

# EXACT PHYLODYNAMICS VIA STRUCTURED MARKOV GENEALOGY PROCESSES

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

ABSTRACT. We derive an algorithm for the exact computation of the likelihood of an observed transmission tree.

## 1. Introduction.

Problem of phylodynamics. Factorization of problem into two subproblems.

Relation to previous work. Existing methods (Volz et al., 2009; Stadler, 2010). Large-population, small sample-size approximations.

Extension of previous results (King et al., 2022). Broader class of state-spaces. Accommodating discrete structure.

Classes of Markov processes. Utility and flexibility of Markov assumptions.

Population process induces Markov history and genealogy processes. Using these, we derive equations for the likelihood of a genealogy conditional on the history. We then integrate out the history to obtain nonlinear filtering equations, the solution of which yields the likelihood. These readily lend themselves to a family of sequential Monte Carlo algorithms for computing the likelihood. We demonstrate with several examples.

In the following, we show a Markov population process of the kind that is a staple in epidemiology induces a Markov process on the space of genealogies. We then show how one can compute the likelihood of a given genealogy.

## 2. Mathematical preliminaries.

**2.1. Notation.** Throughout the paper, we will adopt the convention that a bold-face symbol (e.g.,  $\mathbf{X}$ ), denotes a random element. We will be concerned with a variety of stochastic processes, in both discrete and continuous time. In both cases, we will use a subscript to indicate the time parameter: e.g.,  $\mathbf{X}_t$  or  $Gr_k$ , where  $t$  takes values in the non-negative reals  $\mathbb{R}_+$  and  $k$  in the non-negative integers  $\mathbb{Z}_+$ . In the case of continuous-time processes, we will assume that sample paths are right-continuous with left limits (i.e., càdlàg). We will frequently need to refer to the left-limit of such a process. Accordingly, if  $\mathbf{Z}_t$  is a càdlàg random process, we define

$$\tilde{\mathbf{Z}}_t := \begin{cases} \lim_{t' \uparrow t} \mathbf{Z}_{t'}, & t > 0, \\ \mathbf{Z}_0, & t = 0. \end{cases}$$

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

If  $Zr_t$ ,  $t \in \mathbb{R}_+$  is a pure jump process, knowledge of its sample path is equivalent to knowledge of the number,  $\mathbf{K}_t$ , of jumps it has taken as of time  $t$ , the jump times  $\hat{T}_k$ , and the embedded chain  $\hat{\mathbf{Z}}_k := \mathbf{Z}_{\hat{T}_k}$ ,  $k = 0, \dots, \mathbf{K}_t$ . In particular, if we adopt the convention that  $\hat{T}_0 = 0$  and  $\hat{T}_{\mathbf{K}_t+1} = t$ , then  $\mathbf{Z}_t = \hat{\mathbf{Z}}_k$  for  $t \in [\hat{T}_k, \hat{T}_{k+1})$ ,  $k = 0, \dots, \mathbf{K}_t$ . If  $\mathbf{Z}$  is such a pure jump process, let  $\hat{\mathbf{Z}}_t = \{\hat{T}_1, \dots, \hat{T}_{\mathbf{K}_t}\}$  denote the set of its event times in the interval  $[0, t]$ .

**2.2. Population process.** Motivating examples: compartmental models. Wide variety of models. Linear chain trick. Migration, superspreading, competition between strains.

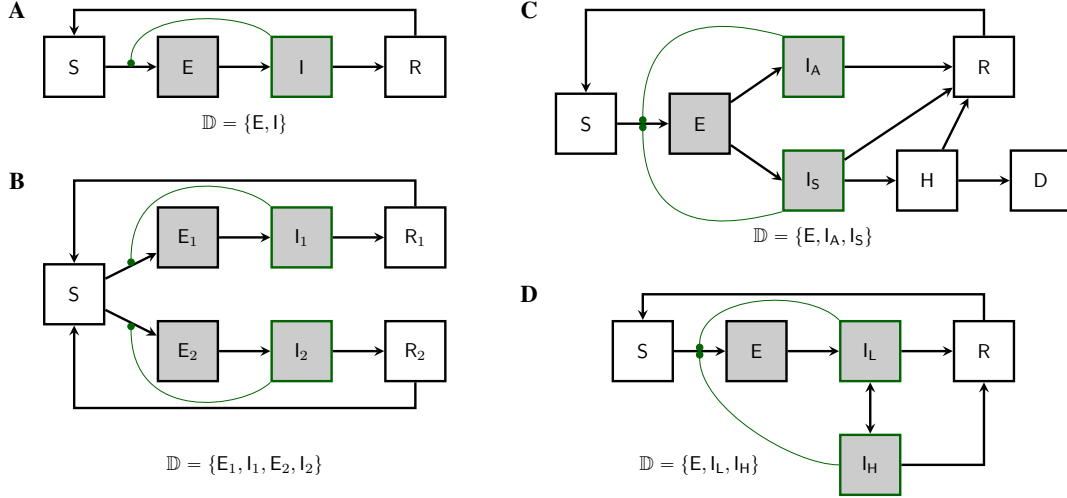


FIGURE 1. Examples of discretely-structured population models. Demes are shaded. Compartments containing infectious hosts are outlined in blue. Curved green lines connect transmission rates with the compartments whose occupancies control their modulation; each such connection gives rise to a nonlinearity in the model. **(A)** An SEIRS model. Susceptible individuals (S), once infected, enter a transient incubation phase (E) before they become infectious (I). Upon recovery (R), individuals experience immunity from reinfection. If this immunity wanes, they re-enter the susceptible compartment. Pathogen lineages are to be found in hosts within the E and I compartments only. Accordingly, there are two demes ( $\mathbb{D} = \{E, I\}$ ). If there is exactly one lineage per host, then the occupancy,  $n(\mathbf{X}_t) = (n_E(\mathbf{X}_t), n_I(\mathbf{X}_t))$ , is the integer 2-vector giving the numbers of hosts in the respective compartments. **(B)** In this four-deme model, two distinct pathogen strains compete for susceptibles. **(C)** A three-deme model for SARS-CoV-2 infection. After an incubation period, individuals may develop asymptomatic infection ( $I_A$ ). If they do not recover, symptomatically infected individuals ( $I_S$ ) can progress to hospitalization (H) and death (D). **(D)** A three-deme model with heterogeneity in transmission behavior. Contagious individuals move randomly between low-transmission ( $I_L$ ) and high-transmission ( $I_H$ ) behaviors.

We will assume that our population process is a time-inhomogeneous Markov jump process,  $\mathbf{X}_t, t \in \mathbb{R}_+$ , taking values in some space  $\mathbb{X}$ . In earlier work (King et al., 2022), we limited ourselves to the case  $\mathbb{X} = \mathbb{Z}^d$ , but here we assume only that  $\mathbb{X}$  is a complete metric space with a countable dense subset, i.e., a Polish space. The population process is completely specified by its initial-state distribution,  $p_0$ , and its transition rates  $\alpha$ . In particular, we suppose that

$$\text{Prob}[\mathbf{X}_0 \in \mathcal{E}] = \int_{\mathcal{E}} p_0(x) dx \quad (1)$$

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

$$\alpha(t, x, x') \geq 0, \quad \int_{\mathbb{X}} \alpha(t, x, x') dx' < \infty,$$

for all  $t \in \mathbb{R}_+$  and  $x, x' \in \mathbb{X}$ . Henceforth, we understand that integrals are taken over all of  $\mathbb{X}$  unless otherwise specified. Let  $\mathbf{K}_t$  be the number of jumps that  $\mathbf{X}$  has taken by time  $t$ . We assume that  $\mathbf{K}_t$  is a simple counting

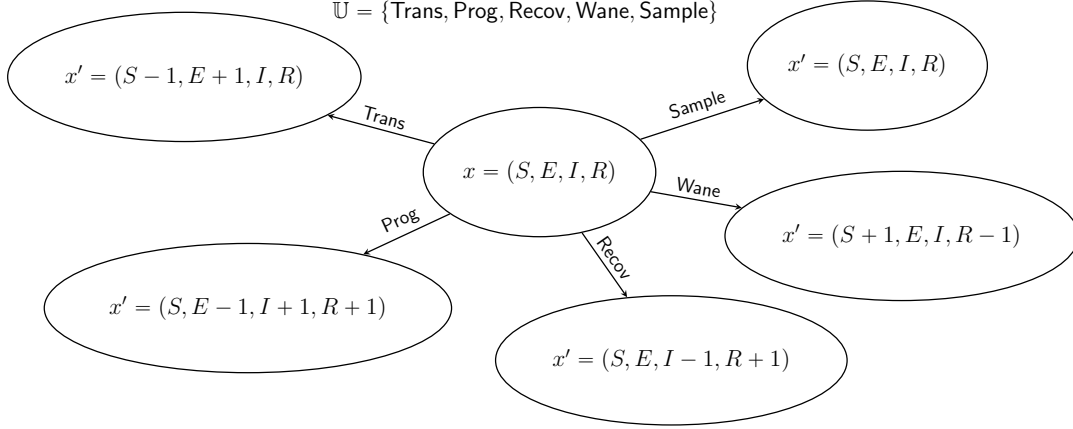


FIGURE 2. Markov state transition diagram for the SEIRS model depicted in Fig. 1A. The state,  $x$ , is characterized by four numbers,  $S$ ,  $E$ ,  $I$ , and  $R$ . From a given state  $x$ , there are five possible kinds of jumps  $x \mapsto x'$ . Accordingly, the set,  $\mathbb{U}$ , of jump marks has five elements. Each of these is of a different type: Trans (transmission) is of birth type, Prog (progression) is of migration type, Recov (recovery) is of death type, Sample (sampling) is of sample type, and Wane (loss or waning of immunity) is of neutral type. See §3.2 for a description of these jump types. Note that, in this formulation, when a sampling event occurs, the state does not change.

process so that

$$\begin{aligned} \text{Prob}[\mathbf{K}_{t+\Delta} = n+1 \mid \mathbf{K}_t = n] &= \Delta \int \alpha(t, x, x') dx' + o(\Delta), \\ \text{Prob}[\mathbf{K}_{t+\Delta} > n+1 \mid \mathbf{K}_t = n] &= o(\Delta), \\ \text{Prob}[\mathbf{X}_{t+\Delta} \in \mathcal{E} \mid \mathbf{X}_t = x, \mathbf{K}_{t+\Delta} - \mathbf{K}_t = 1] &= \frac{\int_{\mathcal{E}} \alpha(t, x, x') dx'}{\int \alpha(t, x, x') dx'}. \end{aligned}$$

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

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

$$\frac{\partial w}{\partial t}(t, x) = \int w(t, x') \alpha(t, x', x) dx' - \int w(t, x) \alpha(t, x, x') dx', \quad (2)$$

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

Another perspective on the Markov processes is to be had from its Markov state transition diagram (Fig. 2).

**2.2.2. Structured populations, demes, and deme occupancy.** In an *unstructured* Markov population process, every lineage is exactly like every other. King et al. (2022) showed how every such process induces an unstructured Markov genealogy process. Here, our aim is to expand the theory considerably by allowing our population of lineages to have discrete structure. In particular, we suppose that there are a countable set of subpopulations that differ in their vital rates, but within each of which, individual lineages are statistically identical. We call these subpopulations *demes*, and use the symbol  $\mathbb{D}$  to denote an index set for them.

For any  $i \in \mathbb{D}$ , we let  $n_i(\mathbf{X}_t)$  denote the number of lineages present in deme  $i$  at time  $t$ , i.e., the *occupancy* of deme  $i$ . Thus  $n(\mathbf{X}_t) \in \mathbb{Z}_+^{\mathbb{D}}$  is the vector of deme occupancies.

**2.2.3. Jump marks.** In the following, it will be useful to break the jumps into distinct categories. For this purpose, we let  $\mathbb{U}$  be a countable set of jump *marks* such that

$$\alpha(t, x, x') = \sum_u \alpha_u(t, x, x').$$

In Fig. 2, we use the marks to distinguish events with distinct effects on genealogical relationships. Here and in the following, sums over  $u$  are taken over the whole of  $\mathbb{U}$  unless otherwise indicated.

Let us define the *jump mark* process,  $\mathbf{U}_t$ , to be the mark of the latest jump as of time  $t$ . As usual, we take the sample paths of  $\mathbf{U}_t$  to be càdlàg. Observe that  $(\mathbf{X}_t, \mathbf{U}_t)$  is a Markov process, though  $\mathbf{U}_t$  is not.

### 2.3. Examples.

#### 2.3.1. SEIRS model.

#### 2.3.2. SIIR model.

#### 2.3.3. Linear birth-death model.

#### 2.3.4. Moran model and the Kingman coalescent.

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

$$\begin{aligned} \pi^{\mathbf{H}}(d\mathbf{H}_t) = & p_0(\hat{\mathbf{X}}_0) d\hat{\mathbf{X}}_0 \prod_{k=1}^{\mathbf{K}_t} \alpha_{\hat{\mathbf{U}}_k}(\hat{\mathbf{T}}_k, \hat{\mathbf{X}}_{k-1}, \hat{\mathbf{X}}_k) d\hat{\mathbf{X}}_k d\hat{\mathbf{T}}_k \\ & \times \exp \left( - \sum_{k=0}^{\mathbf{K}_t} \int_{\hat{\mathbf{T}}_k}^{\hat{\mathbf{T}}_{k+1}} \sum_u \int \alpha_u(t', \hat{\mathbf{X}}_k, x') dx' dt' \right), \end{aligned}$$

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

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

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

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

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

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

## 3. The induced genealogy process.

**3.1. Genealogies.** For our purposes, a *genealogy*,  $G_T$ , is a labeled, time-calibrated tree. It encapsulates the relationships of shared ancestry among a set of lineages that are alive at time  $T \in \mathbb{R}_+$ , and perhaps a set of samples collected at earlier times. There are three distinct kinds of nodes: (i) *tip nodes*, which represent labeled living lineages; (ii) *internal nodes*, which represent events at which lineages diverged and/or moved from one deme to another; and (iii) *sample nodes*, which represent labeled samples. Each node  $a$  is associated with a specific time,  $t(a)$ . In particular, if  $a$  is a tip node in  $G_T$ , then  $t(a) = T$ ; if  $a$  is a sample node, then  $t(a)$  is the time at which the sample was taken. Moreover, if node  $a$  is ancestral to node  $a'$ , then  $t(a) \leq t(a')$  and  $t(a') - t(a)$  is the distance between  $a$  and  $a'$  along the genealogy. We let  $\hat{G}_T$  denote the set of all node-times of the genealogy  $G_T$ . Importantly, a genealogy informs us not only about the shared ancestry of any pair of lineages, but also about

where in the set of demes any given lineage was at all times. Accordingly, we can visualize a genealogy as a tree, the nodes and edges of which are painted, with a distinct color for each deme.

Formally, we define a genealogy,  $G_T$ , to be an ordered pair  $(y, z)$ . The  $z$  element specifies the topological structure, as follows. Let  $L(G_T)$  be the set of all tip and sample nodes in  $G_T$  and define  $P$  to be the set of all partitions of  $L(G_T)$ . The partition *fineness* defines a partial order on  $P$ . Specifically, for  $p, p' \in P$ , we say  $p \preceq p'$  if and only if for every  $c' \in p'$  there is a unique  $c \in p$  such that  $c \supseteq c'$ . The topological structure of  $G_T$  is defined by a map  $z : [0, T] \rightarrow P$  that is càdlàg and monotone in the sense that  $t_1 < t_2$  implies  $z_{t_1} \preceq z_{t_2}$ .

The  $y$  element of  $G_T$  specifies the coloring of branches and location of nodes. In particular, for  $a \in L(G_T)$  and  $t \in [0, T]$ , define  $d_t(a)$  to be the deme in which the lineage of  $a$  was found at time  $t$ . To make this well-defined, let  $d_t(a) = \bar{\partial}$  for  $t \in [t(a), T]$ ; we call  $\bar{\partial}$  the *underdeme* and define  $\bar{\mathbb{D}} := \mathbb{D} \cup \{\bar{\partial}\}$ . Thus  $d : [0, T] \times L(G_T) \rightarrow \bar{\mathbb{D}}$  is a well-defined càdlàg function.

Since not every node along a lineage corresponds to a change of deme, we also count the number of nodes along each lineage. Let  $m : [0, T] \times L(G_T) \rightarrow \mathbb{Z}_+$  be such that  $m_t(a)$  is the number of internal or sample nodes encountered along lineage  $a$  as one goes from time 0 to time  $t \leq T$  along the genealogy. For  $t \in [t(a), T]$ , let  $m_t(a) = m_{t(a)}(a)$ . Clearly,  $m_t(a)$  is a simple counting process, with  $m_0(a) = 0$  for all  $a$ .

Finally, we define the map  $y : [0, T] \times L(G_T) \rightarrow \bar{\mathbb{D}} \times \mathbb{Z}_+$  by  $y_t(a) = (d_t(a), m_t(a))$ . The two càdlàg functions  $y$  and  $z$  specify the genealogy  $G_T$  completely and uniquely. When the genealogy is not clear from context, we write  $y^G$  and  $z^G$  to refer to the respective components of genealogy  $G$ .

Clearly, the topological structure of a tree imposes constraints on the permissible functions  $y$  that describe the painting and node placement along it. In particular,  $y$  must respect the topological structure of  $G_T$ , such that if two lineages share a common ancestor at time  $t$ , both lineages must have been in the same deme at that time and the number of previous events along the lineage must agree. In other words, if  $a, b \in p \in z_t$ , then  $y_t(a) = y_t(b)$ . Let  $\mathbb{Y}^z$  denote the space of all functions that obey this constraint at every  $t$ . Let  $\mathbb{Y}_t^z \subseteq \bar{\mathbb{D}} \times \mathbb{Z}_+$  be the time- $t$  section of  $\mathbb{Y}^z$ , i.e.,  $\mathbb{Y}_t^z = \bigcup_{y \in \mathbb{Y}^z} \{y_t\}$ .

**3.2. Jump types.** We will show how a given population process naturally induces a process in the space of genealogies. Because different kinds of jumps have different effects on the genealogy, we need a way of distinguishing them. In particular, there are five distinct *pure types* of jumps:

- (a) *Birth-type jumps* result in the branching of one or more new lineages, each from some existing lineages. Examples of birth-type events include transmission events, speciations, and actual births. For simplicity, we assume that all new lineages arising from a birth event share the same parent.
- (b) *Death-type jumps* result in the extinction of one or more lineages. Examples include recovery from infection, death of a host, and species extinctions. We allow for the possibility that multiple lineages die simultaneously.
- (c) *Migration-type jumps* result in the movement of a lineage from one deme to another. Spatial movements, changes in host age or behavior, and progression of an infection can all be represented as migration-type events. We permit multiple movements to occur simultaneously.
- (d) *Sample-type jumps* result in the collection of a sample from a lineage but do not in themselves affect the inventory process. We allow for the possibility that multiple samples are collected in a single sample event. In this case, we assume that, in each sample-type event, each extant lineage is sampled at most once.
- (e) *Neutral-type jumps* result in no change to any of the lineages.

Fig. 2 depicts an example with all five of the pure types. It is not necessary that a jump be of a pure type. Compound jumps partake of more than one type. For example, a sample/death-type event, in which a lineage is simultaneously sampled and removed, has been proposed (Leventhal et al., 2014), as have birth/death events in which one lineage reproduces at the same moment that another dies (e.g., the Moran process (Moran, 1958)). The theory presented here places few restrictions on the complexity of the events that can occur.

**3.3. Genealogy process.** We now show how a given population process induces a stochastic process,  $G_t$ , on the space of genealogies. In the case of unstructured population processes (i.e., those having a single deme), King

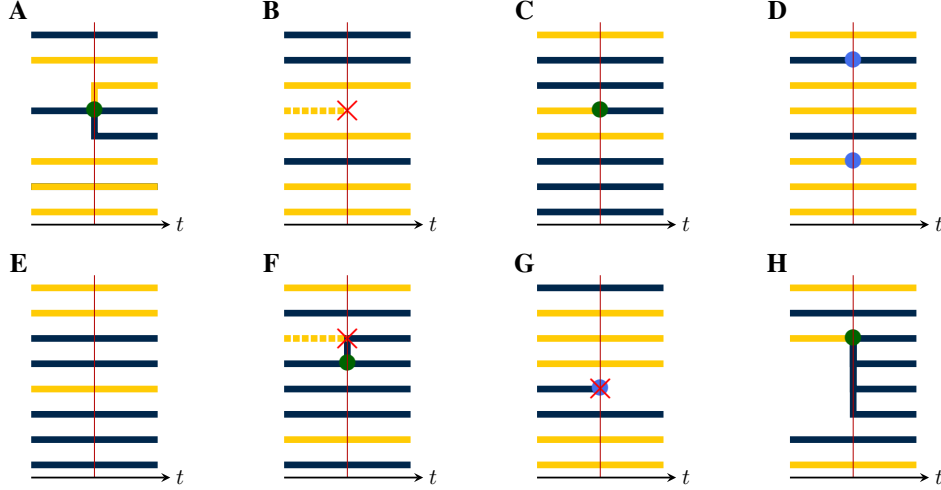


FIGURE 3. Jump types differ by their effects on the genealogy. This can be seen by examining the local structure of the genealogy in a neighborhood of a jump event. **(A)** A birth-type jump results in the branching of one or more child lineages from the parent. There can be only one parent, though the child lineages may be in demes different from those of their parent. Here, a parent of the blue deme sires one child lineage in each of the blue and yellow demes. The production of an event is an integer vector, with one entry for each deme. The production of this event is therefore  $r = (r_{\text{blue}}, r_{\text{yellow}}) = (2, 1)$ . **(B)** A death-type event causes the extinction of a lineage. Since internal node without children are recursively removed, the affected branch disappears. The production of this event is  $r = (0, 0)$ . **(C)** A migration-type event results in the movement of one or more lineages from one deme to another. Here, one lineage moves from the yellow to the blue deme. The production of this event is  $r = (1, 0)$ , i.e., the production is 1 for the blue deme and 0 for the yellow. **(D)** In a sample-type event, one or more sample nodes (blue circles) are inserted. Here, there are two samples, one in each of the blue and yellow demes. Accordingly, its production is  $r = (1, 1)$ . **(E)** A neutral-type event has no effect on the genealogy and zero production in all demes:  $r = (0, 0)$ . **(F)** The theory presented here allows for compound events. As an example, here a birth/death-type event occurs, wherein one yellow lineage is extinguished and a blue lineage simultaneously sires a blue child. The production of this event is  $r = (2, 0)$ . **(G)** Here, a compound sample/death-type event with  $r = (0, 0)$  occurs. A blue lineage is sampled and simultaneously extinguished. Note that recursive removal does not occur, since sample nodes are never removed. **(H)** A compound birth/migration-type event with  $r = (4, 0)$ .

et al. (2022) gave a construction. Although we now treat a more general case, the ideas are much the same. Readers wishing more detail on the construction here should consult the earlier paper.

At each event in the population process, one or more of the following changes happen to the genealogy, according to the type of the event (Fig. 3):

- A birth-type event at time  $t$  (Fig. 3A) results in the creation of one new internal node, call it  $b$ . A tip node,  $a$ , of the appropriate deme is chosen with uniform probability from among those present and  $b$  is inserted so that its ancestor is that of  $a$ , while  $a$  takes  $b$  as its ancestor. One new tip node, of the appropriate deme, is created for each of the children, all of which take  $b$  as their immediate ancestor.
- In a death-type event (Fig. 3B), one or more tip nodes of the appropriate demes are selected with uniform probability from among those present. These are deleted. Next, branch nodes without children are recursively removed. Sample nodes are never removed.
- In a migration-type event (Fig. 3C), the appropriate number of migrating lineages are selected at random, with uniform probability according to deme, from among those present. For each selected lineage, one

new branch node is inserted between the selected tip node and its ancestor. The color of the descendant branch changes accordingly.

- (d) At a sample-type event (Fig. 3D), the appropriate number of sampled lineages are selected at random from among the tip nodes, with uniform probability according to deme. One new sample node is introduced for each selected lineage: each is inserted between a selected tip nodes and its ancestor.
- (e) At a neutral-type event (Fig. 3E), no change is made to the genealogy.
- (f) Finally, events of compound type (e.g., Fig. 3F–H) are accommodated by combining the foregoing rules.

In each of these events, the new node or nodes that are introduced are situated at time  $t$ .

**3.3.1. Emergent lineages and production.** The lineages which descend from an inserted node are said to *emerge* from the event. Thus, after a birth-type event, the emerging lineages include all the new offspring as well as the parent. Likewise, at pure migration- or sample-type events, each migrating or sampled lineage emerges from the event. At pure death-type events, no lineages emerge. In general, at an event of mark  $u$ , there are  $r_i^u$  emergent lineages in deme  $i$ . We require that  $r_i^u$  be a constant, for each  $u$  and  $i$ . Since one is free to expand the set of jump-marks  $\mathbb{U}$  as needed, this is not an important restriction on the models that the theory can accommodate. Thus there is a function  $r : \mathbb{U} \times \mathbb{D} \rightarrow \mathbb{Z}_+$ , such that  $r_i^u$  lineages of deme  $i$  emerge from each event of mark  $u$ . We say  $r^u = (r_i^u)_{i \in \mathbb{D}}$  is the *production* of an event of mark  $u$ . Note that the lineages that die as a result of an event do not count in the production but that all parent lineages, if they survive the event, are counted in the production.

**3.3.2. Conditional independence and exchangeability.** Application of these rules at each jump of  $\mathbf{X}_t$  constructs a chain of genealogies  $\hat{\mathbf{G}}_k$ . In particular, at each jump-time  $\hat{T}_k$ , the genealogy  $\hat{\mathbf{G}}_{k-1}$  is modified according to the jump-mark  $\hat{\mathbf{U}}_k$  to yield  $\hat{\mathbf{G}}_k$ . We view  $\hat{\mathbf{G}}_k$  as the embedded chain of the continuous-time genealogy process  $\mathbf{G}_t$ . It is very important to note that, conditional on  $\hat{\mathbf{U}}_k$ , the number of parents and number of offspring in each deme is determined and the random choice of which lineages die, migrate, are sampled, or sire offspring is independent of the choices at any other times and independent of  $\hat{\mathbf{X}}_k$  for all  $k$ . Moreover, by construction, any lineage within a deme is as likely as any other lineage in that deme to be selected as a parent or for death, sampling, or migration. We refer to this property as the *exchangeability* of lineages within a deme.

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

**3.4.1. Pruned genealogy.** Given a genealogy  $\mathbf{G}_t$ , one obtains the *pruned genealogy*,  $\mathbf{P}_t = \text{prune}(\mathbf{G}_t)$  by first dropping every tip node and then recursively dropping every childless internal node. In a pruned genealogy only internal and sample nodes remain, and sample nodes are found at all of the leaves and possibly some of the interior nodes of the genealogy. Observe that a pruned genealogy is a colored genealogy: it retains information about where among the demes each of its lineages was through time. Note also that a pruned genealogy  $\mathbf{P}_t$  is characterized by functions  $y^{\mathbf{P}_t}$  and  $z^{\mathbf{P}_t}$  just as an un-pruned genealogy is.

**3.4.2. Lineage count and saturation.** In the following, we will find that we need to count the deme-specific numbers of lineages present in a given pruned genealogy at a given time. Accordingly, suppose  $\mathbf{P}_T = (y, z)$  is a pruned genealogy and suppose  $t \in [0, T]$ . Let  $\ell_i$  denote the number of lineages in deme  $i$  at time  $t$  and  $\ell = (\ell_i)_{i \in \mathbb{D}} \in \mathbb{Z}_+^{\mathbb{D}}$ . Clearly,  $\ell$  depends only on the values,  $(y_t, z_t)$ , of  $y$  and  $z$  at time  $t$ . Therefore, we can define  $\ell$  as a function such that, whenever  $\mathbf{P}_T = (y, z)$  is a pruned genealogy,  $\ell(y_t, z_t)$  is the vector of deme-specific lineage counts at time  $t$ . We refer to  $\ell$  as the *lineage-count* function (cf. Fig. 5).

We will also have occasion to refer to the deme-specific number of lineages emerging from a given event. In particular, given a node time  $t$  in a pruned genealogy  $\mathbf{P}_T = (y, z)$ , the number  $s_i$  of lineages of deme  $i$  emerging from all nodes with time  $t$  is well defined and we can write  $s = (s_i)_{i \in \mathbb{D}}$ . Like the lineage-count function,  $s$  depends only on the local structure of  $\mathbf{P}_T$ . However,  $s$  depends not only on  $(y_t, z_t)$ , but also on  $(\tilde{y}_t, \tilde{z}_t)$ . Thus, we can define the *saturation* function such that, whenever  $\mathbf{P}_T = (y, z)$  is a pruned genealogy,  $s(\tilde{y}_t, y_t, \tilde{z}_t, z_t)$  is the integer vector of deme-specific numbers of emerging lineages at time  $t$ . Fig. 5 illustrates.



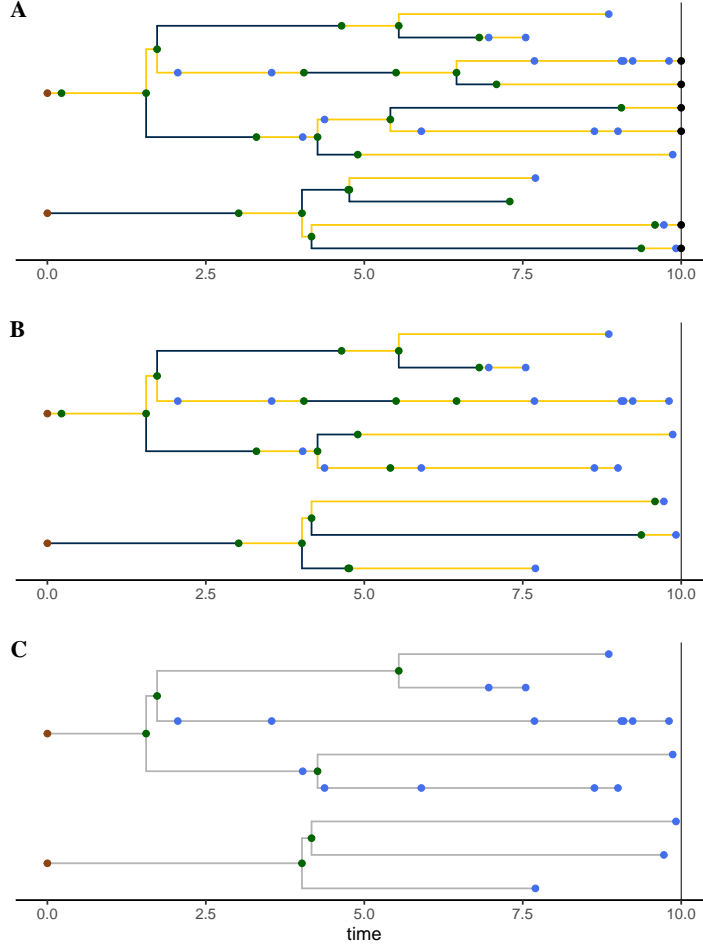


FIGURE 4. Unpruned, pruned, and obscured genealogies from a single realization of the genealogy process induced by the SEIRS model depicted in Figs. 1 and 2. (A) The full genealogy at time  $t = 10$ . Tip nodes, corresponding to lineages alive at time  $t = 10$  are indicated with black points. Blue points represent samples; green points, internal nodes. Branches are colored according to the deme in which the corresponding lineage resided at that point in time: blue denotes E and yellow, I. (B) The full genealogy is *pruned* by removing all tip nodes and then recursively pruning away childless internal nodes. Sample nodes are never removed. (C) A genealogy is *obscured* by removing all deme information from lineage histories: the colors are erased, as are extraneous internal nodes. See the text (§§3.1, 3.4.1, and 3.4.4) for more detail.

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



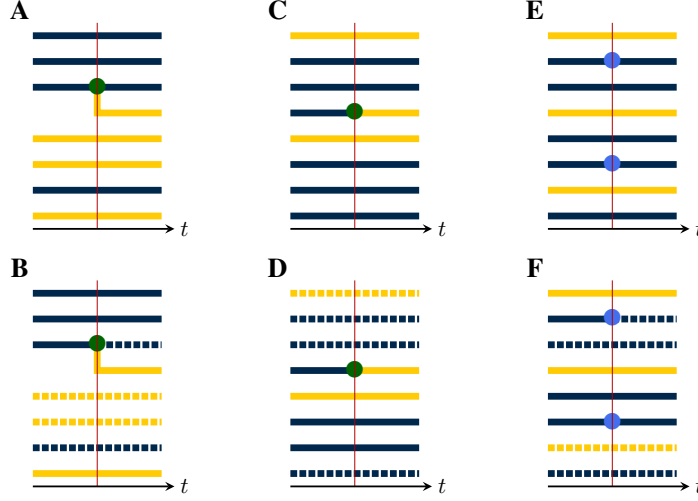


FIGURE 5. **Lineage count and saturation.** Each panel shows the neighborhood of a single event in the full genealogy (top row) and the corresponding pruned genealogy (bottom row). **(A)** A birth-type event with production  $r = (r_{\text{blue}}, r_{\text{yellow}}) = (1, 1)$  occurs. **(B)** Pruning results in the removal of the dashed lineages. The lineage count at this event-time is  $\ell = (\ell_{\text{blue}}, \ell_{\text{yellow}}) = (2, 2)$ . The saturation is  $s = (0, 1)$  since only a single, yellow lineage emerges from the event. **(C)** A migration-type event with production  $r = (0, 1)$  occurs. **(D)** After pruning,  $\ell = (2, 2)$  and  $s = (0, 1)$ . **(E)** A sample-type event occurs in which two blue lineages are sampled (production  $r = (2, 0)$ ). **(F)** After pruning,  $\ell = (2, 2)$  and  $s = (1, 0)$ . Observe that in panels B and D, the local structures of the pruned genealogies are identical, though they arise from events of different type.

**3.4.4. Obscured genealogy.** The *obscured genealogy* is obtained by discarding all information about demes and events not visible from the topology of the tree alone. In particular, if  $\mathbf{P}_T = (y, z)$  is a pruned genealogy, we write  $\text{obs}(\mathbf{P}_T) = z$  to denote the obscured genealogy.

## 4. Results.

**4.1. Likelihood for pruned genealogies.** Our first result will be an expression for the likelihood of a given pruned genealogy  $\mathbf{P}_T^*$  given the history  $\mathbf{H}_T$  of the population process to time  $T$ .

**Theorem 1.** Let  $\mathbf{P}_T^* = (y^*, z^*)$  be a given pruned genealogy and

$$\phi_u(x, y, y', z, z') := \binom{n(x)}{r^u} \frac{\ell(y', z')}{s(y, y', z, z')} Q_u(y, y', z, z'), \quad (3)$$

where  $n$  is the deme occupancy (§2.2.2),  $\ell$  and  $s$  are the lineage-count and saturation functions, respectively (§3.4.2),  $Q$  is the compatibility indicator (§3.4.3), and the binomial ratio is as defined in §2.5. Then

$$\text{Prob}[\mathbf{P}_T = \mathbf{P}_T^* \mid \mathbf{H}_T = \mathbf{H}_T] = \mathbb{1}\{\dot{\mathbf{H}}_T \supseteq \dot{\mathbf{P}}_T^*\} \prod_{t \in \dot{\mathbf{H}}_T} \phi_{U_t}(X_t, \hat{y}_t^*, y_t^*, \hat{z}_t^*, z_t^*).$$

*Proof.* If  $\dot{\mathbf{H}}_T \not\supseteq \dot{\mathbf{P}}_T^*$ , then  $\mathbf{H}_T$  and  $\mathbf{P}_T^*$  are incompatible and  $\text{Prob}[\mathbf{P}_T = \mathbf{P}_T^* \mid \mathbf{H}_T] = 0$ . Similarly, if any event of  $\mathbf{H}_T$  is incompatible with the local structure of  $\mathbf{P}_T^*$  in the sense of §3.4.3, then  $\text{Prob}[\mathbf{P}_T = \mathbf{P}_T^* \mid \mathbf{H}_T] = 0$ . Let us therefore suppose that neither of these conditions hold. Conditional on  $\mathbf{H}_T = \mathbf{H}_T$ , at each time  $t \in \dot{\mathbf{H}}_T$ , a jump of mark  $U_t$  occurred, with a production of  $r^{U_t} = (r_i)_{i \in \mathbb{D}}$ , resulting in a new deme-occupancy of  $n(X_t) = (n_i)_{i \in \mathbb{D}}$ . In  $\mathbf{P}_T^*$ , at time  $t$ , there are  $\ell_i = \ell_i(y_t^*, z_t^*)$  lineages in deme  $i$ , of which  $s_i = s_i(\hat{y}_t^*, y_t^*, \hat{z}_t^*, z_t^*)$  are emergent. By assumption, at each genealogical event, lineages within a deme are exchangeable: each has an identical probability of being involved. This exchangeability implies that each lineage present in a deme at time  $t$  was equally likely

to have been one of the emergent lineages. In particular, at time  $t$ , the probability that  $s_i$  of the  $\ell_i$  deme- $i$  lineages were among the  $r_i$  of  $n_i$  lineages emergent in the inventory process is the same as the probability that, upon drawing  $\ell_i$  balls without replacement from an urn containing  $r_i$  red balls and  $n_i - r_i$  black balls, exactly  $s_i$  of the drawn balls are red, namely

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

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

$$\text{Prob}[\mathbf{P}_T = P_T^* \mid \mathbf{H}_T = H_T] = \prod_{t \in \mathring{H}_T} \binom{n(X_t)}{r^{U_t}} \frac{\ell(y_t^*, z_t^*)}{s(\tilde{y}_t^*, y_t^*, \tilde{z}_t^*, z_t^*)}.$$

Returning to the possibility that  $H_T$  is incompatible with  $P_T^*$ , since  $\text{Prob}[\mathbf{P}_T = P_T^*] = 0$  if either any  $Q_u = 0$  or any event of  $P_T^*$  is not an event of  $H_T$ , we obtain the result.  $\square$

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

**Theorem 2.** *Suppose that  $P_T^* = (y^*, z^*)$  is a pruned genealogy, for  $T > 0$ . If  $w = w(t, x)$  satisfies the initial condition  $w(0, x) = p_0(x)$  and the filter equation*

$$\begin{aligned} \frac{\partial w}{\partial t}(t, x) = & \sum_u \int w(t, x') \alpha_u(t, x', x) \phi_u(x, \tilde{y}_t^*, y_t^*, \tilde{z}_t^*, z_t^*) dx' - \sum_u \int w(t, x) \alpha_u(t, x, x') dx' \\ & + \sum_{e \in \mathring{P}_T^*} \delta(t, e) \left\{ \sum_u \int w(t, x') \alpha_u(t, x', x) \phi_u(x, \tilde{y}_t^*, y_t^*, \tilde{z}_t^*, z_t^*) dx' - w(t, x) \right\}, \end{aligned} \quad (4)$$

where  $\phi$  is defined in Eq. 3, then the likelihood of  $P_T^*$  is

$$\mathcal{L}(P_T^*) = \int w(T, x) dx.$$

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

**Theorem 3.** *Let  $z^* = \text{obs}(P_T^*)$  be a given obscured genealogy. Let  $\pi$  and  $q$  be probability kernels, such that for all  $u \in \mathbb{U}$ ,  $t \in \mathbb{R}_+$ ,  $x, x' \in \mathbb{X}$ ,  $y, y' \in \mathbb{Y}_t^{z^*}$ ,  $\pi_u(t, x, x', y, y') \geq 0$ ,  $\sum_{y' \in \mathbb{Y}_t^{z^*}} \pi_u(t, x, x', y, y') = 1$ ,  $q(x, y) \geq 0$ , and  $\sum_{y \in \mathbb{Y}_0^{z^*}} q(x, y) = 1$ . Suppose moreover that  $\pi_u(t, x, x', y, y') > 0$  whenever  $Q_u(y, y', \tilde{z}_t^*, z_t^*) = 1$  and that  $q(x, y) > 0$  whenever  $\text{Prob}[\mathbf{X}_0 = x, y_0^{\mathbf{P}^T} = y] > 0$ . Then there is a Markov jump process  $\mathbf{y}_t$  on  $\mathbb{Y}_{z^*}$  such that*

$$\text{Prob}[\mathbf{z}_T = z^* \mid \mathbf{H}_T = H_T] = \mathbb{1}\{\mathring{H}_T \supseteq \mathring{z}_T^*\} \mathbb{E} \left[ \frac{1}{q(X_0, \mathbf{y}_0)} \prod_{t \in \mathring{H}_T} \frac{\phi_{U_t}(X_t, \tilde{\mathbf{y}}_t, \mathbf{y}_t, \tilde{z}_t^*, z_t^*)}{\pi_{U_t}(t, \tilde{X}_t, X_t, \tilde{\mathbf{y}}_t, \mathbf{y}_t)} \right],$$

where  $\phi$  is defined in Eq. 3 and the expectation is taken over the sample paths of  $\mathbf{y}_t$ ,  $t \in [0, T]$ .

*Proof.* First, observe that, since  $\text{obs}$  is a deterministic operator,

$$\text{Prob}[\mathbf{z}_T = z^* \mid \mathbf{H}_T] = \mathbb{E}[\mathbb{1}\{\text{obs}(\mathbf{P}_T) = z^*\} \mid \mathbf{H}_T]. \quad (5)$$

Our strategy will be to evaluate Eq. 5 using importance sampling: we will propose pruned genealogies compatible with  $z_t^*$  as sample paths from a Markov process on  $\mathbb{Y}_{z^*}$  and evaluate the the expectation in Eq. 5 by summing

over these paths. Conditional on  $\mathbf{H}_T$ , the initial distribution  $q$  and probability kernel  $\pi$  generate a Markov chain,  $\hat{\mathbf{y}}_k$ , on  $\mathbb{Y}_{z^*}$  such that

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

The required Markov process  $\mathbf{y}_t$  is the unique càdlàg process with event times  $\hat{\mathbf{T}}_k$  and  $\hat{\mathbf{y}}_k$  as its embedded chain. This construction of  $\mathbf{y}_t$  obviously guarantees that  $\hat{\mathbf{H}}_t \supseteq \hat{\mathbf{y}} \supseteq \hat{z}^*$ .

Let us define  $\mathbf{P}(\mathbf{y}) = (\mathbf{y}, z^*)$ . Then, by construction, for every  $\mathbf{y}$ ,  $\text{obs}(\mathbf{P}(\mathbf{y})) = z^*$  and, conversely, for every pruned genealogy  $\mathbf{P}_T$  satisfying  $\text{obs}(\mathbf{P}_T) = z^*$ ,  $\mathbf{y}^{\mathbf{P}_T}$  is a sample path of this process having nonzero probability. We therefore have that

$$\text{Prob}[\mathbf{z}_T = z^* \mid \mathbf{H}_T = \mathbf{H}_T] = \mathbb{E} \left[ \frac{\text{Prob}[\mathbf{P}_T = P(\mathbf{y}) \mid \mathbf{H}_T = \mathbf{H}_T]}{\pi(\mathbf{y} \mid \mathbf{H}_T)} \right]$$

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

$$\pi(\mathbf{y} \mid \mathbf{H}_T) = \prod_{t \in \hat{\mathbf{H}}_T} \pi_{\mathbf{U}_t}(t, \tilde{\mathbf{X}}_t, \mathbf{X}_t, \tilde{\mathbf{y}}_t, \mathbf{y}_t).$$

The result then follows from Theorem 1. □

The final result shows how to compute the likelihood of an obscured genealogy. It is an immediate consequence of Theorem 3 and Lemma A2.

**Theorem 4.** *Let  $z^*$ , be a given obscured genealogy. Let  $\pi$ ,  $q$ , and  $\mathbf{y}_t$  be as in Theorem 3. For  $t \in \mathbb{R}_+$ ,  $x, x' \in \mathbb{X}$ ,  $y, y' \in \mathbb{Y}_{z^*}$ ,  $u \in \mathbb{U}$ , define*

$$\psi_u(t, x, x', y, y') = \frac{\phi_u(x', y, y', \tilde{z}_t^*, z_t^*)}{\pi_u(t, x, x', y, y')},$$

$$\beta_u(t, x, x', y, y') = \alpha_u(t, x, x') \pi_u(t, x, x', y, y'),$$

where  $\phi$  is defined in Eq. 3. Suppose  $w = w(t, x, y)$  satisfies the filter equation

$$\begin{aligned} \frac{\partial w}{\partial t}(t, x, y) = & \sum_{uy'} w(t, x', y') \beta_u(t, x', x, y', y) \psi_u(t, x', x, y', y) dx' \\ & - \sum_{uy'} \int w(t, x, y) \beta_u(t, x, x', y, y') dx' \\ & + \sum_{e \in \hat{z}^*} \delta(t, e) \left\{ \sum_{uy'} \int w(t, x', y') \beta_u(t, x', x, y, y') \psi_u(t, x', x, y', y) dx' - w(t, x, y) \right\}, \end{aligned} \quad (6)$$

with the initial condition  $w(0, x, y) = p_0(x) q(x, y)$ . Then the likelihood of  $V_T^*$  is

$$\mathcal{L}(V_T^*) = \sum_y \int w(T, x, y) dx.$$

## 5. Examples.

**5.1. SEIRS.** Jumps:  $\mathbb{U} = \{\text{Inf}, \text{Prog}, \text{Recov}, \text{Wane}, \text{Birth}, \text{Death}_S, \text{Death}_E, \text{Death}_I, \text{Death}_R, \text{Sample}\}$ .

Demes:  $\mathbb{D} = \{E, I\}$ .

Jump rates:

- $\alpha_{\text{Inf}}(t, x, x') = \beta(t) \frac{x^S x^I}{N(t)} \mathbb{1}\{x' = x + (-1, 1, 0, 0)\}$
- $\alpha_{\text{Prog}}(x, x') = \rho x^E \mathbb{1}\{x' = x + (0, -1, 1, 0)\}$
- $\alpha_{\text{Recov}}(x, x') = \gamma x^I \mathbb{1}\{x' = x + (0, 0, -1, 1)\}$
- $\alpha_{\text{Wane}}(x, x') = v x^R \mathbb{1}\{x' = x + (1, 0, 0, -1)\}$
- $\alpha_{\text{Sample}}(t, x, x') = \psi x^I \mathbb{1}\{x' = x\}$

- $\alpha_{\text{Birth}}(t, x, x') = B(t) \mathbb{1}\{x' = x + (1, 0, 0, 0)\}$
- $\alpha_{\text{Death}_k}(x, x') = \mu x^k \mathbb{1}\{x'^j = x^j - \delta_{jk}\}, k \in \{S, E, I, R\}$

## 6. Discussion.

### References.

- King, A. A., Lin, Q., & Ionides, E. L. (2022) Markov genealogy processes. *Theor. Popul. Biol.* **143**:77–91.  
<https://doi.org/10.1016/j.tpb.2021.11.003>
- Leventhal, G. E., Günthard, H. F., Bonhoeffer, S., & Stadler, T. (2014) Using an epidemiological model for phylogenetic inference reveals density dependence in HIV transmission. *Mol. Biol. Evol.* **31**:6–17.  
<https://doi.org/10.1093/molbev/mst172>
- Moran, P. A. P. (1958) Random processes in genetics. *Math. Proc. Cambridge Philos. Soc.* **54**:60–71.  
<https://doi.org/10.1017/s0305004100033193>
- Stadler, T. (2010) Sampling-through-time in birth-death trees. *J. Theor. Biol.* **267**:396–404.  
<https://doi.org/10.1016/j.jtbi.2010.09.010>
- Volz, E. M., Kosakovsky Pond, S. L., Ward, M. J., Leigh Brown, A. J., & Frost, S. D. W. (2009) Phylodynamics of infectious disease epidemics. *Genetics* **183**:1421–1430.  
<https://doi.org/10.1534/genetics.109.106021>

### Appendix A. Filter equations.

Explicit expressions for the quantities that arise in this paper are not always readily available. Here, we develop tools for manipulating complex expressions that are otherwise cumbersome.

**Definition.** Suppose  $\mathbf{X}_t$  is a continuous-time Markov process with Kolmogorov forward equation (KFE)

$$\frac{\partial u}{\partial t}(t, x) = \int u(t, x') \beta(t, x', x) dx' - \int u(t, x) \beta(t, x, x') dx'. \quad (\text{A1})$$

Suppose that  $B(t, x, x') \geq 0$  and  $\lambda(t, x)$  are given, real-valued, measurable functions. We say that the equation

$$\frac{\partial w}{\partial t}(t, x) = \int w(t, x') \beta(t, x', x) B(t, x', x) dx' - \int w(t, x) \beta(t, x, x') dx' - \lambda(t, x) w(t, x). \quad (\text{A2})$$

is the *filter equation* with *driver*  $\mathbf{X}_t$  (or *generated by*  $\beta$ ), *boost*  $B$ , and *decay*  $\lambda$ .

*Remark 1.* Trivially, a Kolmogorov forward equation is itself a filter equation with boost 1 and decay 0.

The following results show how filter equations allow one to integrate over random histories.

**Lemma A1.** Suppose that  $B : \mathbb{R}_+ \times \mathbb{X}^2 \rightarrow \mathbb{R}_+$  is measurable. If  $\mathbf{V}_t$  is an  $\mathbb{R}_+$ -valued random process satisfying

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

then  $\mathbb{E}[\mathbf{V}_t \mid \mathbf{X}_t = x]$  is a measure-valued deterministic process on  $\mathbb{X}$  with density  $w(t, x)$  that satisfies the filter equation

$$\frac{\partial w}{\partial t} = \int w(t, x') \alpha(t, x', x) B(t, x', x) dx' - \int w(t, x) \alpha(t, x, x') dx'. \quad (\text{A3})$$

Moreover,  $\mathbb{E}[\mathbf{V}_t] = \int w(t, x) dx$ .

*Proof.* The expectation can be broken into three terms, according to whether  $\mathbf{H}_t$  has zero, one, or more than one event in  $(t - \Delta, t]$ . Accordingly, as  $\Delta \downarrow 0$ ,

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

In the limit, we obtain Eq. A3. The last statement follows by definition.  $\square$

**Lemma A2.** Suppose  $\mathcal{M} \subset \mathbb{R}_+$  is finite and let  $\mathcal{M}_t := \mathcal{M} \cap [0, t]$ . Suppose that  $B : \mathbb{R}_+ \times \mathbb{X}^2 \rightarrow \mathbb{R}_+$  is measurable and  $\mathbf{V}_t$  is an  $\mathbb{R}_+$ -valued process satisfying

$$\mathbb{E}[\mathbf{V}_t \mid \mathbf{H}_t = \mathbf{H}_t] = \mathbb{1}\{\dot{\mathbf{H}}_t \supseteq \mathcal{M}_t\} \prod_{e \in \dot{\mathbf{H}}_t} B(e, \tilde{X}_e, X_e).$$

Then  $\mathbb{E}[\mathbf{V}_t \mid \mathbf{X}_t = x]$  is a measure-valued process on  $\mathbb{R}_+^{|\mathcal{M}_t|} \times \mathbb{X}$  with density  $w(t, x)$  that satisfies

$$\begin{aligned} \frac{\partial w}{\partial t}(t, x) &= \int w(t, x') \alpha(t, x', x) B(t, x', x) dx' - \int w(t, x) \alpha(t, x, x') dx', & t \notin \mathcal{M}, \\ w(t, x) &= \int \tilde{w}(t, x') \alpha(t, x', x) B(t, x', x) dx', & t \in \mathcal{M}. \end{aligned} \quad (\text{A4})$$

Moreover,  $\mathbb{E}[\mathbf{V}_t] = \int w(t, x) dx$ .

**Lemma A3.** Showing how Eq. A4 can be stated as a singular filter equation.

$$\begin{aligned} \frac{\partial w}{\partial t}(t, x) &= \int w(t, x') \beta(t, x', x) B(t, x', x) dx' - \int w(t, x) \beta(t, x, x') dx' \\ &+ \sum_e \int w(t, x') \delta(t, e) \pi(t, x', x) C(t, x', x) dx' - \sum_e \int w(t, x) \delta(t, e) \pi(t, x, x') dx' \\ &- \lambda(t, x) w(t, x). \end{aligned} \quad (\text{A5})$$

Filter equations afford a convenient means of computing expectations and likelihoods for pure jump processes. This is facilitated by the following Lemma, the statement of which uses a certain non-standard Dirac delta function. Specifically, let  $\delta(v, v')$  be the right-sided Dirac delta function satisfying

$$\delta(v, v') = 0 \quad \text{for } v \neq v' \text{ and}$$

$$\int_a^b f(v) \delta(v, v') dv = f(v') \mathbb{1}\{v' \in [a, b]\}, \quad \text{whenever } f \text{ is càdlàg and } -\infty \leq a < b \leq \infty.$$

**Lemma A4.** The filter equation (A3) is satisfied by  $w(t, x) = \int_0^\infty v u(t, x, v) dv$ , where  $u(t, x, v)$  satisfies the KFE

$$\begin{aligned} \frac{\partial u}{\partial t} &= \int_0^\infty \int u(t, x', v') \beta(t, x', x) \delta(v, B(t, x', x) v') dx' dv' \\ &- \int_0^\infty \int u(t, x, v) \beta(t, x, x') \delta(v', B(t, x, x') v) dx' dv' + \frac{\partial}{\partial v} [\lambda(t, x) v u(t, x, v)]. \end{aligned} \quad (\text{A6})$$

*Proof.*

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

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

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

which is simplified to obtain Eq. A3.  $\square$

Eq. A6 is recognizable as the KFE of a certain process  $(\mathbf{X}_t, \mathbf{V}_t)$ . In particular,  $\mathbf{X}_t$  is the driver with KFE (A1). The  $\mathbf{V}_t$  is *directed* by  $\mathbf{X}_t$  in the sense that  $\mathbf{V}$  has jumps wherever  $\mathbf{X}$  does: when  $\mathbf{X}$  jumps at time  $t$  from  $x$  to  $x'$ ,  $\mathbf{V}$  jumps by the multiplicative factor  $B(t, x, x') \geq 0$ . Between jumps,  $\mathbf{V}_t$  decays deterministically and exponentially at rate  $\lambda(t, x)$ . If we view  $\mathbf{V}_t$  as a weight, then Lemma A4 tells us how the  $\mathbf{V}_t$ -weighted average of  $\mathbf{X}_t$  evolves in time: this average is simply  $\int w(t, x) dx$ .

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

*Email address:* kingaa@umich.edu

*URL:* <https://kinglab.eeb.lsa.umich.edu/>

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

E. L. IONIDES, DEPARTMENT OF STATISTICS UNIVERSITY OF MICHIGAN, ANN ARBOR, MI 48109 USA