & t \in S_0, \\ \frac{\psi}{n}, & t \in S_1. \end{cases}$$

Integrating Eqs. 13 yields

$$(14) \qquad \log \mathcal{L} - |S| \log \psi = \sum_{t \in S_0} \log \left(1 - \frac{\ell(t)}{n} \right) - |S_1| \log n + |B| \log \frac{\mu}{\binom{n}{2}} - \frac{\mu}{\binom{n}{2}} \sum_{i=1}^{\infty} \binom{i}{2} s_i,$$

in agreement with the result of the very different derivation of King et al. (2020).

4.3.2. Linear birth-death model. In this model, the state variable, $\mathbf{X}_t$, is the size of a population at time t. All individuals face the same per-capita birth and death rates, which are λ and μ, respectively. Accordingly, the Kolmogorov backward equation for this model is

$$-\frac{\partial F}{\partial s} = \lambda\, x\, [F(s, x+1) - F(s,x)] + \mu\, x\, [F(s, x-1) - F(s,x)].$$

Stadler (2010) considered the case where samples are taken through time at a uniform per-capita rate ψ and lineages extant at time T are sampled with probability ρ. Let B denote the set of branch-times in an observed genealogy, let S_0 and S_1 be, respectively, the sets of terminal and inline sample times prior to time T, and suppose n lineages are sampled at time T. Since there is only a single deme, and since nothing depends on the genealogy coloring, in writing the corresponding filter equation, we can apply Corollary 6 and take F to be independent of y. After summing over y, the regular part of the adjoint filter equation then reads

$$(15) \qquad -\frac{\partial F}{\partial s}(s,x) = \lambda\, x\, \left(1 - \frac{\binom{\ell(s)}{2}}{\binom{x+1}{2}} \right) F(s, x+1) + \mu\, x\, F(s, x-1) - (\lambda + \mu + \psi)\, x\, F(s,x),$$

which holds for $x \geqslant \ell(s)$ and $s \notin B \cup S_0 \cup S_1 \cup \{T\}$. The singular part is

(16)
$$F(s,x) = \begin{cases} \frac{2\lambda}{x+1}\, \undertilde{F}(s, x+1), & s \in B, \\[2mm] \psi\,(x - \ell(s))\, \undertilde{F}(s,x), & s \in S_0, \\[2mm] \psi\, \undertilde{F}(s,x), & s \in S_1, \\[2mm] \binom{x}{n}(1-\rho)^{x-n}\,\rho^n, & s = T. \end{cases}$$

We note that $F(s,x) = 0$ when $x < \ell(s)$. Let us make the ansatz

(17)
$$F(s,x) = C(s)\,G(s)^{x-\widetilde{\ell}(s)}\,H(s)^{\widetilde{\ell}(s)}\,K(x, \widetilde{\ell}(s)),$$

where G, H are smooth functions and C is càglàd and piecewise constant, with discontinuities only at genealogical events, so that

$$\undertilde{F}(s,x) = \undertilde{C}(s)\,G(s)^{x-\ell(s)}\,H(s)^{\ell(s)}\,K(x, \ell(s)).$$

Putting these into the first three cases of Eq. 16 yields conditions on C and K that are satisfied when

$$K(x, \ell) = \frac{x!}{(x-\ell)!} \qquad \text{and} \qquad \frac{C(s)}{\undertilde{C}(s)} = \begin{cases} 2\lambda\,H(s), & s \in B, \\[2mm] \psi\,G(s)\,H(s)^{-1}, & s \in S_0, \\[2mm] \psi, & s \in S_1. \end{cases}$$

To satisfy the last case of Eq. 16, we note that $\widetilde{\ell}(T) = n$ and therefore impose the conditions

(18)
$$G(T) = 1 - \rho, \qquad H(T) = 1, \qquad C(T) = \frac{\rho^n}{n!}.$$

Finally, upon substituting the ansatz into Eq. 15, we obtain for $s \notin B \cup S_0 \cup S_1 \cup \{T\}$,

$$-\frac{1}{F}\frac{\partial F}{\partial s} = -(x-\ell)\frac{G'}{G} - \ell\frac{H'}{H}$$
$$= \lambda x\left(1 - \frac{\binom{\ell}{2}}{\binom{x+1}{2}}\right)\frac{x+1}{x+1-\ell}G + \mu x\frac{x-\ell}{x}G^{-1} - (\lambda + \mu + \psi)\,x$$
$$= \lambda\,(x+\ell)\,G + \mu\,(x-\ell)\,G^{-1} - (\lambda + \mu + \psi)\,x.$$

If this is to hold for all x and ℓ, we must have

(19)
$$-\frac{G'}{G} = \lambda\,G + \mu\,G^{-1} - (\lambda + \mu + \psi), \qquad -\frac{H'}{H} = 2\,\lambda\,G - (\lambda + \mu + \psi).$$

In passing, we note that Eqs. 19 are central to the original derivation of Stadler (2010). Eqs. 18 and 19 form a final-value problem that can be solved by elementary means to obtain closed-form expressions for $G(s)$ and $H(s)$. Corollary 6 then states that $\mathcal{L} = \int F(0,x)\,p_0(x)\,\mathrm{d}x$. In particular, conditional on $\mathbf{X}_0 = x$, we have

(20)
$$\mathcal{L} = \frac{x!}{(x-\ell(0))!}\,\frac{\rho^n}{n!}\,G(0)^{x-\ell(0)}\,H(0)^{\ell(0)}\,\psi^{|S_1|}\prod_{e \in B}(2\,\lambda\,H(e))\prod_{e \in S_0}\left(\psi\,G(e)H(e)^{-1}\right).$$

Stadler (2010, Thm. 3.5) derives an expression for this quantity in the special case $x = \ell(0) = 1$, which is equivalent up to the factor of $n!$ explained in Remark 1.

The foregoing represents an independent verification of the correctness of the theory in the case of the simple linear birth-death process with constant sampling through time and one bulk-sampling event. It should be clear, however, that even within the context of the unstructured, linear, birth-death process, the theory presented here readily accommodates complexities such as time-varying sampling rates, multiple bulk-sampling events at specified or random times, and time-varying birth and death rates.

5. Discussion

The theory presented here represents a generalization of the existing coalescent and birth-death-process approaches to phylodynamic inference. Importantly, because it allows computation of the likelihood via strictly forward-in-time computations, it permits consideration of models for which time-reversal arguments are not available. Moreover, inasmuch as the formulae of Theorem 5 can be efficiently computed via sequential Monte Carlo, explicit expressions for transition probabilities are not needed: it is sufficient to be able to simulate from the population process. This feature of the algorithms—known as the *plug-and-play property* (He et al., 2010)—further expands the class of population models that can be confronted with data.

In particular, the theory gives us the freedom to choose models with many demes. For deterministic population models, Volz (2012) and Rasmussen et al. (2014) showed how one could accommodate discrete population structure. Their procedures involve solving a large number of differential equations backward in time, relying on the time-reversibility of deterministic dynamics. In general, this time-reversibility is not available for stochastic processes.

Some existing methods put rather severe limits on the form of the sampling model and, as Volz & Frost (2014) pointed out, misspecification of the sampling model can lead to large inferential biases. With the theory presented here, essentially arbitrary specification of the sampling model is possible. In particular, one can posit sampling at a rate which is an arbitrary function of time and state and include discrete sampling events as well. It is also possible to condition on the existence of samples.

If sequential Monte Carlo algorithms are used to compute the likelihoods of Theorem 5, then it is straightforward to simultaneously assimilate information from both time-series and genealogical data. One can therefore supplement traditional incidence, disease, or mortality time series with genealogical data to improve inference.

A limitation of the theory is that the population models are assumed to be pure jump processes, which allows consideration of demographic stochasticity and environmental stochasticity modeled by jumps involving multiple individuals (Bretó & Ionides, 2011), but disallows stochastic processes with a diffusive component. However, we anticipate that it will be possible to incorporate the full range of Markovian environmental stochasticity via extension of this theory to population models containing both diffusion and jump components.

Similarly, the theory presented here assumes that at most one birth event can occur at a time and that at most one deme can migrate or be sampled at a time (though multiple migrators or samples are permitted). These restrictions are not essential, and the proof of Theorem 1 can be adapted to accommodate relaxations of these assumptions. This will be developed in a sequel.

The price of the theory's flexibility is primarily computational. When sequential Monte Carlo is used to evaluate the likelihood in Theorem 5, the computational effort scales linearly with the number of samples. In its most straightforward implementation—using an event-driven algorithm (e.g., Gillespie, 1977)—it scales nonlinearly with population size in general. However, stochastic simulation schemes are available that scale independently of population size (Higham, 2008). On the other hand, the importance sampling underlying Theorem 5 will in general require effort that is exponential in the number of demes. For models with many demes, therefore, approaches for ameliorating or circumventing this curse of dimensionality may be necessary. Critically, the substantial freedom one has in the choice of the importance-sampling kernel π can be exploited for this purpose. In particular, since it is permissible to "borrow information" from the future by means of the importance sampling, there is hope for highly efficient algorithmic computation.

Acknowledgments

References

Alizon, S. (2024) Phylodynamics. In G. Didier & S. Guindon (eds.), *Models and Methods for Biological Evolution*, pp. 259–282. Hoboken, New Jersey: Wiley.
DOI: 10.1002/9781394284252.ch11

Arulampalam, M. S., Maskell, S., Gordon, N., & Clapp, T. (2002) A tutorial on particle filters for online nonlinear, non-Gaussian Bayesian tracking. *IEEE Trans Signal Process* **50**:174–188.
DOI: 10.1109/78.978374

Barido-Sottani, J., Vaughan, T. G., & Stadler, T. (2020) A multitype birth-death model for Bayesian inference of lineage-specific birth and death rates. *Systematic Biology* **69**:973–986.
DOI: 10.1093/sysbio/syaa016

Boskova, V., Bonhoeffer, S., & Stadler, T. (2014) Inference of epidemiological dynamics based on simulated phylogenies using birth-death and coalescent models. *PLoS Computational Biology* **10**:e1003913.
DOI: 10.1371/journal.pcbi.1003913

Bretó, C. & Ionides, E. L. (2011) Compound Markov counting processes and their applications to modeling infinitesimally over-dispersed systems. *Stochastic Processes and their Applications* **121**:2571–2591.
DOI: 10.1016/j.spa.2011.07.005

Doucet, A., de Freitas, N., & Gordon, N. (eds.) (2001) *Sequential Monte Carlo Methods in Practice*. New York: Springer.
DOI: 10.1007/978-1-4757-3437-9

Drummond, A. J., Rambaut, A., Shapiro, B., & Pybus, O. G. (2005) Bayesian coalescent inference of past population dynamics from molecular sequences. *Molecular Biology and Evolution* **22**:1185–1192.
DOI: 10.1093/molbev/msi103

Felsenstein, J. (2004) *Inferring Phylogenies*. Sunderland, Mass.: Sinauer.

Giesecke, K. & Schwenkler, G. (2018) Filtered likelihood for point processes. *Journal of Econometrics* **204**:33–53.
DOI: 10.1016/j.jeconom.2017.11.011

Gillespie, D. T. (1977) Exact stochastic simulation of coupled chemical reactions. *Journal of Physical Chemistry* **81**:2340–2361.
DOI: 10.1021/j100540a008

Grenfell, B. T., Pybus, O. G., Gog, J. R., Wood, J. L. N., Daly, J. M., Mumford, J. A., & Holmes, E. C. (2004) Unifying the epidemiological and evolutionary dynamics of pathogens. *Science* **303**:327–332.
DOI: 10.1126/science.1090727

Griffiths, R. C. & Tavaré, S. (1994) Sampling theory for neutral alleles in a varying environment. *Philosophical Transactions of the Royal Society of London, Series B* **344**:403–410.
DOI: 10.1098/rstb.1994.0079

He, D., Ionides, E. L., & King, A. A. (2010) Plug-and-play inference for disease dynamics: measles in large and small populations as a case study. *Journal of the Royal Society, Interface* **7**:271–283.
DOI: 10.1098/rsif.2009.0151

Higham, D. J. (2008) Modeling and simulating chemical reactions. *SIAM Review* **50**:347–368.
DOI: 10.1137/060666457

Ionides, E. L., Breto, C., Park, J., Smith, R. A., & King, A. A. (2017) Monte carlo profile confidence intervals for dynamic systems. *Journal of the Royal Society, Interface* **14**:20170126.
DOI: 10.1098/rsif.2017.0126

King, A. A., Lin, Q., & Ionides, E. L. (2020) The sampled Moran genealogy process. *arXiv* 2002.11184.
DOI: 10.48550/ARXIV.2002.11184

King, A. A., Lin, Q., & Ionides, E. L. (2022) Markov genealogy processes. *Theoretical Population Biology* **143**:77–91.
DOI: 10.1016/j.tpb.2021.11.003

Kingman, J. F. C. (1982a) The coalescent. *Stochastic Processes and their Applications* **13**:235–248.
DOI: 10.1016/0304-4149(82)90011-4

Kingman, J. F. C. (1982b) On the genealogy of large populations. *Journal of Applied Probability* **19**:27–43.
DOI: 10.2307/3213548

Kliemann, W. H., Koch, G., & Marchetti, F. (1990) On the unnormalized solution of the filtering problem with counting process observations. *IEEE Transactions on Information Theory* **36**:1415–1425.
    DOI: 10.1109/18.59936

Leventhal, G. E., Günthard, H. F., Bonhoeffer, S., & Stadler, T. (2014) Using an epidemiological model for phylogenetic inference reveals density dependence in HIV transmission. *Molecular Biology and Evolution* **31**:6–17.
    DOI: 10.1093/molbev/mst172

MacPherson, A., Louca, S., McLaughlin, A., Joy, J. B., & Pennell, M. W. (2021) Unifying phylogenetic birth-death models in epidemiology and macroevolution. *Systematic Biology* **71**:172–189.
    DOI: 10.1093/sysbio/syab049

Möhle, M. (2000) Ancestral processes in population genetics—the coalescent. *Journal of Theoretical Biology* **204**:629–638.
    DOI: 10.1006/jtbi.2000.2032

Möller, S., du Plessis, L., & Stadler, T. (2018) Impact of the tree prior on estimating clock rates during epidemic outbreaks. *Proceedings of the National Academy of Sciences* **115**:4200–4205.
    DOI: 10.1073/pnas.1713314115

Moran, P. A. P. (1958) Random processes in genetics. *Mathematical Proceedings of the Cambridge Philosophical Society* **54**:60–71.
    DOI: 10.1017/s0305004100033193

Notohara, M. (1990) The coalescent and the genealogical process in geographically structured population. *Journal of Mathematical Biology* **29**.
    DOI: 10.1007/bf00173909

Ogata, Y. (1978) The asymptotic behaviour of maximum likelihood estimators for stationary point processes. *Annals of the Institute of Statistical Mathematics* **30**:243–261.
    DOI: 10.1007/bf02480216

Puri, M. L. & Tuan, P. D. (1986) Maximum likelihood estimation for stationary point processes. *Proceedings of the National Academy of Sciences* **83**:541–545.
    DOI: 10.1073/pnas.83.3.541

Rasmussen, D. A., Ratmann, O., & Koelle, K. (2011) Inference for nonlinear epidemiological models using genealogies and time series. *PLoS Computational Biology* **7**:e1002136.
    DOI: 10.1371/journal.pcbi.1002136

Rasmussen, D. A., Volz, E. M., & Koelle, K. (2014) Phylodynamic inference for structured epidemiological models. *PLoS Computational Biology* **10**:e1003570.
    DOI: 10.1371/journal.pcbi.1003570

Seidel, S., Stadler, T., & Vaughan, T. G. (2024) Estimating pathogen spread using structured coalescent and birth-death models: A quantitative comparison. *Epidemics* **49**:100795.
    DOI: 10.1016/j.epidem.2024.100795

Stadler, T. (2010) Sampling-through-time in birth-death trees. *Journal of Theoretical Biology* **267**:396–404.
    DOI: 10.1016/j.jtbi.2010.09.010

Takahata, N. (1988) The coalescent in two partially isolated diffusion populations. *Genetical Research* **52**:213–222.
    DOI: 10.1017/s0016672300027683

Vaughan, T. G. & Stadler, T. (2025) Bayesian phylodynamic inference of multitype population trajectories using genomic data. *Molecular Biology and Evolution* **42**:msaf130.
    DOI: 10.1093/molbev/msaf130

Volz, E. M. (2012) Complex population dynamics and the coalescent under neutrality. *Genetics* **190**:187–201.
    DOI: 10.1534/genetics.111.134627

Volz, E. M. & Frost, S. D. W. (2014) Sampling through time and phylodynamic inference with coalescent and birth-death models. *Journal of the Royal Society, Interface* **11**:20140945.
    DOI: 10.1098/rsif.2014.0945

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.

DOI: 10.1534/genetics.109.106021

Volz, E. M. & Siveroni, I. (2018) Bayesian phylodynamic inference with complex models. *PLoS Computational Biology* **14**:e1006546.

DOI: 10.1371/journal.pcbi.1006546

Wakeley, J. (2009) *Coalescent Theory: An Introduction*. New York: W. H. Freeman.

Wills, A. G. & Schön, T. B. (2023) Sequential Monte Carlo: a unified review. *Annual Review of Control, Robotics, and Autonomous Systems* **6**.

DOI: 10.1146/annurev-control-042920-015119

Appendix A. Proof of Theorem 1.

We are given a pruned genealogy $P = (T, Z, Y)$ and a history H and wish to compute $\text{Prob}\left[\mathbf{P}_T = P \mid \mathbf{H}_T = H\right]$. First, note that if $\text{ev}(H) \not\supseteq \text{ev}(P)$, then H and P are incompatible and $\text{Prob}\left[\mathbf{P}_T = P \mid \mathbf{H}_T = H\right] = 0$. Similarly, if any event of H is incompatible with the local structure of P in the sense of §3.4.3, then $\text{Prob}\left[\mathbf{P}_T = P \mid \mathbf{H}_T = H\right] = 0$. In either case, the conclusion of the theorem follows. Let us therefore suppose that neither of these conditions hold.

We can always construct a time-respecting, but otherwise arbitrary ordering of the sample nodes. Let P_j represent the genealogy spanned by the first j samples in this ordering. Since each P_j contains $P_{j'}$ for all $j' < j$, P_j is a Markov chain. Therefore, we have the factorization

$$\text{Prob}\left[P \mid H\right] = \prod_j \text{Prob}\left[P_j \mid P_{j-1}, H\right].$$

Each factor, $\text{Prob}\left[P_j \mid P_{j-1}, H\right]$, can itself be factorized. In particular, suppose H and P_{j-1} are given and suppose that it is the K-th event in H that introduces the j-th sample. The j-th lineage extends backward in time until it either coalesces with P_{j-1} or reaches $t = 0$ (Fig. A1). Moreover, since P is compatible with H, nothing can happen to this lineage between events of H. Therefore, we proceed backward, event by event, as follows. Since the type of event K is given, the color of the j-th lineage just prior to that event is known. At each subsequent event, the lineage either emerges from the event or does not. Moreover, if the lineage does emerge from the event, the color prior to the event is determined and the color remains unchanged if it does not. Suppose that lineage j is in deme i just to the right of event $k < K$ and that the deme-i occupancy and production are n_i, r_i, respectively. Let ℓ_{ij}, s_{ij} be the lineage count and saturation at this event in P_{j-1}. Then the new lineage can be any one of the $n_i - \ell_i$ lineages as yet unaccounted for. Of these, $r_i - s_{ij}$ emerge from the event. Therefore, the probability that lineage j emerges from the event is $(r_i - s_{ij})/(n_i - \ell_{ij})$ and the probability that it does not is $(n_i - r_i - \ell_{ij} + s_{ij})/(n_i - \ell_{ij})$. Let q_{jk} be the former if lineage j does in fact emerge from the event and the latter if it does not. Then $\text{Prob}\left[P_j \mid P_{j-1}, H\right] = \prod_k q_{jk}$ and $\text{Prob}\left[P \mid H\right] = \prod_{jk} q_{jk}$.
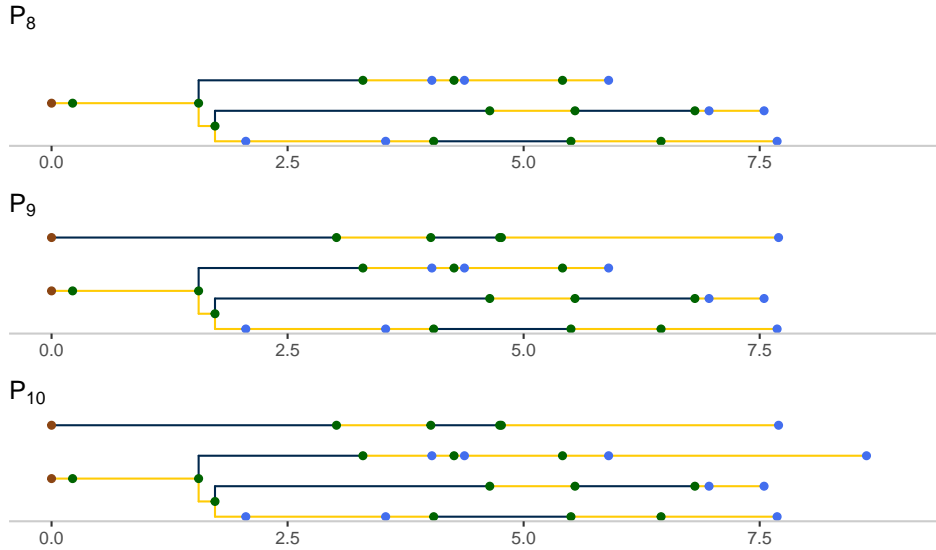
FIGURE A1. **Embedded chain of the pruned genealogy process.** Three successive states of the embedded chain of the pruned genealogy process are shown. At each step, a new lineage is added in a sampling event. This lineage extends backward until it either coalesces with the pre-existing genealogy or reaches $t = 0$. Theorem 1 is derived by computing the probability that it coalesces at each possible point, conditional on the history of the population process.

Now fix k and consider the cumulative product $\prod_{m=1}^{j} q_{mk}$. Clearly $\ell_{1,k} = s_{1,k} = 0$ and, if lineage j has not already coalesced to the right of k, then $\ell_{j+1,k} = \ell_{j,k} + 1$. Moreover, if j emerges from k, then $s_{j+1,k} = s_{j,k} + 1$ and $s_{j+1,k} = s_{j,k}$ if it does not. Therefore, as we accumulate each successive lineage j, we move one step to the right in the following diagram.
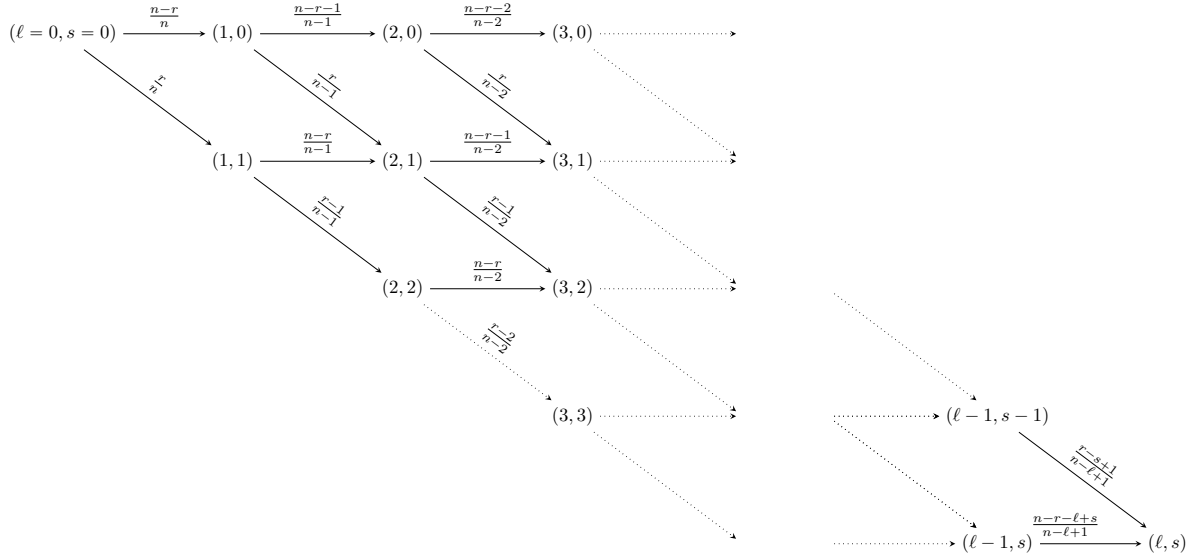
At each step, the lineage count increases by 1 and the saturation either increases by a unit or stays the same. The probability of each option is indicated on the corresponding arrow. Although there are multiple paths by which the process may arrive at the final lineage count and saturation, the product of the probabilities along each path is the same:

$$\frac{\binom{n_i - \ell_i}{r_i - s_i}}{\binom{n_i}{r_i}}.$$

Moreover, since each of the options is independent of the others,

$$\prod_j q_{jk} = \begin{pmatrix} n & \ell \\ r & s \end{pmatrix},$$

the latter being the binomial ratio defined in §3.5.

Alternatively, one can reason as follows. Conditional on $\mathbf{H}_T = H$, at each time $t \in \mathrm{ev}(H)$, a jump of mark U_t occurred, with a production of $r^{\mathrm{U}_t} = (r_i)_{i \in \mathbb{D}}$, resulting in a deme-occupancy of $n(\mathrm{X}_t) = (n_i)_{i \in \mathbb{D}}$. In P, at time t, there are $\ell_i = \ell_i(\mathrm{Y}_t)$ lineages in deme i, of which $s_i = s_i(\widetilde{\mathrm{Y}}_t, \mathrm{Y}_t)$ are emergent. By assumption, at each genealogical event, lineages within a deme are exchangeable: each has an identical probability of being involved. This exchangeability implies that each lineage present in a deme at time t was equally likely to have been one of the emergent lineages. In particular, at time t, the probability that s_i of the ℓ_i deme-i lineages were among the r_i of n_i lineages emergent in the unpruned genealogy process is the same as the probability that, upon drawing ℓ_i balls without replacement from an urn containing r_i red balls and $n_i - r_i$ black balls, exactly s_i of the drawn balls are red, namely

$$\frac{\binom{n_i - \ell_i}{r_i - s_i}\binom{\ell_i}{s_i}}{\binom{n_i}{r_i}}.$$

Because the lineages are labeled, each of the $\binom{\ell_i}{s_i}$ equally probable sets of s_i lineages is distinct; just one of these is the one present in P. Moreover, since, again conditional on $\mathbf{H}_T = H$, the identities of the lineages involved in a genealogical event are random and independent of the identities selected at all other events, we have established

that

$$\mathsf{Prob}\left[\mathbf{P}_\mathrm{T} = \mathrm{P} \mid \mathbf{H}_\mathrm{T} = \mathrm{H}\right] = \prod_{t \in \mathsf{ev}(\mathrm{H})} \begin{pmatrix} n(\mathrm{X}_t) & \ell(\mathrm{Y}_t) \\ r^{\mathrm{U}_t} & s(\widetilde{\mathrm{Y}}_t, \mathrm{Y}_t) \end{pmatrix}.$$

Returning to the possibility that H is incompatible with P, since $\mathsf{Prob}\left[\mathbf{P}_\mathrm{T} = \mathrm{P}\right] = 0$ if either any $Q_{\mathrm{U}_t} = 0$ or $\mathsf{ev}(\mathrm{P}) \nsubseteq \mathsf{ev}(\mathrm{H})$, we obtain the result. $\qquad\square$

Appendix B. Filter equations.

B.1. Definition. The likelihoods that appear in Theorems 2 and 5 are integrals over large sets of histories. As such, explicit expressions for them are not available, and we require mathematical tools to allow us to manipulate these quantities and devise algorithms for their numerical solution. The *filter equations* we introduce here are suitable for these purposes, and we devote this appendix to exposing their essential properties. This extremely convenient formalism has, to our knowledge, not been thoroughly exploited in the context we introduce here, though we note their resemblance to the constructions of Ogata (1978), Puri & Tuan (1986), Kliemann et al. (1990), and Giesecke & Schwenkler (2018).

Definition. Let $\mathbf{X}_t$ be a continuous-time Markov process with KFE

$$(\mathrm{B1}) \qquad \frac{\partial u}{\partial t}(t, x) = \int u(t, x')\,\beta(t, x', x)\,\mathrm{d}x' - \int u(t, x)\,\beta(t, x, x')\,\mathrm{d}x'.$$

Suppose that $B : \mathbb{R}_+ \times \mathbb{X}^2 \to \mathbb{R}_+$ and $\lambda : \mathbb{R}_+ \times \mathbb{X} \to \mathbb{R}$ are are given measurable functions. Let $S \subset \mathbb{R}_+$ be locally finite (i.e., $S \cap [0, t]$ is finite for all $t > 0$). Then the system of equations

$$(\mathrm{B2}) \qquad \frac{\partial w}{\partial t}(t, x) = \int w(t, x')\,\beta(t, x', x)\,B(t, x', x)\,\mathrm{d}x' - \int w(t, x)\,\beta(t, x, x')\,\mathrm{d}x' - \lambda(t, x)\,w(t, x), \qquad t \notin S,$$

$$(\mathrm{B3}) \qquad w(t, x) = \int \widetilde{w}(t, x')\,\beta(t, x', x)\,B(t, x', x)\,\mathrm{d}x', \qquad\qquad t \in S,$$

is called the *filter equation generated by* β, with *boost* B, *decay* λ, and *observed event times* S. The process $\mathbf{X}_t$ is said to be the *driver* of the filter equation. Eq. B2 is the *regular part* of the filter equation; Eq. B3 is known as the *singular part*.

Remark 2. Trivially, a Kolmogorov forward equation is itself a filter equation with boost 1, decay 0, and $S = \varnothing$.

B.2. Properties. The following results show how filter equations allow one to integrate over random histories. First, Lemma B1 shows how one integrates over the full space of histories using a regular filter equation. Lemma B2 builds on this when the set of histories is restricted.

Lemma B1. *Suppose that $B : \mathbb{R}_+ \times \mathbb{X}^2 \to \mathbb{R}_+$ is measurable. Let $\mathbf{V}_t$ be an $\mathbb{R}_+$-valued random process satisfying*

$$\mathbb{E}\left[\mathbf{V}_t \mid \mathbf{H}_t = \mathrm{H}_t\right] = \prod_{e \,\in\, \mathsf{ev}(\mathrm{H}_t)} B(e, \widetilde{\mathrm{X}}_e, \mathrm{X}_e).$$

Let the family of measures λ_t on $\mathbb{X}$ be defined by

$$\lambda_t(\mathcal{E}) = \mathbb{E}\left[\mathbf{V}_t \cdot \mathbb{1}\{\mathbf{X}_t \in \mathcal{E}\}\right],$$

for measurable $\mathcal{E}$, and let $w(t, x)$ be the density of λ_t, i.e., $\lambda_t(\mathrm{d}x) = w(t, x)\,\mathrm{d}x$. In particular, $\mathbb{E}\left[\mathbf{V}_t\right] = \lambda_t(\mathbb{X}) = \int w(t, x)\,\mathrm{d}x$. Then w satisfies the initial condition $w(0, x) = p_0(x)$ and the regular filter equation,

$$(\mathrm{B4}) \qquad \frac{\partial w}{\partial t} = \int w(t, x')\,\alpha(t, x', x)\,B(t, x', x)\,\mathrm{d}x' - \int w(t, x)\,\alpha(t, x, x')\,\mathrm{d}x'.$$

Proof. Since $\mathsf{Prob}\left[\mathbf{V}_0 = 1\right] = 1$, $\lambda_0(\mathcal{E}) = \mathsf{Prob}\left[\mathbf{X}_0 \in \mathcal{E}\right]$, which implies that $w(0, x) = p_0(x)$. For $t > 0$ and $\Delta > 0$, the expectation can be broken into three terms, according to whether H_t has zero, one, or more than one event in $(t, t + \Delta]$. Accordingly, as $\Delta \downarrow 0$,

$$w(t + \Delta, x) = \left(1 - \Delta \int \alpha(t, x, x')\,\mathrm{d}x'\right) w(t, x) + \Delta \int \alpha(t, x', x)\,B(t, x', x)\,w(t, x')\,\mathrm{d}x' + o(\Delta).$$

Forming a difference quotient and taking the limit, we obtain Eq. B4. Note that here we use the assumption that α and B are càdlàg as functions of t. $\qquad\square$

When events are known to have occurred at particular times, it is of interest to integrate over those histories that include an event at each of these times. This leads to singular filter equations, as the next lemma shows. Before we state the lemma, some terminology is needed. Let $\mathbb{S}$ be the space of increasing, locally finite sequences in $\mathbb{R}_+$, with the topology induced by the Skorokhod metric and Lebesgue measure. For $t \in \mathbb{R}_+$ and $S \in \mathbb{S}$, let $S_t := S \cap [0, t]$. Thus if $S \in \mathbb{S}$ and $S_t = (\hat{s}_1, \dots, \hat{s}_K)$, then the infinitesimal element of Lebesgue measure at S_t is $\mathrm{d}S_t = \prod_{n=1}^{K} \mathrm{d}\hat{s}_n$.

Lemma B2. *Suppose that $B : \mathbb{R}_+ \times \mathbb{X}^2 \to \mathbb{R}_+$ is measurable and $\mathbf{V}_t$ is an $\mathbb{R}_+$-valued random process satisfying*

$$\mathbb{E}\left[\mathbf{V}_t \mid \mathbf{H}_t = \mathrm{H}_t\right] = \prod_{e \,\in\, \mathsf{ev}(\mathrm{H}_t)} B(e, \tilde{\mathrm{X}}_e, \mathrm{X}_e).$$

Let λ_t be a family of measures on $\mathbb{X} \times \mathbb{S}$ defined by

$$\lambda_t(\mathcal{E}, \mathcal{S}) = \mathbb{E}\left[\mathbf{V}_t \cdot \mathbb{1}\{\mathbf{X}_t \in \mathcal{E}\} \cdot \mathbb{1}\{\exists S \in \mathcal{S} \text{ s.t. } \mathsf{ev}(\mathbf{H}_t) \supseteq S_t\}\right],$$

whenever $\mathcal{E} \subseteq \mathbb{X}$ and $\mathcal{S} \subseteq \mathbb{S}$ are measurable. Let $w(t, x, S)$ be the density of this measure, i.e.,

$$\lambda_t(\mathrm{d}x\,\mathrm{d}S) = w(t, x, S)\,\mathrm{d}x\,\mathrm{d}S_t.$$

Then w satisfies

(B5) $\quad \dfrac{\partial w}{\partial t}(t, x, S) = \displaystyle\int w(t, x', S)\,\alpha(t, x', x)\,B(t, x', x)\,\mathrm{d}x' - \int w(t, x, S)\,\alpha(t, x, x')\,\mathrm{d}x', \qquad t \notin S,$

(B6) $\quad w(t, x, S) = \displaystyle\int \tilde{w}(t, x', S)\,\alpha(t, x', x)\,B(t, x', x)\,\mathrm{d}x', \qquad\qquad\qquad\qquad\quad t \in S.$

Proof. The proof proceeds by induction on the cardinality of S_t. The base case, for which $S_t = \varnothing$, follows immediately from Lemma B1. Assuming that it holds for $|S_t| < K$, one has only to verify Eq. B6. This can be accomplished by integrating Eq. 6 directly. $\qquad\square$

Remark 3. In the same way that Eqs. 4 and 5 can be represented as a single equation by means of a Dirac delta notation, the Eqs. B5 and B6 can be collapsed into a more compact form if β is allowed to have atoms at a countable set of time-points and the boost B is adjusted appropriately.

B.3. Adjoint filter equation. Given driver, boost, and decay functions, a filter equation determines a family of measures on $\mathbb{X}$. Specifically, suppose that, for $s \leqslant t$ and $S \subset \mathbb{R}$ locally finite, $w(s, x, t, x')$ satisfies the filter equation

(B7) $\quad w(s, x, s, x') = \delta(x, x')$

(B8) $\quad \dfrac{\partial w}{\partial t}(s, x, t, x') = \displaystyle\int w(s, x, t, \xi)\,\beta(t, \xi, x')\,B(t, \xi, x')\,\mathrm{d}\xi$

$$\qquad\qquad\qquad\qquad - \int w(s, x, t, x')\,\beta(t, x', \xi)\,\mathrm{d}\xi - \lambda(t, x')\,w(s, x, t, x'), \qquad t \notin S,$$

(B9) $\quad w(s, x, t, x') = \displaystyle\int \tilde{w}(s, x, t, \xi)\,\beta(t, \xi, x')\,B(t, \xi, x')\,\mathrm{d}\xi, \qquad\qquad\qquad\qquad\quad t \in S.$

Let ω be the measure whose density is w:

$$\omega(s, x, t, \mathcal{E}) := \int_{\mathcal{E}} w(s, x, t, x')\,\mathrm{d}x'.$$

If $f : \mathbb{X} \to \mathbb{R}$ is a measurable function, then one defines its integral with respect to this measure in the usual way:

$$\omega(s, x, t, f) = \int \omega(s, x, t, \mathrm{d}x')\,f(x') = \int w(s, x, t, x')\,f(x')\,\mathrm{d}x'.$$

In particular, $\omega(s, x, t, f)$ is itself a measurable function of x and we have the Chapman-Kolmogorov-like relation

(B10) $\quad \omega(s, x, t, f) = \displaystyle\int \omega(s, x, \tau, \omega(\tau, \xi, t, f))\,\mathrm{d}\xi = \int \omega(s, x, \tau, \mathrm{d}\xi)\,\omega(\tau, \xi, t, f),$

whenever $s \leqslant \tau \leqslant t$. Let us now fix t and f and write $F(s, x) = \omega(s, x, t, f)$. Let us take β, B, and λ to be càdlàg in t and F to be càglàd in s. Applying Eq. B10 we obtain, for $\Delta > 0$ and $s \notin S$,

$$F(s - \Delta, x) = \left(1 - \Delta \int \beta(s - \Delta, x, \xi) \, d\xi - \Delta \, \lambda(s - \Delta, x)\right) F(s, x)$$
$$+ \Delta \int \beta(s - \Delta, x, \xi) \, B(s - \Delta, x, \xi) \, F(s, \xi) \, d\xi + o(\Delta).$$

Taking $\Delta \downarrow 0$ gives

(B11) $$-\frac{\partial F}{\partial s} = \int \widetilde{\beta}(s, x, \xi) \left[\widetilde{B}(s, x, \xi) \, F(s, \xi) - F(s, x)\right] d\xi - \widetilde{\lambda}(s, x) \, F(s, x).$$

For $s \in S$, we have

(B12) $$F(s, x) = \int \omega(s, x, s, d\xi) \, \underset{\sim}{F}(s, \xi) = \int \beta(s, x, \xi) \, B(s, x, \xi) \, \underset{\sim}{F}(s, \xi) \, d\xi,$$

where $\underset{\sim}{F}$ denotes the right-limit of F. Eqs. B11 and B12 constitute the *adjoint form* of Eqs. B7–B9. Together with the final condition $F(t, x) = f(x)$, these determine $\omega(s, x, t, f)$ for all $s \leqslant t$. Note that, when $B = 1$ and $\lambda = 0$, the adjoint filter equation becomes the familiar Kolmogorov backward equation (Eq. 3) for the driving process.

B.4. Solving filter equations by sequential Monte Carlo. Filter equations afford a convenient means of computing expectations and likelihoods for pure jump processes. This is facilitated by the following Lemma, the statement of which uses a one-sided Dirac delta function. Specifically, let $\delta(v, v')$ be the right-sided Dirac delta function satisfying $\delta(v, v') = 0$ for $v \neq v'$ and

$$\int_a^b f(v) \, \delta(v, v') \, dv = f(v') \, \mathbb{1}\{v' \in [a, b)\},$$

whenever f is càdlàg and $-\infty \leqslant a < b \leqslant \infty$.

Lemma B3. *Eqs. B2 and B3 are satisfied by $w(t, x) = \int_0^\infty v \, u(t, x, v) \, dv$, where $u(t, x, v)$ satisfies the KFE*

(B13)
$$\frac{\partial u}{\partial t}(t, x, v) = \frac{\partial}{\partial v} \left[\lambda(t, x) \, v \, u(t, x, v)\right] + \int_0^\infty \int u(t, x', v') \, \beta(t, x', x) \, \delta\big(v, B(t, x', x) \, v'\big) \, dx' \, dv'$$
$$- \int_0^\infty \int u(t, x, v) \, \beta(t, x, x') \, \delta\big(v', B(t, x, x') \, v\big) \, dx' \, dv', \qquad t \notin S,$$

$$u(t, x, v) \, dx = \int_0^\infty \int \widetilde{u}(t, x', v') \, \pi(t, x', dx) \, \delta\big(v, A(t, x') \, B(t, x', x) \, v'\big) \, dx' \, dv', \qquad t \in S.$$

Here, $A(t, x) := \int \beta(t, x, x') \, dx'$ and $\pi(t, x, dx') := \beta(t, x, x') \, dx'/A(t, x)$.

Proof. For each $t \notin S$, we have

$$\frac{\partial w}{\partial t}(t, x) = \int_0^\infty v \, \frac{\partial u}{\partial t}(t, x, v) \, dv$$
$$= \int_0^\infty \int \int_0^\infty v \, u(t, x', v') \, \beta(t, x', x) \, \delta\big(v, B(t, x', x) \, v'\big) \, dv \, dx' \, dv'$$
$$- \int_0^\infty \int \int_0^\infty v \, u(t, x, v) \, \beta(t, x, x') \, \delta\big(v', B(t, x, x') \, v\big) \, dv \, dx' \, dv'$$
$$+ \int_0^\infty v \, \tfrac{\partial}{\partial v} \left[\lambda(t, x) \, v \, u(t, x, v)\right] dv.$$

Here, the non-explosivity assumption guarantees that we can differentiate under the integral sign and exchange the order of integration. Moreover, it ensures that $u \to 0$ as $v \to \infty$. Hence, by evaluating the first integral with respect to v, the second with respect to v', and the third by parts, we obtain

$$\frac{\partial w}{\partial t}(t, x) = \int v' \, u(t, x', v') \, \beta(t, x', x) \, B(t, x', x) \, dv' \, dx' - \int v \, u(t, x, v) \, \beta(t, x, x') \, dv \, dx' - \lambda(t, x) \int v \, u(t, x, v) \, dv,$$

which is simplified to obtain Eq. B2. Similarly, at each $t \in S$, we have

$$
\begin{aligned}
w(t,x)\,\mathrm{d}x &= \int_0^\infty \int \int_0^\infty v\,\widetilde{u}(t,x',v')\,\pi(t,x',\mathrm{d}x)\,\delta\big(v,A(t,x')\,B(t,x',x)\,v'\big)\ \mathrm{d}x'\,\mathrm{d}v'\,\mathrm{d}v \\
&= \int_0^\infty \int v'\,\widetilde{u}(t,x',v')\,\pi(t,x',\mathrm{d}x)\,A(t,x')\,B(t,x',x)\,\mathrm{d}x'\,\mathrm{d}v' \\
&= \int \widetilde{w}(t,x')\,\beta(t,x',x)\,B(t,x',x)\,\mathrm{d}x'\,\mathrm{d}x.
\end{aligned}
$$

which is equivalent to Eq. B3 □

Eqs. B13 are recognizable as the KFE of a certain process $(\mathbf{X}_t,\mathbf{V}_t)$. In particular, the driver $\mathbf{X}_t$ has KFE Eq. B1. $\mathbf{V}_t$ is *directed* by $\mathbf{X}_t$ in the sense that $\mathbf{V}$ has jumps wherever $\mathbf{X}$ does: when $\mathbf{X}$ jumps at time t from x to x', $\mathbf{V}$ jumps by the multiplicative factor $B(t,x,x') \geqslant 0$. Between jumps, $\mathbf{V}_t$ decays deterministically and exponentially at rate $\lambda(t,x)$. At the known times in S, $\mathbf{X}$ jumps according to the probability kernel π and, $\mathbf{V}$ jumps by the factor $A(t,x)\,B(t,x,x')$. If we view $\mathbf{V}_t$ as a weight, then Lemma B3 tells us how the $\mathbf{V}_t$-weighted average of $\mathbf{X}_t$ evolves in time: this average is simply $\int w(t,x)\,\mathrm{d}x$. Thus, Lemma B3 gives a recipe for the integration of Eqs. B5 and B6 in the Monte Carlo sense. A pseudocode representation of this procedure is given in Algorithm B1. This is a simple sequential importance sampling algorithm and as such can be improved upon in many ways (Doucet et al., 2001; Arulampalam et al., 2002; Wills & Schön, 2023). Our purpose here is merely to illustrate the implications of Lemma B3 by showing how the features of Eqs. B2 and B3 translate into computations, in the simplest possible context.

Appendix C.   Examples.

C.1. SIRS model. King et al. (2022) worked out formulae for the exact likelihood of a genealogy induced by an SIRS model. The theory developed in this paper applies, but since there is only one deme in this model, this is a simple case. Its state vector is $x = (S, I, R)$ and its KFE is

$$
\begin{aligned}
\frac{\partial v}{\partial t}(S,I,R) =\ & \frac{\beta\,(S+1)\,(I-1)}{N}\,v(t,S+1,I-1,R) - \frac{\beta\,S\,I}{N}\,v(t,S,I,R) \\
& + \gamma\,(I+1)\,v(t,S,I+1,R-1) - \gamma\,I\,v(t,S,I,R) \\
& + \omega\,(R+1)\,v(t,S-1,I,R+1) - \omega\,R\,v(t,S,I,R).
\end{aligned}
$$

Here $N = S + I + R$ is the host population size. Note that, though the theory allows for time-dependent event rates, this example is time-homogeneous. This model has one deme and occupancy function $n(x) = I$. There are four kinds of jumps: transmission, recovery, waning of immunity, and sampling. Accordingly, the marks are $\mathbb{U} = \{\mathsf{Trans}, \mathsf{Recov}, \mathsf{Wane}, \mathsf{Sample}\}$. Table C1 gives α_u, r^u, and the event type for each of these marks.

TABLE C1.   Elements of the SIRS model pertinent to the genealogy process. The table shows the rate (α_u), jump ($x \mapsto x'$), production (r^u), and event type for each of the model's four marks (u).

u	α_u	$x \mapsto x'$	r^u	Event type
Trans	$\frac{\beta S I}{N}$	$(S,I) \mapsto (S-1,I+1)$	2	pure birth
Recov	γI	$(I,R) \mapsto (I-1,R+1)$	0	pure death
Wane	ωR	$(S,R) \mapsto (S+1,R-1)$	0	neutral
Sample	ψI	$x \mapsto x$	1	pure sample

Given an obscured genealogy Z, let $\mathsf{ev}(Z) = B \cup S_0 \cup S_1$, where B is the set of branch-times, and S_0, S_1 are the sets of sample-times with saturations 0 and 1, respectively. Since there is only one deme, paintings of Z can differ only in number and position of inline, internal nodes along branches. Each of these can only correspond to

Algorithm B1 Sequential Monte Carlo filter. The following is a simple Monte Carlo scheme for the filter given by Eqs. B2 and B3. It takes as input an initial time t_0, a final time T, and a sequence of observed event times $S = \{t_1, \ldots, t_K\}$. We order the latter so that $t_0 \leqslant t_1 < t_2 < \cdots < t_K \leqslant t_{K+1} \coloneqq T$. In addition, we are given event rates β, a boost function B, a decay function λ, and the initial distribution of latent states p_0. As in Lemma B3, we let

$$A(t, x) \coloneqq \int \beta(t, x, x') \, \mathrm{d}x', \qquad \pi(t, x, \mathrm{d}x') \coloneqq \frac{\beta(t, x, x')}{A(t, x)} \, \mathrm{d}x'.$$

For a given, fixed ensemble size, $J \in \mathbb{Z}_+$, the algorithm returns an unbiased Monte Carlo estimate, $\hat{\mathcal{L}}$, of the likelihood.

1: **procedure** SMCFILTER($t_0, T, S, \beta, B, \lambda, p_0, J$)
   Initialize an ensemble of latent states and weights:
2:     $x_j \sim p_0(\cdot)$, for $j = 1, \ldots, J$
3:     $w_j \leftarrow 1$, for $j = 1, \ldots, J$
4:     **for** $k = 0, \ldots, K$ **do**   ▷ loop over observed event times
5:         **for** $j = 1, \ldots, J$ **do**   ▷ loop over the ensemble
6:             $t \leftarrow t_k$
7:             **if** $k > 0$ **then**   ▷ singular part (Eq. B3)
8:                 $x'_j \sim \pi(t, x_j, \cdot)$
9:                 $w_j \leftarrow w_j \, A(t, x_j) \, B(t, x_j, x'_j)$
10:                $x_j \leftarrow x'_j$
11:            **while** $t < t_{k+1}$ **do**   ▷ regular part (Eq. B2)
12:                $\Delta \sim \mathrm{Exp}(A(t, x_j))$ ▷ Sojourn times are exponentially distributed.
13:                **if** $t + \Delta < t_{k+1}$ **then**
14:                    $x'_j \sim \pi(t, x_j, \cdot)$
15:                    $t' \leftarrow t + \Delta$
16:                    $w_j \leftarrow w_j \, B(t', x_j, x'_j) \, \exp\left(-\lambda(t, x_j) \, \Delta\right)$
17:                    $x_j \leftarrow x'_j$
18:                    $t \leftarrow t'$
19:                **else**
20:                    $\Delta \leftarrow t_{k+1} - t$
21:                    $w_j \leftarrow w_j \, \exp\left(-\lambda(t, x_j) \, \Delta\right)$
22:                    $t \leftarrow t_{k+1}$
23:    $\hat{\mathcal{L}} \leftarrow \frac{1}{J} \sum_j w_j$

$u = \mathsf{Trans}$ with $s = 1$. For $t \notin \mathsf{ev}(Z)$, we can take the importance sampling distribution to be

$$\pi_u = \begin{cases} c, & u = \mathsf{Trans}, s = 0, t \notin \mathsf{ev}(Z), \\ \frac{1-c}{\ell}, & u = \mathsf{Trans}, s = 1, t \notin \mathsf{ev}(Z), \\ 0, & \text{otherwise.} \end{cases}$$

Here c is an arbitrary probability that does not affect the computation. The relevant binomial ratios are

$$\binom{n \quad \ell}{r \quad s} = \begin{cases} \frac{(I-\ell)\,(I-\ell-1)}{I\,(I-1)}, & u = \mathsf{Trans}, s = 0, \\ \frac{2\,(I-\ell)}{I\,(I-1)}, & u = \mathsf{Trans}, s = 1, \\ \frac{2}{I\,(I-1)}, & u = \mathsf{Trans}, s = 2, \\ \frac{I-\ell}{I}, & u = \mathsf{Sample}, s = 0, \\ \frac{1}{I}, & u = \mathsf{Sample}, s = 1. \end{cases}$$
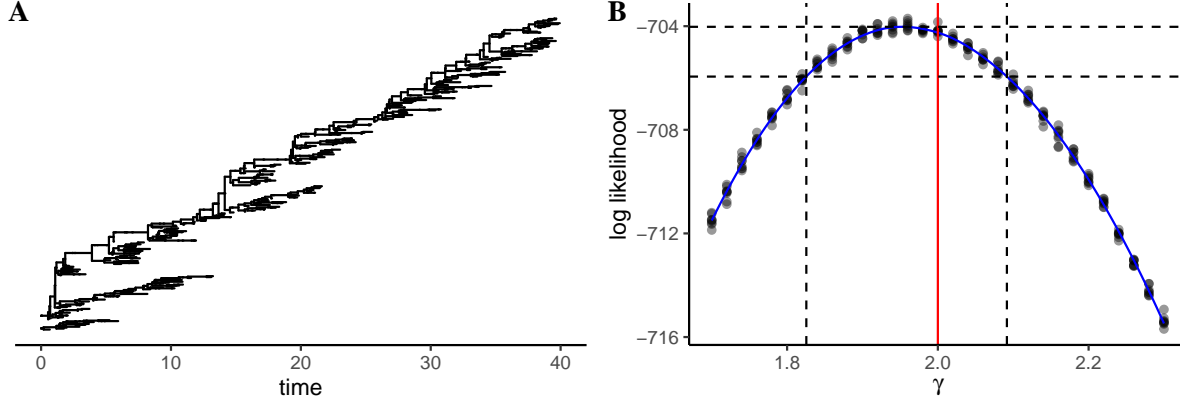
A

B

FIGURE C1.   Likelihood computation for the SIRS model by sequential Monte Carlo. (A) A simulated genealogy for $\beta = 4$, $\gamma = 2$, $\omega = 1$, $\psi = 1$, $(S_0, I_0, R_0) = (97, 3, 0)$. (B) A slice through the likelihood surface at the true parameters in the γ-direction. Each point is a distinct Monte Carlo estimate. The blue curve is a LOESS smooth; the dashed lines bound the Monte Carlo-adjusted 95% confidence interval (Ionides et al., 2017).

This leads, for $t \notin \mathsf{ev}(Z)$, to the following regular part of the filter equation:

$$
\begin{aligned}
\frac{\partial w}{\partial t} =& \frac{\beta\,(S+1)\,(I-1)}{N}\left(1 - \frac{\binom{\ell(t)}{2}}{\binom{I}{2}}\right) w(t, S+1, I-1, R) - \frac{\beta\,S\,I}{N}\,w(t, S, I, R) \\
&+ \gamma\,(I+1)\,w(t, S, I+1, R-1) - \gamma\,I\,w(t, S, I, R) \\
&+ \omega\,(R+1)\,w(t, S-1, I, R+1) - \omega\,R\,w(t, S, I, R) \\
&- \psi\,I\,w(t, S, I, R).
\end{aligned}
$$

Here, we have summed over the various paintings for $u = \mathsf{Trans}$, $s < 2$. Note the presence of the decay term proportional to ψ. At event-times $t \in \mathsf{ev}(Z)$, the singular part of the filter equation reads

$$
w(t, S, I, R) = \begin{cases}
\widetilde{w}(t, S+1, I-1, R)\,\frac{2\,\beta\,(S+1)}{N\,I}, & t \in B, \\
\widetilde{w}(t, S, I, R)\,\psi\,(I - \ell(t)), & t \in S_0, \\
\widetilde{w}(t, S, I, R)\,\psi, & t \in S_1.
\end{cases}
$$

Finally, note that $w(t, S, I, R) = 0$ for all $I < \ell(t)$.

C.2. SEIRS model. A simple, yet interesting, model with more than one deme is the SEIRS model (Fig. 1A). The state space is $\mathbb{R}_+^4$, with the state $x = (S, E, I, R)$ defined by the numbers of hosts in each of the four compartments. The KFE for the population process is

$$
\begin{aligned}
\frac{\partial v}{\partial t}(t, S, E, I, R) =& \frac{\beta\,(S+1)\,I}{N}\,v(t, S+1, E-1, I, R) - \frac{\beta\,S\,I}{N}\,v(t, S, E, I, R) \\
&+ \sigma\,(E+1)\,v(t, S, E+1, I-1, R) - \sigma\,E\,v(t, S, E, I, R) \\
&+ \gamma\,(I+1)\,v(t, S, E, I+1, R-1) - \gamma\,I\,v(t, S, E, I, R) \\
&+ \omega\,(R+1)\,v(t, S-1, E, I, R+1) - \omega\,R\,v(t, S, E, I, R),
\end{aligned}
$$

where $N = S + E + I + R$ is the total population size. Note that the terms associated with sampling cancel each other in the KFE, since, with the state space defined as above, sampling has no effect on the state.

This model has two demes: $\mathbb{D} = \{\mathsf{E}, \mathsf{I}\}$. Its deme occupancy function is $n(x) = (E, I)$. There are five kinds of jumps: transmission, progression, recovery, waning of immunity, and sampling. The corresponding marks are $\mathbb{U} = \{\mathsf{Trans}, \mathsf{Prog}, \mathsf{Recov}, \mathsf{Wane}, \mathsf{Sample}\}$. Table C2 gives α_u, r^u, and the event type for each of these marks.

TABLE C2.   Elements of the SEIRS model pertinent to the genealogy process. The table shows the rate (α_u), jump ($x \mapsto x'$), production (r^u), and event type for each of the model's five marks (u).

u	α_u	$x \mapsto x'$	r^u	Event type
Trans	$\frac{\beta SI}{N}$	$(S,E) \mapsto (S-1, E+1)$	$(1,1)$	pure birth
Prog	σE	$(E,I) \mapsto (E-1, I+1)$	$(0,1)$	pure migration
Recov	γI	$(I,R) \mapsto (I-1, R+1)$	$(0,0)$	pure death
Wane	ωR	$(S,R) \mapsto (S+1, R-1)$	$(0,0)$	neutral
Sample	ψI	$x \mapsto x$	$(0,1)$	pure sample

TABLE C3.   Elements of a scheme for numerically computing the likelihood under the SEIRS model. The regular portion of the filter equation holds in between genealogical events (i.e., for $t \notin B \cup S_0 \cup S_1$); the singular portion describes the effect of these events ($t \in B \cup S_0 \cup S_1$). Each line corresponds to a potential event. An event with mark u and saturation s has the boost given by the binomial ratio shown (third column). The $y \mapsto y'$ column depicts the proposed painting schematically, and π is the probability of that proposal. Blue is used for the E deme and yellow for the I deme. For each line, the filter equation contains m terms. An asterisk ($*$) stands for cases not explicitly mentioned.

	u	s	Q	$\begin{pmatrix} n & \ell \\ r & s \end{pmatrix}$	$y \mapsto y'$	π	m	Line
$t \notin B \cup S_0 \cup S_1$	Trans	$(0,0)$	1	$\left(\frac{E-\ell_E}{E}\right)\left(\frac{I-\ell_I}{I}\right)$		$\frac{I-\ell_I}{I}$	1	1
	Trans	$(1,0)$	1	$\frac{I-\ell_I}{EI}$		$\frac{1}{2I}$	ℓ_I	2
	Trans	$(0,1)$	1	$\frac{E-\ell_E}{EI}$		$\frac{1}{2I}$	ℓ_I	3
	Trans	$(1,1)$	0				0	4
	Prog	$(0,0)$	1	$\frac{I-\ell_I}{I}\,\mathbb{1}\{E \geqslant \ell_E\}$		$\frac{E-\ell_E}{E}\,\mathbb{1}\{E > \ell_E\}$	1	5
	Prog	$(0,1)$	1	$\frac{1}{I}\,\mathbb{1}\{E \geqslant \ell_E\}$		$\frac{1}{E}\,\mathbb{1}\{E > \ell_E\}$	ℓ_E	6
	Recov	$(0,0)$	1	$\mathbb{1}\{I \geqslant \ell_I\}$		$\mathbb{1}\{I > \ell_I\}$	1	7
	Wane	$(0,0)$	1	1		1	1	8
	Sample	$*$	0				0	9
$t \in B$	Trans	$(1,1)$	1	$\frac{1}{EI}$		$\frac{1}{2}$	1	10
	Trans	$(1,1)$	1	$\frac{1}{EI}$		$\frac{1}{2}$	1	11
	Trans	$(1,1)$	0				0	12
	Trans	$(1,1)$	0				0	13
	$*$	$*$	0				0	14
$t \in S_0$	Sample	$(0,0)$	1	$\frac{I-\ell_I}{I}$		1	1	15
	Sample	$(0,0)$	0				0	16
	$*$	$*$	0				0	17
$t \in S_1$	Sample	$(0,1)$	1	$\frac{1}{I}$		1	1	18
	Sample	$(0,1)$	0				0	19
	$*$	$*$	0				0	20

In the application of Theorem 5 and Corollary 6, one has wide latitude in the selection of the kernels π and q. Table C3 gives a particular set of choices that are *naïve* in the sense that they refrain from borrowing information from the future. Such a naïve filter is always available, though it will rarely be optimal.

Using the choices of Table C3, we proceed to write down a filter equation for the SEIRS model. As previously, given an obscured genealogy, let B be the set of its branch times, S_0 be the set of tip-sample times, and S_1 be the set of inline sample times. Then for $t \notin B \cup S_0 \cup S_1$, the filter equation reads:

$$
\begin{aligned}
\frac{\partial w}{\partial t}(t, S, E, I, R, y) =\ & \frac{\beta\,(S+1)\,I}{N}\left(1 - \frac{\ell_E}{E}\right)\left(1 - \frac{\ell_I}{I}\right) w(t, S+1, E-1, I, R, y) \\
& + \sum_{k=1}^{\ell_I} \frac{\beta\,(S+1)\,I}{N}\,\frac{1}{E}\left(1 - \frac{\ell_I}{I}\right) w(t, S+1, E-1, I, R, \widetilde{y}_2^k) \\
& + \sum_{k=1}^{\ell_I} \frac{\beta\,(S+1)\,I}{N}\,\frac{1}{I}\left(1 - \frac{\ell_E}{E}\right) w(t, S+1, E-1, I, R, \widetilde{y}_3^k) - \frac{\beta\,S\,I}{N}\,w(t, S, E, I, R, y) \\
& + \sigma\,(E+1)\,\mathbb{1}\{E \geqslant \ell_E\}\left(1 - \frac{\ell_I}{I}\right) w(t, S, E+1, I-1, R, y) \\
& + \sum_{k=1}^{\ell_E} \sigma\,(E+1)\,\mathbb{1}\{E \geqslant \ell_E\}\,\frac{1}{I}\,w(t, S, E+1, I-1, R, \widetilde{y}_6^k) - \sigma\,E\,w(t, S, E, I, R, y) \\
& + \gamma\,(I+1)\,\mathbb{1}\{I \geqslant \ell_I\}\,w(t, S, E, I+1, R-1, y) - \gamma\,I\,w(t, S, E, I, R, y) \\
& + \omega\,(R+1)\,w(t, S-1, E, I, R+1, y) - \omega\,R\,w(t, S, E, I, R, y) - \psi\,I\,w(t, S, E, I, R, y).
\end{aligned}
$$
(C1)

Here, $\widetilde{y}_j^k$ refers to the coloring of the tree immediately preceding the proposal indicated on line j of Table C3. The integer k specifies the particular branch on which the change occurs.

The singular portion of the filter equation has one component for each distinct type of genealogical event:

(C2)
$$
w(t, S, E, I, R, y) =
\begin{cases}
\dfrac{\beta\,(S+1)\,I}{N}\,\dfrac{1}{EI}\,\widetilde{w}(t, S+1, E-1, I, R, \widetilde{y}_{10}) & \\
\qquad + \dfrac{\beta\,(S+1)\,I}{N}\,\dfrac{1}{EI}\,\widetilde{w}(t, S+1, E-1, I, R, \widetilde{y}_{11}), & t \in B, \\
\psi\,(I - \ell_I)\,\widetilde{w}(t, S, E, I, R, \widetilde{y}_{15}), & t \in S_0, \\
\psi\,\widetilde{w}(t, S, E, I, R, \widetilde{y}_{18}), & t \in S_1.
\end{cases}
$$

In addition to Eqs. C1 and C2, the quantity w should satisfy the condition $w(t, S, E, I, R, y) = 0$ whenever $E < \ell_E$ or $I < \ell_I$.

A variety of importance-sampling kernels q, π are permissible under the terms of Theorem 5. With a particular choice of importance-sampling kernel, the filter equation uniquely specifies a sequential Monte Carlo algorithm for estimating the likelihood. The specific choices made in Table C3 underlie the results displayed in Fig. C2. Some numerical results obtained using this scheme are presented in Fig. C2.

A. A. KING, DEPARTMENT OF ECOLOGY & EVOLUTIONARY BIOLOGY, CENTER FOR THE STUDY OF COMPLEX SYSTEMS, AND DEPARTMENT OF MATHEMATICS, UNIVERSITY OF MICHIGAN, ANN ARBOR, MI 48109 USA *and* SANTA FE INSTITUTE, 1399 HYDE PARK ROAD, SANTA FE, NM 87501 USA

Email address: kingaa@umich.edu

URL: https://kinglab.eeb.lsa.umich.edu/

Q.-Y. LIN, THEORETICAL BIOLOGY AND BIOPHYSICS, LOS ALAMOS NATIONAL LABORATORY, LOS ALAMOS, NM 87545 USA

E. L. IONIDES, DEPARTMENT OF STATISTICS, UNIVERSITY OF MICHIGAN, ANN ARBOR, MI 48109 USA
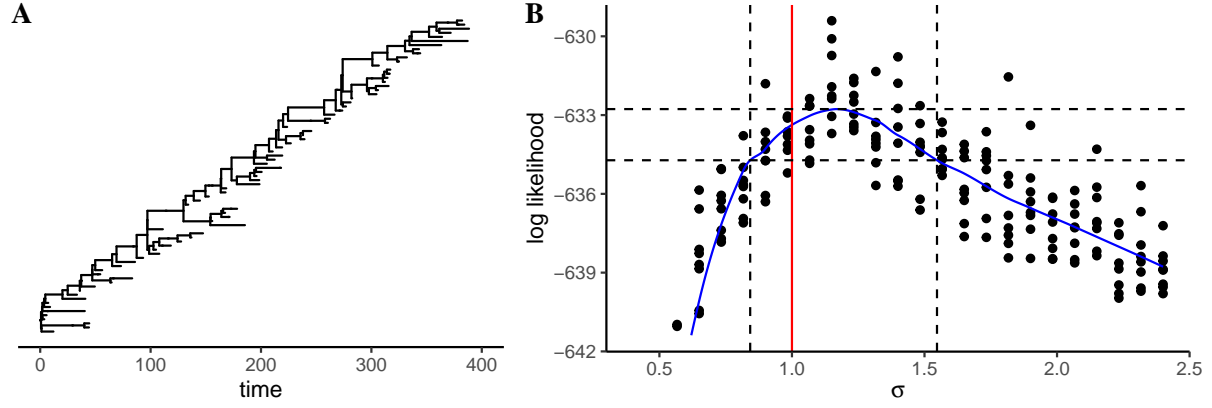
FIGURE C2.   Likelihood computation for the SEIRS model by sequential Monte Carlo, using the scheme of Table C3. (A) Simulated genealogy for $\beta = 3$, $\sigma = 1$, $\gamma = 0.5$, $\psi = 0.02$, $\omega = 0.08$, $(S_0, E_0, I_0, R_0) = (70, 1, 0, 50)$. (B) Likelihood slice in the σ-direction. Each point represents the estimate of an independent sequential Monte Carlo computation. The blue curve shows a LOESS smooth; the dashed vertical lines enclose the Monte Carlo-adjusted 95% confidence interval (Ionides et al., 2017).