MARKOV GENEALOGY PROCESSES

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

ABSTRACT. We construct a family of genealogy-valued Markov processes that are induced by a continuous-time Markov population process. We derive exact expressions for the likelihood of a given genealogy conditional on the history of the underlying population process. These lead to a nonlinear filtering equation which can be used to design efficient Monte Carlo inference algorithms. We demonstrate these calculations with several examples. Existing full-information approaches for phylodynamic inference are special cases of the theory.

1. Introduction.

Phylodynamics is the study of the traces left by the population dynamics of biological organisms in the genomes of their descendants, specifically, in the patterns of genealogical or phylogenetic relatedness among organisms or groups of organisms. The term was first coined with reference to the processes of transmission and pathogen evolution in the context of infectious disease (Grenfell et al., 2004; Frost et al., 2015) and to date many fruitful applications have been in this area (e.g., Koelle et al., 2006; O'Dea & Wilke, 2011; Volz et al., 2013a; Rasmussen et al., 2014; Alizon et al., 2014; Faria et al., 2014; du Plessis & Stadler, 2015; Geoghegan et al., 2015; Vijaykrishna et al., 2015; Biek et al., 2015; Smith et al., 2017; Li et al., 2017; Hadfield et al., 2018; Bedford et al., 2020; Ragonnet-Cronin et al., 2021). In typical applications, one wishes to infer the form and parameterization of a model of pathogen transmission on the basis of information contained in pathogen genome sequences. Similar problems arise outside the realm of infectious disease biology, for example in systematics, comparative biology, cancer, microbiology, and population genetics (Maddison et al., 2007; Gill et al., 2016; Stadler et al., 2021; MacPherson et al., 2021).

A central technical problem in phylodynamics is the precise characterization of the relationship between stochastic population processes and the traces they leave in genomes obtained from a sample of the population. This in turn is usually factored into two subproblems: (i) the relationship between genome sequences and the genealogies or phylogenies that relate them and (ii) the relationship between these genealogies and the population processes that generate them. In this paper, we focus attention on the second subproblem. In particular, we suppose that we have one or more stochastic models that describe the dynamics of infections and recoveries (or births and deaths, or speciations and extinctions) in a population and we desire to estimate parameters and compare these models using data in the form of genealogies that relate sampled infections, individuals, or species.

At present, three broad approaches to the solution of this problem exist. The first compares the reconstructed genealogy with simulated genealogies on the basis of summary statistics designed to capture the important features of the genealogy. Such approaches can be used to estimate parameters and compare models, for example via approximate Bayesian computation (Sisson et al., 2007; Luciani et al., 2009; Ratmann et al., 2012; Poon, 2015) or synthetic likelihood (Wood, 2010; Fasiolo et al., 2016). These methods have the advantage of being applicable for essentially arbitrary models, but cast away some of the information contained in the data. It can be difficult to quantify the amount of information discarded

Date: February 15, 2022.

and to design summary statistics that minimize this information loss. By contrast, the second and third approaches are *full-information* methods, in the sense that they are based on the likelihood. The approach pioneered by Volz, Koelle, and colleagues (Volz et al., 2009; Rasmussen et al., 2011; Koelle & Rasmussen, 2012; Volz et al., 2013b; Dearlove & Wilson, 2013) is based on the Kingman (1982a,b,c) coalescent, which yields the distribution of genealogies as a function of a time-varying *coalescent rate*. These approaches rely on the assumption that the population size is large and the sample size small, so that branching points in the sample lineages are approximately independent of those in the population. Rigorous quantification of the error associated with this approximation exists only for special cases (Fu, 2006). The third approach, associated with Stadler and colleagues (Gernhard, 2008; Stadler, 2010; Stadler et al., 2012; Boskova et al., 2014; Kühnert et al., 2014; MacPherson et al., 2021), is based on birth-death processes. In the case of the linear birth-death process, exact expressions can be obtained, but approximations must be employed to deal with model nonlinearity. In particular, reverse-time arguments that go through in the linear case fail in the nonlinear case.

Recently, Etheridge & Kurtz (2019) described an approach similar in objective to the one described here, but based instead on Fleming-Viot processes and lookdown constructions (Donnelly & Kurtz, 1996, 1999). More closely related to our approach is the work of Vaughan et al. (2019), who recently devised an algorithm for exact phylodynamic likelihood computation for simple jump processes under certain assumptions of time-homogeneity. The arguments in this paper put this algorithm on a firm footing and place it in a much broader context, allowing consideration of a wider class of models and laying the groundwork for more efficient algorithms. We return to this issue in §7.

In this paper, we define the notion of the *genealogy process* that is induced by a population model, i.e., a model of the births and deaths occurring in a specified population. In the general case, these births and deaths will be stochastic, which implies that the induced genealogy process will itself be a stochastic process on the space of genealogies. Accordingly, we are interested in characterizing the probabilistic properties of this process. In practice, information about any real population is obtained through genomic sampling, and we will therefore also be interested in the properties of partial genealogies that relate the samples.

As for the population models, we will restrict our attention to Markov processes. In practice, this is not a major restriction, as most of the models of scientific interest can be readily formulated as, or approximated by, Markov processes. To make the notion of a genealogy definite, it is necessary to conceive of births and deaths as discrete events. It is natural, therefore, to further restrict our attention to Markov jump processes, i.e., continuous-time processes for which births and deaths occur at random times. The approach we describe here can be generalized to a somewhat broader category of Markov processes, but the novel mathematical constructions are already fairly complex, so we postpone these generalizations to a future contribution. In particular, we simplify matters by confining ourselves to the case where birth, death, and sampling events occur one at a time almost surely. We note that some models of theoretical and practical interest do violate this assumption and that our approach can be generalized to accommodate such models. Nevertheless, we defer consideration of these interesting complexities to a sequel.

The remainder of the paper is structured as follows. In §2, we lay mathematical groundwork by constructing a probability space within which we can speak of the genealogies induced by a given population process. In §3, we give several examples, both of processes amenable to treatment within this framework, and of models that motivate further theoretical development. Next, in §4, we define several related Markov genealogy processes induced by a given population model. With these definitions in hand, in §5 we derive our main results. There, we exhibit explicit expressions for the likelihood of an observed genealogy and derive the analogue of the Duncan-Mortensen-Zakai (DMZ) or nonlinear

filtering equation for these processes. In §6, we revisit some of the examples of §3 to demonstrate these likelihood computations more concretely. Finally, in §7, we indicate some of the implications for phylodynamics broadly. In particular, we note that both major existing approaches for likelihood-based phylodynamic inference are special cases. We also point out that, from the DMZ equation, a straight road leads to efficient Monte Carlo inference algorithms capable of accommodating a broader class of models than has heretofore been susceptible to analysis.

2. Mathematical preliminaries.

Markov jump processes on the integer lattice. Suppose we have a non-explosive Markov jump process $\mathcal{X}_t \in \mathbb{Z}^d$, parameterized by $t \in \mathbb{R}_+$, which we think of as a model of the random time-evolution of some kind of population. Henceforth, we refer to $\mathcal{X}_t$ as a *population process*. It is defined by its initial-state distribution and its generator. In particular, we suppose that

$$\mathbb{P}\left[\mathcal{X}_0 = x\right] = p_0(x), \tag{1}$$

for some choice of initial distribution, p_0. The transitions of $\mathcal{X}_t$ are governed by event-rate functions $\alpha_u(t,x) \in \mathbb{R}_+$, for $u, x \in \mathbb{Z}^d$, $t \in \mathbb{R}_+$. In particular, $\alpha_u(t,x)$ is the hazard of a jump from state x to state $x+u$ at time t. These conditions imply that, for any $f \in L^\infty(\mathbb{Z}^d)$, if $F(s,x) \coloneqq \mathbb{E}\left[f(\mathcal{X}_t) \mid \mathcal{X}_s = x\right]$, then

$$\frac{\partial F}{\partial s}(s,x) = -\sum_{u\in\mathbb{Z}^d} \alpha_u(s,x)\,\left[F(s,x+u) - F(s,x)\right], \qquad x \in \mathbb{Z}^d. \tag{2}$$

If we moreover assume that the sample paths of $\mathcal{X}_t$ are right-continuous with left limits (càdlàg), then Eqs. 1 and 2 completely specify $\mathcal{X}_t$. The assumption that $\mathcal{X}_t$ is non-explosive, and the further requirement we make that $\sum_u \alpha_u(t,x) < \infty$ for all t and x restricts the class of allowable rate functions α.

The adjoint of Eq. 2 is the Kolmogorov forward equation (sometimes called the *master equation* in this context):

$$\begin{aligned} \frac{\partial w}{\partial t}(t,x) &= \sum_{u\in\mathbb{Z}^d} \alpha_u(t,x-u)\,w(t,x-u) - \alpha_u(t,x)\,w(t,x), \qquad x \in \mathbb{Z}^d, \\ w(0,x) &= p_0(x). \end{aligned} \tag{3}$$

If $w(t,x)$ satisfies Eqs. 3, then $w(t,x) = \mathbb{P}\left[\mathcal{X}_t = x\right]$.

Definitions. Our goal in this paper is to introduce a family of Markov processes induced by the jump process just described (Fig. 1). While population processes of the kind described above can be constructed in a variety of time-honored ways (e.g., Andersen et al., 1993; Kallenberg, 1997), these classical approaches are not entirely sufficient for the tree-valued Markov processes we will erect on top of the population process. We therefore explicitly construct the probability space that underlies the stochastic processes we subsequently describe. Unavoidably, this leads to technicalities that are necessary for the firm establishment of the properties of these processes but that may distract from the overarching goals. Readers willing to stipulate these properties may skim the remainder of this section, in which we formally construct the probability space and make some definitions that will be needed in the sequel.

Let us define a *jump* to be an ordered triple $(t,u,n) \in \mathbb{R}_+ \times \mathbb{Z}^d \times \mathbb{N}$. We refer to t as the *time* of the jump; u is the *type* of the jump; n is an *auxiliary number* whose use will be made clear below. A *jump sequence* is a countable sequence of jumps at increasing times. That is,

$$\omega = (t_k, u_k, n_k)_{k=0}^K$$

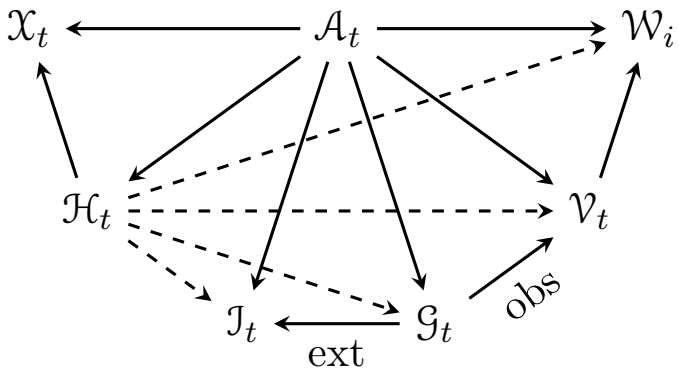

FIGURE 1. Relations among the various Markov processes discussed in the paper. Deterministic maps are indicated with solid arrows; random maps are shown as dashed arrows. All the maps shown commute. $\mathcal{X}_t$ is the *population process*, a model of the dynamics of some system, which we take as a starting point. $\mathcal{H}_t$ is the *history process*, which records the full history of $\mathcal{X}_t$. $\mathcal{I}_t$ is the *inventory process*: at each time t, $\mathcal{I}_t$ is an inventory of all extant individuals in the population, each of which has a globally unique name. $\mathcal{G}_t$ is the *genealogy process*, which captures the precise genealogical relationships among all individuals in $\mathcal{I}_t$, as well as among any samples that have been taken from the population. $\mathcal{V}_t$ is the *visible genealogy process*, which is $\mathcal{G}_t$ pruned so that only relationships among samples remain. Finally $\mathcal{W}_i$ is the *embedded chain of the visible genealogy process*, which is $\mathcal{V}_{s_i}$, s_i being the time of the i-th sample. All of these processes can be obtained via deterministic procedures applied to the *master process* $\mathcal{A}_t$, as described in the text.

is a jump sequence if and only if $K \in \mathbb{N} \cup \{\infty\}$, $u_k \in \mathbb{Z}^d$, $n_k \in \mathbb{N}$ for all k, and $0 = t_0 < t_1 < t_2 < \ldots$. We will take our sample space, Ω, to be the set of all jump sequences. For $\omega \in \Omega$ as above, we write

$$T_k(\omega) \coloneqq t_k, \qquad U_k(\omega) \coloneqq u_k, \qquad N_k(\omega) \coloneqq n_k, \qquad K(\omega) \coloneqq K \tag{4}$$

and also

$$T(\omega) \coloneqq (T_k(\omega))_{k=0}^{K(\omega)}, \qquad U(\omega) \coloneqq (U_k(\omega))_{k=0}^{K(\omega)}, \qquad N(\omega) \coloneqq (N_k(\omega))_{k=0}^{K(\omega)}. \tag{5}$$

We will denote by $\mathring{\Omega}$ the set of all finite jump sequences, i.e., $\mathring{\Omega} \coloneqq \{\omega \in \Omega \mid K(\omega) < \infty\}$.

Note that every element of Ω corresponds to a unique sample path of $\mathcal{X}_t$. In particular, with the convention that $U_0 = \mathcal{X}_0$, we will write

$$\mathcal{X}_t(\omega) = \sum_{k=0}^{K(\omega)} U_k(\omega)\, \mathbb{1}_{[T_k(\omega),\infty)}(t). \tag{6}$$

The sample space Ω has a natural partial order. We write $\omega \preceq \omega'$ if ω' is an extension of ω; that is, if $K(\omega) \leq K(\omega')$, and $(T_k(\omega), U_k(\omega), N_k(\omega)) = (T_k(\omega'), U_k(\omega'), N_k(\omega'))$ for $k = 0, \ldots, K(\omega)$. For $\omega \in \Omega$, the set $\{\omega' \in \Omega \mid \omega' \preceq \omega\}$ is totally ordered. Moreover, for each $\omega \in \mathring{\Omega}$, there is a unique *predecessor*, ω-, such that $\omega\text{-} \prec \omega$ and, for all ω', $\omega\text{-} \preceq \omega' \preceq \omega$ implies that either $\omega' = \omega\text{-}$ or $\omega' = \omega$.

In order to define probabilistic *events*, it is necessary to define the σ-algebra of measurable subsets of Ω. We topologize Ω using a standard approach that makes Ω separable, i.e., possessed of a countable dense subset. Specifically, let d_S represent the Skorokhod metric on the space of càdlàg functions $\mathbb{R}_+ \to \mathbb{Z}^d$

(Kallenberg, 1997; Ethier & Kurtz, 2009). Extend this to a metric d on Ω by

$$d(\omega, \omega') \coloneqq d_S(\mathcal{X}(\omega), \mathcal{X}(\omega')) + \sup_k \left| N_k(\omega) - N_k(\omega') \right|,$$

where the sample-path functions $\mathcal{X}_t$ are defined above (Eq. 6) and it is understood that $N_k(\omega) = 0$ for $k > K(\omega)$. It is straightforward to verify that this is indeed a metric and that, equipped with this metric, Ω is a complete, separable metric space, i.e., a so-called Polish space. We take our event space, $\mathcal{B}$, to be the Borel σ-algebra of Ω.

For $t \in \mathbb{R}_+$ and $\omega \in \Omega$, we define the *restriction*, $\omega|_t \in \Omega$, by

$$\omega|_t \coloneqq \max \left\{ \omega' \in \Omega \mid \omega' \preceq \omega \text{ and } T_{K(\omega')} \leq t \right\}.$$

and let $\Omega|_t$ be the space of t-restrictions. Note that each $\Omega|_t$ is a complete, separable metric space.

We now turn to the probability measure, $\mathbb{P}$, on $\Omega \cong T(\Omega) \times U(\Omega) \times N(\Omega)$. We will specify this by giving its density with respect to a base measure. A natural base measure on $T(\Omega)$ is the probability measure of the rate-μ Poisson point process (Andersen et al., 1993; Kallenberg, 1997); the counting measure serves for the discrete component $U(\Omega) \times N(\Omega)$. Let π_μ denote the product of these two measures. We define a probability measure on Ω by specifying its density with respect to π_μ. In particular, for each $u, x \in \mathbb{Z}^d$, we suppose $\beta_{u,x}$ is a given probability measure on $\mathbb{N}$; these will take specific forms below. For $t \in \mathbb{R}_+$ and $\omega \in \Omega|_t$, define the probability density function

$$\begin{aligned} P_t(\omega) \coloneqq\, & p_0(\mathcal{X}_0) \prod_{k=1}^{K} \left(\frac{1}{\mu}\, \alpha_{U_k}(T_k, \mathcal{X}_{T_k} - U_k)\, \beta_{U_k, \mathcal{X}_{T_k} - U_k}(N_k) \right) \\ & \times \exp\left(\mu\, t - \int_0^t \sum_{u \in \mathbb{Z}^d} \alpha_u(s, \mathcal{X}_s)\, \mathrm{d}s \right). \end{aligned} \tag{7}$$

Here, for the sake of readability, we have suppressed the dependence of the random variables $\mathcal{X}_t$, K, T_k, U_k and N_k on ω. It is readily verified that, for all t, P_t is a probability density with respect to π_μ. One can dispense with the dependence of Eq. 7 on the Poisson rate parameter μ. For example, one can always choose $\mu = 1$ and the factor of e^t, that remains can usually be neglected, it being a mere normalizing constant. However, we preserve the dependence on μ here to remind us of the manner in which the magnitude of P_t depends on the base measure, π_μ.

Master process. We now define a process that serves as the foundation for all that follows. For $t \in \mathbb{R}_+$ and $\omega \in \Omega$, let

$$\mathcal{A}_t(\omega) \coloneqq (t, \omega|_t). \tag{8}$$

We will refer to $\mathcal{A}_t$ as the *master process*. We use the expressions $t(\mathcal{A}_t)$ and $\omega(\mathcal{A}_t)$ to refer to the first and second elements of $\mathcal{A}_t$, respectively.

We define the densities, $P_{\mathcal{A}_{t_1}, \ldots, \mathcal{A}_{t_m}}$, of finite collections of random variables $\{\mathcal{A}_{t_1}, \ldots, \mathcal{A}_{t_m}\}$, $t_1 < t_2 < \cdots < t_m$, in terms of Eq. 7 by defining,

$$P_{\mathcal{A}_{t_1}, \ldots, \mathcal{A}_{t_m}}(a_1, \ldots, a_m) \coloneqq \begin{cases} P_{t_m}(\omega(a_m)), & \text{if } \omega(a_1) \preceq \omega(a_2) \preceq \cdots \preceq \omega(a_m), \\ 0, & \text{otherwise.} \end{cases} \tag{9}$$

Since the densities of Eq. 9 are projective and the state-space is Polish, the Kolmogorov Extension Theorem implies that there is a unique probability measure $\mathbb{P}$ on Ω with these finite-dimensional densities. Thus $(\Omega, \mathcal{B}, \mathbb{P})$ is a probability space. With these definitions, one readily verifies that the master process $\mathcal{A}_t$ is Markov and that the population process $\mathcal{X}_t$, defined by Eq. 6, coincides with the Markov jump

process defined by Eqs. 1 and 2. Moreover, the non-explosion assumption guarantees that $\omega(\mathcal{A}_t) \in \mathring{\Omega}$ for every t.

History process. Next among our constellation of related processes (Fig. 1) is the *history process*, which encapsulates the entire history of $\mathcal{X}$ up to time t. Specifically, for $\omega \in \Omega$, we define

$$\mathcal{H}_t(\omega) := \left(t, (T_k(\omega), U_k(\omega))_{k=0}^{K(\omega|_t)}\right),$$

where T_k, U_k, and K are as in Eq. 4. Thus $\mathcal{H}_t$ contains exactly those elements of ω that are relevant to the construction of $\mathcal{X}_t$. It is trivial to verify that $\mathcal{H}_t$ is Markov and to compute its probability density. In particular, given any history $h_t = \left(t, (t_j, u_j)_{j=0}^{k}\right)$, the marginal density at h_t is obtained by summing Eq. 7 over all possible values of the finite sequence $N(\omega|_t)$, which yields

$$P_{\mathcal{H}_t}(h_t) := p_0(x_0) \prod_{j=1}^{k} \left(\frac{\alpha_{u_j}(t_j, x_{t_j} - u_j)}{\mu}\right) \exp\left(\mu\, t - \int_0^t \sum_{u \in \mathbb{Z}^d} \alpha_u(s, x_s)\, \mathrm{d}s\right),$$

where, according to Eq. 6, $x_s = \sum_{j=0}^{k} u_j\, \mathbb{1}_{[t_j,\infty)}(s)$. Then the probability density of $\mathcal{A}_t$ conditional on $\mathcal{H}_t = h_t$ is

$$P_{\mathcal{A}_t|\mathcal{H}_t}(\mathcal{A}|h_t) := \prod_{j=1}^{k} \beta_{u_j, x_{t_j} - u_j}(N_j(\omega(\mathcal{A}))). \tag{10}$$

That is, conditional on the history process, the auxiliary numbers N_j are independent random variables.

Births, deaths, population size. The population process, $\mathcal{X}_t$ we have defined will, in applications, track the time-evolution of a structured population composed of discrete, exchangeable individuals. In particular, we are interested in the genealogical relationships among members of some focal subpopulation. For example, if we are interested in viral pathogen genealogies, the focal subpopulation might be the population of infected hosts. In the event we are studying species phylogenies, the focal subpopulation might be a group of related species. To facilitate this, we give some additional structure to the probability space we have constructed. In particular, we will suppose there are functions $I, B, D : \mathbb{Z}^d \to \mathbb{N}$ such that

$$\alpha_u(t, x) > 0 \implies I(x + u) - I(x) = B(u) - D(u), \tag{11}$$

for all $x, u \in \mathbb{Z}^d$. For any $x \in \mathbb{Z}^d$, $I(x)$ represents the size of the focal subpopulation when $\mathcal{X}_t = x$. We interpret $B(u)$ as the number of births into our focal subpopulation associated with an event of type u, and $D(u)$ as the number of deaths. Eq. 11 guarantees that I is compatible with B and D: it implies that the difference in population sizes between state x and state y matches the sum of births minus deaths over any possible path from x to y.

Although it is both interesting and possible to treat the general case, in this paper, we assume that births and deaths occur one at a time and never co-occur. In particular, we assume that $B(\mathbb{Z}^d), D(\mathbb{Z}^d) \subseteq \{0, 1\}$. Let $\mathbf{B} := B^{-1}(\{1\})$ and $\mathbf{D} := D^{-1}(\{1\})$ be the sets of event-types associated with births and deaths, respectively. Our insistence that births and deaths not co-occur implies that $\mathbf{B} \cap \mathbf{D} = \emptyset$.

Samples. Similarly, we suppose there is a function $G : \mathbb{Z}^d \to \{0, 1\}$ that we interpret as the number of samples associated with each type of event. That is, $u \in \mathbf{G} := G^{-1}(\{1\})$ implies that an event of type u results in a sample being taken. As with births and deaths, we suppose $\mathbf{G} \cap (\mathbf{B} \cup \mathbf{D}) = \emptyset$. Again, it is possible to relax the assumptions that samples occur singly and do not coincide with births or deaths, but the resulting technical complexities are best handled after the simpler case is in hand.

Absence of structure within the focal subpopulation. Note that we have assumed the existence of a single focal subpopulation. We will need the additional assumption that this focal subpopulation is itself *unstructured*. In particular, we will suppose that the individuals in the focal subpopulation (of size $I(\mathcal{X})$) are exchangeable, in the sense that each is as likely as any other of being parent to the next newborn, of being sampled, or of dying. To enforce this assumption, we postulate that, for $u \in \mathbf{B} \cup \mathbf{D} \cup \mathbf{G}$, $\beta_{u,x}$ is uniform on the first $I(x)$ natural numbers, i.e.,

$$\beta_{u,x}(n) = \frac{\mathbb{1}_{[0,I(x))}(n)}{I(x)}.$$

3. Examples.

Many interesting models fit within the constraints we have so far described. Here, we enumerate a few familiar ones. We also provide some examples of models that do not conform to our assumptions.

SIR and SIRS models. The most basic model of an immunizing infection is the SIR model, whereby susceptible individuals immediately become infectious upon being infected and remain so until they recover or are otherwise removed from the population and recovered individuals are permanently immunized against reinfection. Relaxing the latter assumption, i.e., allowing for waning of immunity, leads to the SIRS model, a modest extension. For these models, we take $d = 4$, so that the state vector is $\mathcal{X} = (s, i, r, g)$, s representing the number of susceptibles in the population; i, the number of infectives; r, the number of recovered and immune hosts; and g the cumulative number of genomic samples collected. There are four kinds of jumps, so that the rate function is

$$\alpha_u = \begin{cases} b(t)\, s\, i, & u = (-1, 1, 0, 0),\ s > 0,\ i > 0, \\ \gamma\, i, & u = (0, -1, 0, 0),\ i > 0, \\ \sigma\, r, & u = (1, 0, -1, 0),\ r > 0, \\ \psi(t, s, i, g)\, i, & u = (0, 0, 0, 1),\ i > 0, \\ 0, & \text{otherwise.} \end{cases}$$

Here, the rates shown are those of infection, recovery, loss of immunity, and sampling, respectively. The parameters γ and σ are in this case constants, but we allow for a time-varying transmission rate, $b(t)$. The sampling rate ψ is an arbitrary function, except for the minimal constraints on the rate functions described in §2. In this case, the population size function is simply $I(\mathcal{X}) = i$, whilst $\mathbf{B} = \{(-1, 1, 0, 0)\}$, $\mathbf{D} = \{(0, -1, 0, 0)\}$, and $\mathbf{G} = \{(0, 0, 0, 1)\}$.

S^{2}IR model. There is no barrier to having more than one susceptible class representing, for example, different risk groups. For example, if we have two different susceptible classes, such that the *per capita* risk of infection in the first is $b_1\, i$ and that in the second is $b_2\, i$, i being the number of infectious individuals, then we have $d = 4$, $\mathcal{X} = (s_1, s_2, i, g)$, where s_1, s_2 are the numbers in each of the two susceptible classes and g is as before. The jump rates are

$$\alpha_u = \begin{cases} b_1(t)\, s_1\, i, & u = (-1, 0, 1, 0),\ s_1 > 0,\ i > 0 \\ b_2(t)\, s_2\, i, & u = (0, -1, 1, 0),\ s_2 > 0,\ i > 0 \\ \gamma\, i, & u = (0, 0, -1, 0),\ i > 0 \\ \psi(t, s_1, s_2, i, g)\, i, & u = (0, 0, 0, 1),\ i > 0 \\ 0, & \text{otherwise.} \end{cases}$$

The four non-zero rates are those of infection from the first class, infection from the second class, recovery, and sampling, respectively. In this case, the population size function is again $I(\mathcal{X}) = i$, whilst $\mathbf{B} = \{(-1, 0, 1, 0), (0, -1, 1, 0)\}$, $\mathbf{D} = \{(0, 0, -1, 0)\}$, and $\mathbf{G} = \{(0, 0, 0, 1)\}$.

Linear birth-death-sampling process. Linear processes have proved useful as models of speciation and extinction (Nee et al., 1994; Maddison et al., 2007; Gernhard, 2008; Tavaré, 2018). For example, if we take $d = 2$, $\mathcal{X} = (n, g)$, $I(\mathcal{X}) = n$, and

$$\alpha_u = \begin{cases} \lambda(t)\, n, & u = (1, 0),\ n > 0, \\ \delta(t)\, n, & u = (-1, 0),\ n > 0, \\ \psi(t)\, n, & u = (0, 1),\ n > 0, \\ 0, & \text{otherwise}, \end{cases}$$

where λ, δ, ψ are the per-capita birth, death, and sampling rates, respectively. If these are constants, then we obtain the linear birth-death process with a constant *per capita* sampling rate treated by Stadler (2010). Of course, our formalism embraces nonlinear birth-death-sampling processes as well.

SEIR and SI^{2}R model. As mentioned above, we will rely on the assumption that the individuals whose genealogies we study are exchangeable, in the sense that each is as likely as any other of giving birth, being sampled, or of dying. This precludes consideration of a number of interesting models, including models with a latent period and models with heterogeneity of infectiousness.

Moran process. We have explicitly ruled out the possibility that birth and death events co-occur, which prevents us from applying the theory we develop here to the classical Moran model which plays such an important role in population genetics (Moran, 1958; Wakeley, 2008). We do so only to avoid distracting technicalities in this initial presentation. In fact, with minor modifications, the theory can be extended to deal with this case as well as situations where sampling events co-occur with death events (as in Leventhal et al., 2014). We postpone consideration of these cases to a later paper.

Superspreading events. Likewise, we defer consideration of jump processes for which multiple birth or death events can occur simultaneously. Such processes deserve consideration in their own right (for example, as models of superspreading events) and as models of overdispersed population processes (Bretó & Ionides, 2011). Again, accommodating such processes requires only a modest extension of the present theory, but the notational complexity introduced thereby recommends postponement of these developments to a forthcoming paper. Thus, the genealogical trees under consideration in this paper will necessarily be binary.

4. Markov genealogy processes.

Inventory process. It will be helpful to introduce another Markov process that tracks the composition of the population through time. We assume that every individual born into our focal subpopulation has a unique name (or more accurately, serial number), so that at every instant, the composition of the population is characterized by an *inventory*, i.e., a list of the names of all extant (i.e., currently living) individuals. When a birth occurs, the inventory is augmented with the new, globally unique name; when a death occurs, one name is struck off the list.

To be precise, we define $\mathcal{I}_t(\omega) := \mathrm{inven}(\omega|_t)$, where inven is a deterministic, recursive procedure as follows. First, define the add and drop operations: if $\mathcal{I} = \{c_0, \ldots, c_{m-1}\} \subset \mathbb{N}$, with $c_0 < \cdots < c_{m-1}$ and $n < m$, then $\mathrm{add}(\mathcal{I}) = \mathcal{I} \cup \{c_{m-1} + 1\}$ and $\mathrm{drop}(\mathcal{I}, n) = \mathcal{I} \setminus \{c_n\}$. With these definitions, we write,

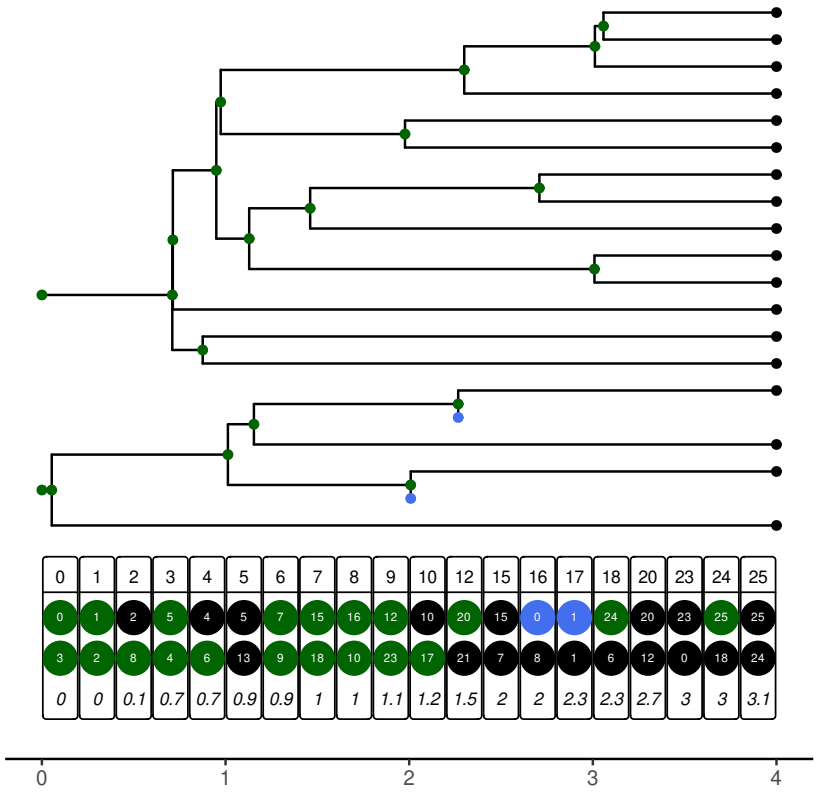

FIGURE 2. **Graphical and diagrammatic representations of a genealogy.** A genealogy can be represented graphically as a forest of binary trees and diagrammatically as a sequence of nodes, as described in the text. In the node-sequence diagram, the nodes (depicted as narrow boxes) are ordered from left to right according to their time (the bottom number in each box). Each node has a name (actually a serial number, the top number in each box) and a *pocket*, which holds two colored balls, one for each of the node's children. *Leaves* are indicated with black points in the tree and black balls in the diagram; these represent living members of the population. *Internal nodes* are indicated with green points in the tree and green balls in the diagram; these correspond to most recent common ancestors of subsets of the extant population. *Samples* are indicated with blue points and balls. The horizontal axis is time. The genealogy depicted has two roots, i.e., two ancestors present at time 0, from whom all living members of the population are descended.

for $\omega \in \mathring{\Omega}$,

$$\mathrm{inven}(\omega) := \begin{cases} \{0, \ldots, I(\mathcal{X}_0(\omega)) - 1\}, & \text{if } K(\omega) = 0; \\ \mathrm{add}(\mathrm{inven}(\omega\text{-})), & \text{if } K(\omega) > 0 \text{ and } U_{K(\omega)} \in \mathbf{B}; \\ \mathrm{drop}(\mathrm{inven}(\omega\text{-}), N_{K(\omega)}(\omega)), & \text{if } K(\omega) > 0 \text{ and } U_{K(\omega)} \in \mathbf{D}; \\ \mathrm{inven}(\omega\text{-}), & \text{otherwise.} \end{cases} \tag{12}$$

In view of Eq. 11, it is clear that, for all t, $|\mathcal{I}_t| = I(\mathcal{X}_t)$. Note that the non-explosiveness assumption is needed to guarantee the existence of ω-.

Genealogies. A genealogy relates the members of a population that are alive at a given time. Specifically, at any instant, every nonempty subset of living individuals has a most recent common ancestor. Moreover, to each such subset is associated a unique time, viz., that at which the most distantly related lineages within the subset diverged. Taken together, the collection of all such subsets and divergence times defines the genealogy, which is most commonly represented as a tree (Fig. 2). However, such representations are not unique and can be difficult to reason about. We seek a representation that is unique and for which it will be straightforward to work out probabilistic properties.

Note that a genealogical tree has at least two kinds of nodes (Fig. 2). *Leaves* represent members of the extant population at some time. *Internal nodes* represent ancestors, each of which is the most recent common ancestor of some subset of the extant population. In addition, if the population is sampled, it is natural to represent the samples as nodes of third type. The horizontal position of nodes in Fig. 2 is significant: each node has an associated time. In the case of leaves, the associated time is that of the extant population, i.e., that of the genealogy itself.

Hence, to characterize a genealogy, it is sufficient to note the time of the leaves and enumerate the internal nodes, noting for each its time and the identities of the nodes that descend immediately from it. Since there are several kinds of nodes, we need some way of distinguishing between them. To accomplish this, we introduce the notion of a *colored ball*, which we define to be an ordered pair $(f, n) \in \mathsf{F} \times \mathbb{N}$, where F is a finite set. We think of f as the *color* of the ball and n as its *name*. Our convention is to associate black with leaves and green with internal nodes. To handle sampling, we will need two additional colors, which we will take to be blue and red. Thus $\mathsf{F} \coloneqq \{\mathrm{green}, \mathrm{black}, \mathrm{blue}, \mathrm{red}\}$.

We define a genealogical *node* to be a triple (n, t, w), where $n \in \mathbb{N}$ is the node's *name*, $t \in \mathbb{R}$ is its *time*, and w is an (unordered) pair of colored balls, which we call the node's *pocket*. Given a node p, we will use $n(\mathsf{p})$, $t(\mathsf{p})$, and $w(\mathsf{p})$ to denote the name, time, and pocket of p, respectively.

A *genealogy* is defined to be an ordered pair, $\mathcal{G} = \left(t, (\mathsf{p}_k)_{k=0}^{K-1}\right)$, where $t \in \mathbb{R}_+$, $K \in \mathbb{N}$, the p_k are nodes, and the following conditions are satisfied:

(i) For all j, $t(\mathsf{p}_j) \leq t(\mathcal{G})$, i.e., no node time is later than that of the genealogy itself.
(ii) $j < k$ implies $t(\mathsf{p}_j) \leq t(\mathsf{p}_k)$, i.e., the nodes are ordered in time.
(iii) $j \neq k$ implies $w(\mathsf{p}_j) \cap w(\mathsf{p}_k) = \emptyset$; the pockets of distinct nodes are disjoint.
(iv) $(\mathrm{green}, n(\mathsf{p}_j)) \in w(\mathsf{p}_k)$ implies $j \geq k$; no green ball is held in the pocket of a later node.
(v) For all j, $(\mathrm{green}, n(\mathsf{p}_j)) \in w(\mathcal{G})$; for every node, there is a green ball bearing the name of that node.

Here, $w(\mathcal{G})$ refers to the contents of the pockets in $\mathcal{G}$ collectively:

$$w(\mathcal{G}) \coloneqq \bigcup_{k=0}^{K-1} w(\mathsf{p}_k).$$

Following our usual convention, we use $t(\mathcal{G})$, $K(\mathcal{G})$, and $\mathsf{p}_k(\mathcal{G})$ to refer to the time, the length, and the k-th node of genealogy $\mathcal{G}$, respectively. We will use $\mathsf{P}(\mathcal{G})$ to refer to the node sequence of genealogy $\mathcal{G}$. In a slight abuse of notation, we will write $\mathsf{p} \in \mathcal{G}$ when p is one node in $\mathsf{P}(\mathcal{G})$.

The black balls serve as pointers to members of the population extant at time t; the blue ones, to samples. The green balls function as pointers to internal nodes. In particular, a green ball held in the pocket of one node signifies that the node whose name matches that of the green ball is the immediate descendant of the first node. Note that we allow a node to hold its own green ball, i.e., it is permissible that $(\mathrm{green}, n(\mathsf{p})) \in w(\mathsf{p})$. Indeed, necessarily $(\mathrm{green}, n(\mathsf{p}_0)) \in w(\mathsf{p}_0)$ and, more generally, any node p

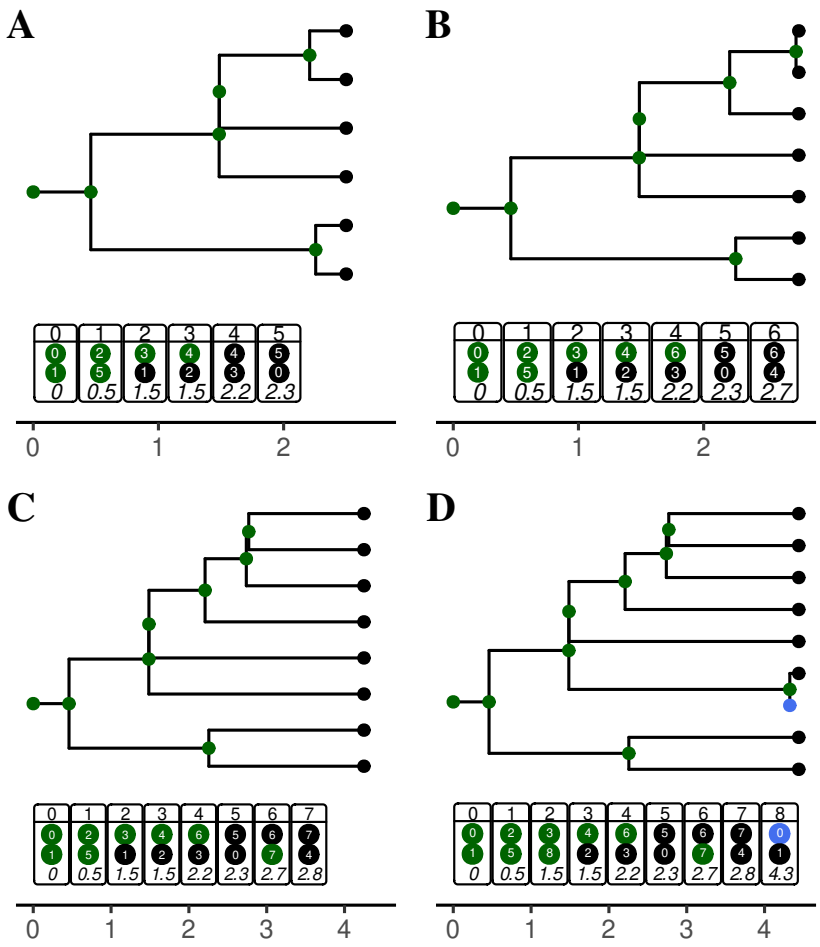

FIGURE 3. **Effects of births and sampling events on a genealogy.** Each panel shows the graphical representation of a genealogy as a tree, together with its representation as a node sequence. The horizontal axis is time. **(A–B)** The change from A to B shows the effect of a birth event on a genealogy. A new node (number 6) is introduced. The green ball corresponding to this node is exchanged for a randomly selected black ball (in this case, one held by node 4). **(C–D)** The change from C to D shows the effect of a sampling event on a genealogy. A new node (number 8), holding a blue ball, is introduced; its green ball is exchanged for a random black ball (in this case, the one held by node 2).

with $(\text{green}, n(\mathsf{p})) \in w(\mathsf{p})$ is a *root* of the genealogy. Note that nothing about our genealogy definition requires that the genealogical tree be connected: multiple roots are allowed.

Effects of births, deaths, and sampling on a genealogy. When births, deaths, or sampling events occur, these lead to changes in the genealogy (Figs. 3 and 4). Here, we describe these changes in detail for each type of event in its turn. In each instance, we assume $\mathcal{G}$ is a genealogy, as defined above, and that the event occurs at time t. Each such event will involve one particular individual in the extant population. Since the extant population is in 1-1 correspondence with the set of black balls in $w(\mathcal{G})$, this amounts to choosing one black ball. In the following, therefore, if $c_0 < c_1 < \cdots < c_{m-1}$ are the names of the m black balls in $w(\mathcal{G})$ and $0 \leq n < m$, then $\mathsf{b} = (\text{black}, c_n)$ will be the selected black ball.

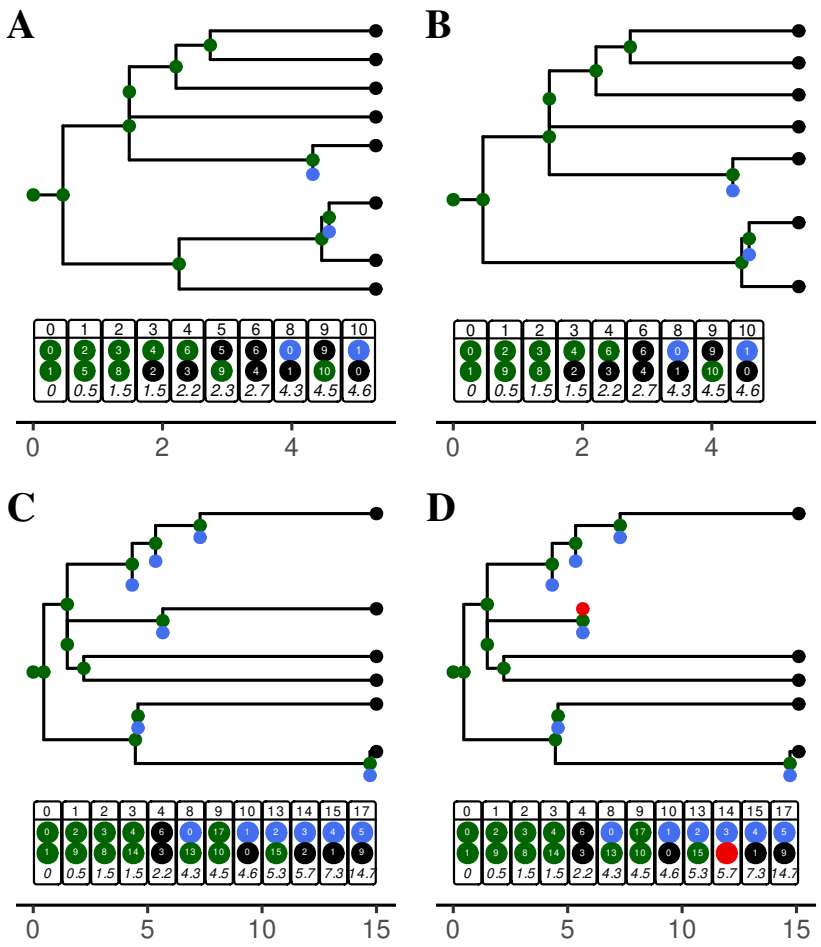

FIGURE 4. **Effect of deaths on a genealogy.** Each panel shows the graphical representation of a genealogy as a tree, together with its representation as a node sequence. The horizontal axis is time. **(A–B)** Passing from A to B, we see one possible effect of a death event on a genealogy. The random black ball chosen in this instance was the one held by node 5 in panel A. Since the other ball held by node 5 was not blue, node 5 is removed after first exchanging this other ball (green ball 9), for its own green ball (previously held by node 1). **(C–D)** The change from C to D shows the other possible effect of a death event. The random black ball chosen here was the one held by node 14. Since its other ball was blue, node 14 is not removed. Rather, it exchanges the selected black ball for a red ball. At any time, nodes with red balls represent samples on lineages that have since died out, while nodes holding one blue and one non-red ball represent samples on lineages that remain alive.

A birth in the population at time t leads to the addition of a new internal node and a new leaf in the genealogy (Fig. 3A–B). Let b be the n-th black ball, as above: this is the parent of the newborn individual. There is a unique node $\mathsf{p} \in \mathcal{G}$ such that $\mathsf{b} \in w(\mathsf{p}) = \{\mathsf{b}, \mathsf{b}'\}$. Let the name of the newborn be $c_m := c_{m-1} + 1$ and let $\mathsf{g} = (\text{green}, c_m)$ and $\mathsf{b}'' = (\text{black}, c_m)$. Construct a new node $\mathsf{p}' = (c_m, t, \{\mathsf{g}, \mathsf{b}''\})$. Now let node p' exchange g with node p for b. Thus, if before the swap we have $w(\mathsf{p}) = \{\mathsf{b}, \mathsf{b}'\}$ and $w(\mathsf{p}') = \{\mathsf{g}, \mathsf{b}''\}$, after the swap we have $w(\mathsf{p}) = \{\mathsf{g}, \mathsf{b}'\}$ and $w(\mathsf{p}') = \{\mathsf{b}, \mathsf{b}''\}$.

Finally, insert p' into the last position in the sequence of nodes. Let $\mathrm{add}(\mathsf{P}(\mathcal{G}), n)$ denote the resulting sequence of nodes. Fig. 3A–B illustrates.

A sample in the population at time t results in the addition of one new internal node equipped with a new blue ball (Fig. 3C–D). Let b be the n-th black ball selected as above: this will be the individual sampled. As before, there is a unique node $\mathsf{p} \in \mathcal{G}$ such that $\mathsf{b} \in w(\mathsf{p}) = \{\mathsf{b}, \mathsf{b}'\}$. Again, we construct a new node by taking $c_m = c_{m-1} + 1$ and letting $\mathsf{p}' = (c_m, t, \{\mathsf{g}, \mathsf{b}''\})$, where $\mathsf{g} = (\mathrm{green}, c_m)$ and $\mathsf{b}'' = (\mathrm{blue}, q)$. Again, we swap b for g between nodes p and p' and insert p' at the last position of the node-sequence. Here we take the name, q, of the new blue ball to be the ordinal number of the sample, $q = |\{\mathsf{b} \in w(\mathcal{G}) \mid \mathsf{b} \text{ is blue}\}|$. We denote the resulting sequence of nodes by $\mathrm{sample}(\mathsf{P}(\mathcal{G}), n)$. Fig. 3C–D illustrates.

A death in the population at time t can lead to the loss of one leaf and one internal node (Fig. 4A–B). The genealogy thus drops all record of the existence of the deceased individual. On the other hand, samples represent recorded events: we do not wish to lose track of them. Therefore, when a death would delete a sample, we prevent this from occurring using a red ball (Fig. 4C–D). To be precise, let b be the n-th black ball selected as above: this is the individual who will die. As usual, there is a unique node $\mathsf{p} \in \mathcal{G}$ such that $\mathsf{b} \in w(\mathsf{p}) = \{\mathsf{b}, \mathsf{b}'\}$. Let $\mathsf{g} = (\mathrm{green}, n(\mathsf{p}))$ and let the unique node holding g be denoted p'. If b' is black, we swap b' for g and then delete p from the node sequence. If b' is blue, we replace b with a red ball, with name matching that of b', leaving everything else intact. We use $\mathrm{drop}(\mathsf{P}(\mathcal{G}), n)$ to denote the resulting sequence of nodes. The two possible effects of sampling are illustrated in Fig. 4.

It is straightforward to verify that whenever $\mathcal{G}$ is a genealogy and $n < m$, $(t(\mathcal{G}), \mathrm{add}(\mathsf{P}(\mathcal{G}), n))$, $(t(\mathcal{G}), \mathrm{drop}(\mathsf{P}(\mathcal{G}), n))$, and $(t(\mathcal{G}), \mathrm{sample}(\mathsf{P}(\mathcal{G}), n))$ as just defined are all valid genealogies. Note also that the add and drop procedures mirror their counterparts for the inventory process. It follows that if $\mathcal{I}$ is an inventory, $\mathcal{G}$ a genealogy, and $\mathcal{I}$, $\mathcal{G}$ have the relation that $(\mathrm{black}, c) \in w(\mathcal{G})$ if and only if $c \in \mathcal{I}$, then the same relation holds between $\mathrm{add}(\mathcal{I})$ and $\mathrm{add}(\mathsf{P}(\mathcal{G}), n)$, $\mathrm{drop}(\mathcal{I}, n)$ and $\mathrm{drop}(\mathsf{P}(\mathcal{G}), n)$, and $\mathcal{I}$ and $\mathrm{sample}(\mathsf{P}(\mathcal{G}), n)$, respectively, for every $n < m$.

Genealogical event times. Given a genealogy $\mathcal{G}$, the set of *genealogical event times*, $\mathsf{E}(\mathcal{G})$, is the set of all node times. Several of its subsets are of interest. In particular, we define

$$
\begin{aligned}
\mathsf{E}(\mathcal{G}) &\coloneqq \{t(\mathsf{p}) \mid \mathsf{p} \in \mathcal{G}\},\\
\mathsf{A}(\mathcal{G}) &\coloneqq \{t(\mathsf{p}) \mid \mathsf{p} \in \mathcal{G}, w(\mathsf{p}) \text{ contains a green ball}\},\\
\mathsf{C}(\mathcal{G}) &\coloneqq \{t(\mathsf{p}) \mid \mathsf{p} \in \mathcal{G}, w(\mathsf{p}) \text{ contains two green balls}\},\\
\mathsf{L}(\mathcal{G}) &\coloneqq \{t(\mathsf{p}) \mid \mathsf{p} \in \mathcal{G}, w(\mathsf{p}) \text{ contains no green balls}\},\\
\mathsf{S}(\mathcal{G}) &\coloneqq \{t(\mathsf{p}) \mid \mathsf{p} \in \mathcal{G}, w(\mathsf{p}) \text{ contains a blue ball}\},\\
\mathsf{D}(\mathcal{G}) &\coloneqq \mathsf{A}(\mathcal{G}) \cap \mathsf{S}(\mathcal{G}).
\end{aligned}
$$

With these definitions, $\mathsf{A}(\mathcal{G})$ comprises the internal node times, $\mathsf{C}(\mathcal{G})$ is the set of branch times, $\mathsf{S}(\mathcal{G})$ holds the sample times, and $\mathsf{D}(\mathcal{G})$ is the set of *direct-descent times*, i.e., the times of samples that are themselves directly ancestral to other samples.

Genealogy process. We now proceed to define the *genealogy process*, $\mathcal{G}_t$. For $t \in \mathbb{R}_+$ and $\omega \in \Omega$, we define $\mathcal{G}_t(\omega) = (t, \mathrm{geneal}(\omega|_t))$, where geneal is a deterministic procedure defined recursively for

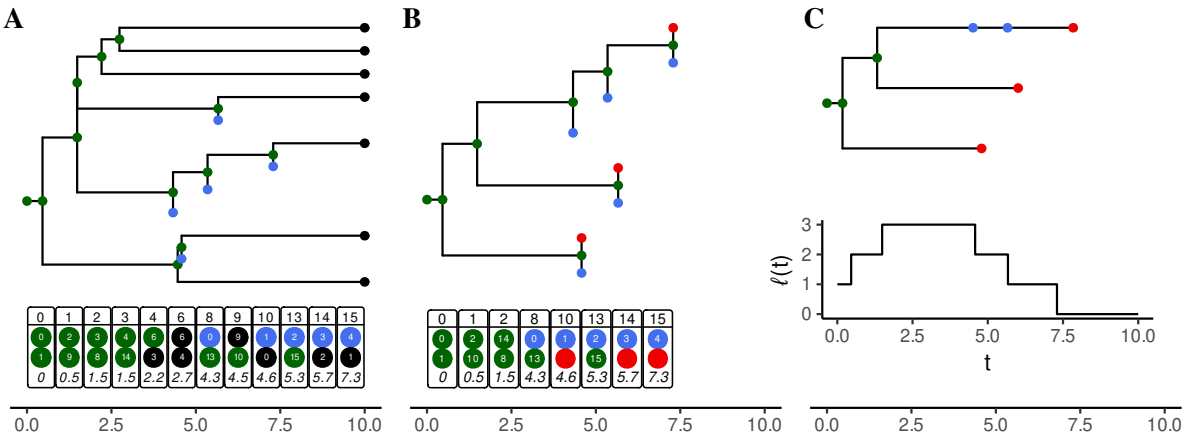

FIGURE 5. **Pruning and the visible genealogy.** The genealogy depicted in **(A)** is pruned, i.e., all black balls are dropped according to the drop procedure described in the text. The resulting *visible genealogy* is represented in **(B)**. A more compact graphical representation is displayed in **(C)**. In it, so-called *red nodes*, i.e., those holding one red and one blue ball; are shown by red points in the graphical representation. Likewise *blue nodes* (those holding one blue and one green ball) and *green nodes* (those holding two green balls) are indicated by blue and green points, respectively. The inset shows the *lineage count*, $\ell(t)$, the number of lineages present in the visible genealogy at time t.

$\omega \in \mathring{\Omega}$ by

$$\text{geneal}(\omega) := \begin{cases} (k, 0, \{(\text{green}, k), (\text{black}, k)\})_{k=0}^{I(\mathfrak{X}_0(\omega))-1}, & \text{if } K(\omega) = 0; \\ \text{add}(\text{geneal}(\omega\text{-}), N_{K(\omega)}(\omega)), & \text{if } K(\omega) > 0 \text{ and } U_{K(\omega)} \in \mathbf{B}; \\ \text{drop}(\text{geneal}(\omega\text{-}), N_{K(\omega)}(\omega)), & \text{if } K(\omega) > 0 \text{ and } U_{K(\omega)} \in \mathbf{D}; \\ \text{sample}(\text{geneal}(\omega\text{-}), N_{K(\omega)}(\omega)), & \text{if } K(\omega) > 0 \text{ and } U_{K(\omega)} \in \mathbf{G}; \\ \text{geneal}(\omega\text{-}), & \text{otherwise.} \end{cases} \tag{13}$$

Thus, $\text{geneal}(\omega)$ is a well defined node sequence for every jump sequence $\omega \in \mathring{\Omega}$. Note that we initialize the genealogy process with a collection of root nodes. The graphical representation of such a genealogy is a forest of single-leaf trees, each of which has zero branch length. At each birth, death, or sample event, we modify the genealogy accordingly. Note that the aforementioned parallelism between the add, drop, and sample procedures ensures that there is a deterministic map, ext, such that $\mathcal{I}_t(\omega) = \text{ext}(\mathcal{G}_t(\omega))$ for all $\omega \in \Omega$.

Visible genealogy process. It is typically impossible to fully sample a population; the genealogical relationships among unsampled lineages remain unobserved. It is therefore of interest to study the genealogy that represents the relationships just among the samples. This is most readily obtained from the full genealogy by a process of *pruning*, which we proceed to describe.

Suppose $\mathcal{G}$ is a genealogy and let $\mathcal{I} = \text{ext}(\mathcal{G})$. Let $\text{obs}(\mathcal{G})$ be the result of iteratively applying the drop operation defined above to $\mathcal{G}$ for each $c \in \mathcal{I}$. Specifically, suppose $\mathcal{I} = \{c_0, \ldots, c_{m-1}\}$. Let $\mathsf{P}_0 = \mathsf{P}(\mathcal{G})$ and $\mathsf{P}_k = \text{drop}(\mathsf{P}_{k-1}, c_{k-1})$, for $k = 1, \ldots, m$. Then $\text{obs}(\mathcal{G}) := (t(\mathcal{G}), \mathsf{P}_m)$. Fig. 5 illustrates the pruning procedure.

For $\omega \in \Omega$, we define the *visible genealogy process*,

$$\mathcal{V}_t(\omega) := \text{obs}(\mathcal{G}_t(\omega)). \tag{14}$$

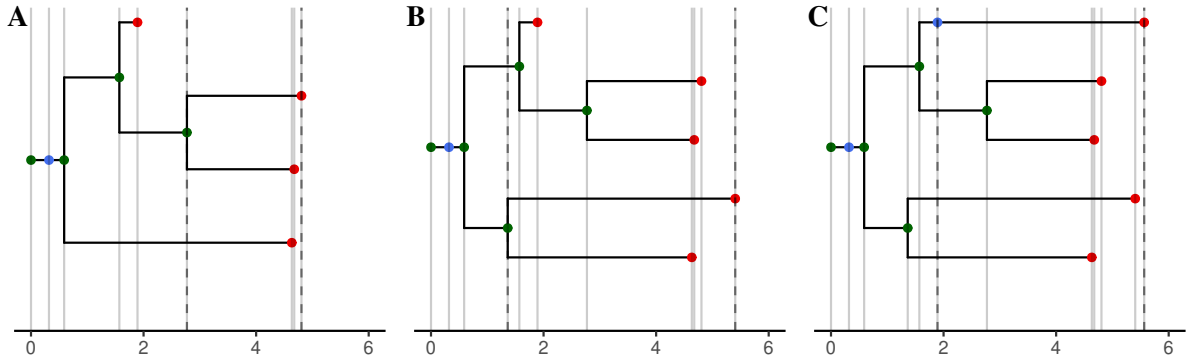

FIGURE 6. **Embedded chain of the visible genealogy process.** Three successive states of the embedded chain, $\mathcal{W}_i$, are shown. In panel A, the visible genealogy $\mathcal{W}_5$ is shown graphically. Note there are 5 samples represented (four red points and one blue one). The grey vertical lines indicate the genealogy event times, $\mathsf{E}(\mathcal{W}_5)$. Panels B and C show $\mathcal{W}_6$ and $\mathcal{W}_7$, respectively. In panel B, the new sample is depicted as the second red point from the bottom; it attaches to $\mathcal{W}_5$ at a green node (a coalescence point). In panel C, the new sample is the topmost red point; it attaches to $\mathcal{W}_6$ at a blue node (a direct-descent event). In each panel, the sample and attachment times of the latest sample are indicated by dashed vertical lines.

Note that $\mathcal{V}_t$ so defined is a itself a genealogy. Fig. 5B depicts one example of $\mathcal{V}_t$ both graphically (as a tree) and diagrammatically (as a node-sequence).

Node color. There are at most three kinds of nodes in a visible genealogy, distinguished by the contents of their pockets: (a) *green nodes*, which have two green balls in their pockets; (b) *blue nodes*, which have one blue and one green ball; (c) *red nodes*, which have one red and one blue ball. This distinction allows a compact graphical representation of the visible genealogy (Fig. 5C) and proves useful in deriving our main results as well. Green nodes correspond to *coalescence points* (branch points) in the visible genealogy; red nodes, to leaves; blue nodes correspond to direct-descent events, i.e., samples which are directly ancestral to other samples. Thus, if $\mathcal{V}$ is a visible genealogy, $\mathsf{E}(\mathcal{V}) = \mathsf{C}(\mathcal{V}) \cup \mathsf{D}(\mathcal{V}) \cup \mathsf{L}(\mathcal{V})$, where $\mathsf{C}(\mathcal{V})$ is the set of times of the green nodes, $\mathsf{D}(\mathcal{V})$ contains the times of the blue nodes, and $\mathsf{L}(\mathcal{V})$ holds the times of the red nodes. Moreover, $\mathsf{S}(\mathcal{V}) = \mathsf{D}(\mathcal{V}) \cup \mathsf{L}(\mathcal{V})$ is the set of sample times and $\mathsf{A}(\mathcal{V}) = \mathsf{D}(\mathcal{V}) \cup \mathsf{C}(\mathcal{V})$ comprises the *attachment times*, i.e., those points at which a sampled lineage attaches to the visible genealogy subtended by earlier samples.

Lineage count. Given a genealogy, $\mathcal{V}$, at each time $t \in \mathbb{R}_+$, there are a finite number, $\ell(t, \mathcal{V})$, of lineages present in the genealogy at that time (Fig. 5C). Evidently, one has

$$\ell(t, \mathcal{V}) := \sum_{e \in \mathsf{C}(\mathcal{V})} \mathbb{1}_{[e,\infty)}(t) - \sum_{e \in \mathsf{L}(\mathcal{V})} \mathbb{1}_{[e,\infty)}(t) = \sum_{e \in \mathsf{A}(\mathcal{V})} \mathbb{1}_{[e,\infty)}(t) - \sum_{e \in \mathsf{S}(\mathcal{V})} \mathbb{1}_{[e,\infty)}(t). \tag{15}$$

With this definition, note that ℓ is a càdlàg function of t.

Embedded chain of the visible genealogy process. Now, for $\omega \in \Omega$, consider the visible genealogy process $\mathcal{V}_t(\omega)$. The sample times, S_i, form an increasing sequence. Let $\mathcal{W}_i(\omega) := \mathcal{V}_{S_i(\omega)}(\omega)$ be the embedded chain of the visible genealogy process. Each genealogy in this chain builds on the previous one precisely in that one additional lineage is added; Fig. 6 illustrates. The terminal point of the new lineage is the latest sample time; it attaches to the preceding genealogy at a random attachment time.

Let A_i be the attachment time of the latest lineage in $\mathcal{W}_i$. Note that the embedded chain is trivially Markov, since each $\mathcal{W}_i$ contains $\mathcal{W}_j$ within it, for $j < i$.

The proof of the following is immediate.

Lemma 1. *Let $s_i = \mathsf{S}(\mathcal{W}_i) \setminus \mathsf{S}(\mathcal{W}_{i-1})$ and $a_i = \mathsf{A}(\mathcal{W}_i) \setminus \mathsf{A}(\mathcal{W}_{i-1})$ be the sample and attachment times, respectively, of the i-th sample lineage. Then, if it is understood that $\ell(t, \mathcal{W}_0) = 0$,*

$$\ell(t, \mathcal{W}_i) = \ell(t, \mathcal{W}_{i-1}) + \mathbb{1}_{[a_i, s_i)}(t), \qquad \forall i > 0.$$

5. Results.

Let $P_{\mathcal{W}_i|\mathcal{H}}$ denote the probability density of $\mathcal{W}_i$ conditional on $\mathcal{H}_{S_i}$. The Markovity of $\mathcal{W}_i$ gives us

$$P_{\mathcal{W}_i|\mathcal{H}} = P_{\mathcal{W}_1|\mathcal{H}} \prod_{j=2}^{i} P_{\mathcal{W}_j|\mathcal{W}_{j-1},\mathcal{H}}.$$

Now, conditional on $\mathcal{H}_{S_1}$, $\mathcal{W}_1$ consists a.s. of one red node and one root node. Thus $P_{\mathcal{W}_1|\mathcal{H}} = 1$.

Let us compute $P_{\mathcal{W}_j|\mathcal{W}_{j-1},\mathcal{H}}$. We fix $A_j = a_j$, $\mathcal{W}_j = w_j$, $S_j = s_j$, and $\mathcal{H}_{s_i} = h = \left(s_i, (t_k, u_k)_{k=0}^K\right)$. We write $x_k = \mathcal{X}_{t_k}$, and define the sets $\Omega_j \coloneqq \{\omega \in \Omega \mid \mathcal{W}_j(\omega) = w_j, \mathcal{H}_{s_i}(\omega) = h\} \subset \Omega$ and $\Omega_{jk} \coloneqq N_k(\Omega_j) \subset \mathbb{N}$, for $j \le i$ and $k \le K$. There is a bijection between Ω_j and $\prod_{k=1}^K \Omega_{jk}$. The conditional independence of the N_k (Eq. 10) allows us to write

$$\begin{aligned} P_{\mathcal{W}_j|\mathcal{W}_{j-1},\mathcal{H}}(w_j|w_{j-1}, h) &= \sum_{\omega\in\Omega_j} P_{\mathcal{A}|\mathcal{H}}(\mathcal{A}(\omega)|h) = \sum_{\omega\in\Omega_j} \prod_{k=1}^{K} \beta_{u_k, x_k - u_k}(N_k(\omega)) \\ &= \prod_{k=1}^{K} \sum_{n_k \in \Omega_{jk}} \beta_{u_k, x_k - u_k}(n_k). \end{aligned}$$

Since the β are uniform, it remains only to count up the number of elements in the sets Ω_{jk}, i.e., the number of choices of N_k that are consistent with $\mathcal{W}_j$, for each j and k.

Define $q_{jk} \coloneqq \sum_{n_k\in\Omega_{jk}} \beta_{u_k, x_k - u_k}(n_k)$ and, for the moment, let $I_k = I(x_k)$, $\ell_{jk} = \ell(t_k, \mathcal{W}_{j-1})$. Note that $\ell_{1k} = 0$ for all k. There are several cases to consider:

(a) If $t_k \notin [a_j, s_j)$, or if $u_k \notin \mathbf{B} \cup \mathbf{G}$, then all choices n_k are compatible, so we have $q_{jk} = 1$.
(b) If $t_k \in \mathsf{A}(\mathcal{W}_{j-1})$, then again, all choices n_k are compatible and $q_{jk} = 1$.
(c) If $t_k \in (a_j, s_j) \setminus \mathsf{A}(\mathcal{W}_{j-1})$ and $u_k \in \mathbf{G}$, then there was a sample event at time t_k but the j-th sample lineage did not directly descend from it. Since $\mathbf{G} \cap (\mathbf{B} \cup \mathbf{D}) = \emptyset$, $\mathcal{I}(t_k-) = \mathcal{I}(t_k)$, i.e., the inventory of the population did not change at time t_k. There were I_k individuals at this time. However, ℓ_{jk} of these could not have been chosen for the sample (or we would have $t_k \in \mathsf{A}(\mathcal{W}_{j-1})$). Only one of these was of the j-th sample lineage. Therefore $q_{jk} = 1 - 1/(I_k - \ell_{jk})$ in this case.
(d) If $t_k = a_j$ and $u_k \in \mathbf{G}$, then the j-th lineage attaches to $\mathcal{W}_{j-1}$ in a direct-descent event. There were $I_k - \ell_{jk}$ individuals who might have been sampled, just one of whom was of the j-th lineage. Therefore, $q_{jk} = 1/(I_k - \ell_{jk})$ in this case.
(e) If $t_k \in (a_j, s_j) \setminus \mathsf{A}(\mathcal{W}_{j-1})$ and $u_k \in \mathbf{B}$, then there was a birth event at time t_k but the j-th sample lineage was not involved in it. Thus the population size just before t_k was $I_k - 1$. The node newly added has a pair of black balls, one corresponding to the newborn, the other to the parent; these are indistinguishable. There are $\binom{I_k}{2}$ such pairs, but $\binom{\ell_{jk}}{2}$ of these can be excluded

because none of the ℓ_{jk} lineages of $\mathcal{W}_{j-1}$ coalesces here. Of the remaining pairs, exactly ℓ_{jk} involve the sole individual of the j-th sample lineage and one individual of one of the lineages in $\mathcal{W}_{j-1}$. Therefore, $q_{jk} = 1 - \ell_{jk}/(\binom{I_k}{2} - \binom{\ell_{jk}}{2})$ in this case.

(f) Finally, we treat the case $t_k = a_j$, $u_k \in \mathbf{B}$. The logic of case e applies, but here the coalescence event did occur. Precisely one of the $\binom{I_k}{2} - \binom{\ell_{jk}}{2}$ pairs that might have coalesced is the one consisting of the individual of the j-th sample lineage and the individual in the sample lineage to which the j-th sample lineage did attach. Therefore, $q_{jk} = 1/(\binom{I_k}{2} - \binom{\ell_{jk}}{2})$ in this case.

To summarize, we have

$$q_{jk} = \begin{cases} 1, & \text{if } t_k \notin [a_j, s_j), \\ 1, & \text{if } u_k \notin \mathbf{B} \cup \mathbf{G}, \\ 1, & \text{if } t_k \in \mathsf{C}(\mathcal{W}_{j-1}) \cup \mathsf{D}(\mathcal{W}_{j-1}), \\ 1 - \frac{1}{I_k - \ell_{jk}}, & \text{if } t_k \in (a_j, s_j) \setminus \mathsf{D}(\mathcal{W}_{j-1}) \text{ and } u_k \in \mathbf{G}, \\ \frac{1}{I_k - \ell_{jk}}, & \text{if } t_k = a_j \in \mathsf{L}(\mathcal{W}_{j-1}), \\ 1 - \frac{\ell_{jk}}{\binom{I_k}{2} - \binom{\ell_{jk}}{2}}, & \text{if } t_k \in (a_j, s_j) \text{ and } u_k \in \mathbf{B}, \\ \frac{1}{\binom{I_k}{2} - \binom{\ell_{jk}}{2}}, & \text{if } t_k = a_j \in \mathsf{C}(\mathcal{W}_j). \end{cases} \tag{16}$$

This establishes the first of our main results, namely:

Theorem 1. *With the definitions as above,* $P_{\mathcal{W}_i|\mathcal{H}}(w|h) = \prod_{j=1}^{i} \prod_{k=1}^{K} q_{jk}$.

Although this result is compactly stated, the computation it suggests is awkward. For example, to compute $P_{\mathcal{W}_i|\mathcal{H}}(w|h)$ via Monte Carlo, one must simulate and store the entire history $\mathcal{H}_t$, in order to compute each of the $\prod_k q_{jk}$. For all but quite small populations, this will prove impracticable. Simply exchanging the order of the product in Theorem 1, however, yields a scheme which requires only simulation and temporary storage of the population process $\mathcal{X}_t$.

Theorem 2. *Suppose $\mathcal{V}_t$ is the visible genealogy process induced by $\mathcal{X}_t$ and $\mathcal{H}_t$ is the history process. Fix $\mathcal{H}_t = h = \left(t, (t_k, u_k)_{k=0}^{K}\right)$ and let $x_t = \sum_{k=0}^{K} u_k \mathbb{1}_{[t_k,\infty)}(t)$. Let $\mathsf{U}(h)$ be the set of unobserved births in history h, i.e., $\mathsf{U}(h) = \{t_k(h) \mid k > 0 \text{ and } u_k(h) \in \mathbf{B}\} \setminus \mathsf{C}(\mathcal{V}_t)$. Then*

$$P_{\mathcal{V}_t|\mathcal{H}_t}(\mathcal{V}_t|h) = \frac{\prod_{e \in \mathsf{U}(h)} \left(1 - \frac{\binom{\ell(e,\mathcal{V}_t)}{2}}{\binom{I(x_e)}{2}}\right) \prod_{e \in \mathsf{L}(\mathcal{V}_t)} \left(1 - \frac{\ell(e,\mathcal{V}_t)}{I(x_e)}\right)}{\prod_{e \in \mathsf{C}(\mathcal{V}_t)} \binom{I(x_e)}{2} \prod_{e \in \mathsf{D}(\mathcal{V}_t)} I(x_e)}.$$

Proof. $\mathcal{V}_t = \mathcal{W}_i$ for some i. We compute $Q_k := \prod_{j=1}^{i} q_{jk}$ for each k. As above, let $I_k = I(x_k)$, $\ell_{jk} = \ell(t_k, \mathcal{W}_{j-1})$.

For the first three cases of Eq. 16, we have $Q_k = 1$.

Suppose $t_k \in \mathsf{L}(\mathcal{W}_i)$. Then $u_k \in \mathbf{G}$, and we have

$$Q_k = \prod_{j=1}^{i} \frac{I_k - \ell_{jk} - 1}{I_k - \ell_{jk}} = \frac{I_k - \ell_{ik} - 1}{I_k - \ell_{1k}} \cdot \prod_{j=1}^{i-1} \frac{I_k - \ell_{jk} - 1}{I_k - \ell_{j+1,k}}.$$

By Lemma 1, $\ell_{j+1,k} = \ell_{jk} + 1$, which implies

$$Q_k = \frac{I_k - \ell_{ik} - 1}{I_k - \ell_{1k}} = 1 - \frac{\ell_{i+1,k}}{I_k}.$$

Now consider $t_k \in \mathsf{D}(\mathcal{W}_i)$. Let m be the unique integer such that $t_k \in \mathsf{D}(\mathcal{W}_m) \setminus \mathsf{D}(\mathcal{W}_{m-1})$. Then we have

$$q_{jk} = \begin{cases} 1 - \frac{1}{I_k - \ell_{jk}}, & j < m, \\ \frac{1}{I_k - \ell_{mk}}, & j = m, \\ 1, & j > m. \end{cases}$$

It follows that

$$Q_k = \frac{1}{I_k - \ell_{mk}} \cdot \prod_{j=1}^{m-1} \frac{I_k - \ell_{jk} - 1}{I_k - \ell_{jk}} = \frac{1}{I_k}.$$

Now consider the case $t_k \notin \mathsf{C}(\mathcal{W}_i)$ but $u_k \in \mathbf{B}$. Again using Lemma 1, we have

$$Q_k = \prod_{j=1}^{i} \left(1 - \frac{\ell_{jk}}{\binom{I_k}{2} - \binom{\ell_{jk}}{2}}\right) = 1 - \frac{\binom{\ell_{i+1,k}}{2}}{\binom{I_k}{2}}.$$

Finally, suppose $t_k \in \mathsf{C}(\mathcal{W}_i)$. Let m be the unique integer such that $t_k \in \mathsf{C}(\mathcal{W}_m) \setminus \mathsf{C}(\mathcal{W}_{m-1})$. Then

$$q_{jk} = \begin{cases} 1 - \frac{\ell_{jk}}{\binom{I_k}{2} - \binom{\ell_{jk}}{2}}, & j < m, \\ \frac{1}{\binom{I_k}{2} - \binom{\ell_{jk}}{2}}, & j = m \\ 1, & j > m. \end{cases}$$

Once again, a straightforward calculation yields

$$Q_k = \frac{1}{\binom{I_k}{2}}.$$

□

The unnormalized nonlinear filter (DMZ) equation. Theorem 2 gives us an expression for the likelihood of $\mathcal{V}_t$ given the history $\mathcal{H}_t$. We now seek to integrate out the dependence on $\mathcal{H}_t$. The form of Theorem 2 implies that this can be done in a sequential fashion, working from earlier times to later ones, thus avoiding the need either to work backward in time or to store the full history process. Indeed, to compute $P_{\mathcal{V}_t|\mathcal{H}_t}$, we progressively accumulate factors, one for each event in the history and one for each event in the visible genealogy. Each such term depends only on the state of the population process, $\mathcal{X}$ at the time of that event. We can therefore integrate out the history by integrating over the possible values of $\mathcal{X}$ at each time.

Given a visible genealogy $\mathcal{V}$, and a time $0 \le t \le t(\mathcal{V})$, let us define the *partial genealogy*, $\mathcal{V}|_t$, to be that portion of the visible genealogy lying to the left of t. For each $t < t(\mathcal{V})$, we will define the *partial weight*, $w(t, x, \mathcal{V})$, in such a way that

$$w(t, x, \mathcal{V}) = P_{\mathcal{V}_t|\mathcal{X}_t}(\mathcal{V} \mid x) \qquad \text{and} \qquad \sum_{x \in \mathbb{Z}^d} w(t, x, \mathcal{V}) = P_{\mathcal{V}_t}(\mathcal{V}). \tag{17}$$

Note that for $t < t(\mathcal{V})$, $w(t, x, \mathcal{V})$ is not itself a likelihood of any subset of the data. However, when $t = t(\mathcal{V})$, the sum (over x) of the partial weights will equal the likelihood of the genealogy, unconditional on the history process. The partial weights are defined by an initial-value problem that they must satisfy. We proceed to derive this now.

Suppose $\mathcal{V}$ is a visible genealogy, $0 \le t < t(\mathcal{V})$, and $\delta t > 0$. Within the interval $[t, t + \delta t)$, the following events are exhaustive and mutually exclusive: (a) nothing occurred, (b) more than one event occurred, (c) an event occurred which was neither a birth nor a sample, and no other event occurred, (d) a birth event, and no other, occurred, and this birth event was not a coalescence event in $\mathcal{V}$, (e) a birth event, which was also a coalescence event, occurred, and no other, (f) a sample event, and no other, occurred, and this sample was a direct-descent event in $\mathcal{V}$, (g) a sample event, which was not a direct-descent event, occurred, and no other. Now, if $t \notin \mathsf{E}(\mathcal{V})$, our non-explosion assumption implies that we can choose δt sufficiently small so that $\mathsf{E}(\mathcal{V}) \cap [t, t + \delta t) = \emptyset$. In this case, the only possible events are a–d. Accordingly, we desire that

$$\begin{aligned} w(t + \delta t, x, \mathcal{V}) = &\left(1 - \sum_{u \in \mathbb{Z}^d} \alpha_u(t, x)\, \delta t\right) w(t, x, \mathcal{V}) \\ &+ \sum_{u \in \mathbb{Z}^d \setminus \mathbf{B} \setminus \mathbf{G}} \alpha_u(t, x - u)\, w(t, x - u, \mathcal{V})\, \delta t \\ &+ \sum_{u \in \mathbf{B}} \alpha_u(t, x - u) \left(1 - \frac{\binom{\ell(t,\mathcal{V})}{2}}{\binom{I(x)}{2}}\right) w(t, x - u, \mathcal{V})\, \delta t + o(\delta t). \end{aligned} \tag{18}$$

Rearranging Eq. 18 and taking $\delta t \downarrow 0$ in the usual way, we obtain, for $t \notin \mathsf{E}(\mathcal{V}) = \mathsf{C}(\mathcal{V}) \cup \mathsf{D}(\mathcal{V}) \cup \mathsf{L}(\mathcal{V})$,

$$\begin{aligned} \frac{\partial w}{\partial t}(t, x, \mathcal{V}) = &\sum_{u \in \mathbb{Z}^d} \left[\alpha_u(t, x - u)\, w(t, x - u, \mathcal{V}) - \alpha_u(t, x)\, w(t, x, \mathcal{V})\right] \\ &- \sum_{u \in \mathbf{G}} \alpha_u(t, x - u)\, w(t, x - u, \mathcal{V}) \\ &- \sum_{u \in \mathbf{B}} \alpha_u(t, x - u)\, \frac{\binom{\ell(t,\mathcal{V})}{2}}{\binom{I(x)}{2}}\, w(t, x - u, \mathcal{V}). \end{aligned} \tag{19}$$

On the other hand, if $t \in \mathsf{E}(\mathcal{V}) = \mathsf{C}(\mathcal{V}) \cup \mathsf{D}(\mathcal{V}) \cup \mathsf{L}(\mathcal{V})$, we desire that

$$w(t, x, \mathcal{V}) = \begin{cases} \displaystyle\sum_{u \in \mathbf{B}} \frac{\alpha_u(t, x - u)}{\mu} \frac{1}{\binom{I(x)}{2}}\, w(t\text{-}, x - u, \mathcal{V}), & t \in \mathsf{C}(\mathcal{V}), \\ \displaystyle\sum_{u \in \mathbf{G}} \frac{\alpha_u(t, x - u)}{\mu} \frac{1}{I(x)}\, w(t\text{-}, x - u, \mathcal{V}), & t \in \mathsf{D}(\mathcal{V}), \\ \displaystyle\sum_{u \in \mathbf{G}} \frac{\alpha_u(t, x - u)}{\mu} \left(1 - \frac{\ell(t, \mathcal{V})}{I(x)}\right) w(t\text{-}, x - u, \mathcal{V}), & t \in \mathsf{L}(\mathcal{V}). \end{cases} \tag{20}$$

Here, t- indicates the left limit.

Making use of the Dirac delta function, $\delta(t)$, we can combine Eqs. 19 and 20 into a single equation, the analogue of the Duncan-Mortensen-Zakai equation (Zakai, 1969) for this problem:

$$\begin{aligned}\frac{\partial w}{\partial t}(t,x,\mathcal{V}) &= \sum_{u\in\mathbb{Z}^d}\left[\alpha_u(t,x-u)\,w(t,x-u,\mathcal{V}) - \alpha_u(t,x)\,w(t,x,\mathcal{V})\right]\\ &\quad - \sum_{u\in\mathbf{G}}\alpha_u(t,x-u)\,w(t,x-u,\mathcal{V})\\ &\quad - \sum_{u\in\mathbf{B}}\alpha_u(t,x-u)\,\frac{\binom{\ell(t,\mathcal{V})}{2}}{\binom{I(x)}{2}}\,w(t,x-u,\mathcal{V})\\ &\quad + \sum_{e\in\mathsf{C}(\mathcal{V})}\delta(t-e)\left\{\sum_{u\in\mathbf{B}}\frac{\alpha_u(t,x-u)}{\mu}\,\frac{1}{\binom{I(x)}{2}}\,w(t,x-u,\mathcal{V})\right\}\\ &\quad + \sum_{e\in\mathsf{D}(\mathcal{V})}\delta(t-e)\left\{\sum_{u\in\mathbf{G}}\frac{\alpha_u(t,x-u)}{\mu}\,\frac{1}{I(x)}\,w(t,x-u,\mathcal{V})\right\}\\ &\quad + \sum_{e\in\mathsf{L}(\mathcal{V})}\delta(t-e)\left\{\sum_{u\in\mathbf{G}}\frac{\alpha_u(t,x-u)}{\mu}\,\left(1-\frac{\ell(t,\mathcal{V})}{I(x)}\right)\,w(t,x-u,\mathcal{V})\right\}\\ &\quad - \sum_{e\in\mathsf{E}(\mathcal{V})}\delta(t-e)\,w(t,x,\mathcal{V}).\end{aligned} \tag{21}$$

The appearance in Eq. 21 of the rate, μ, of the Poisson point process, the probability measure of which is the base measure for our probability densities, serves as a reminder that the numerical values of these densities depend on the choice of time unit.

For Eq. 21 to be valid, we must insist that $I(x) \geq \ell(t,\mathcal{V})$. Moreover, if $I(x) < \ell(t,\mathcal{V})$, then the visible genealogy $\mathcal{V}$ is incompatible with $\mathcal{X}_t = x$. Therefore, we put $w(t,x,\mathcal{V}) = 0$, for all x such that $I(x) < \ell(t,\mathcal{V})$. The proper initial condition is clearly

$$w(0,x,\mathcal{V}_t) = p_0(x). \tag{22}$$

With these definitions, it is a straightforward matter to verify that the unique w satisfying Eqs. 21 and 22 also satisfies Eqs. 17. In particular, the likelihood of a visible genealogy $\mathcal{V}$ is

$$\mathcal{L}_{\mathcal{V}} = \sum_x w(t(\mathcal{V}),x,\mathcal{V}). \tag{23}$$

6. Illustrative examples.

In this section, we return to some of the examples of §3. For each one, we specialize Eq. 21 and perform likelihood calculations on simulated data. Codes for the following (and for all the figures in the paper) are available as a Zenodo digital archive.

Linear birth-death-sampling process. For any given model, Eqs. 19–21 take specific forms. In the case of the linear birth-death-sampling process (§3), if we write $w(t,n)$ for the partial weight associated with population size n, Eq. 19 becomes

$$\frac{\partial w}{\partial t} = \lambda\,(n-1)\,\left(1-\frac{\binom{\ell}{2}}{\binom{n}{2}}\right)\,w(t,n-1) + \delta\,(n+1)\,w(t,n+1) - (\lambda+\delta+\psi)\,n\,w(t,n), \tag{24}$$

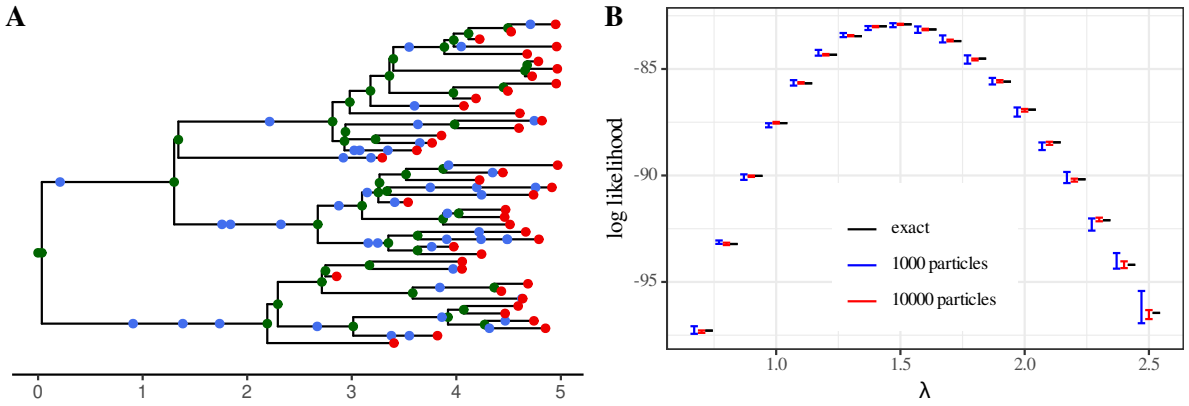

FIGURE 7. **Computing the likelihood for genealogies induced by the linear birth-death-sampling process. (A)** A simulated genealogy for $\lambda = 1.5$, $\mu = 0.8$, $\psi = 1$. As usual, blue and red points correspond to samples; green points represent branch-points. **(B)** Change in the log likelihood as a function of position along a line passing through the true parameter in the λ-direction. The red and blue error bars show the estimates obtained using the particle filter (mean ± 2 s.e.), with different amounts of effort (i.e., number of particles); the black shows the exact log likelihood, which is available in closed form in this case. With increasing computational effort, the estimates converge on the exact value.

which holds for $t \notin \mathsf{E}(\mathcal{V})$. We can integrate Eq. 24 forward in time from each genealogical event to the next. We then adjust w according to the nature of the event, as follows:

$$w(t,n) = \begin{cases} \frac{\lambda\,(n-1)}{\mu}\,\frac{1}{\binom{n}{2}}\,w(t\text{-},n-1), & t \in \mathsf{C}(\mathcal{V}), \\ \frac{\psi}{\mu}\,w(t\text{-},n), & t \in \mathsf{D}(\mathcal{V}), \\ \frac{\psi}{\mu}\,(n-\ell)\,w(t\text{-},n), & t \in \mathsf{L}(\mathcal{V}). \end{cases} \tag{25}$$

In these equations, it is understood that $w(t,n) = 0$ for $n < \ell(t)$.

The astute reader will have noticed that in Eqs. 24 and 25, the only state variable upon which w depends is the population size n, while in §3, we specify a two-dimensional state space for the linear birth-death-sampling process with coordinates (n,g), g being the cumulative number of samples to time t. Since, as is easily verified, solving Eq. 21 results in $w(t,n,g,\mathcal{V}) \neq 0$ if and only if g is precisely equal to the number of samples in $\mathcal{V}|_t$, there is no need to keep track of the the dependence of w on g. The same simplification is used in the other examples we consider below.

Fig. 7 shows the results of a calculation such as might form part of a data analysis. For genealogies induced by the linear birth-death-sampling process, exact expressions for the likelihood are available (Stadler, 2010). Accordingly, we compare the results obtained by integrating Eqs. 24 and 25, using a sequential Monte Carlo integration scheme (King et al., 2016) to these exact results.

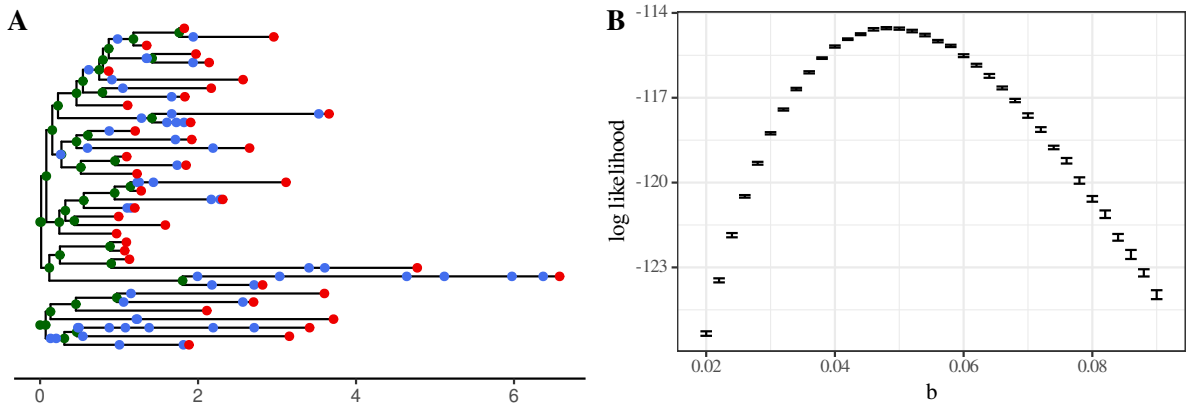

FIGURE 8. **Computing the likelihood for genealogies induced by the SIR process.** **(A)** A simulated genealogy for $b = 0.04$, $\gamma = 1$, $\psi = 1$. The population is of size 100, with 3 infectives at time 0. **(B)** Change in the log likelihood as a function of position along a line passing through the true parameter in the b-direction. The error bars show the estimates obtained using the particle filter (mean ± 2 s.e.).

SIR model. In the case of the SIR model (§3), Eqs. 19 and 20 become, for $t \notin \mathsf{E}(\mathcal{V})$,

$$\begin{aligned}\frac{\partial w}{\partial t} &= b\,(s+1)\,(i-1)\left(1-\frac{\binom{\ell}{2}}{\binom{i}{2}}\right) w(t,s+1,i-1)\\ &\quad + \gamma\,(i+1)\,w(t,s,i+1) - (b\,s\,i+\gamma\,i+\psi\,i)\,w(t,s,i),\end{aligned}$$

while for $t \in \mathsf{E}(\mathcal{V})$,

$$w(t,s,i) = \begin{cases} \frac{b\,(s+1)\,(i-1)}{\mu}\,\frac{1}{\binom{i}{2}}\,w(t\text{-},s+1,i-1), & t \in \mathsf{C}(\mathcal{V}),\\ \frac{\psi}{\mu}\,w(t\text{-},s,i), & t \in \mathsf{D}(\mathcal{V}),\\ \frac{\psi}{\mu}\,(i-\ell)\,w(t\text{-},s,i), & t \in \mathsf{L}(\mathcal{V}).\end{cases}$$

Again, for simplicity, we have taken all parameters to be constant in time and we understand that $w(t,s,i) = 0$ for $i < \ell(t)$. Fig. 8 shows the results of a typical calculation performed using these equations.

SIRS model. The case of the SIRS model (§3) is similar. Again, taking all parameters to be constant for simplicity, Eqs. 19 and 20 assume the following forms. For $t \notin \mathsf{E}(\mathcal{V})$,

$$\begin{aligned}\frac{\partial w}{\partial t} &= b\,(s+1)\,(i-1)\left(1-\frac{\binom{\ell}{2}}{\binom{i}{2}}\right) w(t,s+1,i-1,r) + \gamma\,(i+1)\,w(t,s,i+1,r-1)\\ &\quad + \delta\,(r+1)\,w(t,s-1,i,r+1) - (b\,s\,i+\gamma\,i+\delta\,r+\psi\,i)\,w(t,s,i,r).\end{aligned}$$

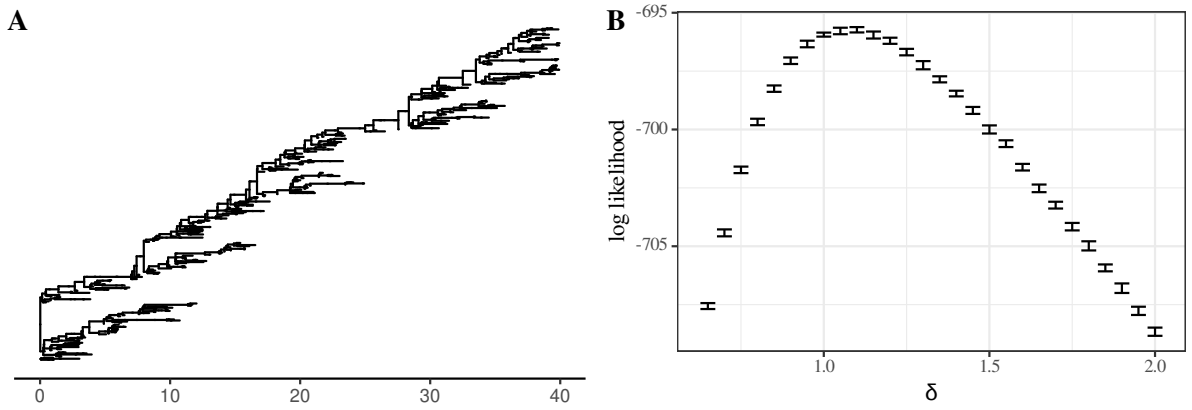

FIGURE 9. **Computing the likelihood for genealogies induced by the SIRS process. (A)** A simulated genealogy for $b = 0.04$, $\gamma = 2$, $\psi = 1$, and $\delta = 1$. The population is of size 100, with 3 infectives at time 0. **(B)** Change in the log likelihood as a function of position along a line passing through the true parameter in the δ-direction. The error bars show the estimates obtained using the particle filter (mean ± 2 s.e.).

For $t \in \mathsf{E}(\mathcal{V})$, we have

$$w(t,s,i,r) = \begin{cases} \frac{b\,(s+1)\,(i-1)}{\mu}\,\frac{1}{\binom{i}{2}}\,w(t\text{-},s+1,i-1,r), & t \in \mathsf{C}(\mathcal{V}), \\ \frac{\psi}{\mu}\,w(t\text{-},s,i,r), & t \in \mathsf{D}(\mathcal{V}), \\ \frac{\psi}{\mu}\,(i-\ell)\,w(t\text{-},s,i,r), & t \in \mathsf{L}(\mathcal{V}). \end{cases}$$

As usual, we have $w(t,s,i,r) = 0$ whenever $i < \ell(t)$. Fig. 9 shows the results of a calculation performed using these equations.

7. Discussion.

Numerical solution of Eq. 21 is readily achieved, as in §6, by applying a Monte Carlo Feynman-Kac approach, i.e., by simulating individual weighted realizations (particles) of the population process between genealogical event times, updating the weights of the particles appropriately at each genealogical event (i.e., according to Eq. 20). Importantly, because the genealogical events correspond to real events in the population process, not only the weights, but also the states, of the particles must be adjusted according to Eq. 20. Thus, this approach leads to a modified version of the standard sequential Monte Carlo (particle filter) algorithm (e.g., King et al., 2016).

Indeed, in the special case that the event rates α are time-homogeneous, the sequential Monte Carlo algorithm for evaluating Eq. 23 just described is precisely that proposed by Vaughan et al. (2019). Our work here therefore shows how this approach can be extended to a much broader class of models. A forthcoming paper will extend the class still farther, to encompass simultaneous birth, death, and sampling events, and will describe several alternative algorithms for the numerical solution of Eq. 21.

One key feature of the algorithms proposed here is that they enjoy the *plug-and-play property* (He et al., 2010). An algorithm for inference on partially observed Markov processes is said to be plug-and-play if it operates without ever needing to evaluate the Markov transition probability densities.

Numerical solution of Eq. 21 requires only that one be able to simulate the population process. Such methods are attractive inasmuch as they allow consideration of models that are scientifically interesting but mathematically inconvenient by other approaches, since it is typically the case that models can be simulated even when their probability density functions are mathematically intractable. The fact that plug-and-play methods avoid the need for method-specific approximations also facilitates objective model comparison, since it puts models on an even footing with respect to inference methodology (He et al., 2010; King et al., 2016).

As we have mentioned at various points above, it is possible to extend the constructions developed here to more general situations. In particular, the state space for the population process $\mathcal{X}$ can be taken to be any separable Banach space, so long as the birth, death, sampling, and population size functions, B, D, G, and I, remain well defined. One can also relax the requirements that B, D, and G have ranges in $\{0, 1\}$ and the assumption that these different kinds of events never coincide. It is necessary to relax these assumptions, for example, if one wishes to entertain models of superspreading (Lloyd-Smith et al., 2005), or more generally to allow for overdispersion in the latent population process, often an important component of well-fitting models (He et al., 2010; Bretó & Ionides, 2011). In accommodating these extensions, the combinatoric arguments of Theorems 1 and 2 are more intricate, but remain tractable. We will describe these extensions in a future paper.

The first line of Eq. 21 resembles the Kolmogorov forward equation (Eqs. 3) for the population process. Accordingly, in the absence of sampling, Eq. 21 preserves the normalization of w, i.e., $\sum_x w(t, x, \mathcal{V}) = 1$. The remaining lines represent the accumulation of evidence supporting each of the alternative hypotheses $\mathcal{X}_t = x$. Note that some of the evidence assimilated into w at any time t is derived from data that are only collected *after* time t. Because of this, the partial weights are not measurable in the filtrations induced by the Markov processes of Fig. 1. One must therefore resist the temptation to over-interpret the partial weights $w(t, x, \mathcal{V})$ for $t < t(\mathcal{V})$: they should be viewed merely as elements of an algorithm that ultimately yields the full likelihood. Some previous full-information phylodynamic approaches have made mistakes of this kind (e.g., Rasmussen et al., 2011; Leventhal et al., 2014). The correct interpretation of Eq. 21 is that a portion of the information in each sample is *referred* to earlier times. This referral is in some sense the reverse of the evolutionary process whereby information about transmission and recoveries (or births and deaths, or speciations and extinctions) is stored in the genome. From this point of view, the genealogy itself can be understood as nothing other than a prescription for this information referral.

Acknowledgments.

The authors gratefully acknowledge useful conversations with Alexandre Bouchard-Côté, Simon Frost, Katia Koelle, Vladimir Minin, Mitchell Newberry, David Rasmussen, Jonathan Terhorst, Erik Volz, and two anonymous reviewers. 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 was supported by a fellowship from the Michigan Institute for Data Science.

References.

Alizon, S., Lion, S., Murall, C. L., & Abbate, J. L. (2014) Quantifying the epidemic spread of Ebola virus (EBOV) in Sierra Leone using phylodynamics. *Virulence* **5**:825–827.

Andersen, P. K., Borgan, Ø., Gill, R. D., & Keiding, N. (1993) *Statistical models based on counting processes*. Springer Series in Statistics. New York: Springer-Verlag.

Bedford, T., Greninger, A. L., Roychoudhury, P., Starita, L. M., Famulare, M., Huang, M.-L., Nalla, A., Pepper, G., Reinhardt, A., Xie, H., Shrestha, L., Nguyen, T. N., Adler, A., Brandstetter, E., Cho, S., Giroux, D., Han, P. D., Fay, K., Frazar, C. D., Ilcisin, M., Lacombe, K., Lee, J., Kiavand, A., Richardson, M., Sibley, T. R., Truong, M., Wolf, C. R., Nickerson, D. A., Rieder, M. J., Englund, J. A., Hadfield, J., Hodcroft, E. B., Huddleston, J., Moncla, L. H., Müller, N. F., Neher, R. A., Deng, X., Gu, W., Federman, S., Chiu, C., Duchin, J. S., Gautom, R., Melly, G., Hiatt, B., Dykema, P., Lindquist, S., Queen, K., Tao, Y., Uehara, A., Tong, S., MacCannell, D., Armstrong, G. L., Baird, G. S., Chu, H. Y., Shendure, J., & Jerome, K. R. (2020) Cryptic transmission of SARS-CoV-2 in Washington state. *Science* **370**:571.

Biek, R., Pybus, O. G., Lloyd-Smith, J. O., & Didelot, X. (2015) Measurably evolving pathogens in the genomic era. *Trends in Ecology & Evolution* **30**:306–313.

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.

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.

Dearlove, B. & Wilson, D. J. (2013) Coalescent inference for infectious disease: meta-analysis of hepatitis C. *Philosophical Transactions of the Royal Society of London, Series B* **368**:20120314.

Donnelly, P. & Kurtz, T. G. (1996) A countable representation of the Fleming-Viot measure-valued diffusion. *Annals of Probability* **24**:698–742.

Donnelly, P. & Kurtz, T. G. (1999) Particle representations for measure-valued population models. *Annals of Probability* **27**:166–205.

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

Etheridge, A. M. & Kurtz, T. G. (2019) Genealogical constructions of population models. *Annals of Probability* **47**:1827–1910.

Ethier, S. N. & Kurtz, T. G. (2009) *Markov processes: characterization and convergence*. John Wiley & Sons.

Faria, N. R., Rambaut, A., Suchard, M. A., Baele, G., Bedford, T., Ward, M. J., Tatem, A. J., Sousa, J. D., Arinaminpathy, N., Pépin, J., Posada, D., Peeters, M., Pybus, O. G., & Lemey, P. (2014) HIV epidemiology. the early spread and epidemic ignition of HIV-1 in human populations. *Science* **346**:56–61.

Fasiolo, M., Pya, N., & Wood, S. N. (2016) A comparison of inferential methods for highly nonlinear state space models in ecology and epidemiology. *Statistical Science* **31**:96–118.

Frost, S. D. W., Pybus, O. G., Gog, J. R., Viboud, C., Bonhoeffer, S., & Bedford, T. (2015) Eight challenges in phylodynamic inference. *Epidemics* **10**:88–92.

Fu, Y. X. (2006) Exact coalescent for the Wright-Fisher model. *Theoretical Population Biology* **69**:385–394.

Geoghegan, J. L., Tan, L. V., Kuhnert, D., Halpin, R. A., Lin, X. D., Simenauer, A., Akopov, A., Das, S. R., Stockwell, T. B., Shrivastava, S., Ngoc, N. M., Uyen, L. T. T., Tuyen, N. T. K., Thanh, T. T., Hang, V. T. T., Qui, P. T., Hung, N. T., Khanh, T. H., Thinh, L. Q., Nhan, L. N. T., Van, H. M. T., Viet, D. C., Tuan, H. M., Viet, H. L., Hien, T. T., Chau, N. V. V., Thwaites, G., Grenfell, B. T., Stadler, T., Wentworth, D. E., Holmes, E. C., & Van Doorn, H. R. (2015) Phylodynamics of enterovirus A71-associated hand, foot, and mouth disease in Viet Nam. *Journal of Virology* **89**:8871–8879.

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

Gill, M. S., Lemey, P., Bennett, S. N., Biek, R., & Suchard, M. A. (2016) Understanding past population dynamics: Bayesian coalescent-based modeling with covariates. *Systematic Biology* **65**:1041–1056.

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.

Hadfield, J., Megill, C., Bell, S. M., Huddleston, J., Potter, B., Callender, C., Sagulenko, P., Bedford, T., & Neher, R. A. (2018) Nextstrain: real-time tracking of pathogen evolution. *Bioinformatics* **34**:4121–4123.

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.

Kallenberg, O. (1997) *Foundations of Modern Probability*. Springer, 2nd edn.

King, A. A., Nguyen, D., & Ionides, E. L. (2016) Statistical inference for partially observed Markov processes via the R package pomp. *Journal of Statistical Software* **69**:1–43.

Kingman, J. F. C. (1982a) The coalescent. *Stochastic Processes and their Applications* **13**:235–248.

Kingman, J. F. C. (1982b) Exchangeability and the evolution of large populations. In G. Koch & F. Spizzichino (eds.), *Exchangeability in Probability and Statistics*, pp. 97–112. North-Holland, Amsterdam.

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

Koelle, K., Cobey, S., Grenfell, B., & Pascual, M. (2006) Epochal evolution shapes the phylodynamics of interpandemic influenza A (H3N2) in humans. *Science* **314**:1898–1903.

Koelle, K. & Rasmussen, D. A. (2012) Rates of coalescence for common epidemiological models at equilibrium. *Journal of the Royal Society, Interface* **9**:997–1007.

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

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.

Li, L. M., Grassly, N. C., & Fraser, C. (2017) Quantifying transmission heterogeneity using both pathogen phylogenies and incidence time series. *Molecular Biology and Evolution* **34**:2982–2995.

Lloyd-Smith, J. O., Schreiber, S. J., Kopp, P. E., & Getz, W. M. (2005) Superspreading and the effect of individual variation on disease emergence. *Nature* **438**:355–359.

Luciani, F., Sisson, S. A., Jiang, H., Francis, A. R., & Tanaka, M. M. (2009) The epidemiological fitness cost of drug resistance in Mycobacterium tuberculosis. *Proceedings of the National Academy of Sciences of the U.S.A.* **106**:14711–14715.

MacPherson, A., Louca, S., McLaughlin, A., Joy, J. B., & Pennell, M. W. (2021) Unifying phylogenetic birth-death models in epidemiology and macroevolution. *Systematic Biology* .

Maddison, W. P., Midford, P. E., & Otto, S. P. (2007) Estimating a binary character's effect on speciation and extinction. *Systematic Biology* **56**:701–710.

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

Nee, S., May, R. M., & Harvey, P. H. (1994) The reconstructed evolutionary process. *Philosophical Transactions of the Royal Society of London, Series B* **344**:305–311.

O'Dea, E. B. & Wilke, C. O. (2011) Contact heterogeneity and phylodynamics: How contact networks shape parasite evolutionary trees. *Interdisciplinary Perspectives on Infectious Diseases* **2011**.

Poon, A. F. Y. (2015) Phylodynamic inference with kernel ABC and its application to HIV epidemiology. *Molecular Biology and Evolution* pp. –.

Ragonnet-Cronin, M., Boyd, O., Geidelberg, L., Jorgensen, D., Nascimento, F. F., Siveroni, I., Johnson, R. A., Baguelin, M., Cucunubá, Z. M., Jauneikaite, E., Mishra, S., Watson, O. J., Ferguson, N., Cori, A., Donnelly, C. A., & Volz, E. (2021) Genetic evidence for the association between COVID-19 epidemic severity and timing of non-pharmaceutical interventions. *Nature Communications* **12**.

Rasmussen, D. A., Boni, M. F., & Koelle, K. (2014) Reconciling phylodynamics with epidemiology: the case of dengue virus in southern Vietnam. *Molecular Biology and Evolution* **31**:258–271.

Rasmussen, D. A., Ratmann, O., & Koelle, K. (2011) Inference for nonlinear epidemiological models using genealogies and time series. *PLOS Computational Biology* **7**:e1002136.

Ratmann, O., Donker, G., Meijer, A., Fraser, C., & Koelle, K. (2012) Phylodynamic inference and model assessment with approximate Bayesian computation: influenza as a case study. *PLOS Computational Biology* **8**:e1002835.

Sisson, S. A., Fan, Y., & Tanaka, M. M. (2007) Sequential Monte Carlo without likelihoods. *Proceedings of the National Academy of Sciences of the U.S.A.* **104**:1760–1765.

Smith, R. A., Ionides, E. L., & King, A. A. (2017) Infectious disease dynamics inferred from genetic data via sequential Monte Carlo. *Molecular Biology and Evolution* **34**:2065–2084.

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

Stadler, T., Kouyos, R., von Wyl, V., Yerly, S., Böni, J., Bürgisser, P., Klimkait, T., Joos, B., Rieder, P., Xie, D., Günthard, H. F., Drummond, A. J., Bonhoeffer, S., & Study, S. H. I. V. C. (2012) Estimating the basic reproductive number from viral sequence data. *Molecular Biology and Evolution* **29**:347–357.

Stadler, T., Pybus, O. G., & Stumpf, M. P. H. (2021) Phylodynamics for cell biologists. *Science* **371**:eaah6266.

Tavaré, S. (2018) The linear birth–death process: an inferential retrospective. *Advances in Applied Probability* **50**:253–269.

Vaughan, T. G., Leventhal, G. E., Rasmussen, D. A., Drummond, A. J., Welch, D., & Stadler, T. (2019) Estimating epidemic incidence and prevalence from genomic data. *Molecular Biology and Evolution* **36**:1804–1816.

Vijaykrishna, D., Holmes, E. C., Joseph, U., Fourment, M., Su, Y. C. F., Halpin, R., Lee, R. T. C., Deng, Y. M., Gunalan, V., Lin, X. D., Stockwell, T. B., Fedorova, N. B., Zhou, B., Spirason, N., Kuhnert, D., Boskova, V., Stadler, T., Costa, A. M., Dwyer, D. E., Huang, Q. S., Jennings, L. C., Rawlinson, W., Sullivan, S. G., Hurt, A. C., Maurer-Stroh, S., Wentworth, D. E., Smith, G. J. D., & Barr, I. G. (2015) The contrasting phylodynamics of human influenza B viruses. *eLife* **4**:e05055.

Volz, E. M., Ionides, E., Romero-Severson, E. O., Brandt, M.-G., Mokotoff, E., & Koopman, J. S. (2013a) HIV-1 transmission during early infection in men who have sex with men: a phylodynamic analysis. *PLOS Medicine* **10**:e1001568.

Volz, E. M., Koelle, K., & Bedford, T. (2013b) Viral phylodynamics. *PLOS Computational Biology* **9**:e1002947.

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.

Wakeley, J. (2008) *Coalescent Theory: An Introduction*. W. H. Freeman.

Wood, S. N. (2010) Statistical inference for noisy nonlinear ecological dynamic systems. *Nature* **466**:1102–1104.

Zakai, M. (1969) On the optimal filtering of diffusion processes. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete* **11**:230–243.

A. A. KING, DEPARTMENT OF ECOLOGY & EVOLUTIONARY BIOLOGY, CENTER FOR THE STUDY OF COMPLEX SYSTEMS, CENTER FOR COMPUTATIONAL MEDICINE & BIOLOGY, AND MICHIGAN INSTITUTE FOR DATA SCIENCE, UNIVERSITY OF MICHIGAN, ANN ARBOR, MI 48109 USA

Email address: kingaa@umich.edu

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

Q.-Y. LIN, MICHIGAN INSTITUTE FOR DATA SCIENCE, UNIVERSITY OF MICHIGAN, ANN ARBOR, MI 48109 USA

E. L. IONIDES, DEPARTMENT OF STATISTICS AND MICHIGAN INSTITUTE FOR DATA SCIENCE, UNIVERSITY OF MICHIGAN, ANN ARBOR, MI 48109 USA