

Birth–death skyline plot reveals temporal changes of epidemic spread in HIV and hepatitis C virus (HCV)

Tanja Stadler^{a,1,2}, Denise Kühnert^{b,c,1}, Sebastian Bonhoeffer^a, and Alexei J. Drummond^{b,c}

^aDepartment of Environmental Systems Science, Eidgenössische Technische Hochschule Zürich, 8092 Zürich, Switzerland; and ^bDepartment of Computer Science and ^cAllan Wilson Centre for Molecular Ecology and Evolution, University of Auckland, Auckland, New Zealand

Edited by Robert M. May, University of Oxford, Oxford, United Kingdom, and approved November 15, 2012 (received for review May 10, 2012)

Phylogenetic trees can be used to infer the processes that generated them. Here, we introduce a model, the Bayesian birth–death skyline plot, which explicitly estimates the rate of transmission, recovery, and sampling and thus allows inference of the effective reproductive number directly from genetic data. Our method allows these parameters to vary through time in a piecewise fashion and is implemented within the BEAST2 software framework. The method is a powerful alternative to the existing coalescent skyline plot, providing insight into the differing roles of incidence and prevalence in an epidemic. We apply this method to data from the United Kingdom HIV-1 epidemic and Egyptian hepatitis C virus (HCV) epidemic. The analysis reveals temporal changes of the effective reproductive number that highlight the effect of past public health interventions.

birth–death prior | epidemiological dynamics | phylodynamics

Measurably evolving populations such as RNA viruses allow the reconstruction of past population dynamics based on molecular sequence data. This reconstruction is frequently performed in a Bayesian framework to account for uncertainty in the underlying phylogenetic tree. The Bayesian reconstruction framework requires specification of a model generating the phylogenetic tree. A popular approach to modeling the phylogenetic tree in this context employs an extension of Kingman's coalescent (1) that allows for sequentially sampled sequences (2).

In particular for RNA viruses, population sizes may fluctuate in complex ways due to a changing host population, seasonal factors, or public health interventions. Thus, the coalescent skyline plot (3–5) was introduced as an extension of the classical coalescent models to account for arbitrary changing population sizes. The coalescent skyline plot assumes stepwise constant population sizes, thus yielding plots of population size vs. time looking like a skyline. The coalescent-based Bayesian skyline plot implemented in BEAST (6) is currently one of the standard models used for reconstructing ancestral dynamics of measurably evolving populations (5).

Recently, limitations of the Bayesian skyline plot, due to the model being based on the coalescent, have been highlighted (7–9). First, the model fails to acknowledge that both incidence and prevalence affect coalescent rates. Second, the coalescent is an approximation of the population dynamics assuming a small random sample from a large background population. However, during epidemic outbreaks and in large cohort studies, the proportion of infections sampled may be very high.

To overcome these two limitations, we introduce here the birth–death skyline model, which is based on a forward-in-time model of transmission, death/recovery, and sampling. We allow the parameters (transmission, death/recovery, and sampling) to change in a piecewise fashion (recall that, under the coalescent skyline plot, effective population size is changed in a piecewise fashion). The probability density of a tree under this birth–death skyline model is determined, such that the Bayesian birth–death skyline plot may be used as an alternative to the Bayesian coalescent skyline plot. The method is implemented as an add-on to the BEAST2 software framework (<http://beast2.cs.auckland.ac.nz>).

The birth–death skyline model essentially combines two previous approaches. Previously, a skyline model was introduced that assumed samples were all taken at one point in time, corresponding to a sample of extant species (10). Earlier work had also described how to model sequential sampling for constant epidemiological rates (a birth–death alternative to the exponential growth coalescent model; see refs. 9 and 11). Combining the skyline model (10) with the sequential sampling model (11), and embedding the result in a Bayesian inference framework (9), yields the approach described in this paper.

We apply the birth–death skyline method to an HIV transmission cluster from the United Kingdom and a sample of hepatitis C virus (HCV) sequences from Egypt to investigate the temporal changes of epidemic spread. We decided to use these two datasets as they are representatives of very different epidemics and data types. The HIV samples correspond to an epidemic outbreak which started about 25 years ago and data were collected over several years. In contrast, the HCV samples were collected at one time point (year 1993), although the epidemic presumably broke out hundreds or thousands of years ago.

Results

Bayesian Birth–Death Skyline Plot. We introduce the birth–death skyline plot as a model for transmission. The birth–death skyline plot is based on a birth–death process, meaning that each infected individual may transmit with a rate λ and eventually becomes noninfectious with a rate δ . Becoming noninfectious may be due to several processes, e.g., recovery, host death, behavior change, or successful treatment. Upon becoming noninfectious, the individual is sampled with probability s and thus included into the dataset (e.g., by enrollment into a cohort). We emphasize that our model assumes that a sampled individual does not remain infectious. Reasons for sampled individuals being noninfectious may be successful treatment [which, for example, reduces transmission risk substantially in HIV (12)] or behavior change. Methodological difficulties in relaxing this assumption are discussed in ref. 9. This model extends the previously introduced birth–death-sampling model (9, 11) by relaxing the assumption of constant transmission rate, constant becoming noninfectious rate, and constant sampling probability, but allowing these parameters to change in a piecewise constant fashion.

In a model extension, we allow individuals to become sampled not only with a rate δs , but also at fixed time points with a probability p . This extension allows the consideration of datasets in

Author contributions: T.S., D.K., S.B., and A.J.D. designed research; T.S., D.K., and A.J.D. performed research; T.S., D.K., and A.J.D. analyzed data; and T.S., D.K., S.B., and A.J.D. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Freely available online through the PNAS open access option.

¹T.S. and D.K. contributed equally to this work.

²To whom correspondence should be addressed. E-mail: tanja.stadler@env.ethz.ch.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1207965110/-DCSupplemental.

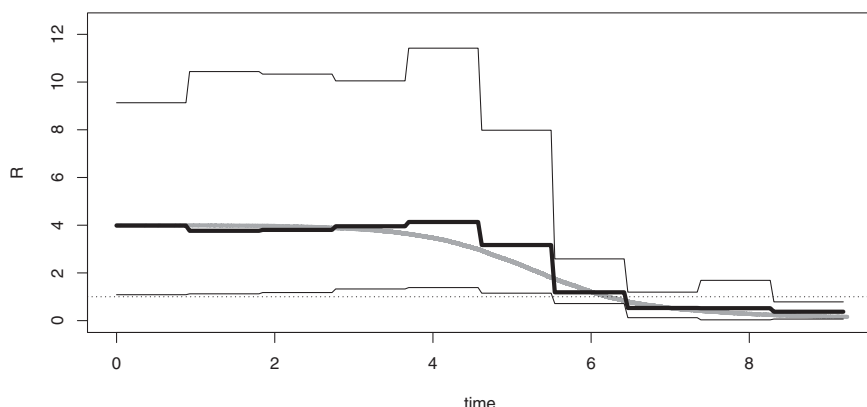


Fig. 1. Simulated (gray) and estimated (black) effective reproductive number R for a tree on 200 tips evolving under the SIR model. We assumed a population size of 5,000, a transmission rate of 2, a becoming noninfectious rate of 0.5, and a sampling proportion of 5%. Gray is the simulated R . The thick black line is the estimated median R , and the thin black lines indicate the HPD intervals.

which several individuals are sampled at the same time point (with probability ρ). In contrast, under a rate δs , sampling several individuals at one point in time has probability zero. We emphasize that if samples are taken at only a single point in time with probability ρ and sampling probability is zero before and after that time point (e.g., sampling in year 1993 in the analyzed HCV dataset), we do not have to assume that sampled individuals become noninfectious, as we do not consider the subsequent epidemic.

We note that an important epidemiological parameter, namely the effective reproductive number $R = \frac{\beta}{\delta}$, is quantified by the model. The effective reproductive number is defined as the number of expected secondary infections of an infected individual. The effective reproductive number is closely related to the basic reproductive number (13): the latter additionally assumes a completely susceptible population, and thus the two quantities are equal at the start of an epidemic outbreak.

We derived an analytic expression for the likelihood of a phylogenetic tree spanning the sampled individuals (Theorem 1 in *Methods*; proof is found in [SI Appendix](#)) and included this likelihood function into the Bayesian software package BEAST2, as an alternative to the previous coalescent skyline plot (*Methods*). In [SI Appendix](#), we show that, in a phylogenetic framework, one parameter of the model correlates with the others and thus cannot be estimated independently. However, with few prior assumptions on the parameters, one can circumvent this drawback. We circumvented the problem in our empirical analysis by constraining the sampling probability.

Simulation Study. We performed a simulation study to investigate accuracy of our method. We simulated trees under the birth–death skyline model as well as under a discrete-state continuous-time version (14, 15) of a popular epidemiological model, the susceptible–infected–recovered (SIR) model (16). Testing our method on this simulated data showed that parameters can be estimated reliably using the birth–death skyline method; an example is given in Fig. 1 and Table 1. More details about the simulation study are provided in *Methods* and *SI Appendix*, and simulation results for more parameter combinations are provided in *SI Appendix*, Fig. S1 and Tables S1–S4.

HIV-1 Type B in the United Kingdom. One of the great successes in modern medicine is the development of multidrug therapy for treatment of HIV infection. The birth–death skyline model can be applied to examine whether introduction of the therapy had an impact on the epidemiological dynamics. Allowing rate changes over time, we can observe the impact of prevention or therapy measures on R throughout an epidemic. Exemplarily, we analyze an HIV-1 subtype B dataset from the United Kingdom.

consisting of 62 sequences of a genetic cluster formerly analyzed in ref. 17. The samples are taken between 1999 and 2003.

Fig. 2 shows the birth–death skyline rate estimates as well as the Bayesian skyline plot reconstruction of the effective population size N_e . Although the 95% highest posterior density (HPD) interval for R gets broader the further we go into the past, we have strong evidence that R is larger than one at the time of the most recent common ancestor. In fact, the median R is 1.87 in the oldest interval starting from the origin of the cluster epidemic. Around 1992, R starts declining, dropping below 1 around 1998, indicating a declining epidemic. In fact, there is substantial evidence (Bayes factor, 5.2) that R has decreased over the time spanned by this dataset, possibly a reflection of the introduction of antiretroviral drug treatment in 1987 and further advances in multidrug therapy in the mid-1990s. $R > 1$ indicates a growing epidemic, whereas $R < 1$ indicates a declining epidemic. The expected behavior for $R \geq 1$ is reflected in the plot of the effective population size N_e obtained from the independent Bayesian coalescent skyline plot analysis, whereas the recent decline predicted by $R < 1$ is not reflected in the coalescent skyline plot.

The most recent common ancestor was estimated to be in 1984 (median, 1984.44). The median rate to become noninfectious δ was estimated to be 0.26 per lineage per year in the mid-1990s, which implies an infectious period of roughly 4 years, i.e., on average it took 4 years for an infected individual to either be sampled or become noninfectious due to other reasons (behavior change, successful treatment, death). Toward the present, the becoming noninfectious rate grows larger. This pattern is clearly driven by the sampling scheme, i.e., much effort is undertaken to sample individuals (who then become noninfectious due to, for

Table 1. Posterior parameter estimates and accuracy obtained from 100 simulated trees with 100 tips sampled sequentially through time

Par	True	Median	Error	Bias	HPD width	95% HPD accuracy, %
R_1	2	2.06	0.08	0.03	1.84	100
R_2	0.75	0.74	0.10	-0.02	0.37	96
δ_1	4	3.77	0.09	-0.06	4.57	100
δ_2	4	3.80	0.10	-0.05	3.18	100
s_1	0.25	0.26	0.14	0.04	0.42	100
s_2	0.25	0.26	0.18	0.03	0.43	100
t_1	1	1.02	0.03	0.02	0.17	95

The parameters (Par) at the start of the epidemic are with index 1; after time t_1 , the parameters change to the values with index 2, i.e., R drops from above 1 to below 1. For each parameter, the median over the 100 medians (1 for each tree)/errors/biases/HPD widths/HPD accuracies is provided.

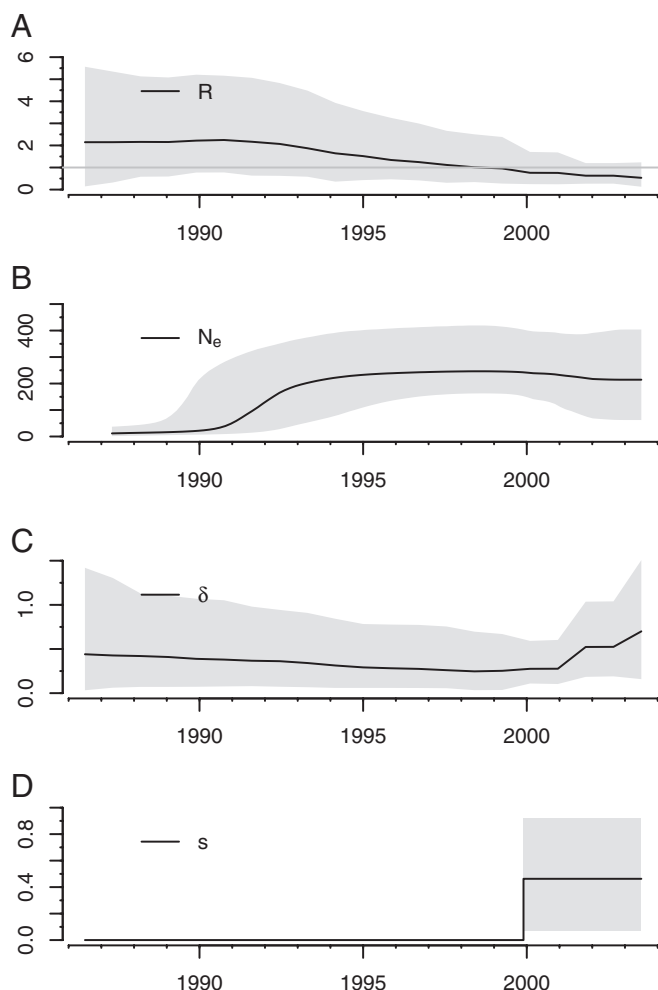


Fig. 2. HIV-1 in the United Kingdom: Birth-death skyline plot (A, C, and D) and Bayesian skyline plot (B) median estimates and HPD intervals for the (A) effective reproductive number, (B) effective population size, (C) becoming noninfectious rate, and (D) constrained sampling proportion.

example, successful treatment), and thus the time of being infectious is shortened. Within the period in which sampling occurred (from 1999), the sampling proportion s was estimated to be 50%.

HCV in Egypt. HCV has only been discovered recently; its existence was noticed in the 1970s but only confirmed in the late 1980s. The largest HCV epidemic circulates in Egypt and is believed to have been caused by viral contamination of injections treating bilharziosis (18). Former studies indicate that the period of the antischistosomal injection campaigns indeed coincides with an increase in the effective number of HCV infections. We applied the birth-death skyline model to the same HCV dataset used to demonstrate the coalescent-based Bayesian skyline plot (5), to see whether we can confirm this hypothesis.

The analysis of 63 Egyptian HCV sequences again confirms a rapid escalation of HCV infections at the time when the antischistosomal injection campaigns started. The R estimate rises significantly at the beginning of the 20th century, staying clearly above 1 until 1960 when it decreases to just above 1 (median, 1.08) (Fig. 3). In the 1970s, antischistosomal injection therapy was gradually replaced by oral therapy (18). We have decisive evidence (Bayes factor, >100) that R is larger than 1 from the early 19th century through to 1993, the time of sampling. Again, for the time spans with $R \sim 1$, the coalescent skyline plot analysis inferred

a roughly constant N_e , as expected, whereas periods with $R > 1$ coincide with an increase in N_e .

Although the virus has only been discovered late in the 20th century, it has been estimated to be hundreds of years old. In fact, HCV type 4 is characterized by long-term localized endemic infection (19). Here, the phylogeny's most recent common ancestor was estimated to be about 410 years old, that is, the origin dates back to the 16th century. However, this point estimate comes with a large HPD interval.

Discussion

The birth-death skyline model described here combines previous approaches, yielding a sound method to analyze serially as well as contemporaneously sampled data arising from an evolutionary process in which epidemiological parameters vary through time. Although we showed in ref. 9 that it is possible to estimate key epidemiological parameters such as R only based on sequence data, the birth-death skyline model additionally enables the accurate reconstruction of changes in the epidemiological parameters through time only based on sequence data, i.e., without any additional clinical data.

The framework provides interesting insight into the phylodynamics of two datasets previously studied using coalescent-based techniques. Previous coalescent-based analysis of the United Kingdom HIV-1 cluster indicated that prevalence followed a logistic curve, with high but stable recent prevalence. Our birth-death skyline model can additionally reconstruct incidence rates through time, supporting the early exponential expansion of the clade, and furthermore clearly uncovering a signature of a recent declining epidemic. Interestingly, the estimated R drops to below 1 after the introduction of highly active antiretroviral treatment in the mid-1990s, indicating that the availability of successful treatment for the individual patient also had a positive impact on the whole epidemic (by removing the treated individual from the infectious class). Applied to the contemporaneously sampled HCV dataset, our method recovers an increase in R around 1920, coinciding with the beginning of antischistosomal injection campaigns in Egypt (18), which supports the hypothesis of the injection campaigns having caused the HCV epidemic in Egypt. Our results confirm previous studies, such as the coalescent-based Bayesian skyline plot analysis (5), with an increase in prevalence around 1920.

Comparing the coalescent-based Bayesian skyline plots of the two analyzed datasets with our R estimates, we notice a difference in the confidence of the estimates. As expected, the HPD intervals from the birth-death skyline model grow wider the further we go into the past. That is not the case for the coalescent-based Bayesian skyline plot. In fact, for the HCV dataset with an estimated most recent common ancestor dating hundreds of years back in time, the effective population size HPD width decreases to a quite narrow interval from around 1900 all of the way back to the most recent common ancestor. This is most likely due to the fact that the skyline plot uses coalescent events as boundaries for the steps in the population function, so that each step has approximately the same amount of data about population size. Whereas equidistant intervals tend to lead to a reduction of coalescent events per interval going back in time and thus greater uncertainty in parameter estimates near the root. The choice between these two approaches reflects a trade-off between precision about the timing of parameter changes versus precision about the parameter estimates themselves. We cannot have both.

This birth-death skyline approach is a powerful alternative to the coalescent-based skyline plot. As already mentioned, classical coalescent-based approaches like the coalescent skyline plot (5) cannot distinguish between incidence and prevalence, whereas the birth-death skyline model can distinguish by parameterizing a separate birth rate and death rate instead of only parameterizing

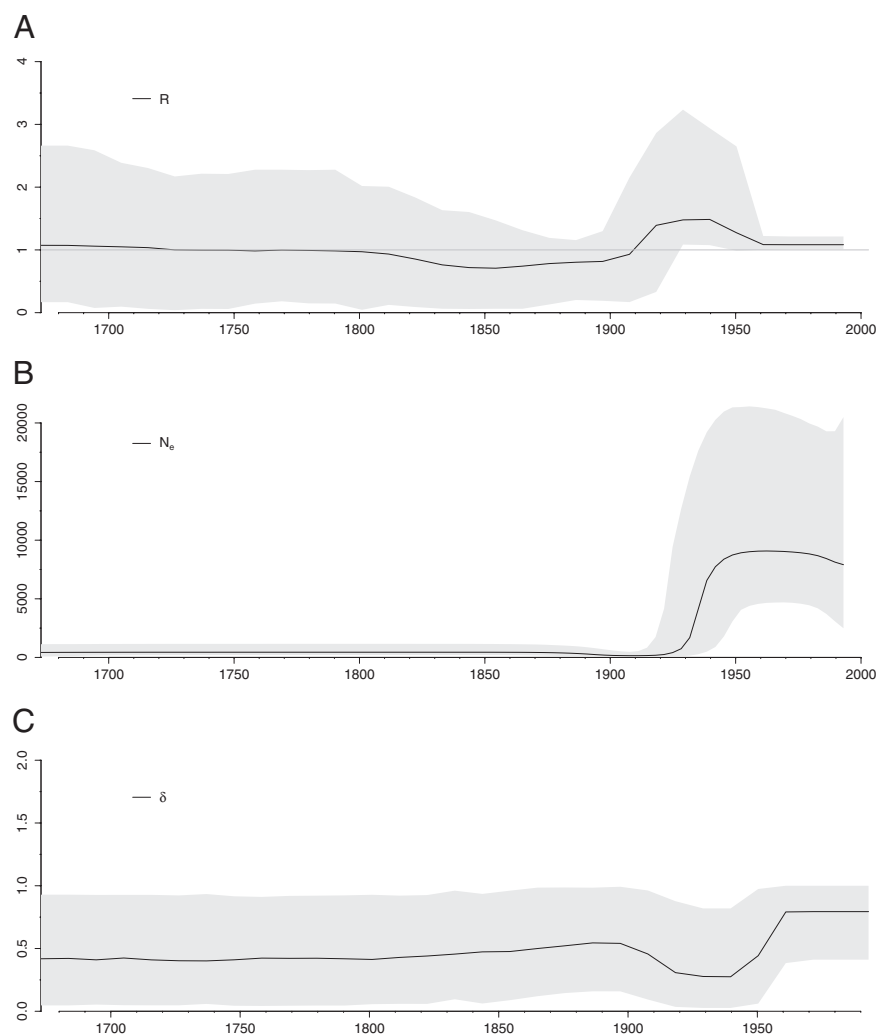


Fig. 3. HCV in Egypt: Birth-death skyline plot (A and C) and Bayesian skyline plot (B) median estimates and highest posterior density intervals for the (A) effective reproductive number, (B) effective population size, and (C) becoming noninfectious rate.

population size. Recent advances were made to account for incidences and prevalences in coalescent approaches (7, 20, 21), and it will be interesting to compare the performance of the coalescent-based approaches to our birth-death approach.

We expect our approach to be superior in the case of large sample sizes and early epidemic outbreaks. The coalescent is an approximation of population dynamics assuming small sample sizes, whereas the birth-death skyline model accounts for any sampling intensity by parameterizing directly the sampling probability. The coalescent assumption is violated when using data from large cohort studies (e.g., some HIV cohorts have enrolled more than one-half of the infected individuals), but also when sparse sampling is performed in a scenario of increasing population size (e.g., epidemic outbreak): the genealogy may contain almost all lineages back in time when population size was small. The coalescent, assuming a deterministic (potentially changing) population size through time, is overconfident when the number of lineages in the tree is close to the population size, as stochastic population size changes are ignored. We observed this phenomenon in simulations of epidemic outbreak scenarios (9) as well as in the analyzed empirical data (see above). We want to emphasize that we did not see a bias of the parameters estimated by the coalescent; the problem is that the estimated HPD intervals were too narrow to contain the true parameter values with high probability.

Overall, the birth-death skyline model provides a different lens on phylodynamic tree shapes and thus past population

dynamics. Future work would aim to incorporate geographical and contact network structure into the tree prior, thus enabling analysis of epidemics that stretch over larger geographic regions, or involve complex contact network structure.

Methods

Birth-Death Skyline Plot. Model definition. For modeling piecewise constant transmission, death/recovery and sampling rates, we define a birth-death skyline model as follows. The process starts at time $0 = t_0$ with a single individual and is stopped at time t_m .

The vector $t = (t_1, \dots, t_{m-1})$, with $t_0 < t_1 < \dots < t_{m-1} < t_m$, determines the times of rate shift events. The vector $\lambda = (\lambda_1, \dots, \lambda_m)$, where $\lambda_i > 0$ ($i = 1, \dots, m$), defines the transmission rates. λ_i is the transmission rate for each infected individual in time interval $[t_{i-1}, t_i]$. The vector $\mu = (\mu_1, \dots, \mu_m)$, where $\mu_i \geq 0$ ($i = 1, \dots, m$), defines the death rates. Note that the death rates may be larger than the birth rates. μ_i is the death rate for each infected individual in time interval $[t_{i-1}, t_i]$. The vector $\psi = (\psi_1, \dots, \psi_m)$, where $\psi_i \geq 0$ ($i = 1, \dots, m$), defines the sampling rates. ψ_i is the sampling rate for each infected individual in time interval $[t_{i-1}, t_i]$.

At time point t , where $t_{i-1} \leq t < t_i$, each individual transmits with rate λ_i ; the two subtending lineages are denoted "left" and "right" such that we can distinguish between them. Each individual dies with rate μ_i . Dying means that an individual is not infectious any more, which might be caused by a number of factors: death, recovery, behavior change, or successful treatment. Each individual is sampled with rate ψ_i . Once an individual is sampled, it is removed from the population (i.e., sampling is followed by becoming noninfectious). This assumption is valid when considering epidemics in which sampling is followed by, for example, effective treatment, thus reducing the risk of further transmission significantly. For epidemics in which transmission

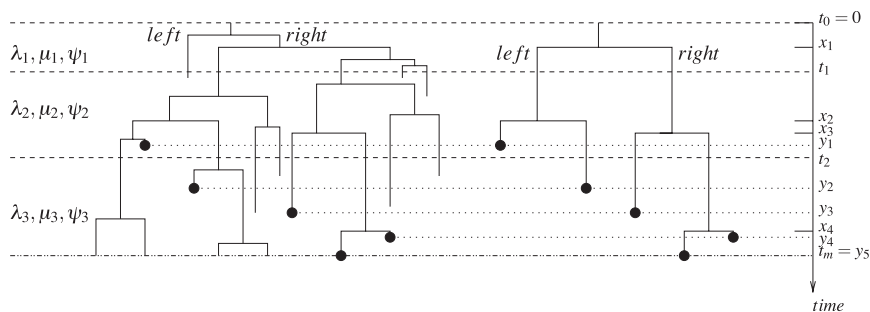


Fig. 4. Notation under the birth-death skyline model. (Left) A tree that evolved under the birth-death skyline model with two rate shifts (at time t_1 and t_2); the sampled individuals are marked with a black circle. (Right) Corresponding sampled tree, i.e., all lineages with no sampled descendants are pruned.

after sampling is common, we ideally want to assume a probability r of becoming noninfectious after transmission. However, even though the likelihood of a tree under such a model is known (9), such a model cannot be used directly in BEAST: for $r < 1$, sampled individuals may occur along branches, whereas in BEAST the sampled individuals must be the tips of the tree.

We point out that the birth-death skyline model parameterizes important epidemiological quantities: The effective reproductive number is specified by the following:

$$R = \frac{\lambda}{\mu + \psi},$$

the duration of infection ($\tau = \frac{1}{\delta}$) is determined by the total rate of becoming noninfectious,

$$\delta = \mu + \psi,$$

and the probability of an individual being sampled is as follows:

$$s = \frac{\psi}{\mu + \psi},$$

where sampling happens at the time of becoming noninfectious. From the three epidemiological parameters R , δ , s , we can calculate the birth-death skyline parameters through $\lambda = R\delta$, $\mu = \delta - s\delta$, $\psi = s\delta$.

In an extension of the birth-death skyline model, we allow for special sampling efforts at the times t_i . Each infected individual at time t_i is sampled (and becomes thus noninfectious) with rate ρ_i . Note that a process with $\psi = 0$ and $\rho_1, \dots, \rho_{m-1} = 0$, $\rho_m > 0$ induces a tree of only contemporaneous tips. Such a model may be used for species phylogenies (10) or virus data collected at one point in time. Note that, under this model, we do not have to make an assumption about the infectiousness of sampled individuals, as we do not consider the epidemic postsampling (so our estimates are independent from the dynamics after sampling).

A birth-death skyline model induces a tree (Fig. 4, Left). Suppressing all extinct and all nonsampled lineages yields the sampled tree (Fig. 4, Right). In the sampled tree, we subdivide the edge from t_s to t_e with time $t_s < t_i < t_e$ at time t_i , which yields a degree-two vertex at time t_i .

Let the number of sequentially sampled leaves (i.e., the leaves sampled at times different from $t = t_1, \dots, t_m$) in a sampled tree be n . Let the number of leaves sampled at time t_i be N_i , and let $\sum_{i=1}^m N_i = N$. Let the number of degree-two vertices at time t_i be n_i , with $n_m = 0$. The $N + n - 1$ transmission times in a sampled tree with $N + n$ sampled individuals are $x_1 < x_2 < \dots < x_{N+n-1}$, the sampling times of the sequentially sampled tips are $y_1 < \dots < y_n$.

Note that a phylogeny inferred from data does not have the orientation labels "left" and "right," but each leaf has a unique label. However, we consider oriented trees and calculate their probability density because of convenient notation. The probability density of a labeled tree with n leaves (each leaf is assigned a unique label, and the orientations left and right are dropped) follows from the equations for an oriented tree by multiplying with $\frac{2^{n-1}}{n!}$, as each of the $n!$ labelings and each of the 2^{n-1} orientations is equally likely (22). Note that the factor $\frac{2^{n-1}}{n!}$ is constant for a given sample of size n , and therefore it can be discarded. Throughout the rest of the paper, we will only consider oriented trees.

Probability density of a sampled tree. We will now provide the probability density of a sampled tree \mathcal{T} under the birth-death skyline model. This probability density will be used as a prior distribution for trees in the software package BEAST2 as an alternative to the coalescent-based Bayesian

skyline plot. We show in SI Appendix, Theorem S4, the following key mathematical result.

Theorem 1. The probability density of a sampled tree, \mathcal{T} , conditioned on at least one sampled individual (S), is as follows:

$$f[\mathcal{T} | \lambda, \mu, \psi, \rho, t, S] = \frac{q_1(0)}{1 - \rho_1(0)} \prod_{i=1}^{N+n-1} \lambda_{l(x_i)} q_{l(x_i)}(x_i) \cdot \prod_{i=1}^n \frac{\psi_{l(y_i)}}{q_{l(y_i)}(y_i)} \prod_{i=1}^m \rho_i^{N_i} ((1 - \rho_i) q_{i+1}(t_i))^{N_i},$$

with $l(t) = i$ iff $t_{i-1} \leq t < t_i$ and with $(t_{i-1} \leq t < t_i, i = 1, \dots, m)$,

$$A_i = \sqrt{(\lambda_i - \mu_i - \psi_i)^2 + 4\lambda_i\psi_i},$$

$$B_i = \frac{(1 - 2(1 - \rho_i)\rho_{i+1}(t_i))\lambda_i + \mu_i + \psi_i}{A_i},$$

$$p_i(t) = \frac{\lambda_i + \mu_i + \psi_i - A_i \frac{e^{A_i(t-t_i)}(1 + B_i) - (1 - B_i)}{e^{A_i(t-t_i)}(1 + B_i) + (1 - B_i)}}{2\lambda_i},$$

$$q_i(t) = \frac{4e^{-A_i(t-t_i)}}{(e^{-A_i(t-t_i)}(1 + B_i) + (1 - B_i))^2},$$

and $\rho_{m+1}(t_m) = 1$.

Note that $p_i(t)$ is the probability that an infected individual at time t has no sampled descendants when the process is stopped (i.e., at time t_m), with $t_{i-1} \leq t < t_i$ ($i = 1, \dots, m$) (SI Appendix, Theorem S1). Furthermore, $q_i(t)$ is the probability density of an individual at time t giving rise to an edge between time t and t_i (SI Appendix, Theorem S2) with $q_i(t_i) = 1$.

We typically condition the tree probability density on observing at least one sampled individual (S), motivated by the fact that we only consider datasets with at least one sample. However, if the tree probability density shall not be conditioned on observing at least one sample, $f[\mathcal{T} | \lambda, \mu, \psi, \rho, t]$ is provided in SI Appendix, Theorem S3.

Procedure for calculating $f[\mathcal{T} | \lambda, \mu, \psi, \rho, t, (S)]$. Theorem 1 allows us to evaluate the probability density of a tree efficiently, which is required in each Markov chain Monte Carlo (MCMC) step as follows:

Calculate A_i for $i = m, \dots, 1$.

Calculate B_i and $p_i(t_{i-1})$ recursively for $i = m, \dots, 1$.

Calculate $q_i(t)$ for $i = m, \dots, 1$.

Calculate the tree probability density $f[\mathcal{T} | \lambda, \mu, \psi, \rho, t, S]$ (if not conditioning on S , use SI Appendix, Theorem S3, for calculating $f[\mathcal{T} | \lambda, \mu, \psi, \rho, t]$).

Bayesian Inference Implementation. A Bayesian inference method is implemented, which takes a multiple sequence alignment D and the sampling times y as the input data. We form the posterior distribution over the product space of trees and birth-death skyline parameters based on the multiple sequence alignment D and the sampling times y as follows. \mathcal{T} is a tree with sampling times y . Let the event C be that the most recent sampling event is at present (t_m). We assume that in our data we observe C , i.e., we assume $y_{N+n-1} = t_m$, which reduces the number of parameters by one (as we specify $t_m - y_{N+n-1} = 0$). S is the optional conditioning on sampling at least one individual in the tree. The posterior can be rewritten as follows using Bayes' theorem:

$$\begin{aligned}
 f[T, \lambda, \mu, \psi, \rho, t, \nu | D, y, C; (S)] &= \frac{f[T, \lambda, \mu, \psi, \rho, t, \nu | D, y; (S)]}{f[C | D, y; (S)]} \\
 &= \frac{\mathbb{P}[D | \nu T] f[T | \lambda, \mu, \psi, \rho, t; (S)] f[\lambda, \mu, \psi, \rho, t, \nu | (S)]}{f[D, y, C; (S)]} \\
 &\propto \mathbb{P}[D | \nu T] f[T | \lambda, \mu, \psi, \rho, t; (S)] f[\lambda, \mu, \psi, \rho, t, \nu | (S)],
 \end{aligned}$$

where ν is a vector of substitutions rates (one per branch) that convert the time tree into a phylogenetic tree with branch lengths in units of substitutions per site and $\mathbb{P}[D | \nu T]$ is the phylogenetic likelihood (23).

Note that D, y, C is part of the data. If conditioning on S , we consider S as fixed, and thus S is influencing the prior distributions for $\lambda, \mu, \psi, \rho, t, \nu, T$.

Metropolis–Hastings MCMC (24–26) is used to sample $(T, \lambda, \mu, \psi, \rho, t, \nu) \sim f[T, \lambda, \mu, \psi, \rho, t, \nu | D, y, C; (S)]$. Our implementation in BEAST2 uses the equation for $f[T | \lambda, \mu, \psi, \rho, t, S]$ given in Theorem 1 (alternatively, also the equation in *SI Appendix, Theorem S3*, for $f[T | \lambda, \mu, \psi, \rho, t]$ can be used). The expressions are evaluated as explained above (Procedure for calculating $f[T | \lambda, \mu, \psi, \rho, t, S]$).

The choice of the number of intervals m depends on the aim of the analysis. It should be large enough to capture all rate changes, but small enough to avoid overparameterization. In the empirical analyses and the SIR simulations, we chose $m = 10$ and chose equidistant intervals, meaning the interval width changes whenever the distance from origin to last tip changes. For simulation scenarios 1–4 (*SI Appendix*), we chose $m = 2$ and estimated the interval sizes. A future improvement of the method would be to select the number of intervals and their locations using nonparametric approaches such as have been developed for the coalescent (5, 27–30).

In our analyses, we used as parameter priors for R and δ gamma or log-normal distributions, and a beta distribution for s . All our prior choices for the different simulation and empirical data analyses are summarized in *SI Appendix, Table S5*.

We show in *SI Appendix* that we can vary at least one parameter in the birth–death skyline model with changing the others in a certain way, without changing the likelihood of the sampled tree. For inference, this means that we need to put a strong prior on at least one parameter to obtain confined HPD intervals. In most of our analyses, we constrained the sampling proportion to be constant through time.

The model accommodates special cases in which not all model parameters are allowed to change. For example, a researcher might know that δ is constant while R may change over time.

Simulation Study. In our simulation study, we first simulated under the birth–death skyline model, i.e., assuming piecewise constant rates. We show that

we can accurately reestimate the parameters using our birth–death skyline inference method (Table 1 and *SI Appendix, Tables S1–S4*). Second, we simulated under a discrete-state continuous-time version (14, 15) of the SIR model (16) and showed that the simulated trajectory of R is accurately recovered by our birth–death skyline inference method (Fig. 1 and *SI Appendix, Fig. S1*), choosing the number of intervals to be $m = 10$ with equidistant intervals per step. The chosen prior distributions for the different analyses are summarized in *SI Appendix, Table S5*. Details about the simulations are provided in *SI Appendix*.

Empirical Data Analysis. We demonstrate the model on two example datasets: (i) a serially sampled HIV-1 dataset from the United Kingdom and (ii) a contemporaneous HCV dataset from Egypt. In both analyses, we apply log-normal priors to the effective reproductive number R and the total death rate δ (for a summary of the priors, see *SI Appendix, Table S5*), and we condition on at least one sampled individual. We choose the number of intervals to be $m = 10$ with equidistant intervals per step. Setting the number of intervals to 10 allowed up to nine rate changes throughout the time between the cluster's origin and the latest sampling date.

HIV-1 type B in the United Kingdom. In this analysis of 62 HIV-1 sequences, we constrain the sampling probability s to be zero before 1999.9 and constant afterward to avoid sampling bias, because no sampling was taken before 1999.9. Sampling dates are known to the day for all but two of the samples, for which only the year was known. The analysis used a GTR + Γ + I substitution model, and its parameters were coestimated alongside the birth–death skyline model parameters and the two missing tip dates (31). We fixed the substitution rate to 2.55×10^{-3} (as in ref. 17) and compared the birth–death skyline reconstruction with the Bayesian skyline plot analysis (Fig. 2). The prior and posterior distributions for the individual parameters are shown in *SI Appendix, Figs. S3–S5*. When the rate is estimated simultaneously, the HPD intervals are slightly broader (*SI Appendix, Fig. S2*).

HCV in Egypt. This dataset, consisting of 63 Egyptian sequences from 1993, was analyzed with a GTR + Γ + I substitution model under a strict clock branch rate model with the clock rate fixed to 0.79×10^{-3} substitutions per site per year (19), and the remaining parameters were coestimated alongside the birth–death skyline model parameters (Fig. 3). The prior and posterior distributions for the individual parameters are shown in *SI Appendix, Figs. S6 and S7*.

ACKNOWLEDGMENTS. T.S. and S.B. thank the Swiss National Science Foundation for funding. D.K. and A.J.D. were supported by Marsden Grant UOA0809 from the Royal Society of New Zealand.

- Kingman JFC (1982) On the genealogy of large populations. *J Appl Probab* 19A:27–43.
- Rodrigo A, Felsenstein J (1999) Coalescent approaches to HIV population genetics. *The Evolution of HIV* (Johns Hopkins Univ Press, Baltimore), pp 233–272.
- Pybus OG, Rambaut A, Harvey PH (2000) An integrated framework for the inference of viral population history from reconstructed genealogies. *Genetics* 155(3):1429–1437.
- Strimmer K, Pybus OG (2001) Exploring the demographic history of DNA sequences using the generalized skyline plot. *Mol Biol Evol* 18(12):2298–2305.
- Drummond AJ, Rambaut A, Shapiro B, Pybus OG (2005) Bayesian coalescent inference of past population dynamics from molecular sequences. *Mol Biol Evol* 22(5):1185–1192.
- Drummond AJ, Rambaut A (2007) BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol Biol* 7:214.
- Volz EM, Kosakovsky Pond SL, Ward MJ, Leigh Brown AJ, Frost SD (2009) Phylodynamics of infectious disease epidemics. *Genetics* 183(4):1421–1430.
- Rasmussen DA, Ratmann O, Koelle K (2011) Inference for nonlinear epidemiological models using genealogies and time series. *PLoS Comput Biol* 7(8):e1002136.
- Stadler T, et al. (2012) Estimating the basic reproductive number from viral sequence data. *Mol Biol Evol* 29(1):347–357.
- Stadler T (2011) Mammalian phylogeny reveals recent diversification rate shifts. *Proc Natl Acad Sci USA* 108(15):6187–6192.
- Stadler T (2010) Sampling-through-time in birth–death trees. *J Theor Biol* 267(3):396–404.
- Cohen MS, et al. (2011) Prevention of HIV-1 infection with early antiretroviral therapy. *N Engl J Med* 365(6):493–505.
- Anderson RM, May RM (1979) Population biology of infectious diseases: Part I. *Nature* 280(5721):361–367.
- Chen WY, Bokka S (2005) Stochastic modeling of nonlinear epidemiology. *J Theor Biol* 234(4):455–470.
- Kühnert D, Wu CH, Drummond AJ (2011) Phylogenetic and epidemic modeling of rapidly evolving infectious diseases. *Infect Genet Evol* 11(8):1825–1841.
- Kermack WO, McKendrick AG (1927) Contributions to the mathematical theory of epidemics–I. *Proc R Soc Lond A Math Phys Sci* 115:700–721.
- Hué S, Pillay D, Clewley JP, Pybus OG (2005) Genetic analysis reveals the complex structure of HIV-1 transmission within defined risk groups. *Proc Natl Acad Sci USA* 102(12):4425–4429.
- Frank C, et al. (2000) The role of parenteral antischistosomal therapy in the spread of hepatitis C virus in Egypt. *Lancet* 355(9207):887–891.
- Pybus OG, et al. (2001) The epidemic behavior of the hepatitis C virus. *Science* 292(5525):2323–2325.
- Volz EM (2012) Complex population dynamics and the coalescent under neutrality. *Genetics* 190(1):187–201.
- Koelle K, Rasmussen D (2012) Rates of coalescence for common epidemiological models at equilibrium. *J R Soc Interface* 9(70):997–1007.
- Ford D, Matsen FA, Stadler T (2009) A method for investigating relative timing information on phylogenetic trees. *Syst Biol* 58(2):167–183.
- Felsenstein J (1981) Evolutionary trees from DNA sequences: A maximum likelihood approach. *J Mol Evol* 17(6):368–376.
- Metropolis N, et al. (1953) Equation of state calculations by fast computing machines. *J Chem Phys* 21(6):1087–1092.
- Hastings W (1970) Monte-Carlo sampling methods using Markov chains and their applications. *Biometrika* 57(1):97–109.
- Drummond AJ, Nicholls GK, Rodrigo AG, Solomon W (2002) Estimating mutation parameters, population history and genealogy simultaneously from temporally spaced sequence data. *Genetics* 161(3):1307–1320.
- Oppen-Rhein R, Fahrmeir L, Strimmer K (2005) Inference of demographic history from genealogical trees using reversible jump Markov chain Monte Carlo. *BMC Evol Biol* 5(1):6.
- Minin VN, Bloomquist EW, Suchard MA (2008) Smooth skyride through a rough skyline: Bayesian coalescent-based inference of population dynamics. *Mol Biol Evol* 25(7):1459–1471.
- Palacios J, Minin V (2011) Gaussian process-based bayesian nonparametric inference of population trajectories from gene genealogies. arXiv:1112.4138.
- Palacios J, Minin V (2012) Integrated nested laplace approximation for bayesian nonparametric phylodynamics. *Population (Paris)* 40:60.
- Shapiro B, et al. (2011) A Bayesian phylogenetic method to estimate unknown sequence ages. *Mol Biol Evol* 28(2):879–887.