EXACT PHYLODYNAMICS VIA STRUCTURED MARKOV GENEALOGY PROCESSES

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

ABSTRACT. We show how each member of a broad class of Markov population models uniquely induces a stochastic process on the space of genealogies. We then derive exact expressions for the likelihood of an observed genealogy in terms of filtering equations, which can be integrated using standard Monte Carlo methods. The class of population models that can be accommodated includes essentially all compartmental models.

1. Introduction.

When an infectious agent mutates quickly enough so that information on the history of its passage through hosts accumulates in its genome, yet not so quickly that this information is dissipated, it is possible to extract this information from sampled genomes to gain insight into key determinants of infectious disease transmission. With sufficiently dense sampling, it can be possible to reconstruct the detailed history of who infected whom, which can aid identification of the correlates of transmission and immunity and the properties of infections. More typically, one can use relatively sparse samples to learn about these features via the estimation of a parsimonious model of the transmission process. In particular, one can formalize one or more mathematical models of transmission, estimate their parameters, and compare their ability to explain data, following standard statistical paradigms. This is known as *phylodynamic* inference, the term being due to Grenfell et al. (2004). Alizon (2024) gives a good review of the history of the subject.

The most common approach to phylodynamic inference rests upon 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 because it is maximally efficient (i.e., loses the least information), it is desirable to be able to compute the likelihood function for models of interest. This is simply the probability density of a given genealogy conditional on a given model, viewed as a function of the parameters of that model. In particular, if S 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 likelihood of S is

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

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

In particular, existing approaches to the phylodynamic likelihood have been based on one of two mathematical idealizations. The first is the Kingman (1982) coalescent, by which likelihood of a given genealogy is computed via a reverse-time argument. This computation is exact when the population size is constant. 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 exact expressions for the likelihood are available (Stadler, 2010).

Date: May 16, 2024.

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 makes them attractive, 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 unutilized 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 afford exact computation of the phylodynamic likelihood, so that comparisons among parameterizations and models could be made on a sound basis. Second, because nonlinearity, nonstationarity, noise, and measurement error are so prominent and ubiquitous in epidemiology, it would accommodate nonlinear, time-inhomogeneous, stochastic transmission models. Third, because many of the most important uncertainties concern heterogeneities in transmission rates and the susceptibility, behavior, age, and location of hosts, it would accommodate host populations structured by these factors. While some structuring factors (e.g., age, spatial location) are most naturally expressed in terms of continuous variables, discretely structured models have repeatedly proved their value in epidemiology. In particular, compartmental models are extremely flexible and have often been used as approximations when continuous structure leads to uncomfortably high model dimension. Finally, because there is typically uncertainty not only in the parameters, but also in the structure, of a host-pathogen system, an improved phylodynamic inference methodology would place minimal restrictions on the form of the models that it can accommodate. This paper demonstrates how these desiderata can be achieved—at least for models with discrete structure—including arbitrary nonlinear compartmental models with a countable number of compartments.

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 indeed 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.

In the following, we take as our starting point a transmission model in the form of a discretely structured, Markov process. We show how such a process uniquely induces each of several stochastic processes in the space of genealogies. We go on to derive expressions for the exact likelihoods of these genealogies.

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 assume that sample paths are càdlàg i.e., right-continuous with left limits. We will frequently need to refer to the left-limit of such a process. Accordingly, if $\mathbf{\Phi}_t$ is a càdlàg random process, we define

$$\widetilde{\mathbf{\Phi}}_t := \begin{cases} \lim\limits_{s \uparrow t} \mathbf{\Phi}_s, & t > 0, \\ \mathbf{\Phi}_0, & t = 0. \end{cases}$$

Note that $\widetilde{\mathbf{\Phi}}_t$ is thus left-continuous with right limits.

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 $\hat{\mathbf{\Phi}}_k := \mathbf{\Phi}_{\hat{\mathbf{T}}_k}$, $k = 0, \ldots, \mathbf{K}_t$. 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 $t \in \left[\hat{\mathbf{T}}_k, \hat{\mathbf{T}}_{k+1}\right)$, $k = 0, \ldots, \mathbf{K}_t$.

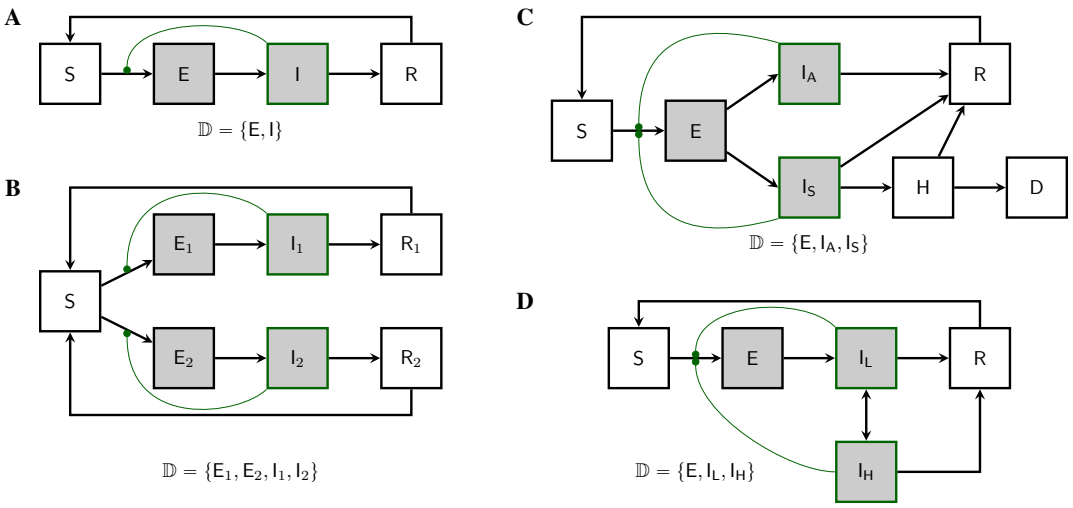

FIGURE 1. Examples of discretely-structured population models. Demes are shaded. Compartments containing infectious hosts are outlined in green. Curved green lines connect transmission rates with the compartments whose occupancies control their modulation; each such connection gives rise to a nonlinearity 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: $\mathbb{D} = \{\mathsf{E}, \mathsf{I}\}$. If there is exactly one lineage per host, then the occupancy, $n(\mathbf{X}_t) = (n_{\mathsf{E}}(\mathbf{X}_t), n_{\mathsf{I}}(\mathbf{X}_t))$, 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.

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

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

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 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

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

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

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

We further assume that $\alpha(t, x, x')$ is càdlàg as a function of time for all $x, x' \in \mathbb{X}$ and that the number of jumps that occur in a finite time-interval is finite, i.e., $\mathsf{Prob}\left[\mathbf{K}_t < \infty\right] = 1$ for all t.

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

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

and if, moreover, $v(0, x) = p_0(x)$, then $\int_{\mathcal{E}} v(t, x)\, \mathrm{d}x = \mathsf{Prob}\left[\mathbf{X}_t \in \mathcal{E}\right]$ for every measurable $\mathcal{E} \subseteq \mathbb{X}$. Eq. 2 is sometimes called the *master equation* for $\mathbf{X}_t$.

2.4. Inclusion of jumps at deterministic times. For modeling purposes, it is sometimes desirable to insist that certain events occur at known 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 $S = \{s_1, s_2, \ldots,\} \subset \mathbb{Z}_+$ is a sequence of event 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 the state just before the jump being x. With this notation, the KFE for the process becomes

$$\frac{\partial v}{\partial t}(t, x) = \int v(t, x')\, \alpha(t, x', x)\, \mathrm{d}x' - \int v(t, x)\, \alpha(t, x, x')\, \mathrm{d}x', \qquad t \notin S, \tag{3}$$

$$v(t, x)\, \mathrm{d}x = \int \tilde{v}(t, x')\, \pi(t, x', \mathrm{d}x)\, \mathrm{d}x', \qquad t \in S. \tag{4}$$

Note that the Eq. 3 is identical to Eq. 2; we call this the *regular part* of the KFE. We refer to Eq. 4 as the *singular part* of the KFE.

As a matter of notation, one can represent Eqs. 3 and 4 as a single equation in the form of Eq. 2. In particular, if in Eq. 2 we make the substitution

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

we obtain an equation which we can view as shorthand for Eqs. 3 and 4. Here, $\delta(t, s)$ is a Dirac delta function and $\mathrm{d}\pi/\mathrm{d}x'$ denotes the density (i.e., Radon-Nikodym derivative) of π with respect to the measure on $\mathbb{X}$.

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.

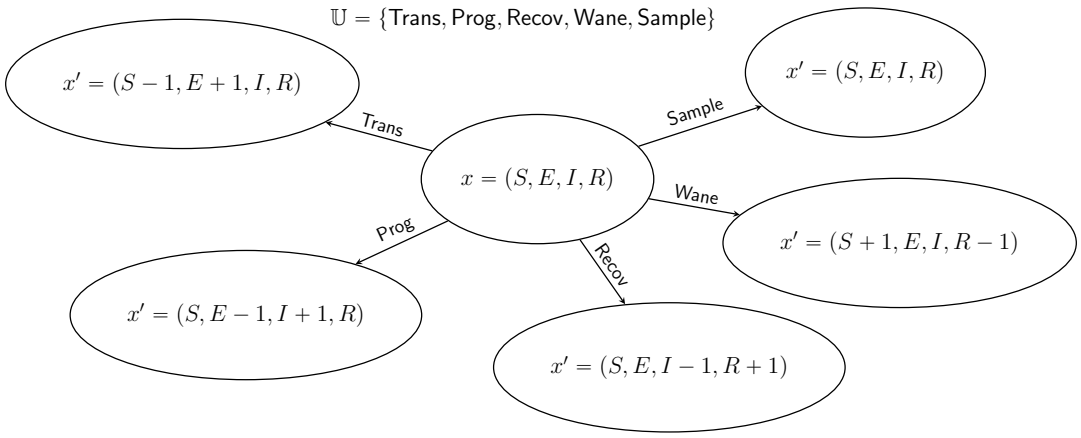

FIGURE 2. Markov state transition diagram for the SEIRS model depicted in Fig. 1A. The state, x, is characterized by four numbers, S, E, I, and R. From a given state x, there are five possible kinds of jumps $x \mapsto x'$. Accordingly, the set, $\mathbb{U}$, of jump marks has five elements. Each of 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. See §3.1 for a description of these jump types. Note that, in this formulation, when a sampling event occurs, the state does not change.

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, though $\mathbf{X}_t$ and $(\mathbf{X}_t, \mathbf{U}_t)$ are Markov processes, $\mathbf{U}_t$ is not.

2.6. Demes and deme occupancy. Our first goal in this paper is to show how a given population process induces a unique 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 use the symbol $\mathbb{D}$ to denote an index set for them. Fig. 1 illustrates this concept in the context of several compartmental 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 indeed. 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. Appendix B works out the theory for each of these examples.

SIRS model. King et al. (2022) worked out formulas 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.

SEIRS model. A simple, yet interesting, model with more than one deme is the SEIRS model (Fig. 1A). The state space is $\mathbb{Z}_+^4$, with the state $x = (S, E, I, R)$ defined by the numbers of hosts in each of the four compartments. It has two demes: $\mathbb{D} = \{\mathsf{E}, \mathsf{I}\}$. The deme occupancy function in this case is $n(x) = (E, I)$. Note that the terms associated with sampling cancel each other in the KFE, since, in this model, sampling has no effect on the state.

Two-strain competition model. A simple model for the competition of two strains for susceptible hosts 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 $n(x) = (E_1, E_2, I_1, I_2)$.

Superspreading model. Fig. 1D depicts a model of superspreading. There are three demes ($\mathbb{D} = \{\mathsf{E}, \mathsf{I}_\mathsf{L}, \mathsf{I}_\mathsf{H}\}$).

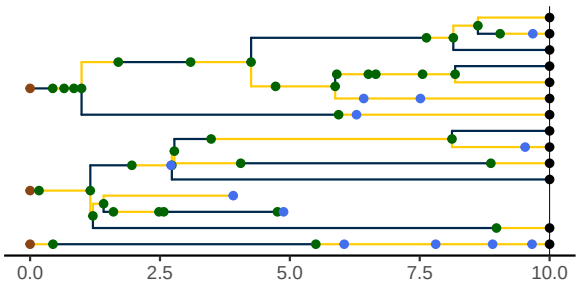

FIGURE 3. A genealogy, G, specifies the relationships of shared ancestry (via its tree-structure) and deme occupancy histories (via the coloring of its branches) of a set of lineages extant at some time $\mathsf{t}(G)$, as well as some samples gathered at earlier times. Here, $\mathsf{t}(G) = 10$ and 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.

Linear birth-death model. The linear birth-death process, a mainstay of existing phylodynamic methods, is 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 $n(X_t) = X_t$.

Moran model and the Kingman coalescent. The (Kingman, 1982) coalescent is another workhorse in existing phylodynamic approaches. 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.

2.8. History. 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 think of $\mathbf{H}_t$ as consisting of 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:

$$\mathsf{Prob}\left[\mathrm{dH}_t\right] = p_0(\hat{\mathrm{X}}_0)\,\mathrm{d}\hat{\mathrm{X}}_0 \prod_{k=1}^{K_t} \alpha_{\hat{\mathrm{U}}_k}\left(\hat{\mathrm{T}}_k, \hat{\mathrm{X}}_{k-1}, \hat{\mathrm{X}}_k\right)\,\mathrm{d}\hat{\mathrm{X}}_k\,\mathrm{d}\hat{\mathrm{T}}_k\,\exp\left(-\sum_{k=0}^{K_t}\int_{\hat{\mathrm{T}}_k}^{\hat{\mathrm{T}}_{k+1}}\sum_u \alpha_u(t', \hat{\mathrm{X}}_k, x')\,\mathrm{d}x'\,\mathrm{d}t'\right), \tag{5}$$

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

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

2.9. 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 tree, 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 and let $\mathsf{partit}(\mathbb{L})$ be the set of all collections of finite, mutually-disjoint subsets of $\mathbb{L}$. That is, an element $\zeta \in \mathsf{partit}(\mathbb{L})$ is a partition of the finite set $\bigcup \zeta \subseteq \mathbb{L}$. Partition *fineness* defines a partial order on $\mathsf{partit}(\mathbb{L})$. Specifically, for $\zeta, \zeta' \in \mathsf{partit}(\mathbb{L})$, we say $\zeta \preccurlyeq \zeta'$ if and only if for every $b' \in \zeta'$ there is $b \in \zeta$ such that $b \supseteq b'$. The tree structure of G is defined by 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 inline and root 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 $\mathsf{Y}_t(Z) \coloneqq \{Y_t \mid Y \in \mathsf{Y}(Z)\}$. 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 a genealogy $G = (\mathsf{t}(G), G^{\mathsf{Z}}, G^{\mathsf{Y}})$.

2.10. Binomial ratio. For $n, r, \ell, s \in \mathbb{Z}_+^{\mathbb{D}}$, define the *binomial ratio*

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

Observe that $\binom{n \quad \ell}{r \quad s} \in [0, 1]$. Moreover, in consequence of the Chu-Vandermonde identity, we have

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

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

3. The induced genealogy process.

3.1. Event types. We now show how a given population process naturally induces a process in the space of genealogies. Specifically, 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, each from some 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.
(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 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. We permit multiple lineages to move simultaneously.
(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.
(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). The theory presented here places few restrictions on the complexity of the events that can occur by combining events of the various pure types.

3.2. 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, 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 demes. 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 nodes 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.

3.2.1. Emergent lineages and production. The lineages which descend from an inserted node 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

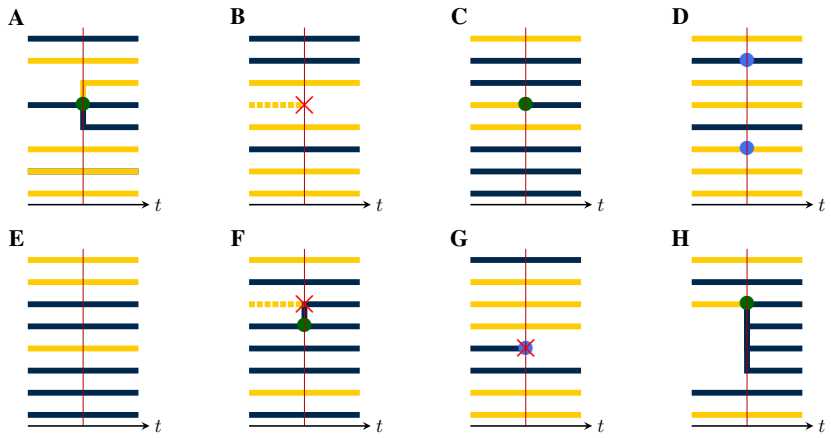

FIGURE 4. Event types differ by their effects on the genealogy. This can be seen by examining the local structure of the genealogy in the neighborhood of a jump. **(A)** A birth-type jump results in the branching of one or more child lineages from the parent. There can be only one parent, though the 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 $r = (r_{\mathsf{blue}}, r_{\mathsf{yellow}}) = (2, 1)$. 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 $n = (n_{\mathsf{blue}}, n_{\mathsf{yellow}}) = (3, 5)$. **(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 $n = (6, 2)$.

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.

3.2.2. 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 $\hat{\mathbf{G}}_k$ 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.

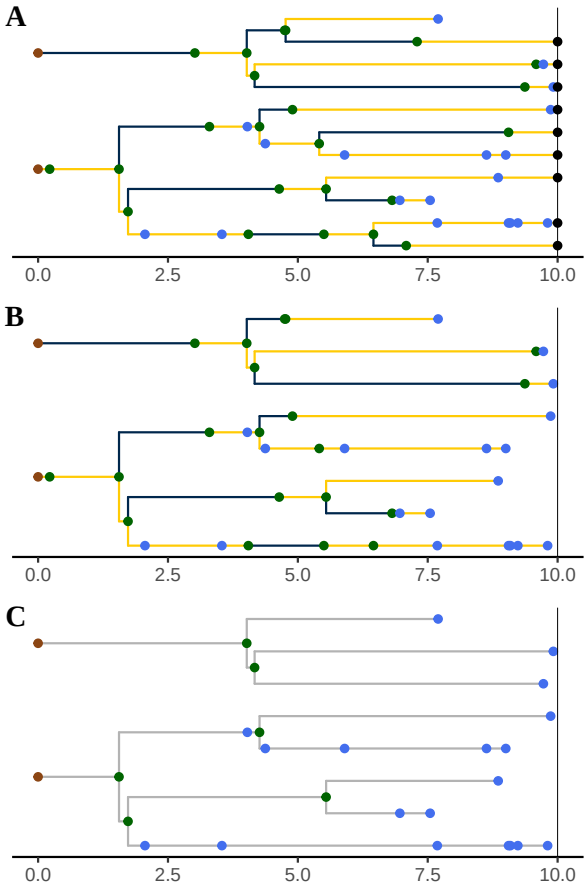

FIGURE 5. Unpruned, pruned, and obscured genealogies from a single realization of the genealogy process induced by the SEIRS model depicted in Figs. 1 and 2. **(A)** A realization of the unpruned genealogy process $\mathbf{G}_t$ is shown at $t = 10$. Tip nodes, corresponding to lineages alive at time $t = 10$ 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 (§§2.9, 3.3.1, and 3.3.4) for more detail.

3.3. 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).

3.3.1. Pruned genealogy. Given a genealogy G, one obtains the *pruned genealogy*, $P = \mathsf{prune}(G)$ 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 un-pruned genealogy is.

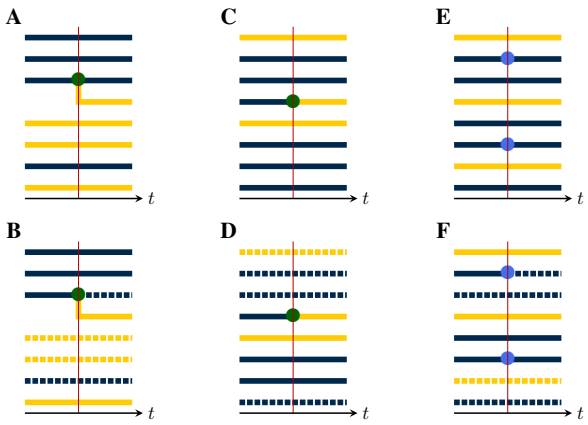

FIGURE 6. **Lineage count and saturation.** Each panel shows the neighborhood of a single event in the unpruned genealogy (top row) and the corresponding pruned genealogy (bottom row). Pruning consists of the removal of all branches that are not ancestral to some sample. In the bottom row of panels, 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.

Finally, observe that, 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.

3.3.2. 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.

3.3.3. Compatibility. Suppose P is a pruned genealogy, with $\mathsf{t}(P) = T$ and $t \in \mathsf{ev}(P)$. The local structure of P at t is, in general, compatible with only a subset of the possible jumps $\mathbb{U}$. For example, if the event in P at t is a branch node or a sample node, then it is compatible only with birth-type or sample-type jumps, respectively. Similarly, if the node in P at time t is one at which a lineage moves from deme i to deme i', then u must be either of $i \to i'$ migration type or of a birth type with parent in i and $r^u_{i'} > 0$. 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 after t—is compatible with an event of type u and $Q = 0$ otherwise. That is, $Q_u(\eta, \eta') = 1$ if and only if there is a feasible genealogy, $\mathrm{G} = (\mathrm{T}, \mathrm{Z}, \mathrm{Y})$, and history, H, and a $t \in [0, \mathrm{T}]$

such that, given $\mathbf{G}_{\mathrm{T}} = \mathrm{G}$ and $\mathbf{H}_{\mathrm{T}} = \mathrm{H}$, we have $\mathrm{U}_t = u$, $\widetilde{\mathrm{Y}}_t = \eta$, and $\mathrm{Y}_t = \eta'$. We refer to Q as the *compatibility indicator*.

3.3.4. 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.

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* $\mathrm{P} = (\mathrm{T}, \mathrm{Z}, \mathrm{Y})$ *is a given pruned genealogy. Define*

$$\phi_u(x, y, y') \coloneqq \begin{pmatrix} n(x) & \ell(y') \\ r^u & s(y, y') \end{pmatrix} Q_u(y, y'), \tag{6}$$

where n *is the deme occupancy (§2.6),* r^u *is the production (§3.2.1),* ℓ *and* s *are the lineage-count and saturation functions, respectively (§3.3.2),* Q *is the compatibility indicator (§3.3.3), and the binomial ratio is as defined in §2.10. Then*

$$\mathsf{Prob}\left[\mathbf{P}_{\mathrm{T}} = \mathrm{P} \mid \mathbf{H}_{\mathrm{T}} = \mathrm{H}\right] = \mathbb{1}\{\mathsf{ev}(\mathrm{H}) \supseteq \mathsf{ev}(\mathrm{P})\} \prod_{t \in \mathsf{ev}(\mathrm{H})} \phi_{\mathrm{U}_t}(\mathrm{X}_t, \widetilde{\mathrm{Y}}_t, \mathrm{Y}_t).$$

Proof. If $\mathsf{ev}(\mathrm{H}) \not\supseteq \mathsf{ev}(\mathrm{P})$, then H and P are incompatible and $\mathsf{Prob}\left[\mathbf{P}_{\mathrm{T}} = \mathrm{P} \mid \mathbf{H}_{\mathrm{T}} = \mathrm{H}\right] = 0$. Similarly, if any event of H is incompatible with the local structure of P in the sense of §3.3.3, then $\mathsf{Prob}\left[\mathbf{P}_{\mathrm{T}} = \mathrm{P} \mid \mathbf{H}_{\mathrm{T}} = \mathrm{H}\right] = 0$. Let us therefore suppose that neither of these conditions hold. Conditional on $\mathbf{H}_{\mathrm{T}} = \mathrm{H}$, at each time $t \in \mathsf{ev}(\mathrm{H})$, a jump of mark U_t occured, 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 our lineages are labelled, 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 $\mathbf{H}_{\mathrm{T}} = \mathrm{H}$, 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 $\mathsf{ev}(\mathrm{P}) \not\subseteq \mathsf{ev}(\mathrm{H})$, we obtain the result. □

Next, we show how the likelihood of a pruned genealogies, unconditional on the history, can be computed. For this, we use the filter equation technology developed in Appendix A. In particular, the following theorem follows immediately from Lemma A2.

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

$$\begin{aligned} \frac{\partial w}{\partial t}(t,x) &= \sum_u \int w(t,x')\,\alpha_u(t,x',x)\,\phi_u(x,\widetilde{\mathrm{Y}}_t,\mathrm{Y}_t)\,\mathrm{d}x' - \sum_u \int w(t,x)\,\alpha_u(t,x,x')\,\mathrm{d}x', \quad t \notin \mathsf{ev}(\mathrm{P}), \\ w(t,x) &= \sum_u \int \widetilde{w}(t,x')\,\alpha_u(t,x',x)\,\phi_u(x,\widetilde{\mathrm{Y}}_t,\mathrm{Y}_t)\,\mathrm{d}x', \qquad\qquad\qquad\qquad\qquad\quad t \in \mathsf{ev}(\mathrm{P}), \end{aligned} \tag{7}$$

where ϕ *is defined in Eq. 6. Then the likelihood of* P *is*

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

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

Theorem 3. *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* $y \in \mathsf{Y}_0(\mathrm{Z})$,

$$q(x,y) \geqslant 0, \qquad \sum_{y \in \mathsf{Y}_0(\mathrm{Z})} q(x,y) = 1,$$

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

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

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* $\mathsf{Prob}\left[\mathbf{P}_0^{\mathsf{Y}} = y \,\middle|\, \mathbf{X}_0 = x\right] > 0$. *Then there is a stochastic jump process* $\mathbf{y}_t$ *with sample paths in* $\mathsf{Y}(\mathrm{Z})$ *such that* $(\mathbf{X}_t, \mathbf{U}_t, \mathbf{y}_t)$ *is Markov and*

$$\mathsf{Prob}\left[\mathbf{P}_{\mathrm{T}}^{\mathsf{Z}} = \mathrm{Z} \,\middle|\, \mathbf{H}_{\mathrm{T}} = \mathrm{H}\right] = \mathbb{1}\{\mathsf{ev}(\mathrm{H}) \supseteq \mathsf{ev}(\mathrm{Z})\}\,\mathbb{E}\left[\frac{1}{q(X_0,\mathbf{y}_0)} \prod_{t \in \mathsf{ev}(\mathrm{H})} \frac{\phi_{\mathrm{U}_t}(\mathrm{X}_t,\widetilde{\mathbf{y}}_t,\mathbf{y}_t)}{\pi_{\mathrm{U}_t}(t,\widetilde{\mathrm{X}}_t,\mathrm{X}_t,\widetilde{\mathbf{y}}_t,\mathbf{y}_t)}\right],$$

where ϕ *is defined in Eq. 6 and the expectation is taken over the sample paths of* $\mathbf{y}_t$.

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

$$\mathsf{Prob}\left[\mathbf{P}_{\mathrm{T}}^{\mathsf{Z}} = \mathrm{Z} \,\middle|\, \mathbf{H}_{\mathrm{T}} = \mathrm{H}\right] = \mathbb{E}\left[\mathbb{1}\{\mathbf{P}_{\mathrm{T}}^{\mathsf{Z}} = \mathrm{Z}\} \,\middle|\, \mathbf{H}_{\mathrm{T}} = \mathrm{H}\right]. \tag{8}$$

Our strategy will be to evaluate Eq. 8 using importance sampling: 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. 8 by summing over these paths. Conditional on $\mathbf{H}_{\mathrm{T}} = \mathrm{H}$, the initial distribution q and probability kernel π generate a Markov chain, $\hat{\mathbf{y}}_k$ such that

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

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

Now, for $\mathrm{y} \in \mathsf{Y}(\mathrm{Z})$, let us define $C(\mathrm{y}) = (\mathrm{T}, \mathrm{Z}, \mathrm{y})$. Then, by construction, $\mathsf{obs}(C(\mathrm{y})) = (\mathrm{T}, \mathrm{Z})$ and, conversely, for every pruned genealogy P satisfying $\mathsf{t}(\mathrm{P}) = \mathrm{T}$ and $\mathrm{P}^{\mathsf{Z}} = \mathrm{Z}$, $C(\mathrm{P}^{\mathsf{Y}}) = \mathrm{P}$. 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 $\mathsf{Prob}\left[\mathbf{y} = \mathrm{P}^{\mathsf{Y}} \,\middle|\, \mathbf{H}_{\mathrm{T}} = \mathrm{H}\right] > 0$. 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. □

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 3 and Lemma A2.

Theorem 4. *Let $V = (\mathrm{T}, \mathrm{Z})$ be a given obscured genealogy. Then there are probability kernels q and π as in Theorem 3 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

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

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

then the likelihood of V is

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

Lemma A3 shows how this can be computed via Sequential Monte Carlo.

5. Discussion.

Importantly, because the theory presented here 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 4 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*—expands the set of population models that can be confronted with data still further.

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 a property of 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 4, 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 in an inferential exercise.

The theory presented here represents a strict generalization of the existing coalescent and birth-death process approaches to phylodynamic inference. Both of the latter processes represent special cases of the genealogical processes constructed here. In Appendix B, this is demonstrated.

A limitation of the theory presented here is that the population models are assumed to be pure jump processes, which allows consideration of demographic stochasticity and some kinds of environmental stochasticity. It should be possible to incorporate of the full range of Markovian environmental stochasticity via extension of this theory to population models containing both diffusion and jump components.

The price of the theory's flexibility is primarily computational: Monte Carlo variability in the likelihood estimate depends on computational effort in the usual way. However, the substantial freedom one has in the choice of the

importance-sampling distribution π (Theorem 4) can be exploited to diminish Monte Carlo error. 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.

This work was supported by grants from the U.S. National Institutes of Health, (Grant #1R01AI143852 to AAK, #1U54GM111274 to AAK and ELI) and a grant from the Interface program, jointly operated by the U.S. National Science Foundation and the National Institutes of Health (Grant #1761603 to ELI and AAK). QL acknowledges the support of the Michigan Institute for Data Science.

References.

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

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.
https://doi.org/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.
https://doi.org/10.1016/j.jeconom.2017.11.011

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.
https://doi.org/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.
https://doi.org/10.1098/rstb.1994.0079

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

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

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.
https://doi.org/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.
https://doi.org/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.
https://doi.org/10.1093/sysbio/syab049

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.
https://doi.org/10.1073/pnas.1713314115

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

Ogata, Y. (1978) The asymptotic behaviour of maximum likelihood estimators for stationary point processes. *Annals of the Institute of Statistical Mathematics* **30**:243–261.
https://doi.org/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. https://doi.org/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. https://doi.org/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. https://doi.org/10.1371/journal.pcbi.1003570

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

Volz, E. M. (2012) Complex population dynamics and the coalescent under neutrality. *Genetics* **190**:187–201. https://doi.org/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. https://doi.org/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. https://doi.org/10.1534/genetics.109.106021

Volz, E. M. & Siveroni, I. (2018) Bayesian phylodynamic inference with complex models. *PLoS Computational Biology* **14**:e1006546. https://doi.org/10.1371/journal.pcbi.1006546

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

Appendix A. Filter equations.

The likelihoods that appear in Theorems 2 and 4 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 *filtering 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, 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

$$\frac{\partial u}{\partial t}(t,x) = \int u(t,x')\,\beta(t,x',x)\,\mathrm{d}x' - \int u(t,x)\,\beta(t,x,x')\,\mathrm{d}x'. \tag{A1}$$

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 countable and locally finite (i.e., $S \cap [0,t]$ is finite for all $t > 0$). Then the system of equations

$$\frac{\partial w}{\partial t}(t,x) = \int w(t,x')\,\beta(t,x',x)\,B(t,x',x)\,\mathrm{d}x' - \int w(t,x)\,\beta(t,x,x')\,\mathrm{d}x' - \lambda(t,x)\,w(t,x), \qquad t \notin S, \tag{A2}$$

$$w(t,x) = \int \tilde{w}(t,x')\,\beta(t,x',x)\,B(t,x',x)\,\mathrm{d}x', \qquad t \in S, \tag{A3}$$

is called the *filter equation generated by* β, with *boost* B, *decay* λ, and *observation times* S. The process $\mathbf{X}_t$ is said to be the *driver* of the filter equation. Eq. A2 is the *regular part* of the filter equation; Eq. A3 is known as the *singular part*.

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

The following results show how filter equations allow one to integrate over random histories. First, Lemma A1 shows how one integrates over the full space of histories using a regular filter equation. Lemma A2 builds on this when the set of histories is restricted.

Lemma A1. *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,

$$\frac{\partial w}{\partial t} = \int w(t,x')\,\alpha(t,x',x)\,B(t,x',x)\,\mathrm{d}x' - \int w(t,x)\,\alpha(t,x,x')\,\mathrm{d}x'. \tag{A4}$$

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 $\Delta > 0$ sufficiently small, the expectation can be broken into three terms, according to whether H_t has zero, one, or more than one event in $(t-\Delta, t]$. Accordingly, as $\Delta \downarrow 0$,

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

In the limit, we obtain Eq. A4, the regular filtering equation generated by α, with boost B and zero decay. □

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 $s_t \coloneqq s \cap [0,t]$. Thus if $s \in \mathbb{S}$ and $s_t = (\hat{s}_1, \ldots, \hat{s}_K)$, 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 A2. *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*

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

Let λ_t be a family 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

$$\frac{\partial w}{\partial t}(t,x,s) = \int w(t,x',s)\,\alpha(t,x',x)\,B(t,x',x)\,\mathrm{d}x' - \int w(t,x,s)\,\alpha(t,x,x')\,\mathrm{d}x', \qquad t \notin s, \tag{A5}$$

$$w(t,x,s) = \int \widetilde{w}(t,x',s)\,\alpha(t,x',x)\,B(t,x',x)\,\mathrm{d}x', \qquad t \in s. \tag{A6}$$

Proof. The proof proceeds by induction on the cardinality of s_t. The base case, for which $s_t = \varnothing$, follows immediately from Lemma A1. Assuming that it holds for $|s_t| < K$, one has only to verify Eq. A6. This can be accomplished by integrating Eq. 5 directly. □

Remark. In the same way that Eqs. 3 and 4 can be represented as a single equation by means of a Dirac delta notation, the Eqs. A5 and A6 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.

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

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

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

Lemma A3. *Eqs. A2 and A3 are satisfied by* $w(t,x) = \int_0^\infty v\,u(t,x,v)\,\mathrm{d}v$*, where* $u(t,x,v)$ *satisfies the KFE*

$$\begin{aligned}
\frac{\partial u}{\partial t}(t,x,v) =& \frac{\partial}{\partial v}\left[\lambda(t,x)\,v\,u(t,x,v)\right] + \int_0^\infty \int u(t,x',v')\,\beta(t,x',x)\,\delta\big(v, B(t,x',x)\,v'\big)\,\mathrm{d}x'\,\mathrm{d}v' \\
&- \int_0^\infty \int u(t,x,v)\,\beta(t,x,x')\,\delta\big(v', B(t,x,x')\,v\big)\,\mathrm{d}x'\,\mathrm{d}v', && t \notin S, \qquad \text{(A7)} \\
u(t,x,v)\,\mathrm{d}x =& \int_0^\infty \int \tilde{u}(t,x',v')\,\pi(t,x',\mathrm{d}x)\,\delta\big(v, A(t,x')\,B(t,x',x)\,v'\big)\,\mathrm{d}x'\,\mathrm{d}v', && t \in S.
\end{aligned}$$

Here, $A(t,x) \coloneqq \int \beta(t,x,x')\,\mathrm{d}x'$ *and* $\pi(t,x,\mathrm{d}x') \coloneqq \beta(t,x,x')\,\mathrm{d}x'/A(t,x)$.

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

$$\begin{aligned}
\frac{\partial w}{\partial t}(t,x) &= \int_0^\infty v\,\frac{\partial u}{\partial t}(t,x,v)\,\mathrm{d}v \\
&= \int_0^\infty \int \int_0^\infty v\,u(t,x',v')\,\beta(t,x',x)\,\delta\big(v, B(t,x',x)\,v'\big)\,\mathrm{d}v\,\mathrm{d}x'\,\mathrm{d}v' \\
&\quad - \int_0^\infty \int \int_0^\infty v\,u(t,x,v)\,\beta(t,x,x')\,\delta\big(v', B(t,x,x')\,v\big)\,\mathrm{d}v\,\mathrm{d}x'\,\mathrm{d}v' \\
&\quad + \int_0^\infty v\,\tfrac{\partial}{\partial v}\left[\lambda(t,x)\,v\,u(t,x,v)\right]\mathrm{d}v.
\end{aligned}$$

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

$$\begin{aligned}
\frac{\partial w}{\partial t}(t,x) =& \int v'\,u(t,x',v')\,\beta(t,x',x)\,B(t,x',x)\,\mathrm{d}v'\,\mathrm{d}x' - \int v\,u(t,x,v)\,\beta(t,x,x')\,\mathrm{d}v\,\mathrm{d}x' \\
&- \lambda(t,x)\int v\,u(t,x,v)\,\mathrm{d}v,
\end{aligned}$$

which is simplified to obtain Eq. A2. Similarly, at each $t \in S$, we have

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

which is equivalent to Eq. A3 □

Remark. Eqs. A7 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. A1. $\mathbf{V}_t$ 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 A3 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 A3 shows how to integrate Eqs. A5 and A6 in the Monte Carlo sense.

Appendix B. Examples.

B.1. SIRS model. King et al. (2022) worked out formulas 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

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

Here $N = S + I + R$ is the host population size. Note that we have here allowed for the possibility that the transmission rate, β, depends on time.

B.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. It has two demes ($\mathbb{D} = \{\mathsf{E}, \mathsf{I}\}$) and the KFE

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

where $N = S+E+I+R$ is the total population size. The deme occupancy function in this case is $n(x) = (E, I)$. Note that the terms associated with sampling cancel each other in the KFE, since, in this model, sampling has no effect on the state.

B.3. Two-strain competition model. A simple model for the competition of two strains for susceptible hosts is depicted in Fig. 1B. This example will be completed in a forthcoming draft.

B.4. Superspreading model. Fig. 1D depicts a model of superspreading. This example will be completed in a forthcoming draft.

B.5. Linear birth-death model. In this model, the state variable is the size, N_t, of a population at time t. All individuals face the same per-capita birth and death rates, which are λ and μ, respectively. The KFE is

$$\frac{\partial v}{\partial t} = \lambda\,(n-1)\,v(t,n-1) - \lambda\,n\,v(t,n) + \mu\,(n+1)\,v(t,n+1) - \mu\,n\,v(t,n)$$

Stadler (2010) considered the case where samples are taken through time at a uniform per-capita rate ψ. In this case, since there is only one deme, in the filter equation, w can be taken to be independent of y. If B is the set of branch-times and S_0, S_1 are the sets of terminal and inline samples, respectively, then the regular part of the filter equation is

$$\begin{aligned}\frac{\partial w}{\partial t}(t,n) =& \lambda\,(n-1)\left(1 - \frac{\binom{\ell(t)}{2}}{\binom{n}{2}}\right) w(t,n-1) - \lambda\,n\,w(t,n)\\ &+\mu\,(n+1)\,w(t,n+1) - \mu\,n\,w(t,n) - \psi\,n\,w(t,n), \quad n \geqslant \ell(t), \quad t \notin B \cup S_0 \cup S_1\end{aligned} \tag{B8}$$

and the singular part is

$$\begin{aligned}w(t,n) &= \frac{\lambda\,(n-1)}{\binom{n}{2}}\,\widetilde{w}(t,n-1), \quad t \in B,\\ w(t,n) = \psi\,n\left(1 - \frac{\ell(t)}{n}\right)\widetilde{w}(t,n), \quad t \in S_0, &\qquad w(t,n) = \psi\,\widetilde{w}(t,n), \quad t \in S_1.\end{aligned} \tag{B9}$$

Eqs. B8 and B9 are supplemented by the ancillary condition $w(t,n) = 0$ for $n < \ell(t)$.

B.6. Moran model and the Kingman coalescent. In the Moran model, events occur according to a rate-μ Poisson process. At each event, a compound birth-death jump (cf. Fig. 4F) occurs so that the population size, n, remains constant. If we let X_t be the number of events that have occurred by time t, then X_t is a simple counting

process, which we can use to define the state of the population process. Its KFE is then

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

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

In the classical case (Kingman, 1982), 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, the filter equation reads

$$w(0) = 1, \qquad \frac{\partial w}{\partial t} = \mu\,w(t)\left(1 - \frac{\binom{\ell(t)}{2}}{\binom{n}{2}}\right) - \mu\,w(t), \quad t \notin B, \qquad w(t) = \frac{\mu}{\binom{n}{2}}\,\widetilde{w}(t), \quad t \in B. \tag{B10}$$

Integrating Eqs. B10 and taking logarithms yields

$$\log w(T) = k\,\log\frac{\mu}{\binom{n}{2}} - \frac{\mu}{\binom{n}{2}} \sum_{i=m-k}^{m} \binom{i}{2}\,s_i, \tag{B11}$$

where $k = |B|$ is the number of branch-points in $[0,T]$ and the $s_i \coloneqq \int \mathbb{1}\{\ell(t) = i\}\,\mathrm{d}t$ are the durations of the *coalescent intervals*, i.e., intervals between successive branch-points. We recognize Eq. B11 as the expression for the Kingman (1982) coalescent (e.g., Wakeley, 2009).

More generally, if in addition samples are taken according to a rate-ν 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

$$\begin{gathered} w(0) = 1, \qquad \frac{\partial w}{\partial t} = -\mu\,\frac{\binom{\ell(t)}{2}}{\binom{n}{2}}\,w(t), \quad t \notin S \cup B, \qquad w(t) = \frac{\mu}{\binom{n}{2}}\,\widetilde{w}(t), \quad t \in B, \\ w(t) = \nu\left(1 - \frac{\ell(t)}{n}\right)\widetilde{w}(t), \quad t \in S_0, \qquad w(t) = \frac{\nu}{n}\,\widetilde{w}(t), \quad t \in S_1. \end{gathered} \tag{B12}$$

Integrating Eqs. B12 yields

$$\log w(T) - |S|\,\log\nu = \sum_{t\in S_0} \log\left(1 - \frac{\ell(t)}{n}\right) - |S_1|\,\log n + |B|\,\log\frac{\mu}{\binom{n}{2}} - \frac{\mu}{\binom{n}{2}} \sum_{i=1}^{\infty} \binom{i}{2}\,s_i. \tag{B13}$$

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