

Fundamental identifiability limits in molecular epidemiology

Stilianos Louca^{1,2*}, Angela McLaughlin^{3,4}, Ailene MacPherson^{5,6,7}, Jeffrey B. Joy^{3,4,8} & Matthew W. Pennell^{5,6*}

¹*Department of Biology, University of Oregon, USA*

²*Institute of Ecology and Evolution, University of Oregon, USA*

³*British Columbia Centre for Excellence in HIV/AIDS, Vancouver, Canada*

⁴*Bioinformatics, University of British Columbia, Vancouver, Canada*

⁵*Biodiversity Research Centre, University of British Columbia, Vancouver, Canada*

⁶*Department of Zoology, University of British Columbia, Vancouver, Canada*

⁷*Department of Ecology and Evolutionary Biology, University of Toronto, Toronto, Canada*

⁸*Department of Medicine, University of British Columbia, Vancouver, Canada*

*Correspondence should be addressed to SL (louca.research@gmail.com) and MWP (pennell@zoology.ubc.ca)

Abstract

Viral phylogenies provide crucial information on the spread of infectious diseases, and many studies fit mathematical models to phylogenetic data to estimate epidemiological parameters such as the effective reproduction ratio (R_e) over time. Such phylodynamic inferences often complement or even substitute for conventional surveillance data, particularly when sampling is poor or delayed. It remains generally unknown, however, how robust phylodynamic epidemiological inferences are, especially when there is uncertainty regarding pathogen prevalence and sampling intensity. Here we use recently developed mathematical techniques to fully characterize the information that can possibly be extracted from serially collected viral phylogenetic data, in the context of the commonly used birth-death-sampling model. We show that for any candidate epidemiological scenario, there exists a myriad of alternative, markedly different and yet plausible “congruent” scenarios that cannot be distinguished using phylogenetic data alone, no matter how large the dataset. In the absence of strong constraints or rate priors across the entire study period, neither maximum-likelihood fitting nor Bayesian inference can reliably reconstruct the true epidemiological dynamics from phylogenetic data alone; rather, estimators can only converge to the “congruence class” of the true dynamics. We propose concrete and feasible strategies for making more robust epidemiological inferences from viral phylogenetic data.

Keywords: epidemiology; phylogenetics; statistical inference; birth-death-sampling model

For rapidly evolving pathogens, such as RNA viruses, genetic diversity accumulates on the same timescale as transmission [1]. Thus, pathogen genealogies reconstructed from patient samples can provide valuable information on the transmission dynamics of diseases [2, 3]. As sequencing technology and computational methods continue to improve, such phylogenetic approaches [4, 5] are increasingly being used to help inform public health policy during ongoing epidemics, such as during the 2013-2016 Ebolavirus outbreak [6], the 2015-2016 expansion of Zika virus [7], and the SARS-CoV-2 pandemic that began in 2019 [8]. One of the most popular mathematical frameworks used for such phylodynamic inferences is the birth-death (BD) model [9–11], variants of which are also used to reconstruct macroevolutionary dynamics [12]. BD models are typically either fitted to a given time-calibrated phylogeny (henceforth timetree) or jointly estimated with the timetree from molecular sequences, to obtain estimates of the birth or speciation rate (λ , corresponding to transmission between hosts in epidemiology or speciation in macroevolution), the death or extinction rate (μ , host death or recovery in epidemiology; extinction in macroevolution), and the sampling rate (ψ , number of pathogen lineages sampled per time and per extant lineage) through time. From these rates one can calculate critical epidemiological parameters such as the effective reproduction ratio $R_e = \lambda/(\mu + \psi)$ [9, 10].

Despite the increasing importance of phylodynamic estimates to public health policy, it is generally unknown precisely what information we can hope to extract from phylogenetic data and how robust these estimates are expected to be, particularly when all rates exhibit temporal variation and are unknown a priori. In macroevolutionary BD models, where only extant lineages are sampled, much work has been done to understand identifiability limits [13–18]. That work has shown that if the birth rate λ and death rate μ vary through time, there are a vast number of alternative plausible combinations of λ and μ that could explain any extant timetree (i.e., comprising only extant species) equally well. Such “congruent” scenarios cannot be distinguished from one another statistically, even in the presence of an infinitely large and completely sampled timetree — in other words, it is impossible to design asymptotically consistent estimation methods for λ and μ based on extant timetrees alone and without strong prior assumptions [18].

It is largely unknown to what extent and in what form such congruency issues exist in epidemiological BD models, i.e., with continuous sampling through time. While some relatively minor parameter correlations have been known for special cases [10, 19], an understanding of general identifiability limits is lacking, and the macroevolutionary case has taught us that parameter correlations known for specific cases might severely underestimate the full extent of the problem. This question is non-trivial: while epidemiological BD models appear similar to macroevolutionary BD models, they are more complex because pathogen sequences are typically not sampled at a single final time point. Samples obtained serially through time provide additional information on an epidemic, however new uncertainty is introduced when the sampling rate (ψ) is unknown and, especially, when it varies over time.

Here we provide a definite answer to the above questions and demonstrate that, similar to the macroevolutionary case, there are fundamental limits to how much information can be gleaned from timetrees sampled through time in the absence of strong additional constraints. Specifically, we prove mathematically that for any one hypothesized birth-death-sampling (BDS) scenario — i.e., with specific time-varying λ , μ and ψ , there exists an infinite number of alternative, markedly different and yet plausible BDS scenarios that are statistically indistinguishable from the hypothesized scenario, even with infinitely large phylogenetic datasets. Using simulations and real sequence data from an HIV outbreak in Northern Alberta, Canada we demonstrate that this identifiability issue means that many epidemiological inferences from phylogenetic data alone may not be as well-supported as previously thought. Fortunately, and in contrast to the macroevolutionary case, we are able to identify concrete and feasible approaches towards resolving these issues in practice.

Identifiability of general birth-death-sampling models

Our starting point is the general birth-death-sampling (BDS) model with arbitrary time-dependent birth rate λ , death rate μ and sampling rate ψ , where we make the common assumptions that sampled lineages (tips in the phylogeny) are immediately removed from the pool of extant lineages and that branching events correspond to transmission events [20]. We use the term ‘‘BDS scenario’’ (or ‘‘epidemiological scenario’’) to refer to a specific choice of profiles over time for the parameters λ , μ and ψ . Using mathematical techniques similar to those developed for macroevolutionary models [18], we find that the likelihood of any timetree under a given epidemiological scenario is entirely determined by only two model parameters, called the pulled birth rate (denoted $\tilde{\lambda}$) and pulled sampling rate (denoted $\tilde{\psi}$). Here $\tilde{\lambda}$ is equal to the birth rate, λ , multiplied by the probability that a lineage is included in the phylogeny, while $\tilde{\psi}$ is equal to the sampling rate, ψ , divided by the probability that a lineage is included in the phylogeny (overview of symbols in Table 1). The $\tilde{\lambda}$ and $\tilde{\psi}$ are thus the expected occurrence rate of internal nodes and tips, respectively, over time when divided over the current number of lineages in the tree and in the limit of infinitely large trees (proof in Supplement S.1.3). Note that $\tilde{\lambda}$ and $\tilde{\psi}$ are purely theoretical properties of the BDS scenario that can be calculated from λ , μ and ψ , and do not depend on any particular data set. We henceforth call any two BDS scenarios congruent if they have the same pulled birth rate $\tilde{\lambda}$ and the same pulled sampling rate $\tilde{\psi}$. By extension, the congruence class of any BDS scenario henceforth refers to the set of all congruent BDS scenarios. Any two congruent BDS scenarios generate timetrees with the same probability distribution (Supplements S.1.3 and S.1.4), and will yield identical likelihoods for any given timetree. This means that there is no way to distinguish between two congruent scenarios solely based on the properties of sampled timetrees, no matter how large. This result is analogous to that of macroevolutionary BD models (i.e., when all the tips are contemporaneous), where the probability distribution of generated timetrees is entirely determined by the pulled birth rate, and any two scenarios with identical pulled birth rates are statistically indistinguishable. This result is also analogous to identifiability issues discovered for population demographic models, where markedly different population demographics can generate the same sample frequency spectrum for arbitrarily large sample sizes [21, 22].

Several questions follow from this: For any given BDS scenario, (i) how can one easily determine if another scenario is congruent?; (ii) how many congruent scenarios are there?; and (iii) how different can the epidemiological implications of these congruent scenarios be? To answer these questions, it is useful to consider a number of alternative model parameters, the first of which is called the pulled diversification rate and defined as:

$$\tilde{r} = \lambda - \mu - \psi + \frac{1}{\lambda} \frac{d\lambda}{d\tau}, \quad (1)$$

where τ denotes age or time before present (Table 1). The pulled diversification rate is equal to the net diversification rate $r = \lambda - \mu - \psi$ (or net exponential growth rate in the case of an infectious disease) when the birth rate λ is constant over time, but differs from r otherwise. As we prove in Supplement S.1.3, two BDS scenarios are congruent if and only if they have the same pulled diversification rate \tilde{r} and the same pulled birth rate $\tilde{\lambda}$. Hence, to check if two BDS scenarios are congruent one can simply compare their \tilde{r} and $\tilde{\lambda}$. Equivalently, two BDS scenarios are congruent if and only if they exhibit the same deterministic branching density $\tilde{\beta}$ and the same deterministic sampling density $\tilde{\sigma}$, which can be interpreted as rescaled probability densities over time of any randomly chosen observed branching event or sampling event, respectively, in the limit of infinitely large trees (precise definitions in Supplement S.1.2). Hence, the shapes of $\tilde{\beta}$ and $\tilde{\sigma}$ contain easily interpretable information about the temporal distribution of branching and sampling events in the tree. Similarly to $\tilde{\lambda}$ and $\tilde{\psi}$, or $\tilde{\lambda}$ and \tilde{r} , so $\tilde{\beta}$ and $\tilde{\sigma}$ constitute an alternative parameterization of congruence classes, however we stress that all of these parameters do not contain sufficient information for recovering the original model parameters λ , μ and ψ .

Having introduced the above new parameters, it becomes easy to answer questions (ii) and (iii). For any given scenario (λ, μ, ψ) and any given alternative death rate μ^* , one can find a corresponding λ^* and ψ^* such that the new scenario $(\lambda^*, \mu^*, \psi^*)$ has the same pulled diversification rate and the same pulled birth rate, i.e., such that the new scenario is congruent to the first one (Supplement S.1.5.2). Indeed, one just needs to solve the differential equation:

$$\frac{d\lambda^*}{d\tau} = \lambda^* \cdot (\tilde{r} + \mu^* - \lambda^*) + \lambda\psi \quad (2)$$

with any initial condition, and choose $\psi^* = \lambda\psi/\lambda^*$. Many other analogous ways exist for creating congruent scenarios: For example, one can first specify an arbitrary sampling rate ψ^* and then adjust λ^* and μ^* accordingly, or first specify an arbitrary birth rate λ^* and then adjust ψ^* and μ^* accordingly, or first specify an arbitrary effective reproduction ratio R_e^* and adjust λ^* , μ^* and ψ^* (details in Supplement S.1.5). We note that some scenarios in a congruence class may have negative birth, death, or sampling rates and are therefore biologically irrelevant [23]. Since μ^* (or ψ^* , or λ^* , or R_e^*) can be chosen nearly arbitrarily and can depend on an arbitrary number of free parameters, the space of congruent BDS scenarios is infinitely large and infinite-dimensional. Many congruent scenarios can appear similarly plausible and similarly complex, and yet exhibit markedly different features, including very different values and opposite trends in λ , μ , ψ , or R_e (examples in Fig. 1).

This ambiguity limits the identifiability of epidemiological scenarios when based solely on phylogenetic data, even for infinitely large datasets. Indeed, whatever the true epidemiological history was, there will always exist an infinite number of congruent epidemiological histories. When fitting specific functional forms for λ , μ and ψ , such as piecewise constant profiles known as skyline models [10], estimators will generally converge to the scenario closest to the congruence class of the true epidemiological history, but not necessarily to the scenario closest to the true epidemiological history itself (Fig. 2). Depending on the particular functional forms chosen for λ , μ and ψ , this can yield markedly different rate profiles over time that may not even approximately resemble the true epidemiological history, including wrong trends and major spurious features (e.g., Fig. 2C). This can occur even if the fitted functional forms are flexible enough to reasonably approximate the true epidemiological history (e.g., Fig. 2B). In other words, even if one can closely identify the congruence class that best explains the data (for example, in terms of the $\tilde{\lambda}$ and $\tilde{\psi}$, or in terms of the $\tilde{\beta}$ and $\tilde{\sigma}$), without additional constraints or information one cannot determine which member of the congruence class generated the data. We emphasize that it is not necessary for multiple congruent scenarios to exist within the class of fitted functional forms in order to run into these issues. Since the scenarios in a congruent class are indistinguishable for any dataset of any size, with some generally appearing simpler and others more complex than the true epidemiological history, model selection techniques based on parsimony that do not incorporate additional information (e.g., from previous or independent studies) cannot alone resolve this issue (detailed discussion in Supplement S.2). In a Bayesian context, the existence of vastly different congruent scenarios means that the uncertainty in the estimated λ , μ , and ψ is even more sensitive to the choice of priors than would be apparent from comparing the posterior to the prior distributions of the parameters of some fitted model class. These issues are also expected to affect hypothesis testing approaches that rely on model fitting, for example fitting a linear profile for λ and examining whether the fitted slope is statistically significant to see if λ has been changing in a specific direction. Following the above arguments, the fitted slope need not at all reflect the true trend in λ (e.g., it could point in the opposite direction), and for a sufficiently large dataset this wrong slope will inevitably be statistically significant. The above issues also apply to other equivalent model parameterizations used in epidemiology, for example based on R_e , the sampling proportion ($S = \psi/(\mu + \psi)$) and the removal rate ($\delta = \mu + \psi$, also known as “become uninfected rate”) [10]. Note that, in contrast, $\tilde{\lambda}$, $\tilde{\psi}$, \tilde{r} , $\tilde{\beta}$ and $\tilde{\sigma}$ are asymptotically estimateable, i.e., given sufficiently large trees (Supplement S.1.2).

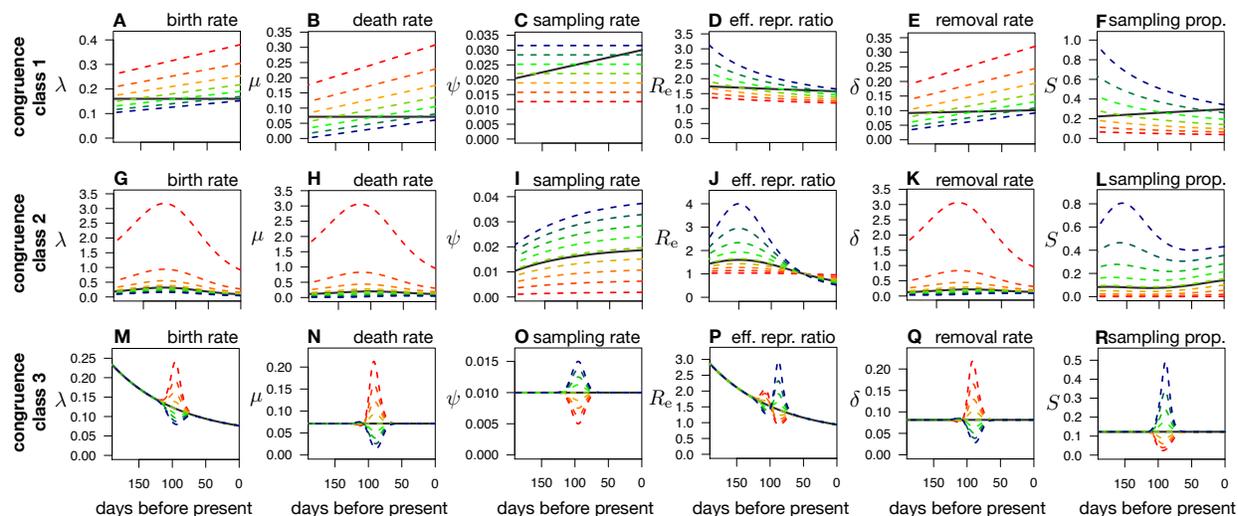


Figure 1: Examples of congruent epidemiological scenarios. (A–F) Birth (or speciation) rate (A), death (or extinction) rate (B), sampling rate (C), effective reproduction ratio $R_e = \lambda/(\mu + \psi)$ (D), removal rate $\delta = \mu + \psi$ (E), and sampling proportion $S = \psi/(\mu + \psi)$ (F) of a specific epidemiological scenario (thick black curves), compared to various alternative congruent (i.e., statistically indistinguishable) scenarios (dashed curves). Similarly colored curves across sub-figures correspond to a specific diversification scenario. No viral phylogeny, no matter how large, could possibly distinguish between these (and in fact a myriad of other) scenarios. (G–L) Similar to A–F, but showing scenarios congruent to a different reference scenario. (M–R) Similar to A–F, but showing scenarios congruent to a different reference scenario.

It was previously demonstrated that for constant-rate scenarios (i.e., where the parameters λ , μ and ψ do not vary with time) it is impossible to simultaneously estimate λ , μ and ψ from timetrees alone, because alternative combinations of constant λ , μ and ψ yield the same likelihood function [10, 19]. It is also known that when fitting BDS skyline models, at least one of the parameters must be fixed in at least one of the time intervals to eliminate correlations between parameter estimates [10]. Our findings are a generalization of these special cases and reveal that these underestimate the true extent of the issue. As soon as one considers that epidemiological parameters could vary over time in an unknown fashion [24–26], even if they can in principle be approximated by the considered model type (e.g., Fig. 2B), much more information is needed to “collapse” the congruence class and accurately reconstruct the true epidemiological dynamics. In particular, fitting generic BDS skyline models to real molecular data without constraints (or only constraining one parameter in one time interval), as is common practice [10, 27–34], cannot be expected to yield an accurate reconstruction, no matter how large the dataset.

Table 1: Overview of main BDS parameters discussed. A parameter is called congruence-invariant if it is identical across congruent BDS scenarios (and thus asymptotically identifiable). Non-congruence-invariant parameters cannot possibly be estimated from phylogenies alone (no matter how large) in the absence of strong additional constraints. Note that each parameter may be time-dependent, and that τ denotes age (time before present). Definitions are provided for non-standard BDS parameters. Birth and death rates refer to pathogen lineages, not the hosts.

symbol	description	definition	congr.-invariant
λ	birth (speciation) rate	-	no
μ	death (extinction) rate	-	no
ψ	sampling rate	-	no
R_e	effective reproduction ratio	$\lambda/(\mu + \psi)$	no
δ	removal (become uninfected) rate	$\mu + \psi$	no
S	sampling proportion	$\psi/(\mu + \psi)$	no
E	probability of a lineage missing from the phylogeny	Eq. (3) in S.1.1	no
$\tilde{\lambda}$	pulled birth rate	$(1 - E) \cdot \lambda$	yes
$\tilde{\psi}$	pulled sampling rate	$\psi/(1 - E)$	yes
\tilde{r}	pulled diversification rate	$\lambda - \mu - \psi + (1/\lambda)d\lambda/d\tau$	yes
$\tilde{\beta}$	deterministic branching density	Eq. (16) in S.1.2	yes
$\tilde{\sigma}$	deterministic sampling density	Eq. (17) in S.1.2	yes
\tilde{M}	normalized deterministic LTT	Eq. (15) in S.1.2	yes

The above results are analogous to the macroevolutionary case, where the pulled birth rate is asymptotically identifiable, but does not contain sufficient information for recovering the true birth and death rates [18]. Further, as in the macroevolutionary case, so here two congruent BDS scenarios have identical deterministic lineages-through-time (dLTT) curves (i.e., the number of lineages one would expect in the tree over time based on a deterministic interpretation of the rates λ , μ and ψ) when conditioned on the number of lineages at some given time point. In contrast to the macroevolutionary case, however, the reverse need no longer be true, i.e., two BDS scenarios with identical dLTTs need not necessarily be congruent. Moreover, while in the macroevolutionary case a congruence class could be uniquely described by a single time-dependent parameter (e.g., the dLTT, or the pulled birth rate over time), here a BDS congruence class is determined by two time-dependent parameters (e.g., the pulled birth rate and pulled diversification rate).

Simulation examples

To demonstrate the challenges for epidemiological inference stemming from the existence of model congruencies, we simulated various hypothetical but realistic epidemiological scenarios and then used two alternative well-established approaches for reconstructing the original dynamics from the generated data. In the first approach, we used the true timetree as an input and estimated the epidemiological dynamics via maximum-likelihood fitting. In the second approach we used nucleotide sequences simulated along the timetree as input, and jointly estimated the timetree together with the epidemiological dynamics using Bayesian MCMC. The latter approach resembles the common situation in molecular epidemiology where the phylogeny is not a priori known, thus introducing additional uncertainty in the reconstruction of the epidemic's dynamics. To avoid introducing our own biases, for example in the choice of priors, the Bayesian analysis was performed in a blinded way, with some members of our team conducting the simulations and others conducting the Bayesian inference.

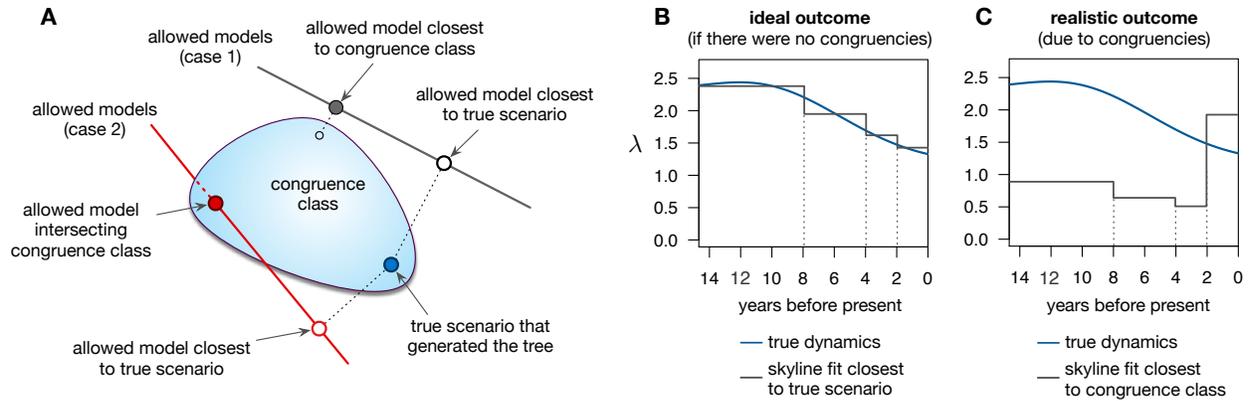


Figure 2: Conceptual illustration of the effects of model congruencies on epidemiological reconstruction. (A) The large light-blue “balloon” represents the congruence class of a single true epidemiological scenario (dark-blue circle), in the space of all biologically plausible epidemiological histories. Each straight continuous line represents a limited set of models or functional forms (e.g., skyline) fitted to data generated by the true scenario, for example via maximum likelihood, in an effort of approximately reconstructing the true scenario. The specific member (i.e., with a specific parameterization) chosen among each model set will be the one closest to the congruence class (filled grey circle) or may even intersect the congruence class (filled red circle), but will not necessarily be the one closest to the true scenario (open circles). This issue persists even for infinitely large datasets. (B and C) Hypothetical illustration of a skyline model (grey curve) fit to data generated by a hypothetical scenario (blue curve, here only showing λ). While no practical model perfectly matches reality, in the absence of model congruencies one would nevertheless expect to obtain a fit approximately resembling reality, roughly as shown in B. Instead, due to model congruencies one can easily obtain a fit that very poorly resembles the true scenario, as in C, since the fit closest to the congruence class is not necessarily the fit closest to the true scenario. See Fig. 4 for a real example.

For the maximum-likelihood inference we simulated timetrees with >50000 tips and fitted generic piecewise-linear profiles for λ , μ and ψ to each timetree, while selecting the optimal number of inflection points using AIC. We used such massive timetrees to avoid errors stemming from small sample sizes, thus focusing on identifiability issues. The fitted models matched the LTTs of the timetrees and the deterministic LTTs of the true scenarios nearly perfectly (Fig. 3D and Supplemental Fig. S1G). The fitted models also adequately explained the timetrees based on 3 different statistical tests performed via parametric bootstrapping [35] (Kolmogorov-Smirnov tests on the distributions of node ages, tip ages, and edge lengths, $P > 0.05$ in all cases). Despite the large sizes of the datasets and the adequacy of the models in explaining the data, the corresponding estimated λ , μ , ψ , R_e , δ and S were nearly always very different from the true profiles used in the simulations, sometimes even exhibiting opposite trends (Figs. 3A–C and Supplemental Figs. S1A–F and S2A–F). Importantly, the fitted models nearly exactly reproduced the deterministic branching density $\tilde{\beta}$ and deterministic sampling density $\tilde{\sigma}$ of the true scenario ($R^2 > 0.99$ in all cases), which, as explained earlier, implies that the fitted models came very close to the true scenario’s congruence class (Figs. 3E,F). This confirms our expectation that fitting yields an estimate of the true epidemiological history’s congruence class, but not necessarily of the true epidemiological history itself. These issues are expected to be even more pronounced with the smaller datasets typical in epidemiology, due to elevated stochasticity.

For the Bayesian inference we considered two alternative epidemiological scenarios to simulate timetrees of sizes typical in molecular epidemiology (590 and 540 tips, spanning about 15 years, Supplemental Fig. S3). The epidemiological scenarios and the nucleotide substitution models exhibited parameters typically reported for HIV-1 [36, 36–38]. The profiles of λ , μ , ψ , R_e , δ , and S exhibited moderate variation over time that could be well-approximated using skyline models with three to four intervals. For each of the two

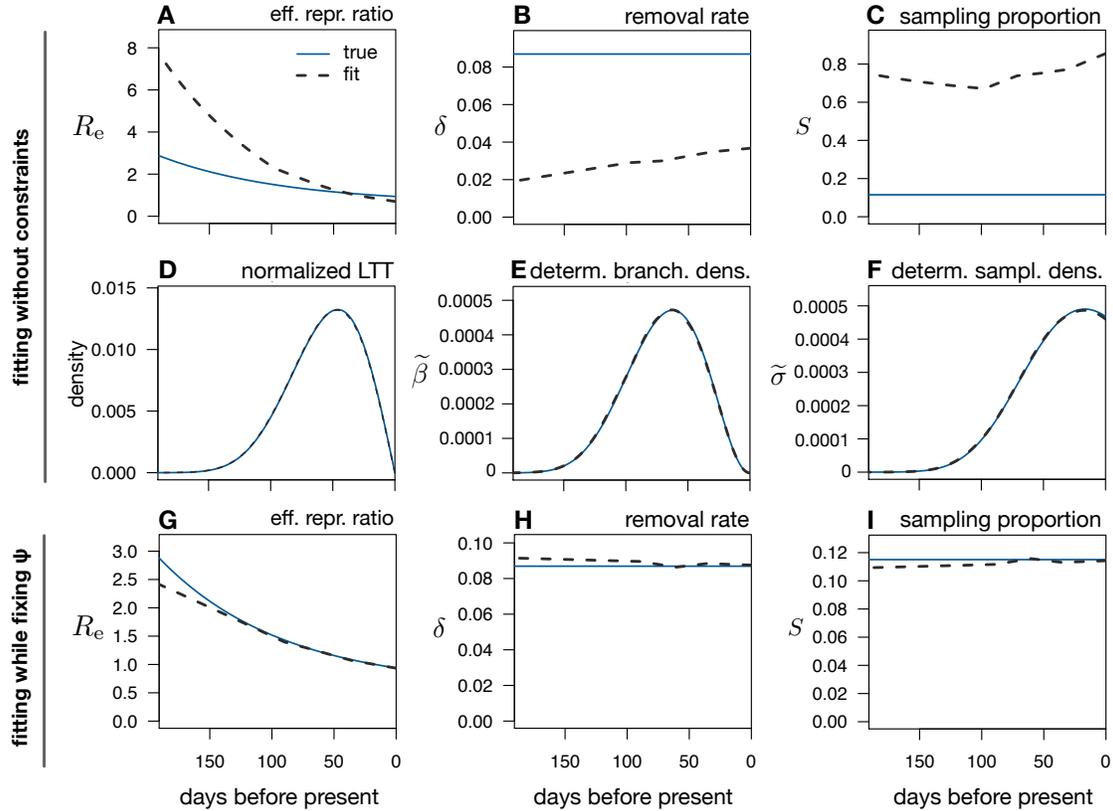


Figure 3: Limits to reconstructing epidemic's dynamics via maximum likelihood. (A–C) Maximum-likelihood estimates (grey dashed curves) of the effective reproduction ratio (R_e), removal rate ($\delta = \mu + \psi$), and sampling proportion ($S = \psi/(\mu + \psi)$) over time, based on a timetree with 175,440 tips simulated under a hypothetical birth-death-sampling scenario (blue continuous curves). Rates are in day^{-1} . Model adequacy was confirmed via predictive posterior simulations with multiple test statistics. Observe the poor agreement between the estimated and true profiles. (D–F) Maximum-likelihood estimates of the deterministic lineages-through-time curve (normalized to have unit area under the curve), deterministic branching density ($\tilde{\beta}$), and deterministic sampling density ($\tilde{\psi}$), corresponding to the same fitted model as in A–C, compared to their true profiles. The good agreement between the inferred and true profiles shows that the fitted model converged towards the true epidemiological scenario's congruence class but not the true scenario itself. (G–I) Maximum-likelihood estimates of R_e , δ , and S inferred from the same data as in A–F, while fixing the sampling rate ψ to its true profile. For additional BDS parameters see Supplemental Fig. S1. For a statistical analysis of estimation accuracies across trees simulated from many random scenarios, see Supplemental Figs. S17 and S19.

scenarios we then conducted a BDS skyline model inference in BEAST2 [5] using the sampled sequences as input data. As indicated above, this inference was internally blinded, i.e., the team members performing the inference had no knowledge of the true epidemiological scenarios used in the simulations. The sole information provided was (a) that model parameters were within the typical ranges known from HIV-1, (b) that four time intervals are sufficient for reasonably approximating the true dynamics with a skyline model (thus avoiding complications in the selection of the number of intervals), (c) the nucleotide substitution model and number of rate categories used, and (d) the value of the present-day sampling proportion, to account for known identifiability issues within skyline models (one parameter in one time interval must be fixed to eliminate correlations between parameters according to Stadler *et al.* [10]). Each of the parameters R_e , δ , and S varied over time with rate shifts 2, 4, and 8 years before the present, and the sampling proportion

in the present was fixed to its known value. The adequacy of the posterior models to explaining the true tree was fully confirmed via predictive posterior simulations [35] based on the same statistical tests as used for the maximum-likelihood fits. The molecular clock rate and the parameters of the nucleotide substitution model were also largely accurately estimated (Supplemental Figs. S10 and S12).

In contrast, in both scenarios the estimated λ , μ , ψ , R_e , δ , and S deviated substantially from their true profiles, in many cases exhibiting different trends over time or differing by more than an order of magnitude (for example, when considering the maximum-posterior or median-posterior profiles, Fig. 4A–C and Supplemental Figs. S5 and S7), consistent with our expectations. Moreover, the posterior 95% equal-tailed credible intervals barely overlapped with the true profiles of these parameters, suggesting that the inferred posterior distributions severely underestimate the true uncertainty of the parameters. This is not surprising, since these posteriors only reflect the uncertainty stemming from finite data sizes, but do not account for congruencies. Importantly, in all cases the posterior distributions of the deterministic branching and sampling densities ($\tilde{\beta}$ and $\tilde{\sigma}$) predicted by BEAST2 closely resembled their true profiles (Fig. 4D–F). This implies that BEAST2 accurately inferred the congruence class of the true epidemiological history, just not the true epidemiological history itself (again, consistent with our predictions). To further test this interpretation, we repeated the Bayesian inference while fixing the removal rate to its true profile (approximated by a piecewise constant curve for compatibility with the skyline model). In both scenarios the inferred BDS model parameters much more closely resembled their true profiles (Supplemental Figs. S6 and S8), consistent with the fact that fixing the removal rate profile and the present-day sampling proportion collapses the congruence class to a single scenario (mathematical proof in Supplement S.1.5.7). Together, these results confirm our expectations that any tool attempting to reconstruct an epidemic’s dynamics based on phylogenetic data alone and without strong additional constraints, no matter how large the dataset, can only reconstruct the congruence class of those dynamics rather than the true dynamics.

To further confirm that the above issues are likely to occur in practice, we performed multiple simulations of random epidemiological scenarios (i.e., with λ , μ and ψ having randomly constructed but plausible profiles over time), and examined the accuracy of maximum-likelihood fitted BDS models with generic piecewise linear profiles (the grid size was chosen according to AIC [39]). Trees comprised between 500 and 50 000 tips, and we only considered cases where the fitted model adequately described the data, based on the same Kolmogorov-Smirnov tests as above (details in Supplement S.3). In the vast majority of cases, and even for the largest trees, the estimated parameters λ , μ , ψ , R_e , S and δ poorly reflected the true profiles used in the simulations, often exhibiting opposite long-term trends and spurious major features, and sometimes even deviated by orders of magnitude from the truth (overview in Supplemental Figs. S17 and S19, examples in Supplemental Figs. S18 and S20). For example, the coefficient of determination (R^2) between the true R_e and the fitted R_e was almost always below 0 (Figs. S17A and S19A). In contrast, the deterministic branching and sampling densities ($\tilde{\beta}$ and $\tilde{\sigma}$) were generally accurately estimated (typical $R^2 > 0.9$, Figs. S17D,E and S19D,E). Similarly, the mean error relative to the average was about an order of magnitude lower for $\tilde{\beta}$ and $\tilde{\sigma}$ (average $\approx 7\%$) than for R_e (average $\approx 80\%$). These findings show that most fitted models did converge towards the true congruence classes, but not to the actual true scenarios.

Illustration for an HIV epidemic

To further illustrate the practical implications of model congruencies, we considered the dynamics of HIV-1 subtype B in Northern Alberta, Canada, over the course of roughly 20 years, reconstructed from 563 molecular sequences using Bayesian BDS skyline models in BEAST2 [5, 10]. We assumed that R_e , δ , and S varied over time and shifted in 1998 (when triple antiretroviral therapy became available) and in 2010 (to

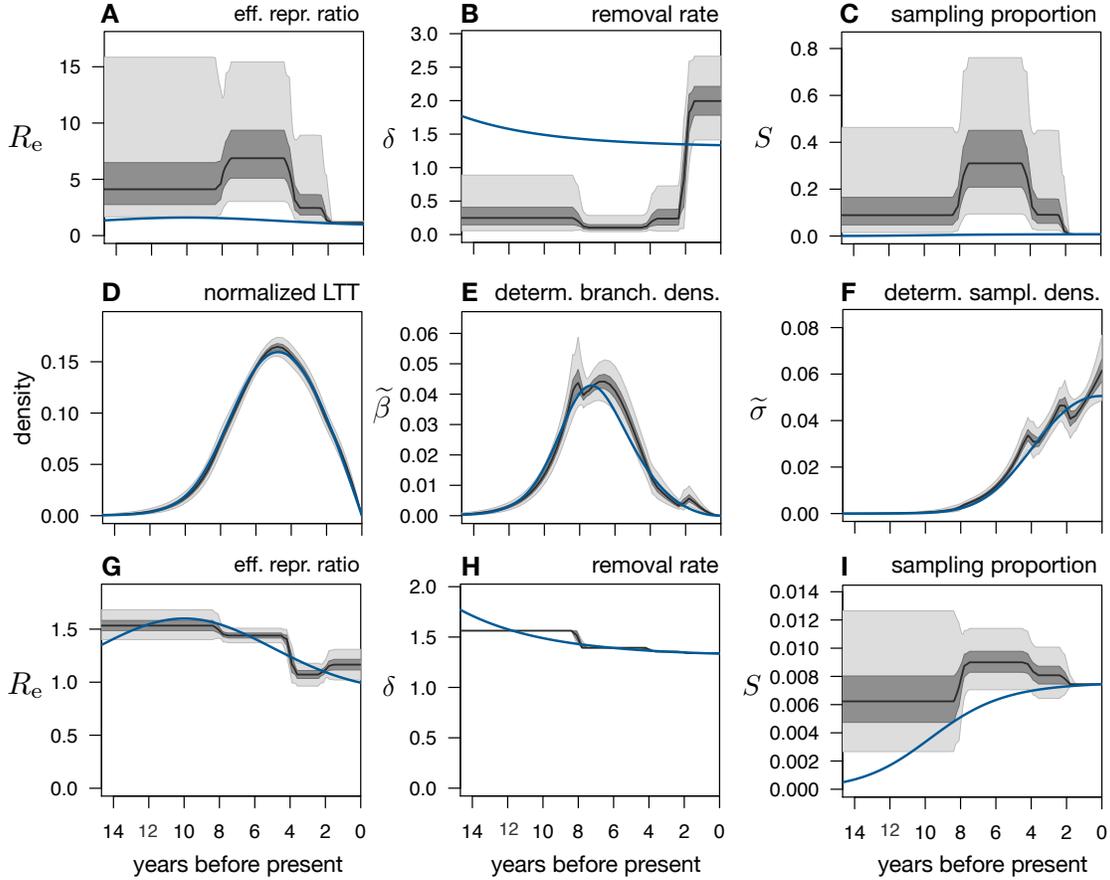


Figure 4: Limits to reconstructing an epidemic’s dynamics in a Bayesian framework. (A–C) Posterior distributions of the effective reproduction ratio ($R_e = \lambda/(\mu + \psi)$), removal rate ($\delta = \mu + \psi$), and sampling proportion ($S = \psi/(\mu + \psi)$), as inferred from 590 sequences simulated under a hypothetical birth-death-sampling scenario (blue curves) using BEAST2. Black curves show posterior median, dark and light shades represent equal-tailed 50%- and 95%-credible intervals of the posterior. All rates are in yr^{-1} . The present-day sampling proportion was fixed to its true value during fitting to account for previously reported identifiability issues in skyline models [10]. Model adequacy was confirmed using predictive posterior simulations with multiple test statistics. Observe the poor agreement between the posterior predictions and the true profiles. For additional epidemiological parameters (λ , μ , ψ) see Supplemental Fig. S5. For the molecular evolution parameters see Supplemental Fig. S10. (D–F) Distributions of the deterministic lineages-through-time curves (normalized to have unit area under the curve), deterministic branching densities ($\tilde{\beta}$), and deterministic sampling densities ($\tilde{\psi}$), corresponding to the same posterior models as in A–C, compared to their true profiles (blue curves). The relatively good agreement between the inferred and true profiles shows that BEAST2 closely reconstructed the true epidemiological history’s congruence class, but not the true epidemiological history itself. (G–I) Posterior distributions of R_e , δ , and S inferred from the same data, while fixing the present-day sampling proportion and the removal rate’s profile to their true values. For additional parameters see Supplemental Figs. S6 and S11.

achieve a roughly balanced partitioning of sampling dates). Rate priors were chosen conservatively to reflect the general uncertainty across HIV outbreaks (see Methods) and an uncorrelated log normal relaxed clock was supported over a strict clock based on nested sampling [40]. We found that the posterior distribution of BDS models (Fig. 5E–H) strongly suggests a decline of R_e over time, a stabilization of the sampling rate in the last two intervals, and a dramatic increase in transmission and recovery rates when comparing the first to the last time interval. The narrow posterior 95% equal-tailed credible intervals for R_e in the second and

third time intervals (Fig. 5H) suggest that R_e is well-constrained.

However, this posterior is misleading because the inferred credible intervals do not properly capture the ambiguities stemming from model congruencies. Indeed, recall that what we are really estimating is the congruence class of the true epidemiological history, and not the true epidemiological history itself. For illustration, consider the sampled model with maximum posterior probability, shown in Fig. 5I–L. This representative scenario is congruent to a myriad of alternative and markedly different epidemiological scenarios, many of which are similarly complex and a priori similarly plausible (examples in Fig. 5I–L). All of these alternative scenarios are equally likely to have generated the data at hand, and this would be true for any other phylogenetic dataset as well. Ruling out some of these congruent scenarios in favor of others requires additional knowledge, such as strong priors on the parameters. Hence, the true uncertainty in the inferred epidemiological parameters is largely determined by the imposed priors, rather than by the computed posterior densities. However, many of the congruent scenarios are not in strong contrast to our priors (examples in Fig. 5A–D), which are typical in the epidemiological literature. In other words, much stronger priors would be needed to collapse the congruence class down to a practical size (e.g., suitably precise for policy decisions), even with massive phylogenetic datasets.

Ways forward

In our recent analysis of macroevolutionary BD models [18] we proposed that researchers could develop methods to draw insight from asymptotically identifiable variables (i.e., those which are identical between congruent scenarios); in the epidemiological case such quantities include $\tilde{\lambda}$, $\tilde{\psi}$, \tilde{r} , $\tilde{\beta}$, and $\tilde{\sigma}$. Indeed, for the macroevolutionary case such identifiable variables do contain useful information about historical diversification dynamics [41]. A similar strategy could potentially be fruitful for epidemiological data, although we do not further explore that possibility here. Instead, in the following we discuss a number of ways, some of which build upon current practices in the field, to robustly reconstruct typical epidemiological variables of interest such as R_e .

First, following on what is sometimes done in practice, one can use additional clinical or surveillance data to constrain specific epidemiological parameters. While it is generally recognized that parameter estimation benefits from the use of available constraints, the precise effects of constraints in phylodynamics remained poorly understood and their importance severely underestimated. Our results precisely clarify the amount of information necessary to make an epidemiological scenario asymptotically identifiable. For example, if one of λ or ψ are known beforehand, then the remaining variables become asymptotically identifiable (details in Supplement S.1.5). This is because in any congruence class there exists at most one scenario with a specific λ or ψ . Indeed, when we fixed the sampling rate ψ to its true profile in our earlier maximum-likelihood fitting tests, the fitted BDS models closely reproduced the true scenarios (Supplemental Figs. S1H–N and S2H–N). Similarly, if μ is somehow independently known and either λ or ψ is known on at least one time point, then the full scenario again becomes asymptotically identifiable. Similar arguments can be made for R_e , δ , or S (Supplements S.1.5.5, S.1.5.6 and S.1.5.8). Such constraints might be obtained in a variety of ways. For example, in some situations it might be assumed that nearly all people are diagnosed and sampled, in which case the sampling proportion might be fixed to 1. Alternatively, one may estimate the true prevalence of a disease through occasional serological surveys of randomly chosen individuals [42, 43] and then divide the “background rate” of disease detection by that estimate to obtain ψ . Further, clinical data may be used to estimate the rate of host death or recovery (i.e., μ). During maximum-likelihood fitting, one can fix the independently known parameters. In a Bayesian framework, one can impose appropriate data-driven priors, i.e., based on independent information, to constrain the known parameters. However, we stress that

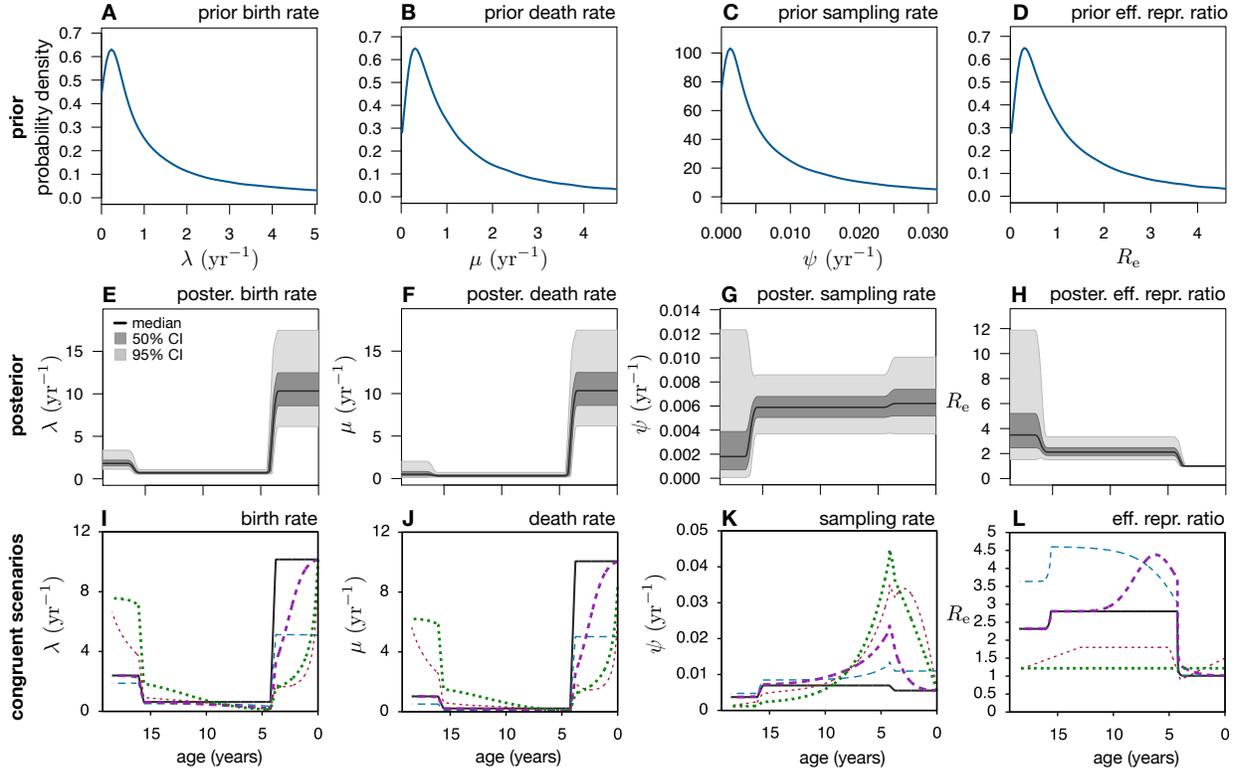


Figure 5: Bayesian reconstruction of HIV spread is compromised by model congruencies. (A–D) Specified priors for BDS (skyline) model parameters of HIV-1 subtype B in Northern Alberta, reflecting our a priori knowledge of the plausible range of these parameters. (E–H) Distribution of BDS parameters over time, based on models sampled from the posterior distribution by BEAST2. At each time point, the black curve shows the median value of a parameter across all posterior-sampled models, while the dark and light shadings show 50% and 95% equal-tailed highest posterior density intervals, respectively. (I–L) Maximum posterior probability BDS “reference” scenario (continuous black curves) compared to multiple alternative “congruent” scenarios (dashed curves). Each congruent scenario would generate timetrees with the same probability distribution as the reference scenario, and is thus statistically indistinguishable from the latter. For the posterior distributions of molecular evolution parameters see Supplemental Fig. S23.

in order to eliminate most congruencies and accurately reconstruct the remaining parameters these priors will need to be much more restrictive than in typical studies [44]. For example, in the second simulation of our Bayesian inference tests described earlier the true removal rate was nearly constant; when we used this information to constrain the removal rate in BEAST2 (i.e., demanding that it is constant across all time-intervals), the parameters λ , μ , ψ , R_e , δ , and S were estimated much more accurately than in the absence of this constraint (Supplemental Fig. S9). Their accuracy improved further when we fixed the removal rate to its true profile (Supplemental Fig. S8). While constraints such as the above are sometimes included in molecular epidemiological studies, many studies still attempt to estimate the full epidemiological dynamics (λ , μ and ψ) solely from phylogenies [10, 45, 46]. In contrast, multiple sources of information are commonly utilized when fitting mechanistic epidemiological models, such as differential equation models, or in non-parametric analyses of surveillance data over time. For example, it is common for surveillance data to be combined with independently determined infectivity profiles [47], which are essentially a generalization of λ , or with estimates of the serial interval [48], to estimate R_e over time. Our mathematical results clarify that using additional information beyond just molecular data, e.g., from clinical trials or serological surveys, is not just a means to increase the number of data points — it is essential for ensuring the identifiability of

temporally variable epidemiological histories. Developing new phylodynamic models that readily integrate appropriate additional data sources would greatly facilitate this practice [49, 50] (but see Supplement S.5 on the use of occurrence data).

Second, while molecular epidemiological studies are typically purely observational — i.e., the sampling and data analysis are done independently — this does not need to be the case. Indeed, a properly designed sampling scheme can help reduce identifiability issues. Concretely, short, high-intensity “concentrated” or “contemporaneous” sampling attempts (CSAs), where many individuals are randomly sampled (in addition to sampling symptomatic individuals), can yield valuable information for reconstructing the dynamics of an epidemic and for partly resolving congruencies (details in Supplements S.1.6 and S.1.7). The main requirements are that, first, these CSAs are much shorter than the current expected birth and death times (i.e., much shorter than $1/\lambda$ and $1/\mu$), second, λ does not differ substantially between the beginning and end of the CSA, and third, the number of lineages sampled during the CSA is much greater than the number of birth or death events occurring in that time period. Such a sampling strategy is not just a hypothetical possibility: during 2020, multiple governments reportedly conducted such CSAs to estimate the seroprevalence of SARS-CoV-2 [51, 52]. If such sampling attempts are performed repeatedly over time and at sufficient temporal resolution, and/or they are combined with other local sequencing data, they can enable an accurate reconstruction of λ , ψ , and consequently μ over time. For example, when we re-simulated the two hypothetical BDS scenarios used earlier for maximum-likelihood estimation while including 3 CSAs, the subsequently re-fitted models reproduced the true BDS scenarios much more accurately (Supplemental Figs. S10–U and S20–U). Note that CSAs differ from the common approach of testing individuals only upon the appearance of symptoms, as in these cases the number of infections detected during any given period tends to be smaller than the number of new infections occurring during that period. The importance of optimal sampling design to improve identifiability has previously been recognized for coalescent models in epidemiology [53, 54]. Notably, Stack *et al.* [53] concluded that sampling sequences at specific time points tends to improve inferential power compared to less focused sampling protocols, resembling our conclusions above for BDS models.

A third potential approach could be to only fit profiles for λ , μ , and ψ with a strong mechanistic justification, i.e. derived from models for infectious disease dynamics [11, 20, 55–57], rather than generic profiles (e.g., skyline models). Notably, birth-death SIR and SIS models [11, 57, 58] are increasingly used and a good start in this direction, although for longer time periods or spatially structured epidemics more complex models will generally be needed. Whether this approach is effective in avoiding the issues stemming from congruencies in practice is unknown and warrants future investigation. If the true epidemiological history was indeed perfectly described by a given mechanistic model, then fitting that model to a timetree will probably yield accurate parameter estimates, provided of course a sufficiently large dataset. However, nature rarely exactly follows our mechanistic models, and this is certainly true for epidemics. In this situation, as when fitting generic profiles (e.g., skyline models), fitting a mechanistic model will generally merely yield a scenario close to the congruence class of the true epidemiological history; in the case of a complex mechanistic model exhibiting a variety of qualitatively different behaviors (depending on parameter choice) this might yield a scenario far from the true history itself, even for large datasets (see conceptual Fig. 2 and earlier discussion).

We stress that coalescent analysis, an alternative popular framework for reconstructing an epidemic’s effective size (N_e) over time based on pathogen sequences [59, 60], cannot resolve the issues discussed here. While N_e is in theory asymptotically identifiable provided sufficiently large trees, provided that all assumptions of coalescent theory are satisfied and provided that the generation time T (or serial interval, which is analogous to $1/\lambda$) are independently known, the issue of identifiability is essentially being replaced by strong and questionable assumptions about the sampling and transmission process [61]. For example, using solely phylogenetic data coalescent theory can a priori only yield information on the product $N_e T$, and additional assumptions or independent information are needed about T in order to actually obtain N_e and R_e [62]. In

the general situation where λ (and hence T) can vary arbitrarily through time and is unknown, one is thus faced with similar identifiability issues as with BDS models.

Conclusions

Our results highlight the limitations of epidemiological inference using phylogenetic data alone. The reported identifiability issues are particularly serious for cases where a reconstruction of historical dynamics is attempted based solely on phylogenetic data and without any additional strong constraints. In such situations, it is generally impossible to reliably reconstruct key epidemiological variables, such as R_e , over time, no matter how large the data. We stress that these issues are separate from the well-recognized errors due to small datasets [55], since any two congruent BDS scenarios remain statistically indistinguishable even for infinitely large phylogenies and hence large phylogenetic datasets alone cannot possibly resolve these issues. Instead, additional data sources beyond just phylogenies, such as from clinical experiments or seroprevalence surveys, are necessary for accurate reconstruction. On a more positive note, we have fully resolved the informational connections between major epidemiological parameters and provide tools for determining their identifiability based on phylogenetic data. In particular, we provide code for exploring the full extent of congruent BDS scenarios and for assessing which epidemiological scenarios can in principle (i.e., with a sufficiently large dataset) be distinguished from one another (Supplement S.4). Our results can help guide proper experimental and sampling designs optimized for epidemiological reconstruction.

Methods

Simulations for maximum-likelihood inferences

To demonstrate the implications of model congruencies using simulated trees, we proceeded as follows. Timetrees were generated under two alternative BDS scenarios using the function `generate_tree_hbds` in the R package `castor` v1.6.7 [63] (with options “`include_extant=FALSE`, `include_extinct=FALSE`, `no_full_extinction=TRUE`”). In the first scenario, μ and ψ were assumed to be constant over time while R_e was exponentially decreasing over time (profiles in Supplemental Fig. S1A–F) and the simulation was halted after 200 days, resulting in a tree with 175,440 tips. In the second scenario, λ and μ were constant over time while ψ increased continuously towards the present (profiles in Supplemental Fig. S2A–F) and the simulation was halted after 200 days, resulting in a tree with 55,934 tips. To each tree, we fitted BDS models with a priori unknown λ , μ and ψ , each defined as a piecewise linear function of time with inflections at fixed time points (chosen such that their density is approximately proportional to the square root of the tree’s LTT). The optimal number of time-grid points (i.e., the number of fitted parameters for each of λ , μ and ψ) was chosen by minimizing the AIC of the fitted model [39]. For any given grid size, fitting was done via maximum-likelihood using the `castor` function `fit_hbds_model_on_grid`, with options “`condition='auto'`, `max_start_attempts=100`, `Ntrials=100`”. Additional epidemiological variables of the fitted models (such as R_e and the LTT) were computed using the `castor` function `simulate_deterministic_hbds`. To demonstrate the effects of fixing ψ to its true value during fitting, we repeated the above fitting process while fixing ψ to its true time profile. To demonstrate the effects of concentrated sampling attempts (CSAs), discussed in the main article, we also simulated trees under BDS scenarios similar to the above but modified to include CSAs at 3 discrete time points. We then used these new trees to fit BDS models with unknown λ , μ and ψ , defined as piecewise linear functions, while also accounting for the added CSAs in the computation of

the likelihood [20]. As before, fitting was performed via maximum-likelihood and by choosing the time-grid size according to the AIC. During fitting, the times and intensities (i.e., sampling probabilities) of the CSAs were fixed to their true values (through options “CSA_ages” and “fixed_CSA_probs”), reflecting general-population randomized seroprevalence surveys in which these properties can be independently determined. The fitted models are shown in Supplemental Figs. S1 and S2. We mention that throughout this article “age” refers to time before present, and “present” refers to the time at which the sampling process was halted.

Simulations for Bayesian inference

To explore the ability of Bayesian MCMC sampling to recover the true epidemiological history with varying levels of constraints, we conducted an internally-blinded BEAST2 birth-death skyline inference using sampled sequences generated from realistic HIV epidemic simulations. Specifically, group member A simulated timetrees and nucleotide sequences under two different BDS scenarios lasting approximately 15 years, resulting in trees with 590 and 540 tips, respectively (Supplemental Fig. S3). The parameters λ , μ , ψ , R_e , δ , and S were all within reasonable ranges, i.e., with values within the priors that would typically be specified for an HIV model, with moderate variation through time that can be well-approximated using a skyline model with 3–4 time intervals (Supplemental Figs. S5A–F and S7A–F). Nucleotide sequences of length 1000 bp were simulated along the timetree under an independent-sites HKY substitution model [64] with transition/transversion ratio 5 [38] and stationary base frequencies A:0.4, C:0.17, G:0.21, T:0.22 [36]. The root sequence was chosen randomly according to the stationary base frequencies. Nucleotides were randomly assigned to one of 4 strict-clock substitution rate categories, whose rates were chosen according to a discretized gamma distribution as described by [65], with shape parameter $\alpha = 0.5$ [36–38] and mean rate $2 \times 10^{-3} \text{ yr}^{-1}$.

The simulated sequence alignments and their sampling dates were provided to a second team (“B”) as input data for reconstructing the epidemiological dynamics over time using serial skyline (i.e., piecewise constant) birth-death-sampling models in BEAST2 v2.6.2 with BEAGLE v4.1 [5, 10, 66]. As mentioned earlier, team B did not have knowledge of the true epidemiological parameters, and was initially only provided with the following information: the nucleotide substitution model used (HKY with independent sites), the fact that there were four rate categories according to a discretized gamma distribution, the fact that four time intervals would be sufficient for reasonably approximating the two epidemic histories, the fact that all parameters were chosen within ranges typical for HIV-1, and the present-day sampling proportion. Team B confirmed that there was adequate temporal signal in the sequence data by evaluating the distribution of pairwise patristic distances and divergence over time on a preliminary approximate maximum likelihood tree made in FastTree2 [67] and rooted using residual mean square fitting in Tempest v1.5.3 [68]. Based on the distribution of sampling and branching dates (Supplemental Figs. S3C–F and S4), and to ensure similar amounts of data in each interval, the rate shifts in these models were specified to occur at 2, 4, and 8 years before the present. Skyline models were parameterized in terms of R_e , removal rate δ , and sampling proportion S , each of which could vary independently in each time interval. In all cases, the present-day sampling proportion was fixed to its true value, to account for known correlations between skyline model parameters [10]. Runs were set up on two independent MCMC chains for 100–200 million states, sampled every 10 000 states (overview of priors in Supplemental Table S1). For each run, log files from both chains were combined using LogCombiner [5] after confirming convergence in Tracer and removing 10% burn-in. We refer to these two runs as U1 and U2 (“unconstrained” 1 and 2). For each model drawn from the posterior distribution, we calculated the deterministic LTT, the deterministic branching density, and deterministic sampling density using the castor function `simulate_deterministic_hbds`. Equal-tailed credible intervals of various model parameters were calculated for the posterior distribution of scenarios using the `quantile` function in R. To investigate

how the parameter estimates would improve if one were to provide sufficient constraints to collapse the congruence class, we repeated the BEAST2 runs while fixing the removal rate over the entire time period to its true profile (approximated by a piecewise constant curve for compatibility with the skyline model) and while fixing the present-day sampling proportion to its true value (as before); we refer to these new runs as F1 and F2, respectively (“fixed” 1 and 2). Lastly, for the 2nd scenario where the removal rate was nearly constant over time, we investigated how this information might improve parameter estimates, by constraining the removal rate to be constant across all time intervals (with unknown value); we refer to this run as C2 (“constrained” 2). Posterior distributions of epidemiological parameters from runs U1, F1, U2, F2 and C2 are shown in Supplemental Figs. S5, S6, S7, S8 and S9, respectively. The corresponding posterior distributions of molecular evolution parameters are shown in Supplemental Figs. S10, S11, S12, S14 and S13. MCMC traces of runs U1 and U2 are shown in Supplemental Figs. S15 and S16, respectively. The above analyses were also repeated for sequences simulated under a strict molecular clock model with a single discretized rate category from the gamma distribution, yielding similar results.

Model adequacy tests

To verify that each maximum-likelihood-fitted model was adequate for explaining the timetree, i.e., that parameter estimates were not due to bad model fits, we used parametric bootstrapping to compare various properties of the tree to those expected under the fitted model [35, 69], as follows. For any given tree and fitted model, we simulated 1000 random trees using the model from the root the present-day using the function `generate_tree_hbd_reverse` in the R package `castor` [70]. We then compared the distribution of tip ages generated by the fitted model to the original tree using a Kolmogorov-Smirnov test. Specifically, for every simulated tree we calculated the empirical cumulative distribution function (CDF) of the tip ages (denoted F , and evaluated at the original tree’s tip ages via linear interpolation), and then calculated the average of those CDFs, hence obtaining an estimate for the CDF of tip ages generated by the model (denoted \bar{F}). The Kolmogorov-Smirnov (KS) distance between a tree’s CDF F and \bar{F} , denoted $D(F, \bar{F})$, is the maximum distance between F and \bar{F} at any age. The statistical significance (P) of the original tree’s KS distance $D(F_o, \bar{F})$ was calculated as the fraction of simulated trees for which $D(F, \bar{F})$ was larger than $D(F_o, \bar{F})$. Hence, a small P means that the original tree’s distribution of node ages differs substantially from that expected under the fitted model. A similar approach was followed for comparing the model’s and tree’s distribution of node ages or edge lengths. To perform analogous model adequacy tests for our Bayesian analysis, we compared the true timetree (generated during the simulation of the hypothetical epidemic) to the scenarios drawn from the posterior distribution. The methodology was nearly identical to that described above for the maximum-likelihood fits, with the only substantial difference being that each random tree was generated by a scenario randomly chosen from the posterior distribution. The above statistical tests are conveniently implemented in the `castor` function `model_adequacy_hbds` [63].

Empirical HIV analysis

For the empirical HIV analysis we used publicly available HIV-1 sequences from Northern Alberta, Canada, collected between 2007 and 2013 and previously described by Vrancken *et al.* [71]. Of the 1055 partial pol sequences (consisting of full *protease* and the first 240 or 400 codons of *reverse transcriptase*) available on Genbank, the analysis was restricted to 809 subtype B sequences, as determined by Vrancken *et al.* based on a maximum likelihood phylogeny of Alberta sequences alongside Los Alamos HIV database sequences (<http://hiv.lanl.gov>), confirmed using Comet [72]. Four sequences were removed because they were duplicates and one sequence was removed because it had >0.05 ambiguous nucleotides, leaving

804 sequences. Sequences were aligned using mafft v7.402 [73] and known drug resistance mutation sites relative to HXB2 reference were removed [74]. Similarly to Vrancken *et al.*, we identified a weak temporal signal within an approximate maximum likelihood tree of all subtype B sequences inferred using FastTree v2.1.11 [67] and rooted by residual mean squared (rms) regression fit in Tempest [68]. The subtype B phylogeny consisted of two deeply split clades herein denoted B.1 (n=624) and B.2 (n=185). After splitting B.1 and B.2 at their MRCA into two trees, re-rooting using rms, and retaining only sequences with residuals < 0.02 substitutions/site (B.1, n=563; B.2, n=164), this yielded an increase in the correlation coefficient of the molecular clock rate fit from 0.21 to 0.33 and 0.34, respectively. Here we focused our analyses on the larger B.1 clade (n=563).

The B.1 alignment was used to jointly infer a time-calibrated phylogeny and fit a birth-death skyline serial model in BEAST2 [5, 10]. Model selection consisted of comparing strict and relaxed uncorrelated log normal (UCLN) clocks with free and fixed mean clock rates, as well as multiple rate intervals for R_e , S , and $/delta$ using nested sampling [40]. For all the models compared, site model averaging was conducted using bModelTest [75]; additional priors are summarized in S2. UCLN models with free mean clock rates were more well-supported than their fixed mean clock rate or strict clock model equivalents; and models with rate shifts occurring in 1998 and 2010 had higher likelihoods than their equivalents with equally spaced intervals from origin to the most recent sample. For every model, two parallel MCMC chains of 500 million steps were combined after confirming each run converged, as assessed by effective sample sizes greater than 200 following 10% burn-in. The probability densities of BDS model parameters, based on samples drawn by BEAST from the posterior distribution, are shown in Supplemental Fig. S22. The posterior distributions of molecular evolution parameters are shown in Supplemental Fig. S23. MCMC trace plots are shown in Supplemental Fig. S24.

Data availability

No raw data were generated for this manuscript. The sequence data from the HIV epidemic in Alberta was described previously Vrancken *et al.* [71] and is publicly available (see Methods for details). The simulated trees and sequences (used for maximum-likelihood fitting and Bayesian analyses) are given in Supplemental File 1.

Code availability

All analyses were performed using freely available software, indicated throughout the manuscript where appropriate. Novel code for fitting BDS models, examining BDS model congruencies and checking BDS model adequacy (summarized in Supplement S.4) is provided as part of the R package *castor*, available on The Comprehensive R Archive Network. The BEAST2 configuration XML file for the HIV analysis is provided as Supplemental File 2. The BEAST2 configuration XML files for the analyses of simulated sequences are provided as Supplemental File 3.

Acknowledgments

We thank Luke Harmon and Josef Uyeda for thoughtful discussions of this work. This project was supported by a GCRC grant from the University of British Columbia. SL and MWP were supported by a US National Science Foundation RAPID grant (#2028986); MWP was also supported by a NSERC Discovery Grant. AML was supported by a CIHR Canada Graduate Scholarships Doctoral award. AM was supported by the EEB department Postdoctoral Fellowship from the University of Toronto. JBJ was supported by a Genome Canada Bioinformatics and Computational Biology grant (287PHY), a Canadian Institutes of Health Research Corona Virus Rapid Response Grant (440371) and the BC Centre for Excellence in HIV/AIDS.

Author contributions

SL performed the mathematical derivations, numerical simulations and maximum-likelihood fitting, and developed the code for examining BDS model congruencies. AML performed all BEAST analyses and all analyses of the HIV epidemic. All authors contributed to the project's conception and to the writing of the manuscript.

References

- [1] Drummond, A. J., Pybus, O. G., Rambaut, A., Forsberg, R. & Rodrigo, A. G. Measurably evolving populations. *Trends in Ecology & Evolution* **18**, 481–488 (2003).
- [2] Pybus, O. G. & Rambaut, A. Evolutionary analysis of the dynamics of viral infectious disease. *Nature Reviews Genetics* **10**, 540–550 (2009).
- [3] Volz, E. M., Koelle, K. & Bedford, T. Viral phylodynamics. *PLoS Comput Biol* **9**, e1002947 (2013).
- [4] Grenfell, B. T. *et al.* Unifying the epidemiological and evolutionary dynamics of pathogens. *science* **303**, 327–332 (2004).
- [5] Bouckaert, R. *et al.* BEAST 2.5: An advanced software platform for Bayesian evolutionary analysis. *PLoS computational biology* **15**, e1006650 (2019).
- [6] Holmes, E. C., Dudas, G., Rambaut, A. & Andersen, K. G. The evolution of Ebola virus: Insights from the 2013–2016 epidemic. *Nature* **538**, 193–200 (2016).
- [7] Faria, N. R. *et al.* Zika virus in the americas: early epidemiological and genetic findings. *Science* **352**, 345–349 (2016).
- [8] Worobey, M. *et al.* The emergence of SARS-CoV-2 in Europe and North America. *Science* **370**, 564–570 (2020).
- [9] Stadler, T. *et al.* Estimating the basic reproductive number from viral sequence data. *Molecular Biology and Evolution* **29**, 347–357 (2012).
- [10] Stadler, T., Kühnert, D., Bonhoeffer, S. & Drummond, A. J. Birth–death skyline plot reveals temporal changes of epidemic spread in HIV and hepatitis C virus (HCV). *Proceedings of the National Academy of Sciences* **110**, 228–233 (2013).
- [11] Kühnert, D., Stadler, T., Vaughan, T. G. & Drummond, A. J. 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 (2014).
- [12] Morlon, H. Phylogenetic approaches for studying diversification. *Ecology Letters* **17**, 508–525 (2014).
- [13] Kendall, D. G. *et al.* On the generalized “birth-and-death” process. *The Annals of Mathematical Statistics* **19**, 1–15 (1948).
- [14] Nee, S., Holmes, E. C., May, R. M. & Harvey, P. H. Extinction rates can be estimated from molecular phylogenies. *Philosophical Transactions: Biological Sciences* **344**, 77–82 (1994).
- [15] Nee, S., May, R. M. & Harvey, P. H. The reconstructed evolutionary process. *Philosophical Transactions of the Royal Society of London B: Biological Sciences* **344**, 305–311 (1994).
- [16] Stadler, T. On incomplete sampling under birth–death models and connections to the sampling-based coalescent. *Journal of Theoretical Biology* **261**, 58–66 (2009).
- [17] Stadler, T. & Steel, M. Swapping birth and death: Symmetries and transformations in phylodynamic models. *Systematic Biology* **68**, 852–858 (2019).

- [18] Louca, S. & Pennell, M. W. Extant timetrees are consistent with a myriad of diversification histories. *Nature* **580**, 502–505 (2020).
- [19] Gavryushkina, A., Welch, D., Stadler, T. & Drummond, A. J. Bayesian inference of sampled ancestor trees for epidemiology and fossil calibration. *PLoS Comput Biol* **10**, e1003919 (2014).
- [20] MacPherson, A., Louca, S., McLaughlin, A., Joy, J. B. & Pennell, M. W. A general birth-death-sampling model for epidemiology and macroevolution. *bioRxiv* 2020.10.10.334383 (2020).
- [21] Myers, S., Fefferman, C. & Patterson, N. Can one learn history from the allelic spectrum? *Theoretical Population Biology* **73**, 342–348 (2008).
- [22] Bhaskar, A. & Song, Y. S. Descartes’ rule of signs and the identifiability of population demographic models from genomic variation data. *The Annals of Statistics* **42**, 2469–2493 (2014).
- [23] Louca, S. & Pennell, M. W. Why extinction estimates from extant phylogenies are so often zero. *bioRxiv* (2021).
- [24] Grassly, N. C. & Fraser, C. Mathematical models of infectious disease transmission. *Nature Reviews Microbiology* **6**, 477–487 (2008).
- [25] Cowling, B. J. *et al.* The effective reproduction number of pandemic Influenza: Prospective estimation. *Epidemiology* **21**, 842–846 (2010).
- [26] Nishiura, H. & Chowell, G. The effective reproduction number as a prelude to statistical estimation of time-dependent epidemic trends. In Chowell, G., Hyman, J. M., Bettencourt, L. M. A. & Castillo-Chavez, C. (eds.) *Mathematical and Statistical Estimation Approaches in Epidemiology*, 103–121 (Springer Netherlands, Dordrecht, 2009).
- [27] Kühnert, D., Stadler, T., Vaughan, T. G. & Drummond, A. J. Phylodynamics with migration: A computational framework to quantify population structure from genomic data. *Molecular Biology and Evolution* **33**, 2102–2116 (2016).
- [28] Kühnert, D. *et al.* Tuberculosis outbreak investigation using phylodynamic analysis. *Epidemics* **25**, 47–53 (2018).
- [29] Faria, N. R. *et al.* Genomic and epidemiological characterisation of a dengue virus outbreak among blood donors in Brazil. *Scientific Reports* **7**, 15216 (2017).
- [30] Mir, D. *et al.* Inferring population dynamics of HIV-1 subtype C epidemics in Eastern Africa and Southern Brazil applying different Bayesian phylodynamics approaches. *Scientific Reports* **8**, 8778 (2018).
- [31] Vasylyeva, T. I. *et al.* Tracing the impact of public health interventions on HIV-1 transmission in Portugal using molecular epidemiology. *The Journal of Infectious Diseases* **220**, 233–243 (2019).
- [32] Müller, N. F. *et al.* Characterising the epidemic spread of influenza A/H3N2 within a city through phylogenetics. *PLOS Pathogens* **16**, e1008984 (2020).
- [33] Nie, Q. *et al.* Phylogenetic and phylodynamic analyses of SARS-CoV-2. *Virus Research* **287**, 198098 (2020).
- [34] Resende, P. C. *et al.* Evolutionary dynamics and dissemination pattern of the SARS-CoV-2 lineage B.1.1.33 during the early pandemic phase in Brazil. *Frontiers in Microbiology* **11**, 3565 (2021).

- [35] Brown, J. M. & Thomson, R. C. Evaluating model performance in evolutionary biology. *Annual Review of Ecology, Evolution, and Systematics* **49**, 95–114 (2018).
- [36] Posada, D. & Crandall, K. A. Selecting models of nucleotide substitution: an application to human immunodeficiency virus 1 (HIV-1). *Molecular biology and evolution* **18**, 897–906 (2001).
- [37] Leitner, T., Kumar, S. & Albert, J. Tempo and mode of nucleotide substitutions in gag and env gene fragments in human immunodeficiency virus type 1 populations with a known transmission history. *Journal of Virology* **71**, 4761–4770 (1997).
- [38] Duchêne, S., Ho, S. Y. & Holmes, E. C. Declining transition/transversion ratios through time reveal limitations to the accuracy of nucleotide substitution models. *BMC Evolutionary Biology* **15**, 36 (2015).
- [39] Akaike, H. Likelihood of a model and information criteria. *Journal of Econometrics* **16**, 3–14 (1981).
- [40] Russel, P. M., Brewer, B. J., Klaere, S. & Bouckaert, R. R. Model selection and parameter inference in phylogenetics using nested sampling. *Systematic Biology* **68**, 219–233 (2019).
- [41] Louca, S. *et al.* Bacterial diversification through geological time. *Nature Ecology & Evolution* **2**, 1458–1467 (2018).
- [42] Farrington, C. P. & Whitaker, H. J. Estimation of effective reproduction numbers for infectious diseases using serological survey data. *Biostatistics* **4**, 621–632 (2003).
- [43] Lai, C.-C., Wang, J.-H. & Hsueh, P.-R. Population-based seroprevalence surveys of anti-SARS-CoV-2 antibody: An up-to-date review. *International Journal of Infectious Diseases* (2020).
- [44] Nadeau, S. A., Vaughan, T. G., Sciré, J., Huisman, J. S. & Stadler, T. The origin and early spread of SARS-CoV-2 in Europe. *medRxiv* (2020).
- [45] Paraskevis, D. *et al.* Enhanced HIV-1 surveillance using molecular epidemiology to study and monitor HIV-1 outbreaks among intravenous drug users (IDUs) in Athens and Bucharest. *Infection, Genetics and Evolution* **35**, 109–121 (2015).
- [46] Lai, A., Bergna, A., Acciarri, C., Galli, M. & Zehender, G. Early phylogenetic estimate of the effective reproduction number of SARS-CoV-2. *Journal of Medical Virology* **92**, 675–679 (2020).
- [47] Cori, A., Ferguson, N. M., Fraser, C. & Cauchemez, S. A new framework and software to estimate time-varying reproduction numbers during epidemics. *American Journal of Epidemiology* **178**, 1505–1512 (2013).
- [48] Najafi, F. *et al.* Serial interval and time-varying reproduction number estimation for COVID-19 in western Iran. *New Microbes and New Infections* **36**, 100715 (2020).
- [49] Gupta, A., Manceau, M., Vaughan, T., Khammash, M. & Stadler, T. The probability distribution of the reconstructed phylogenetic tree with occurrence data. *Journal of Theoretical Biology* **488**, 110115 (2020).
- [50] Manceau, M., Gupta, A., Vaughan, T. & Stadler, T. The probability distribution of the ancestral population size conditioned on the reconstructed phylogenetic tree with occurrence data. *Journal of Theoretical Biology* **509**, 110400 (2021).
- [51] Pollán, M. *et al.* Prevalence of SARS-CoV-2 in Spain (ENE-COVID): a nationwide, population-based seroepidemiological study. *The Lancet* **396**, 535–544 (2020).

- [52] Menachemi, N. *et al.* Population point prevalence of SARS-CoV-2 infection based on a statewide random sample - Indiana, April 25-29, 2020. *MMWR. Morbidity and mortality weekly report* **69**, 960–964 (2020).
- [53] Stack, J. C., Welch, J. D., Ferrari, M. J., Shapiro, B. U. & Grenfell, B. T. Protocols for sampling viral sequences to study epidemic dynamics. *Journal of the Royal Society Interface* **7**, 1119–1127 (2010).
- [54] Parag, K. V. & Pybus, O. G. Robust design for coalescent model inference. *Systematic Biology* **68**, 730–743 (2019).
- [55] Rasmussen, D. A., Volz, E. M. & Koelle, K. Phylodynamic inference for structured epidemiological models. *PLOS Computational Biology* **10**, e1003570 (2014).
- [56] Frost, S. D. W. *et al.* Eight challenges in phylodynamic inference. *Epidemics* **10**, 88–92 (2015).
- [57] Vaughan, T. G. *et al.* Estimating epidemic incidence and prevalence from genomic data. *Molecular Biology and Evolution* **36**, 1804–1816 (2019).
- [58] Leventhal, G. E., Günthard, H. F., Bonhoeffer, S. & Stadler, T. Using an epidemiological model for phylogenetic inference reveals density dependence in hiv transmission. *Molecular Biology and Evolution* **31**, 6–17 (2014).
- [59] Hill, V. & Baele, G. Bayesian estimation of past population dynamics in BEAST 1.10 using the skygrid coalescent model. *Molecular Biology and Evolution* **36**, 2620–2628 (2019).
- [60] Lambert, A. & Stadler, T. Birth–death models and coalescent point processes: The shape and probability of reconstructed phylogenies. *Theoretical Population Biology* **90**, 113–128 (2013).
- [61] Boskova, V., Bonhoeffer, S. & Stadler, T. Inference of epidemiological dynamics based on simulated phylogenies using birth-death and coalescent models. *PLOS Computational Biology* **10**, e1003913 (2014).
- [62] Hall, M. D., Woolhouse, M. E. J. & Rambaut, A. The effects of sampling strategy on the quality of reconstruction of viral population dynamics using Bayesian skyline family coalescent methods: A simulation study. *Virus Evolution* **2** (2016).
- [63] Louca, S. & Doebeli, M. Efficient comparative phylogenetics on large trees. *Bioinformatics* **34**, 1053–1055 (2018).
- [64] Hasegawa, M., Kishino, H. & Yano, T.-a. Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. *Journal of Molecular Evolution* **22**, 160–174 (1985).
- [65] Yang, Z. Maximum likelihood phylogenetic estimation from DNA sequences with variable rates over sites: approximate methods. *J Mol Evol* **39** (1994).
- [66] Ayres, D. L. *et al.* Beagle: An application programming interface and high-performance computing library for statistical phylogenetics. *Systematic Biology* **61**, 170–173 (2012).
- [67] Price, M. N., Dehal, P. S. & Arkin, A. P. FastTree 2: Approximately maximum-likelihood trees for large alignments. *PLoS ONE* **5**, e9490 (2010).
- [68] Rambaut, A., Lam, T. T., Max Carvalho, L. & Pybus, O. G. Exploring the temporal structure of heterochronous sequences using TempEst (formerly Path-O-Gen). *Virus Evolution* **2** (2016).

- [69] Schwery, O. M. *The formation of diversity - The role of environment and biogeography in dung beetle species richness, and the adequacy of current diversification models*. Ph.D. thesis, University of Tennessee, Knoxville (2019).
- [70] Louca, S. Simulating trees with millions of species. *Bioinformatics* **36**, 2907–2908 (2020).
- [71] Vrancken, B. *et al.* The multi-faceted dynamics of HIV-1 transmission in Northern Alberta: A combined analysis of virus genetic and public health data. *Infection, Genetics and Evolution* **52**, 100–105 (2017).
- [72] Struck, D., Lawyer, G., Ternes, A.-M., Schmit, J.-C. & Bercoff, D. P. COMET: adaptive context-based modeling for ultrafast HIV-1 subtype identification. *Nucleic Acids Research* **42**, e144–e144 (2014).
- [73] Katoh, K., Kuma, K.-i., Toh, H. & Miyata, T. MAFFT version 5: improvement in accuracy of multiple sequence alignment. *Nucleic acids research* **33**, 511–518 (2005).
- [74] Shafer, R. W. Rationale and uses of a public HIV drug-resistance database. *The Journal of Infectious Diseases* **194**, S51–S58 (2006).
- [75] Bouckaert, R. R. & Drummond, A. J. bModelTest: Bayesian phylogenetic site model averaging and model comparison. *BMC evolutionary biology* **17**, 42 (2017).

Fundamental identifiability limits in molecular epidemiology - Supplemental Information -

Stilianos Louca^{1,2*}, Angela McLaughlin^{3,4}, Ailene MacPherson^{5,6,7}, Jeffrey B. Joy^{3,4,8} & Matthew W. Pennell^{5,6*}

¹*Department of Biology, University of Oregon, USA*

²*Institute of Ecology and Evolution, University of Oregon, USA*

³*British Columbia Centre for Excellence in HIV/AIDS, Vancouver, Canada*

⁴*Bioinformatics, University of British Columbia, Vancouver, Canada*

⁵*Biodiversity Research Centre, University of British Columbia, Vancouver, Canada*

⁶*Department of Zoology, University of British Columbia, Vancouver, Canada*

⁷*Department of Ecology and Evolutionary Biology, University of Toronto, Toronto, Canada*

⁸*Department of Medicine, University of British Columbia, Vancouver, Canada*

*Correspondence should be addressed to SL and MWP

Contents

S.1 Mathematical derivations	3
S.1.1 General model	3
S.1.2 Pulled parameters	4
S.1.3 Congruent models	6
S.1.4 Congruent models have identical distributions of tree sizes	8
S.1.5 Constructing congruent scenarios	9
S.1.5.1 Congruent scenarios by specifying ψ	9
S.1.5.2 Congruent scenarios by specifying μ and $\lambda(\tau_o)$	10
S.1.5.3 Congruent scenarios by specifying μ and $\psi(\tau_o)$	10
S.1.5.4 Congruent scenarios by specifying λ	10
S.1.5.5 Congruent scenarios by specifying R_e and $\lambda(\tau_o)$	11
S.1.5.6 Congruent scenarios by specifying $\mu + \psi$ and $\lambda(\tau_o)$	11
S.1.5.7 Congruent scenarios by specifying $\mu + \psi$ and $S(\tau_o)$	12
S.1.5.8 Congruent scenarios by specifying $\psi/(\mu + \psi)$ and $\lambda(\tau_o)$	13
S.1.6 The information content of concentrated sampling attempts	13
S.1.7 Infinitesimally short concentrated sampling attempts	15
S.2 Why common model selection methods cannot resolve congruencies	16
S.2.1 General considerations (big picture)	16
S.2.2 Regularization	17
S.2.3 AIC and BIC	17
S.3 Statistical evaluation of multiple simulated scenarios	18
S.4 Overview of computer code	19
S.5 On the use of occurrence data with phylogenies	20

S.1 Mathematical derivations

S.1.1 General model

The general cladogenic model considered here, and recently formally described by MacPherson *et al.* [1], is a stochastic birth-death-sampling model, where lineages split (“speciate”) stochastically at some birth rate λ , disappear (“go extinct”) at some death rate μ , and are sampled (detected) at rate ψ . In an epidemiological context, λ corresponds to the transmission rate, μ to the rate of recovery or death of infected individuals, and ψ to the rate at which infected individuals are detected and their pathogen population sequenced. Note that, unless stated otherwise, throughout this document the term “birth” refers to the emergence of a new pathogen lineage (via transmission) and “death” refers to the disappearance of a pathogen lineage without sampling (due to host recovery or death), and should not be confused with host birth or death. The parameters λ , μ , and ψ can each depend arbitrarily on time. We consider the distribution of time-calibrated reconstructed phylogenies (“timetrees”) obtained through this process, i.e., connecting all sampled variants, after some time τ_{or} . Note that this cladogenic process is a generalization of the birth-death-skyline model by Stadler *et al.* [2] to the case where λ , μ , and ψ depend arbitrarily on time rather than in a piecewise constant manner, in the absence of concentrated sampling attempts.

In the following, we count time backward from tips to root, i.e., all variables are expressed as functions of “age” (time before present), with age zero corresponding to the time at which the sampling process is halted. It was recently shown [1] that the likelihood (density) of a given bifurcating timetree generated by a specific BDS scenario starting at age τ_{or} , conditioned upon the survival and sampling of at least one lineage, is given by:

$$L_{\text{or}} = \frac{\Phi(\tau_{\text{or}})}{1 - E(\tau_{\text{or}})} \prod_{i=1}^{n-1} \lambda(x_i) \Phi(x_i) \prod_{j=1}^n \frac{\psi(y_j)}{\Phi(y_j)}, \quad (1)$$

where n is the number of tips in the tree, i iterates over all branching events (i.e., internal nodes), j iterates over all sampling events (i.e., tips), x_i is the age of the i -th branching event, y_j is the age of the j -th sampling event, $E(\tau)$ is the probability that a lineage alive at age τ will not be sampled by the present-day (i.e., is not included in the tree), and Φ is the “flow” of the model [3], defined as:

$$\Phi(\tau) = \exp \left[\int_0^\tau [2\lambda E - (\lambda + \mu + \psi)] ds \right]. \quad (2)$$

The probability E satisfies the differential equation:

$$\frac{dE}{d\tau} = -(\lambda + \mu + \psi)E + \lambda E^2 + \mu, \quad (3)$$

with initial condition $E(0) = 1$. Note that one may alternatively consider the likelihood of the tree conditioned on the age of the tree’s root τ_r , i.e., a splitting occurring at age τ_r and the eventual sampling of both child lineages:

$$L_r = \frac{\Phi(\tau_r)}{\lambda(\tau_r) \cdot (1 - E(\tau_r))^2} \prod_{i=1}^{n-1} \lambda(x_i) \Phi(x_i) \prod_{j=1}^n \frac{\psi(y_j)}{\Phi(y_j)}. \quad (4)$$

In the limit of infinitely large trees, i.e., when stochastic effects on the LTT curve become negligible, the deterministic LTT curve (dLTT) predicted by the model, denoted M , is equal to $M = N \cdot (1 - E)$, where

$N(\tau)$ is the total number of lineages alive at age τ . Using the fact that $dN/dt = (\lambda - \mu - \psi) \cdot N$ as well as Eq. (3), it is straightforward to derive the following differential equation for the model's dLTT:

$$\frac{dM}{dt} = -M \cdot \left[(1 - E) \cdot \lambda - \frac{\psi}{1 - E} \right]. \quad (5)$$

Note that to calculate the dLTT one also needs to specify an initial condition, i.e., conditioned on the number of lineages at some specific age $\tau > 0$ ($M(0)$ will always be zero).

S.1.2 Pulled parameters

The following composite parameters will be particularly useful in our derivations. We define the ‘‘pulled birth rate’’ $\tilde{\lambda}$ as follows:

$$\tilde{\lambda} = (1 - E) \cdot \lambda. \quad (6)$$

Note that $\tilde{\lambda}$ is equal to λ if and only if $E = 0$, i.e., if all lineages ever alive are represented in the tree. In the presence of deaths or incomplete sampling, however, $\tilde{\lambda}$ is ‘‘pulled’’ downwards compared to λ . Observe that $\tilde{\lambda}$ is the expected rate at which internal nodes occur in the tree over time, normalized by the current number of lineages in the tree (M) and in the limit of infinitely large trees (see Eq. 5). For sufficiently large trees $\tilde{\lambda}$ becomes identifiable, since it can just be ‘‘read off’’ the tree’s LTT. We also define the ‘‘pulled sampling rate’’ $\tilde{\psi}$ as follows:

$$\tilde{\psi} = \frac{\psi}{1 - E}. \quad (7)$$

Note that $\tilde{\psi}$ is equal to ψ if and only if $E = 0$; otherwise, $\tilde{\psi}$ is pulled upward compared to ψ . Observe that $\tilde{\psi}$ is the expected rate at which tips occur in the tree over time, normalized by the current number of lineages in the tree (M) and in the limit of infinitely large trees (see Eq. 5). For sufficiently large trees $\tilde{\psi}$ becomes identifiable, since it can just be ‘‘read off’’ the tree’s LTT. We also define the following composite parameter:

$$\tilde{\zeta} = \lambda\psi. \quad (8)$$

Note that for $\tau > 0$ we have $\tilde{\zeta}(\tau) = \tilde{\lambda}(\tau)\tilde{\psi}(\tau)$. In contrast to $\tilde{\psi}$, which becomes infinite at present day (i.e., at age 0), the parameter $\tilde{\zeta}$ remains finite. Since $\tilde{\lambda}$ and $\tilde{\psi}$ are asymptotically identifiable (i.e., in the limit of infinitely large trees), the same also holds for $\tilde{\zeta}$. We also define the ‘‘pulled diversification rate’’ \tilde{r} as follows:

$$\tilde{r} = \lambda - \mu - \psi + \frac{1}{\lambda} \frac{d\lambda}{d\tau}. \quad (9)$$

Note that \tilde{r} is equal to the net diversification rate $r = \lambda - \mu - \psi$ only if λ is constant over time. If λ varies over time, \tilde{r} is pulled upwards or downwards compared to r , depending on how λ varies with time. Taking the derivative on both sides of Eq. (6) and using Eqs. (3) and (8) it is straightforward to show that $\tilde{\lambda}$ satisfies the differential equation:

$$\frac{d\tilde{\lambda}}{d\tau} = \tilde{\lambda} \cdot (\tilde{r} - \tilde{\lambda} + \tilde{\psi}). \quad (10)$$

Note that we can also write Eq. (10) as:

$$\frac{d\tilde{\lambda}}{d\tau} = \tilde{\lambda}\tilde{r} - \tilde{\lambda}^2 + \tilde{\zeta}. \quad (11)$$

Hence, $\tilde{\lambda}$ is fully determined by \tilde{r} and $\tilde{\zeta}$ (note that $\tilde{\lambda}$ always satisfies the initial condition $\tilde{\lambda}(0) = 0$). Solving Eq. (10) for \tilde{r} yields:

$$\tilde{r} = \frac{1}{\tilde{\lambda}} \frac{d\tilde{\lambda}}{d\tau} + \tilde{\lambda} - \tilde{\psi}. \quad (12)$$

Since $\tilde{\lambda}$ and $\tilde{\psi}$ are asymptotically identifiable, the same also holds for \tilde{r} . Solving Eq. (10) for $\tilde{\psi}$ yields:

$$\tilde{\psi} = \frac{1}{\tilde{\lambda}} \frac{d\tilde{\lambda}}{d\tau} + \tilde{\lambda} - \tilde{r}. \quad (13)$$

Thus, $\tilde{\psi}$ is fully determined by $\tilde{\lambda}$ and \tilde{r} . Multiplying Eq. (13) by $\tilde{\lambda}$ yields:

$$\tilde{\zeta} = \tilde{\psi}\tilde{\lambda} = \frac{d\tilde{\lambda}}{d\tau} + \tilde{\lambda}^2 - \tilde{r}\tilde{\lambda}. \quad (14)$$

Hence, $\tilde{\zeta}$ is fully determined by $\tilde{\lambda}$ and \tilde{r} . It is also worth introducing three additional parameters, the “normalized dLTT”:

$$\tilde{M} := \frac{M}{\int_0^\infty M(s) ds}, \quad (15)$$

the “deterministic branching density”:

$$\tilde{\beta} := \tilde{\lambda}\tilde{M} = \frac{\tilde{\lambda}M}{\int_0^\infty M(s) ds}, \quad (16)$$

and the “deterministic sampling density”:

$$\tilde{\sigma} := \tilde{\psi}\tilde{M} = \frac{\tilde{\psi}M}{\int_0^\infty M(s) ds}. \quad (17)$$

Note that neither \tilde{M} , $\tilde{\beta}$ nor $\tilde{\sigma}$ depend on the specific dLTT, as long as the latter solves the differential equation (5), in other words \tilde{M} , $\tilde{\beta}$ and $\tilde{\sigma}$ are independent of the specific initial condition chosen for M . Hence, \tilde{M} , $\tilde{\beta}$ and $\tilde{\sigma}$ are fully determined by $\tilde{\lambda}$ and $\tilde{\psi}$, and are thus asymptotically identifiable. Reciprocally, \tilde{M} , $\tilde{\lambda}$ and $\tilde{\psi}$ are fully determined by $\tilde{\beta}$ and $\tilde{\sigma}$. Indeed, from Eq. (5) we have:

$$\frac{d\tilde{M}}{d\tau} = -\tilde{M} \cdot (\tilde{\lambda} - \tilde{\psi}) = \tilde{\sigma} - \tilde{\beta}, \quad (18)$$

with initial condition $X(0) = 0$. Hence:

$$\tilde{M} = \int_0^\tau (\tilde{\sigma}(s) - \tilde{\beta}(s)) ds, \quad (19)$$

and:

$$\tilde{\lambda} = \frac{\tilde{\beta}}{\tilde{M}}, \quad (20)$$

and:

$$\tilde{\psi} = \frac{\tilde{\sigma}}{\tilde{M}}. \quad (21)$$

The normalized dLTT is simply the dLTT rescaled so that its area-under-the-curve is equal to 1; hence, the normalized dLTT is a property of the particular BDS scenario that does not depend on the specific initial condition for M , and which can be interpreted as a probability density over time. The numerator $\tilde{\lambda}(\tau)M(\tau)$ on the right-hand-side of Eq. (16) is the expected rate of branching events observed in the tree at age τ , while the denominator $\int_0^\infty M(s)ds$ merely acts as a rescaling. Hence the normalized version of $\tilde{\beta}$ (i.e., if divided by its integral $\int_0^\infty \tilde{\beta}(s)ds$) can be interpreted as the probability density on $[0, \infty)$ for any randomly chosen observed branching event. The normalization of $\tilde{\beta}$, however, is such that the integral $\int_0^\infty \tilde{\beta}(s)ds$ is equal to the average pulled birth rate, where the average is calculated based on the normalized dLTT:

$$\int_0^\infty \tilde{\beta}(\tau)d\tau = \int_0^\infty \tilde{\lambda}(s)\tilde{M}(s) ds. \quad (22)$$

In particular, $\tilde{\beta}$ contains information not just about the shape of the branching rate over time, but also about the absolute magnitude of the pulled birth rate. A similar interpretation exists for the deterministic sampling density $\tilde{\sigma}$, i.e. $\tilde{\psi}$ is the (non-normalized) probability density on $[0, \infty)$ for any randomly chosen observed tip. In contrast to $\tilde{\psi}$, the sampling density $\tilde{\sigma}$ stays finite at age 0:

$$\tilde{\sigma} = \frac{\tilde{\psi}N}{\int_0^\infty M(s) ds}. \quad (23)$$

Note that the interpretations of some of the above parameters are strictly speaking only valid in the limit of infinitely large trees, however the parameters themselves are simply properties of the BDS scenario and thus well-defined regardless of the size of the tree.

S.1.3 Congruent models

From Eq. (1) and Eq. (4) it becomes clear that the likelihood of a timetree only depends on the sampling and branching ages in the tree, but not the precise tree topology itself. As we show below, a similar statement can also be made about the model's role, as follows: Any two birth-death-sampling scenarios with the same pulled birth rate $\tilde{\lambda}$ and the same pulled sampling rate $\tilde{\psi}$ will always yield the same likelihoods. We thus henceforth call two birth-death-sampling scenarios “congruent” if they have the same $\tilde{\lambda}$ and the same $\tilde{\psi}$. Equivalently, following Supplement S.1.2, two scenarios are congruent if and only if they have the same $\tilde{\lambda}$ and the same \tilde{r} , or equivalently, if and only if they have the same $\tilde{\lambda}$ and the same $\tilde{\zeta}$, or equivalently, if and only if they have the same \tilde{r} and the same $\tilde{\zeta}$, or equivalently, if and only if they have the same deterministic branching density $\tilde{\beta}$ and the same deterministic sampling density $\tilde{\sigma}$.

To prove that congruent BDS scenarios have the same likelihoods regardless of the data, we proceed as

follows. From Eq. (2) we have:

$$\begin{aligned}
\Phi(\tau) &= \exp \left[\int_0^\tau [2\lambda E - (\lambda + \mu + \psi)] ds \right] \\
&= \exp \left[\int_0^\tau [2\lambda - 2\tilde{\lambda} - \lambda - \mu - \psi] ds \right] \\
&= e^{-2\tilde{\Lambda}(\tau)} \exp \left[\int_0^\tau [\lambda - \mu - \psi] ds \right],
\end{aligned} \tag{24}$$

where we defined:

$$\tilde{\Lambda}(\tau) := \int_0^\tau \tilde{\lambda}(s) ds. \tag{25}$$

Using Eq. (9) we can further write:

$$\begin{aligned}
\Phi(\tau) &= e^{-2\tilde{\Lambda}(\tau)} \exp \left[\int_0^\tau \left[\tilde{r} - \frac{1}{\lambda} \frac{d\lambda}{ds} \right] ds \right] \\
&= e^{-2\tilde{\Lambda}(t) + \tilde{R}(t)} \cdot \exp \left[- \int_0^\tau \frac{d \ln \lambda}{ds} ds \right] \\
&= e^{-2\tilde{\Lambda}(t) + \tilde{R}(t)} \cdot \frac{\lambda(0)}{\lambda(\tau)},
\end{aligned} \tag{26}$$

where we defined:

$$\tilde{R}(\tau) := \int_0^\tau \tilde{r}(s) ds. \tag{27}$$

Using Eq. (26) in the formula for the likelihood, Eq. (1), yields:

$$\begin{aligned}
L_{\text{or}} &= \frac{\Phi(\tau_{\text{or}})}{1 - E(\tau_{\text{or}})} \prod_{i=1}^{n-1} \lambda(x_i) \Phi(x_i) \prod_{j=1}^n \frac{\psi(y_j)}{\Phi(y_j)} \\
&= \frac{e^{-2\tilde{\Lambda}(\tau_{\text{or}}) + \tilde{R}(\tau_{\text{or}})} \lambda(0)}{\lambda(\tau_{\text{or}})(1 - E(\tau_{\text{or}}))} \prod_{i=1}^{n-1} \lambda(x_i) e^{-2\tilde{\Lambda}(x_i) + \tilde{R}(x_i)} \frac{\lambda(0)}{\lambda(x_i)} \prod_{j=1}^n \psi(y_j) e^{2\tilde{\Lambda}(y_j) - \tilde{R}(y_j)} \frac{\lambda(y_j)}{\lambda(0)} \\
&= \frac{e^{-2\tilde{\Lambda}(\tau_{\text{or}}) + \tilde{R}(\tau_{\text{or}})}}{\tilde{\lambda}(\tau_{\text{or}})} \prod_{i=1}^{n-1} e^{-2\tilde{\Lambda}(x_i) + \tilde{R}(x_i)} \prod_{j=1}^n e^{2\tilde{\Lambda}(y_j) - \tilde{R}(y_j)} \psi(y_j) \lambda(y_j).
\end{aligned} \tag{28}$$

Since $\lambda\psi = \tilde{\zeta}$, we can further write:

$$L_{\text{or}} = \frac{e^{-2\tilde{\Lambda}(\tau_{\text{or}}) + \tilde{R}(\tau_{\text{or}})}}{\tilde{\lambda}(\tau_{\text{or}})} \prod_{i=1}^{n-1} e^{-2\tilde{\Lambda}(x_i) + \tilde{R}(x_i)} \prod_{j=1}^n e^{2\tilde{\Lambda}(y_j) - \tilde{R}(y_j)} \tilde{\zeta}(y_j). \tag{29}$$

Note that the likelihood in Eq. (29) depends solely on the variables $\tilde{\lambda}$, \tilde{r} , $\tilde{\zeta}$, $\tilde{\Lambda}$ and \tilde{R} , all of which are fully determined by the pulled birth rate $\tilde{\lambda}$ and pulled sampling rate $\tilde{\psi}$.

Using a similar approach one can also write the likelihood L_{r} , i.e., conditioned on the root age, in a format

that only depends on pulled variables:

$$L_r = \frac{e^{-2\tilde{\Lambda}(\tau_r) + \tilde{R}(\tau_r)}}{\tilde{\lambda}^2(\tau_r)} \prod_{i=1}^{n-1} e^{-2\tilde{\Lambda}(x_i) + \tilde{R}(x_i)} \prod_{j=1}^n e^{2\tilde{\Lambda}(y_j) - \tilde{R}(y_j)} \tilde{\zeta}(y_j). \quad (30)$$

□

S.1.4 Congruent models have identical distributions of tree sizes

In the following we show that any two congruent BDS scenarios will generate trees whose size (i.e., number of tips) has the same probability distributions, when conditioned on the age of the stem and the sampling of at least one tip (or the age of the root, and the sampling of both daughter lineages). Let $P_n(\tau)$ be the probability that a lineage alive at age τ will have exactly n sampled descendants in the final tree (i.e., at present-day, where the process ends). It is straightforward to show that P_n (for any $n \in \{0, 1, 2, \dots\}$) satisfies the differential equation:

$$\frac{dP_n}{d\tau} = \lambda \sum_{k=0}^n P_k P_{n-k} + \mu \delta_{n,0} + \psi \delta_{n,1} - (\lambda + \mu + \psi) P_n, \quad (31)$$

with initial condition:

$$P_n(0) = \delta_{n,0}, \quad (32)$$

where τ denotes age and δ denotes the Kronecker-delta symbol, i.e. $\delta_{n,k}$ is 1 if $k = n$ and 0 otherwise. Denote $\tilde{P}_n := P_n/(1 - E)$. We wish to show that \tilde{P}_n is the same for any two congruent BDS scenarios, for any $n > 0$. We have:

$$\begin{aligned} \frac{d\tilde{P}_n}{d\tau} &= \frac{1}{1-E} \frac{dP_n}{d\tau} + \frac{P_n}{(1-E)^2} \frac{dE}{d\tau} \\ &\stackrel{(31)}{=} \frac{1}{(1-E)^2} \left[(1-E) \left(\lambda \sum_{k=0}^n P_k P_{n-k} + \psi \delta_{n,1} - (\lambda + \mu + \psi) P_n \right) \right. \\ &\quad \left. + (\lambda E^2 + \mu - (\lambda + \mu + \psi) E) P_n \right] \\ &= \frac{1}{(1-E)^2} \left[\underbrace{\lambda(1-E)}_{\tilde{\lambda}} \sum_{k=1}^{n-1} P_k P_{n-k} + 2\lambda E P_n + (1-E)\psi \delta_{n,1} - (\lambda + \psi) P_n - \lambda E^2 P_n \right] \quad (33) \\ &= \tilde{\lambda} \sum_{k=1}^{n-1} \tilde{P}_k \tilde{P}_{n-k} + \underbrace{\frac{\psi}{1-E}}_{\tilde{\psi}} \delta_{n,1} - \underbrace{\frac{\psi}{1-E}}_{\tilde{\psi}} \cdot \underbrace{\frac{P_n}{1-E}}_{\tilde{P}_n} - \underbrace{\frac{P_n}{1-E}}_{\tilde{P}_n} \cdot \underbrace{(1-E)\lambda}_{\tilde{\lambda}} \\ &= \tilde{\lambda} \sum_{k=1}^{n-1} \tilde{P}_k \tilde{P}_{n-k} + \tilde{\psi} \delta_{n,1} - \tilde{\psi} \tilde{P}_n - \tilde{\lambda} \tilde{P}_n. \end{aligned}$$

Further, according to L'Hopital's rule one has:

$$\lim_{\tau \rightarrow 0^+} \tilde{P}_n(\tau) = \lim_{\tau \rightarrow 0^+} \frac{P'_n(\tau)}{1 - E'(\tau)} = \lim_{\tau \rightarrow 0^+} \frac{\psi(0)\delta_{n,1}}{\psi(0)} = \delta_{n,1}. \quad (34)$$

Since any two congruent scenarios have the same $\tilde{\lambda}$ and the same $\tilde{\psi}$, each of their \tilde{P}_n (with $n \geq 1$) will satisfy the same differential equation (33) with the same initial condition (34). We thus conclude that any two congruent scenarios will have the same conditional probabilities \tilde{P}_n .

A similar conclusion can be drawn for the case where we condition on the age of the root and the survival of its two child lineages, as follows. Denote by $\hat{P}_n(\tau)$ the conditional probability that the tree has size n , given that the root split at age τ and that both of its child lineages survived. Then for any $n \geq 2$ we have:

$$\hat{P}_n = \frac{1}{(1 - E)^2} \sum_{k=1}^{n-1} P_k P_{n-k} = \sum_{k=1}^{n-1} \tilde{P}_k \tilde{P}_{n-k}. \quad (35)$$

Since the \tilde{P}_n are identical between congruent scenarios (for any $n \geq 1$), we conclude that \hat{P}_n are also identical between congruent scenarios (for any $n \geq 2$). □

S.1.5 Constructing congruent scenarios

For any given BDS scenario (λ, μ, ψ) there are various approaches towards constructing alternative congruent scenarios, described below. All of the presented methods have in common that one first calculates the “pulled variables” of the original scenario (which fully define the scenario's congruence class), and then provides sufficient additional constraints to obtain a new single scenario within the congruence class. The discussion below also reveals how scenarios can become fully identifiable when additional constraints are provided.

S.1.5.1 Congruent scenarios by specifying ψ

In the first approach, we specify a new sampling rate ψ^* and then adjust the remaining model variables to obtain a congruent scenario. Specifically, let $\tilde{\lambda}$, \tilde{r} , and $\tilde{\zeta}$ be the pulled variables of the scenario as defined in Supplement S.1.2. Let $\psi^* > 0$ be some arbitrary alternative sampling rate. Define $\lambda^* := \tilde{\zeta}/\psi^*$ and:

$$\mu^* := \lambda^* - \psi^* + \frac{1}{\lambda^*} \frac{d\lambda^*}{d\tau} - \tilde{r} \quad (36)$$

Then the pulled diversification rate of the new scenario $(\lambda^*, \mu^*, \psi^*)$ is:

$$\tilde{r}^* = \lambda^* - \mu^* - \psi^* + \frac{1}{\lambda^*} \frac{d\lambda^*}{d\tau} \stackrel{(36)}{=} \lambda^* - \lambda^* + \psi^* - \frac{1}{\lambda^*} \frac{d\lambda^*}{d\tau} + \tilde{r} - \psi^* + \frac{1}{\lambda^*} \frac{d\lambda^*}{d\tau} = \tilde{r}. \quad (37)$$

Further, $\tilde{\zeta}^* = \lambda^* \psi^* = \tilde{\zeta}$. Hence, the new scenario has the same \tilde{r} and the same $\tilde{\zeta}$ as the original scenario. By Supplement S.1.3, the two scenarios are thus congruent. Hence, simply by choosing an alternative sampling rate ψ^* , one can find a corresponding congruent scenario $(\lambda^*, \mu^*, \psi^*)$, as long as the resulting μ^* is physically meaningful (i.e., non-negative). Reciprocally, if the congruence class of a diversification/epidemiological scenario is known (e.g., estimated via maximum-likelihood), and in addition ψ is somehow independently

estimated, then the full scenario can be reconstructed.

S.1.5.2 Congruent scenarios by specifying μ and $\lambda(\tau_o)$

Instead of first specifying an alternative ψ^* , one could construct congruent scenarios by first specifying μ^* and the birth rate $\lambda^*(\tau_o)$ at some age τ_o , as follows. Let \tilde{r} and $\tilde{\zeta}$ be the pulled variables of the scenario (λ, μ, ψ) as defined in Supplement S.1.2. For any given death rate profile μ^* , age τ_o and birth rate $\lambda_o^* \geq 0$ at age τ_o , choose λ^* as the solution of the following differential equation:

$$\frac{d\lambda^*}{d\tau} = \lambda^* \cdot [\tilde{r} + \mu^* - \lambda^*] + \tilde{\zeta}, \quad (38)$$

with condition $\lambda^*(\tau_o) = \lambda_o^*$, and set $\psi^* := \tilde{\zeta}/\lambda^*$. The pulled diversification rate of the new scenario $(\lambda^*, \mu^*, \psi^*)$ is given by:

$$\tilde{r}^* = \lambda^* - \mu^* - \psi^* + \frac{1}{\lambda^*} \frac{d\lambda^*}{d\tau} \stackrel{(38)}{=} \lambda^* - \mu^* - \psi^* - [\lambda^* - \tilde{r} - \mu^*] + \frac{\tilde{\zeta}}{\lambda^*} = \tilde{r}. \quad (39)$$

Further, $\tilde{\zeta}^* = \lambda^* \psi^* = \tilde{\zeta}$. Hence, the new scenario has the same \tilde{r} and the same $\tilde{\zeta}$ as the original scenario. By Supplement S.1.3, the two scenarios are thus congruent. Hence, simply by choosing an alternative sampling rate μ^* and an arbitrary initial condition $\lambda^*(\tau_o)$ at some age τ_o , one can find a corresponding congruent scenario $(\lambda^*, \mu^*, \psi^*)$, provided that the resulting λ^* is physically meaningful (i.e., non-negative). Reciprocally, if the congruence class of a diversification/epidemiological scenario is known (e.g., estimated via maximum-likelihood), and in addition μ and $\lambda(\tau_o)$ are somehow independently obtained (e.g., from clinical experiments), then the full scenario can be reconstructed.

Note that, as long as $\lambda\psi > 0$ and $\lambda^*(0) > 0$, and as long as \tilde{r} , $\tilde{\zeta}$ and μ^* are continuous, the corresponding λ^* will always be strictly positive. To see why this is the case, observe that if λ^* was zero or negative at some point, there would exist an age τ_z where $\lambda^*(\tau_z) = 0$ and $\lambda^*(\tau) > 0$ for all $\tau < \tau_z$. Since $\tilde{\zeta}(\tau_z) = \lambda(\tau_z)\psi(\tau_z) > 0$ and $\lambda^*(\tau_z) \cdot [\tilde{r}(\tau_z) + \mu^*(\tau_z) - \lambda^*(\tau_z)] = 0$, by Eq. (38) the derivative $d\lambda^*/d\tau$ must be strictly positive at τ_z , and in fact strictly positive in a small neighborhood of τ_z , which would mean λ^* could not reach zero at τ_z — a contradiction.

S.1.5.3 Congruent scenarios by specifying μ and $\psi(\tau_o)$

Alternatively to specifying the profile μ^* and the value $\lambda^*(\tau_o)$ at some age τ_o , one could also specify the profile μ^* and the value $\psi^*(\tau_o)$. In that case, simply choose $\lambda^*(\tau_o) = \tilde{\zeta}(\tau_o)/\psi^*(\tau_o)$, and apply the procedure in Supplement S.1.5.2 to obtain a congruent scenario $(\lambda^*, \mu^*, \psi^*)$.

S.1.5.4 Congruent scenarios by specifying λ

Another way of constructing congruent scenarios is to first specify λ^* and then choose μ^* and ψ^* accordingly, as follows. Let \tilde{r} and $\tilde{\zeta}$ be the pulled variables of the scenario (λ, μ, ψ) as defined in Supplement S.1.2. For any given λ^* , choose $\psi^* := \tilde{\zeta}/\lambda^*$ and:

$$\mu^* := \lambda^* - \psi^* + \frac{1}{\lambda^*} \frac{d\lambda^*}{d\tau} - \tilde{r}. \quad (40)$$

Then the pulled diversification rate of the new scenario $(\lambda^*, \mu^*, \psi^*)$ is:

$$\tilde{r}^* = \lambda^* - \mu^* - \psi^* + \frac{1}{\lambda^*} \frac{d\lambda^*}{d\tau} \stackrel{(40)}{=} \lambda^* - \lambda^* + \psi^* - \frac{1}{\lambda^*} \frac{d\lambda^*}{d\tau} + \tilde{r} - \psi^* + \frac{1}{\lambda^*} \frac{d\lambda^*}{d\tau} = \tilde{r}. \quad (41)$$

Hence, the new scenario has the same \tilde{r} and the same $\tilde{\zeta}$ as the original scenario. The two scenarios are thus congruent. Hence, simply by choosing an alternative sampling rate λ^* , one can find a corresponding congruent scenario $(\lambda^*, \mu^*, \psi^*)$, as long as the resulting μ^* is physically meaningful (i.e., non-negative).

S.1.5.5 Congruent scenarios by specifying R_e and $\lambda(\tau_o)$

Congruent scenarios can also be constructed by first choosing the effective reproduction ratio profile and the birth rate $\lambda^*(\tau_o)$ at some age τ_o , as follows. Let \tilde{r} and $\tilde{\zeta}$ be the pulled variables of the scenario (λ, μ, ψ) as defined in Supplement S.1.2. For any given basic reproduction ratio profile R_e^* , age τ_o and birth rate $\lambda_o^* \geq 0$ at age τ_o , choose λ^* as the solution of the following differential equation:

$$\frac{d\lambda^*}{d\tau} = \lambda^* \cdot \left[\lambda^* \cdot \left(\frac{1}{R_e} - 1 \right) + \tilde{r} \right], \quad (42)$$

with condition $\lambda^*(\tau_o) = \lambda_o^*$, and set $\psi^* := \tilde{\zeta}/\lambda^*$ and $\mu^* := \lambda^*/R_e - \psi^*$. The pulled diversification rate of the new scenario $(\lambda^*, \mu^*, \psi^*)$ is given by:

$$\tilde{r}^* = \lambda^* - \mu^* - \psi^* + \frac{1}{\lambda^*} \frac{d\lambda^*}{d\tau} = \lambda^* - \frac{\lambda^*}{R_e} + \psi^* - \psi^* + \lambda^* \cdot \left(\frac{1}{R_e} - 1 \right) + \tilde{r} = \tilde{r}. \quad (43)$$

Further, $\tilde{\zeta}^* = \lambda^* \psi^* = \tilde{\zeta}$. Hence, the new scenario has the same \tilde{r} and the same $\tilde{\zeta}$ as the original scenario. By Supplement S.1.3, the two scenarios are thus congruent. Hence, simply by choosing an alternative R_e^* and an arbitrary condition $\lambda^*(\tau_o) = \lambda_o^*$ at some age τ_o , one can find a corresponding congruent scenario $(\lambda^*, \mu^*, \psi^*)$, provided that the resulting λ^* and μ^* are physically meaningful (i.e., non-negative). Reciprocally, if the congruence class of a diversification/epidemiological scenario is known (e.g., estimated via maximum-likelihood), and in addition R_e and $\lambda(\tau_o)$ are somehow independently obtained, then the full scenario can be reconstructed.

Note that the differential equation (42) is of Bernoulli type, and can thus be solved analytically:

$$\lambda^*(\tau) = \frac{\lambda_o^* e^{\tilde{R}(\tau)}}{e^{\tilde{R}(\tau_o)} + \left(1 - \frac{1}{R_e^*}\right) \lambda_o^* \int_{\tau_o}^{\tau} e^{\tilde{R}(s)} ds}, \quad (44)$$

where \tilde{R} is defined as in Eq. (27).

S.1.5.6 Congruent scenarios by specifying $\mu + \psi$ and $\lambda(\tau_o)$

Congruent scenarios can also be constructed by first choosing the total removal rate $\delta := \mu + \psi$ and the birth rate $\lambda^*(\tau_o)$ at some age τ_o , as follows. Let \tilde{r} and $\tilde{\zeta}$ be the pulled variables of the scenario (λ, μ, ψ) as defined in Supplement S.1.2. For any given profile $\delta^* \geq 0$, age τ_o and birth rate $\lambda_o^* \geq 0$ at age τ_o , choose λ^* as the

solution of the following differential equation:

$$\frac{d\lambda^*}{d\tau} = \lambda^* \cdot (\delta^* + \tilde{r} - \lambda^*), \quad (45)$$

with condition $\lambda^*(\tau_o) = \lambda_o^*$, and set $\psi^* := \tilde{\zeta}/\lambda^*$ and $\mu^* := \delta^* - \psi^*$. The pulled diversification rate of the new scenario $(\lambda^*, \mu^*, \psi^*)$ is given by:

$$\tilde{r}^* = \lambda^* - \mu^* - \psi^* + \frac{1}{\lambda^*} \frac{d\lambda^*}{d\tau} = \lambda^* - \delta^* + \psi^* - \psi^* + (\delta^* + \tilde{r} - \lambda^*) = \tilde{r}. \quad (46)$$

Further, $\tilde{\zeta}^* = \lambda^* \psi^* = \tilde{\zeta}$. Hence, the new scenario has the same \tilde{r} and the same $\tilde{\zeta}$ as the original scenario. By Supplement S.1.3, the two scenarios are thus congruent. Hence, simply by choosing an alternative δ^* and an arbitrary condition $\lambda^*(\tau_o) = \lambda_o^*$ at some age τ_o , one can find a corresponding congruent scenario $(\lambda^*, \mu^*, \psi^*)$, provided that the resulting λ^* and μ^* are physically meaningful (i.e., non-negative). Reciprocally, if the congruence class of a diversification/epidemiological scenario is known (e.g., estimated via maximum-likelihood), and in addition δ and $\lambda(\tau_o)$ are somehow independently obtained, then the full scenario can be reconstructed.

Note that the differential equation (45) is of Bernoulli type, and can thus be solved analytically:

$$\lambda^*(\tau) = \frac{\lambda_o^* e^{\tilde{R}(\tau) + D^*(\tau)}}{e^{\tilde{R}(\tau_o) + D^*(\tau_o)} + \lambda_o^* \int_{\tau_o}^{\tau} e^{\tilde{R}(s) + D^*(s)} ds}, \quad (47)$$

where \tilde{R} is defined as in Eq. (27) and D^* is defined as:

$$D^*(\tau) := \int_0^{\tau} \delta^*(s) ds. \quad (48)$$

From Eq. (47) it becomes clear that if $\tau_o = 0$ and $\lambda_o^* \geq 0$, then $\lambda^*(\tau)$ will be non-negative for all $\tau \geq 0$ (i.e., physically meaningful).

S.1.5.7 Congruent scenarios by specifying $\mu + \psi$ and $S(\tau_o)$

Congruent scenarios can also be constructed by first choosing the total removal rate $\delta := \mu + \psi$ and the sampling proportion $S^*(\tau_o)$ at some age τ_o , as follows. Let \tilde{r} and $\tilde{\zeta}$ be the pulled variables of the scenario (λ, μ, ψ) as defined in Supplement S.1.2. For any given profile $\delta^* > 0$, age τ_o and sampling proportion $S_o^* \in (0, 1]$ at age τ_o , choose λ^* as the solution of the following differential equation:

$$\frac{d\lambda^*}{d\tau} = \lambda^* \cdot (\delta^* + \tilde{r} - \lambda^*), \quad (49)$$

with initial condition $\lambda^*(\tau_o) = \tilde{\zeta}(\tau_o)/(S_o^* \cdot \delta^*(\tau_o))$, and set $\psi^* := \tilde{\zeta}/\lambda^*$ and $\mu^* := \delta^* - \psi^*$. The pulled diversification rate of the new scenario $(\lambda^*, \mu^*, \psi^*)$ is given by:

$$\tilde{r}^* = \lambda^* - \mu^* - \psi^* + \frac{1}{\lambda^*} \frac{d\lambda^*}{d\tau} = \lambda^* - \delta^* + \psi^* - \psi^* + (\delta^* + \tilde{r} - \lambda^*) = \tilde{r}. \quad (50)$$

Further, $\tilde{\zeta}^* = \lambda^* \psi^* = \tilde{\zeta}$. Hence, the new scenario has the same \tilde{r} and the same $\tilde{\zeta}$ as the original scenario. By Supplement S.1.3, the two scenarios are thus congruent. Hence, simply by choosing an alternative δ^*

and an arbitrary condition $S^*(\tau_o) = S_o^*$ at some age τ_o , one can find a corresponding congruent scenario $(\lambda^*, \mu^*, \psi^*)$, provided that the resulting λ^* and μ^* are physically meaningful (i.e., non-negative). Reciprocally, if the congruence class of a diversification/epidemiological scenario is known (e.g., estimated via maximum-likelihood), and in addition δ and $S(\tau_o)$ are somehow independently obtained, then the full scenario can be reconstructed.

S.1.5.8 Congruent scenarios by specifying $\psi/(\mu + \psi)$ and $\lambda(\tau_o)$

Congruent scenarios can also be constructed by first choosing the ‘‘sampling proportion’’ $S := \psi/(\mu + \psi)$ and the birth rate $\lambda^*(\tau_o)$ at some age τ_o , as follows. Let \tilde{r} and $\tilde{\zeta}$ be the pulled variables of the scenario (λ, μ, ψ) as defined in Supplement S.1.2. For any given profile $S^* > 0$, age τ_o and birth rate $\lambda_o^* \geq 0$ at age τ_o , choose λ^* as the solution of the following differential equation:

$$\frac{d\lambda^*}{d\tau} = \frac{\tilde{\zeta}}{S^*} + \lambda^* \cdot (\tilde{r} - \lambda^*) \quad (51)$$

with condition $\lambda^*(\tau_o) = \lambda_o^*$, and set $\psi^* := \tilde{\zeta}/\lambda^*$ and $\mu^* := \psi^*/S^* - \psi^*$. The pulled diversification rate of the new scenario $(\lambda^*, \mu^*, \psi^*)$ is given by:

$$\tilde{r}^* = \lambda^* - \mu^* - \psi^* + \frac{1}{\lambda^*} \frac{d\lambda^*}{d\tau} = \lambda^* - \frac{\psi^*}{S^*} + \psi^* - \psi^* + \frac{\tilde{\zeta}}{S^*\lambda^*} + (\tilde{r} - \lambda^*) = \tilde{r}. \quad (52)$$

Further, $\tilde{\zeta}^* = \lambda^*\psi^* = \tilde{\zeta}$. Hence, the new scenario has the same \tilde{r} and the same $\tilde{\zeta}$ as the original scenario. By Supplement S.1.3, the two scenarios are thus congruent. Hence, simply by choosing an alternative S^* and an arbitrary condition $\lambda^*(\tau_o) = \lambda_o^*$ at some age τ_o , one can find a corresponding congruent scenario $(\lambda^*, \mu^*, \psi^*)$, provided that the resulting λ^* and μ^* are physically meaningful (i.e., non-negative). Reciprocally, if the congruence class of a diversification/epidemiological scenario is known (e.g., estimated via maximum-likelihood), and in addition S and $\lambda(\tau_o)$ are somehow independently obtained, then the full scenario can be reconstructed.

S.1.6 The information content of concentrated sampling attempts

In the following we describe how samples obtained during concentrated sampling attempts can yield valuable insight into an epidemic. We define a concentrated sampling attempt (CSA) as a short period of sampling, taking place between two ages $\tau_1 > \tau_2 > 0$, and satisfying the following assumptions:

- A. $|\tau_1 - \tau_2|$ is much smaller than $1/\lambda$ and $1/\mu$ throughout the CSA.
- B. λ does not vary substantially between the times τ_1 and τ_2 , i.e., we assume that $\lambda(\tau_1) \approx \lambda(\tau_2)$.
- C. The probability that a lineage alive at τ_1 will be sampled before age τ_2 , denoted ρ , is sufficiently high so that during the CSA the number of lineages sampled is much larger than the number of birth and death events, i.e. $\rho \gg \lambda \cdot |\tau_1 - \tau_2|$ and $\rho \gg \mu \cdot |\tau_1 - \tau_2|$.
- D. In the exterior vicinity of the open interval (τ_1, τ_2) , $\tilde{\lambda}$ and $\tilde{\psi}$ are continuous, so that $\tilde{\lambda}(\tau_1)$, $\tilde{\lambda}(\tau_2)$, $\tilde{\psi}(\tau_1)$ and $\tilde{\psi}(\tau_2)$ can be asymptotically identified.

We will show that in the limit of infinitely large trees, $\lambda(\tau_1)$, $\psi(\tau_1)$ and ρ can be accurately identified. Assuming that the CSA is so short that birth and death are unlikely to occur in any given lineage (assumption A), we have:

$$1 - E(\tau_1) \approx \rho + (1 - \rho) \cdot [1 - E(\tau_2)], \quad (53)$$

which is equivalent to:

$$E(\tau_1) \approx E(\tau_2) \cdot (1 - \rho). \quad (54)$$

In other words, the probability that a lineage alive at age τ_1 will eventually be sampled is approximately equal to the probability of being sampled during the CSA plus the probability of not being sampled during CSA multiplied by the probability that a lineage alive at age τ_2 will eventually be sampled. In the absence of births and deaths, the above approximation formula would be exact. Assuming that $\lambda(\tau_1) \approx \lambda(\tau_2)$ (assumption B) and using Eq. (54), we have:

$$\begin{aligned} \tilde{\lambda}(\tau_1) &= \lambda(\tau_1) \cdot [1 - E(\tau_1)] \stackrel{(54)}{\approx} \lambda(\tau_1) \cdot [1 - E(\tau_2)(1 - \rho)] \\ &= \lambda(\tau_1) \cdot [1 - E(\tau_2)] + \lambda(\tau_1)\rho E(\tau_2) \stackrel{B}{\approx} \lambda(\tau_2) \cdot [1 - E(\tau_2)] + \lambda(\tau_1)\rho E(\tau_2) \\ &= \tilde{\lambda}(\tau_2) + \lambda(\tau_1)\rho E(\tau_2). \end{aligned} \quad (55)$$

We mention that, strictly speaking, we only need to assume that $\lambda(\tau_1) \approx \lambda(\tau_2)$, i.e. in principle λ could vary during the CSA as long as it returns to its immediate pre-CSA value. The number of lineages sampled during the age interval $[\tau_1, \tau_2]$, denoted S , can be accurately read off the tree, if the tree is sufficiently large. Assuming that births and deaths within any given lineage are negligible compared to the sampling effort (i.e., changes in the total number of extant lineages are mostly due to sampling, assumption C), and for sufficiently large trees, we have:

$$S \approx N(\tau_1)\rho \approx \frac{M(\tau_1)\rho}{1 - E(\tau_1)}, \quad (56)$$

where N is the total number of extant lineages and M is the tree's LTT. Denoting:

$$\tilde{\rho} := \frac{\rho}{1 - E(\tau_1)}, \quad (57)$$

we see that $\tilde{\rho}$ is asymptotically identifiable, since $\tilde{\rho} \approx S/M(\tau_1)$. Note that:

$$\lambda(\tau_1)\rho = \tilde{\lambda}(\tau_1)\tilde{\rho}. \quad (58)$$

Using Eq. (55) and Eq. (58) we obtain:

$$\tilde{\lambda}(\tau_1) = \tilde{\lambda}(\tau_2) + \tilde{\lambda}(\tau_1)\tilde{\rho}E(\tau_2), \quad (59)$$

and hence:

$$E(\tau_2) = \frac{\tilde{\lambda}(\tau_1) - \tilde{\lambda}(\tau_2)}{\tilde{\rho}\tilde{\lambda}(\tau_1)}. \quad (60)$$

Since $\tilde{\lambda}(\tau_1)$ and $\tilde{\lambda}(\tau_2)$ are asymptotically identifiable (assumption D), and $\tilde{\rho}$ is also asymptotically identifiable, it follows that $E(\tau_2)$ is also asymptotically identifiable. Note that the condition $\tau_2 > 0$ is necessary for ensuring that $\tilde{\lambda}(\tau_2)$ is asymptotically identifiable, since we need samples younger than τ_2 for estimating $\tilde{\lambda}(\tau_2)$. Solving Eq. (54) and Eq. (57) for ρ and $E(\tau_1)$ leads to:

$$\rho \approx \frac{\tilde{\rho} \cdot [1 - E(\tau_2)]}{1 - \tilde{\rho}E(\tau_2)}. \quad (61)$$

Hence, ρ is also asymptotically identifiable. Combined with Eq. (54), this implies that $E(\tau_1)$ is also asymptotically identifiable. Solving Eq. (58) for $\lambda(\tau_1)$ leads to:

$$\lambda(\tau_1) \approx \frac{\tilde{\lambda}(\tau_1)\tilde{\rho}}{\rho}. \quad (62)$$

Hence, $\lambda(\tau_1)$ is also asymptotically identifiable. Lastly, since $\tilde{\psi}(\tau_1)$ and $\tilde{\psi}(\tau_2)$ are asymptotically identifiable (assumption D), and since $\psi = \tilde{\psi} \cdot (1 - E)$, we conclude that $\psi(\tau_1)$ and $\psi(\tau_2)$ are also asymptotically identifiable. □

S.1.7 Infinitesimally short concentrated sampling attempts

From a numerical perspective, it might be more practical to describe very short CSAs as separate instantaneous sampling processes with designated parameters (e.g., the time of the CSA and the probability of sampling a lineage during the CSA), rather than trying to approximate CSAs as very sharp peaks in ψ . The resulting model structure would then resemble the one described previously by Stadler *et al.* [2]. This is also the approach taken in the R package *castor* when fitting BDS models to timetrees: CSAs are parameterized separately from Poissonian sampling [4]. Formally, one can derive the corresponding likelihood starting from the BDS likelihood introduced earlier (Eqs. (1) or (4)) by considering a modified sampling rate of the form:

$$\bar{\psi} = \psi + \sum_{k=1}^m w_k \delta(\tau - s_k), \quad (63)$$

where δ is the Dirac distribution, m is the number of CSAs, s_k is the age of the k -th CSA ($0 \leq s_1 < \dots < s_m \leq \tau_{\text{or}}$), $w_k := -\ln(1 - \rho_k)$ is a weight, $\rho_k \in (0, 1]$ is the sampling probability during the k -th CSA and ψ is the “background” (Poissonian) sampling rate outside of the CSAs. For a derivation of the corresponding likelihood of a bifurcating timetree in terms of λ , μ , ψ , the s_k and the ρ_k see [1]. For convenience, we provide the likelihood below using the notation of the present manuscript. We denote by N_k the number of tips sampled during and due to the k -th CSA, and by n_k the number of lineages “crossing” over s_k in the timetree. The likelihood density of the tree, when conditioned on the age of the origin τ_{or} and the survival of at least one lineage, is then given by:

$$L_{\text{or}} = \frac{\Phi_m(\tau_{\text{or}})}{1 - E(\tau_{\text{or}})} \prod_{i=1}^{n-1} \lambda(x_i) \Phi_{m_{x_i}}(x_i) \prod_{k=1}^m [(1 - \rho_k) \Phi_{k-1}(s_k)]^{n_k} \rho_k^{N_k} \prod_{j=1}^n \frac{\psi(y_j)}{\Phi_{m_{y_j}}(y_j)}, \quad (64)$$

where as before $E(\tau)$ is the probability that a lineage alive at age τ is not included in the timetree, m_τ is the largest possible integer for which $s_{m_\tau} \leq \tau$, and Φ_0, \dots, Φ_m are the sub-flows [3] between CSAs, defined as:

$$\Phi_k(\tau) := \exp \left[\int_{s_k}^{\tau} [2\lambda(s)E(s) - \lambda(s) - \mu(s) - \psi(s)] ds \right], \quad k = 0, \dots, m, \quad (65)$$

where for notational simplicity we defined $s_0 := 0$. Note that for each $k = 0, \dots, m$ the probability E satisfies within the age interval $[s_k, s_{k+1})$ the differential equation:

$$\frac{dE}{d\tau} = -(\lambda + \mu + \psi)E + \lambda E^2 + \mu, \quad (66)$$

with initial condition $E(s_k) = E(s_k^-) \cdot (1 - \rho_k)$ and $E(0^-) = 1$. Similarly, one can derive the likelihood of a bifurcating timetree conditioned on the age of the tree's root τ_r , i.e., a splitting occurring at age τ_r and the eventual sampling of both child lineages:

$$L_r = \frac{\Phi_m(\tau_r)}{\lambda(\tau_r) \cdot (1 - E(\tau_r))^2} \prod_{i=1}^{n-1} \lambda(x_i) \Phi_{m_{x_i}}(x_i) \prod_{k=1}^m [(1 - \rho_k) \Phi_{k-1}(s_k)]^{n_k} \rho_k^{N_k} \prod_{j=1}^n \frac{\psi(y_j)}{\Phi_{m_{y_j}}(y_j)}. \quad (67)$$

S.2 Why common model selection methods cannot resolve congruencies

In this section we clarify why model selection methods that do not integrate additional independent information and are entirely based on parsimony principles, such as AIC [5], BIC [6] or regularization [7], are not suitable for resolving model congruencies.

S.2.1 General considerations (big picture)

Parsimony-based model selection methods such as AIC, BIC and regularization essentially penalize excessive model complexity and are designed to prevent overfitting to finite datasets, i.e., to avoid inferring spurious complexities when the data at hand can also be adequately explained by simpler models. Indeed, it is common for multiple models to explain a dataset similarly well and in this case it is prudent to select the model with fewer free parameters since this tends to decrease the prediction error. These approaches do not, by any means, guarantee or even suggest that for any given dataset the selected (simpler) model is the one closest to the truth, as is often erroneously assumed. These methods merely constitute a methodological principle for gradually and conservatively choosing more complex models as new data appear, under the presumption that as new data accumulate the selected models will eventually converge towards the truth.

When modelling pathogen epidemics, however, one must decide between congruent scenarios with differing levels of complexity, and no phylogenetic data, no matter how large, can distinguish between their proximity to the truth. In fact, all scenarios within a congruence class generate phylogenies with the same probability distribution, and hence they also have identical expected “prediction errors” (which, for example, AIC was designed to minimize). Hence, merely selecting the simpler model will almost always lead to an erroneous scenario, no matter how much sequence data we add. The very promise of conventional parsimony-based model selection — that selected models eventually converge to the truth as one adds more data — simply does not hold in this case.

One may of course compare scenarios belonging to different congruence classes since these will generally

have different likelihoods for a particular dataset, and will certainly be distinguishable as the size of the dataset becomes very large. However, this does not solve the problem at hand, since one is not really selecting the best model for the data but in fact merely the best congruence class. Even if one of the candidate scenarios in the candidate set is entirely adequate for explaining the data at hand, or in the extreme case, is in the exact same congruence class as the true historical epidemiological dynamics, it does not follow that this scenario will resemble, even qualitatively, the true historical dynamics.

In conclusion, model selection methods that do not integrate additional independent information can at most be used to find a congruence class that "most efficiently" reproduces the data, or that has the lowest prediction error. In that sense, fitting congruence classes (e.g., by directly fitting pulled variables) is a parsimonious approach towards describing a phylogenetic dataset, while fitting full birth-death-sampling models adds redundant complexity that cannot possibly be identified using phylogenetic inference alone.

S.2.2 Regularization

Let us consider specifically the example of fitting piecewise linear (or piecewise constant) profiles for λ , μ and ψ on a grid over time, combined with a regularization approach that penalizes excessive oscillations or excessively large rate estimates (an example being Tikhonov regularization [8, 9]). To see why such an approach cannot possibly resolve model congruencies, suppose that a specific epidemiological history occurred, and consider the hypothetical limit where a phylogeny generated by that epidemic becomes infinitely large while the considered resolution of the fitted profiles (i.e., the number of grid points) remains fixed. In the absence of model congruencies, i.e., in typical inverse problems where regularization is commonly applied, one would expect that the fitted profiles would eventually approach the true historical profiles, provided of course that the grid resolution is high enough to approximately capture the true historical profiles. But this is clearly not the case here. Indeed, in the limit of an infinitely large phylogeny regularization becomes irrelevant, since for any given regularization parameter (also known as "smoothing parameter") minimizing deviations from the data becomes infinitely more important than reducing the regularization penalties; in other words, for very large trees regularized profile fitting becomes equivalent to maximum-likelihood fitting. As we have demonstrated using simulations (Supplemental Figs. S1 and S2), maximum-likelihood-fitted profiles can be completely wrong even when using massive phylogenies with tens of thousands of tips and for relatively simple epidemiological histories. The situation can only be worse for smaller datasets.

S.2.3 AIC and BIC

Alternatively, let us consider the example of fitting various functional forms for λ , μ and ψ via maximum-likelihood, with the "best" functional form being selected using AIC [5] or BIC [6]. For any given choice of functional forms, the maximum-likelihood-fitted model will a priori tend to be the one closest to the congruence class of the true epidemiological history, rather than the true epidemiological history itself. Choosing the functional forms that minimize AIC or BIC would only yield a model that balances the number of parameters against the goodness of fit to the congruence class, but not against the goodness of fit to the true epidemiological history. There is little reason to expect that the fitted model selected via AIC or BIC will happen to actually be close to the true history, even for massive datasets (examples in Supplemental Figs. S1 and S2, as well as in Supplement S.3). The situation can only be worse for small datasets.

S.3 Statistical evaluation of multiple simulated scenarios

In the following we describe our statistical analysis of models fitted to multiple trees from simulated epidemiological BDS scenarios (overview of results in Supplemental Figs. S17, S19, S18, S20). All rates are given in units yr^{-1} . All simulations were run for 10 years, and only those that generated trees with 500–50 000 tips were kept. We considered two different classes of BDS scenarios. In the first class (henceforth “exponential”), λ (and similarly μ and ψ) was of the general form:

$$\lambda(t) = A + B \cdot e^{\alpha t}, \quad (68)$$

where t is time and A , B and α are parameters chosen randomly as follows. Denote $t_s = 0$ the start and $t_e = 10$ yr the end of a simulation. We randomly and uniformly chose $\lambda(t_s)$ and $\lambda(t_e)$ between 1 and 10, and randomly and uniformly chose α in $[-0.5, -0.1] \cup [0.1, 0.5]$. Based on these three constraints, we algebraically determined the appropriate A and B . A similar approach was taken for constructing μ and ψ , with the difference that we randomly and uniformly chose each of the ratios $\mu(t_s)/\lambda(t_s)$ and $\mu(t_e)/\lambda(t_e)$ between 0.1 and 1, and randomly chose each of $\psi(t_s)$ and $\psi(t_e)$ between 0.01 and 1 (uniformly on a logarithmic scale). For any particular choice of λ , μ and ψ , we simulated a tree for 10 years using the function `generate_tree_hbds` in the R package `castor` v1.6.7 [4]. If the resulting tree had too few tips (<500) or too many tips (>50 000), we chose an entirely new set of λ , μ and ψ and repeated the simulation until obtaining an acceptable tree. In the 2nd class (henceforth “OU”) the time-profile for λ (and similarly μ and ψ) was generated according to an Ornstein-Uhlenbeck stochastic process [10]. This approach was chosen in order to cover a wide range of epidemiological scenarios with realistic temporal complexity. An Ornstein-Uhlenbeck process is characterized by its stationary or asymptotic expectation (λ_{ex}), its stationary or asymptotic standard deviation (λ_{dev}) and its decay rate α . We chose the decay rate randomly and uniformly between 0.05 and 0.2, the stationary expectation λ_{ex} randomly and uniformly between 1 and 10, and fixed the stationary standard deviation to $\lambda_{\text{dev}} = 0.5 \cdot \lambda_{\text{ex}}$. The actual profile λ was then generated randomly according to the Ornstein-Uhlenbeck process evaluated at 6-month intervals, restricted (capped) to above $0.1 \cdot \lambda_{\text{ex}}$, and linearly interpolated between points. A similar approach was taken for constructing μ and ψ , with the difference that for μ we chose the stationary expectation μ_{ex} randomly and uniformly between $0.1 \cdot \lambda_{\text{ex}}$ and λ_{ex} , and for ψ we chose the stationary expectation ψ_{ex} randomly between 0.01 and 1 (uniformly on a logarithmic scale). Similarly to the exponential scenarios, we simulated a tree for 10 years under a OU scenario using the function `generate_tree_hbds`, and repeated the process until a tree comprising between 500 and 50000 tips was obtained. In total we generated 100 trees from exponential scenarios and 100 trees from OU scenarios.

To each generated tree we fitted a BDS skyline model (i.e., with piecewise constant λ , μ and ψ) and a BDS model with piecewise-linear λ , μ and ψ , via maximum-likelihood using the `castor` function `fit_hbds_model_on_grid`. Fitting was done agnostically of the true scenario used in the simulations, except for the constraint that the present-day ψ was fixed to its true value to account for previously known identifiability issues in BDS skyline models [2]. To avoid local non-global likelihood maxima, we repeated the fitting using 100 random starting points (option `Ntrials=100`). The grid points were placed non-uniformly so that their density was roughly proportional to the square root of the tree’s LTT, using the `castor` function `get_inhomogeneous_grid_1D`; this was done so that points tend to be concentrated in regions with higher information content. The best grid size was chosen according to the AIC [5]. Since in almost all cases the best fitted piecewise linear model had a better (lower) AIC value than the best fitted skyline model, we henceforth focused on piecewise linear models for simplicity. To avoid erroneous parameter estimates stemming from bad model fits, we focused on those trees where the best fitted piecewise linear model (henceforth simply “fitted model”) adequately explained the input tree, based on the same Kolmogorov-Smirnov tests for the node ages, tip ages and edge lengths described in the main article (implemented in the `castor` function

model_adequacy_hbds). In total we thus considered 92 exponential and 66 OU simulations with adequate model fits.

For every considered tree and every parameter (such as λ), we calculated the coefficient of determination between the true profile and the fitted profile, R^2 , as follows:

$$R^2 = 1 - \frac{\sum_{i=1}^N (\lambda(t_i) - \hat{\lambda}(t_i))^2}{N \cdot \mathbb{V}(\lambda)}, \quad (69)$$

where t_1, \dots, t_N is a fine time grid, λ is the true birth rate, $\hat{\lambda}$ is the estimated birth rate, and $\mathbb{V}(\lambda)$ is the variance of the true birth rate (evaluated on the same time grid t_1, \dots, t_N). Hence, a negative R^2 corresponds to a worse model fit than if one just used the mean as a prediction. We calculated the mean modulus of the normalized error (MMNE) as follows:

$$\text{MMNE} = \frac{1}{N} \sum_{i=1}^N \frac{|\lambda(t_i) - \hat{\lambda}(t_i)|}{\mathbb{E}(\lambda)}, \quad (70)$$

where $\mathbb{E}(\lambda)$ is the mean value of the true birth rate. To calculate the linear trend of a model parameter (such as λ) over time, we used least-squares linear regression with an intercept, applied to the true or estimated profile of the parameter. The R^2 , MMNE and true and estimated linear trends of selected model parameters for all considered simulations are shown in Supplemental Figs. S17 (exponential) and S19 (OU). Specific examples of fitted R_e are shown in Supplemental Figs. S18 and S20. As can be seen, the vast majority of fitted models poorly reconstructed the true epidemiological scenario, including in particular the R_e . However, the fitted models did generally come close to the true scenario's congruence class, as seen from the accurately reconstructed deterministic branching and sampling densities ($\tilde{\beta}$ and $\tilde{\sigma}$).

S.4 Overview of computer code

This section provides an overview of computational tools developed as part of this manuscript, and publicly available in the R package `castor` v1.6.5 [4], for working with BDS models and congruencies. The functions described below have been designed with efficiency in mind, and can typically scale well to phylogenies with hundreds of thousands of tips. For detailed instructions consult the `castor` user manual.

generate_tree_hbds: Generate a random timetree according to a time-dependent BDS model with arbitrary rates λ , μ , ψ through time. The rates are specified as piecewise linear curves (or quadratic splines, or cubic splines) on a discrete time grid. By choosing a sufficiently fine time grid, any arbitrary functional forms can in principle be accommodated to arbitrary accuracy.

simulate_deterministic_hbds: Simulate a time-dependent BDS model with arbitrary rates λ , μ , ψ through time, in the deterministic limit (i.e., use differential equations rather than a stochastic process). This function can be used to calculate various alternative parameters of a BDS model, such as the R_e , the removal rate δ , the sampling proportion S , the LTT, the pulled birth rate $\tilde{\lambda}$ and so on. In particular, this function can be used to check if two BDS models are congruent, by comparing their pulled birth rate ($\tilde{\lambda}$) and event density ($\lambda\psi$).

fit_hbds_model_on_grid: Fit a time-dependent BDS model to a timetree via maximum-likelihood. The λ , μ and ψ can be assumed to be either piecewise constant, piecewise linear, quadratic splines or cubic splines

defined on a discrete time grid; the values of these parameters are thus fitted at each grid point. This function uses parametric bootstrapping to calculate confidence intervals in the estimated parameters.

fit_hbds_model_parametric: Fit a BDS model to a timetree via maximum-likelihood. The profiles of λ , μ and ψ are described by user-specified functions that depend on a finite number of scalar parameters to be fitted. This function can thus be used to fit arbitrary functional forms. The function uses parametric bootstrapping to calculate confidence intervals in the estimated parameters.

congruent_hbds_model: Construct new BDS models by providing information on the congruence class (such as the pulled birth rate) and any additional constraints needed for identifying a specific member of the congruence class. This function can be used to explore the set of models congruent to some reference model. It was used, for example, to generate the scenarios in Fig. 1 in the main article.

model_adequacy_hbds: Test if a given BDS model (or a distribution of BDS models) adequately describes a given timetree in terms of various statistics. For example, this function can perform a Kolmogorov-Smirnov test to examine if the distribution of node ages (or tip ages, or edge lengths) in the timetree deviate significantly from those expected under the model(s). Note that, due to the existence of model congruencies, even if a BDS model adequately describes a given timetree this does not mean that the model is even close to the original BDS scenario that generated the tree. If a BDS model adequately describes a timetree, it merely means that any of the myriad of members of the model’s congruence class would generate trees similar to the tree at hand.

S.5 On the use of occurrence data with phylogenies

While additional data is generally useful for increasing estimation accuracy, we would like to clarify our thoughts on the potential use of “occurrence” data, commonly encountered in epidemiology, in combination with molecular/phylogenetic data, for resolving the identifiability issues uncovered in this paper. Specifically, we caution that the addition of occurrence data, i.e., reports of positive cases detected at specific times without sequencing [11–13], need not necessarily resolve congruencies between epidemiological scenarios, if the rate at which such detections occur (denoted ϑ) is a priori unknown and time-dependent. The reason is that such data contain a strict subset of the information that would be contained in detected cases with sequences (and thus included in the phylogeny), and hence phylogenetic data combined with occurrence data can at most yield as much information as a larger phylogenetic dataset in which each detected case is also sequenced. As we have shown in this paper, in the absence of further information or constraints, the rates λ , μ and ψ cannot be reconstructed from phylogenetic data alone. Thus, in the absence of further information or constraints, λ , μ , ψ , ϑ or $\psi + \vartheta$ cannot possibly be reconstructed using phylogenetic and occurrence data alone.

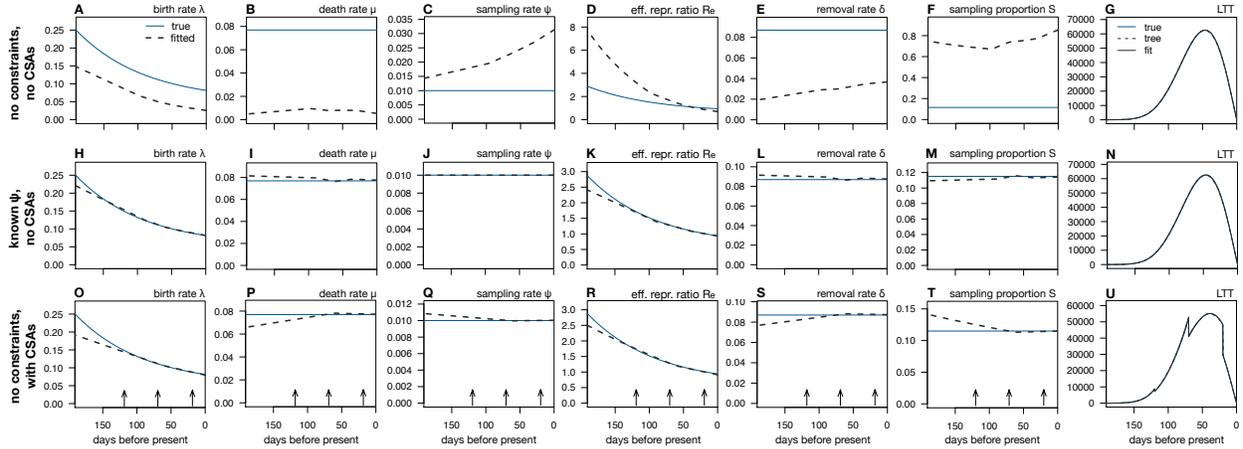


Figure S1: Limits to reconstructing an epidemic's dynamics via maximum-likelihood (scenario 1). (A–F) Maximum-likelihood estimates (grey dashed curves) of the birth rate (λ), death rate (μ), sampling rate (ψ), effective reproduction ratio (R_e), removal rate ($\delta = \mu + \psi$) and sampling proportion ($S = \psi / (\mu + \psi)$) over time, based on a timetree with 175,440 tips simulated under a hypothetical birth-death-sampling scenario (blue continuous curves) and without any additional constraints. All rates are in day^{-1} . Model adequacy was confirmed using predictive posterior simulations with multiple tests. Observe the poor agreement between the estimated and true profiles. (G) Deterministic LTT (dLTT) of the fitted model, compared to the true scenario's dLTT and the LTT of the timetree (note that all curves are nearly identical). (H–N) Similar to A–G, but for a model fitted to the same data as in A–G while fixing the sampling rate to its true profile. Observe the improved agreement between estimated and true profiles. (O–U) BDS parameters and dLTT for a model fitted to a timetree with 172,888 tips, generated under nearly the same BDS scenario as in A–G but with 3 added concentrated sampling attempts (CSAs, times indicated by vertical arrows). No additional constraints were used during fitting.

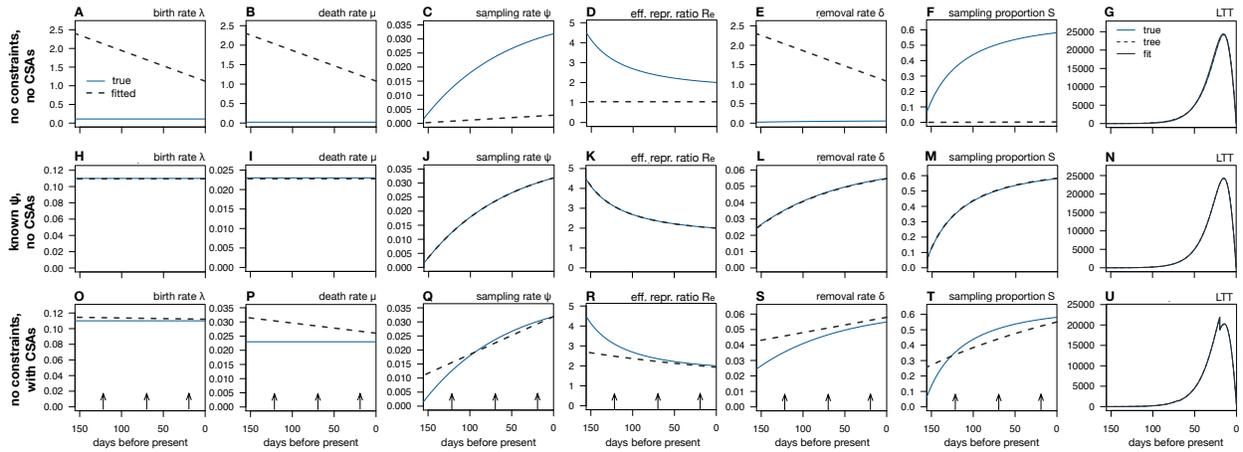


Figure S2: Limits to reconstructing an epidemic’s dynamics via maximum-likelihood (scenario 2). (A–F) Maximum-likelihood estimates (grey dashed curves) of the birth rate (λ), death rate (μ), sampling rate (ψ), effective reproduction ratio (R_e), removal rate ($\delta = \mu + \psi$) and sampling proportion ($S = \psi / (\mu + \psi)$) over time, based on a timetree with 55,934 tips simulated under a hypothetical birth-death-sampling scenario (blue continuous curves) and without any additional constraints. All rates are in day^{-1} . Model adequacy was confirmed using predictive posterior simulations with multiple tests. Observe the poor agreement between the estimated and true profiles. (G) Deterministic LTT (dLTT) of the fitted model, compared to the true scenario’s dLTT and the LTT of the timetree (note that all curves are nearly identical). (H–N) Similar to A–G, but for a model fitted to the same data as in A–G while fixing the sampling rate to its true profile. Observe the improved agreement between estimated and true profiles. (O–U) BDS parameters and dLTT for a model fitted to a timetree with 51,619 tips, generated under nearly the same BDS scenario as in A–G but with 3 added concentrated sampling attempts (CSAs, times indicated by vertical arrows). No additional constraints were used during fitting.

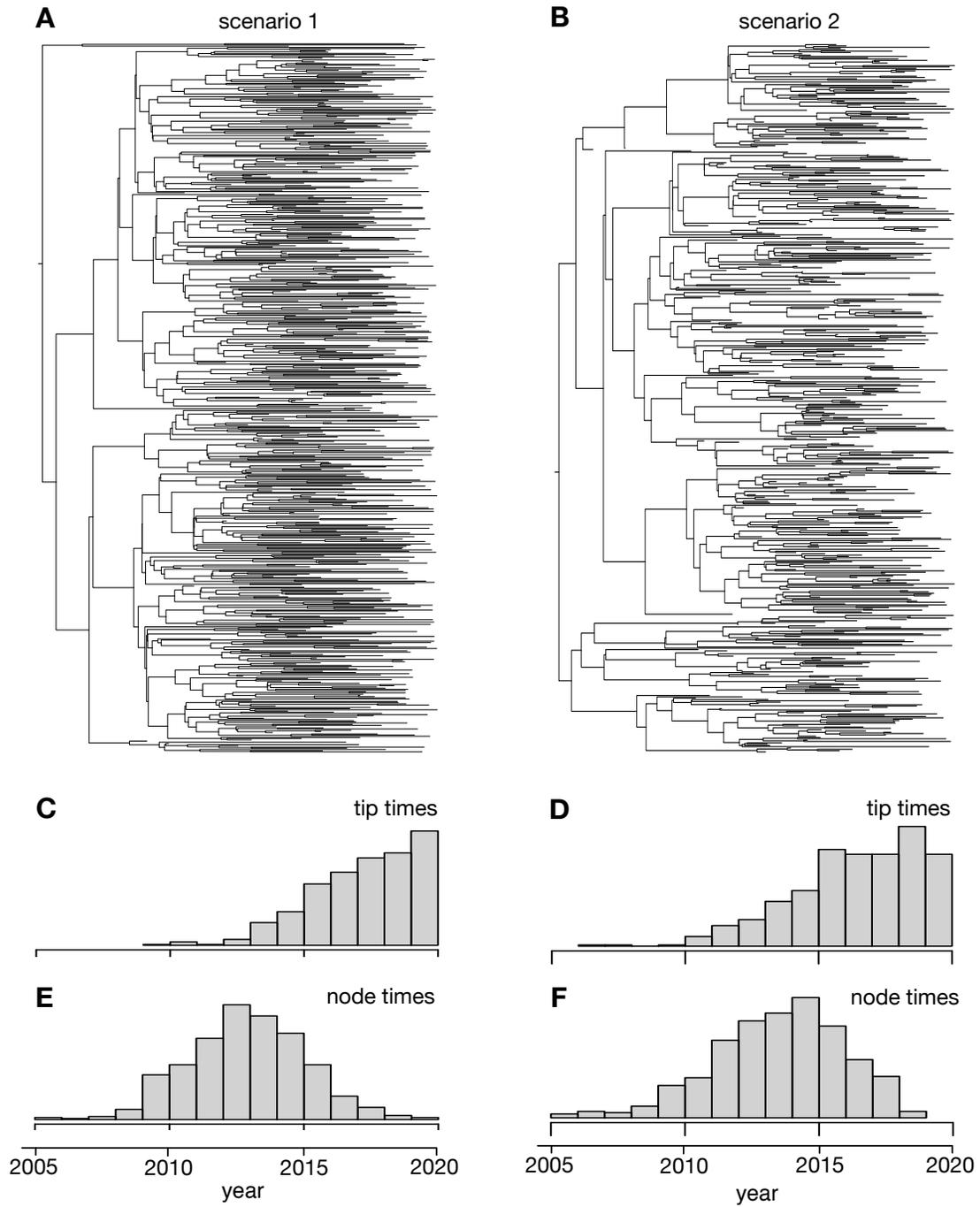


Figure S3: Phylogenetic trees simulated for the BEAST analysis. (A–B) Phylogenetic trees simulated under two hypothetical epidemiological scenarios, for Bayesian inference. (C, E) Histograms of tip ages and node ages in tree A (bar heights are in arbitrary units). (D, F) Histograms of tip ages and node ages in tree B (bar heights are in arbitrary units).

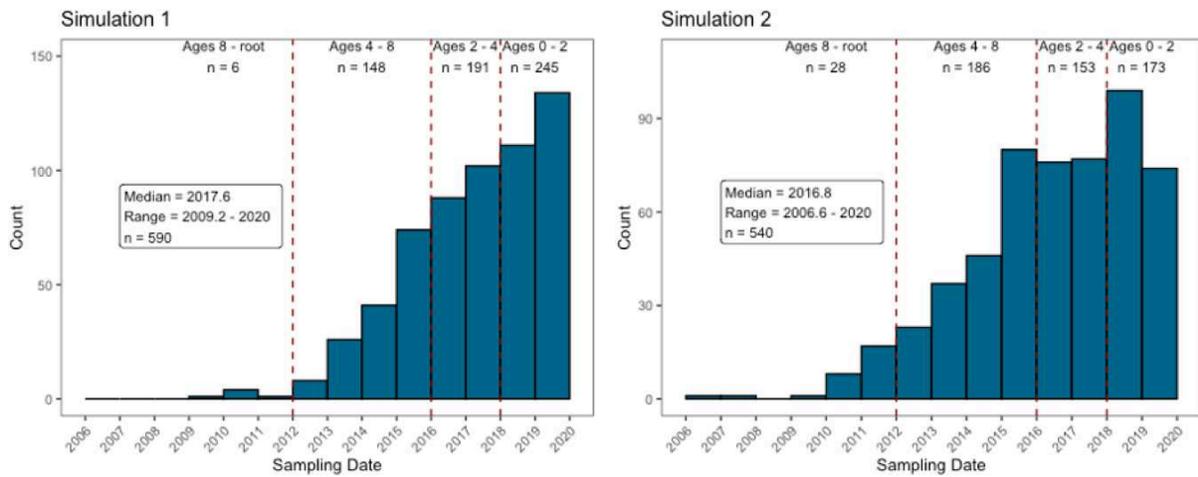


Figure S4: Choosing rate shift times for the BEAST simulation analysis. The distribution of sampling dates in the simulated trees informed the timing of specified rate shifts in fitted skyline models, to partially equalize the number of samples informing each interval and to ensure that the earliest interval had greater than zero samples.

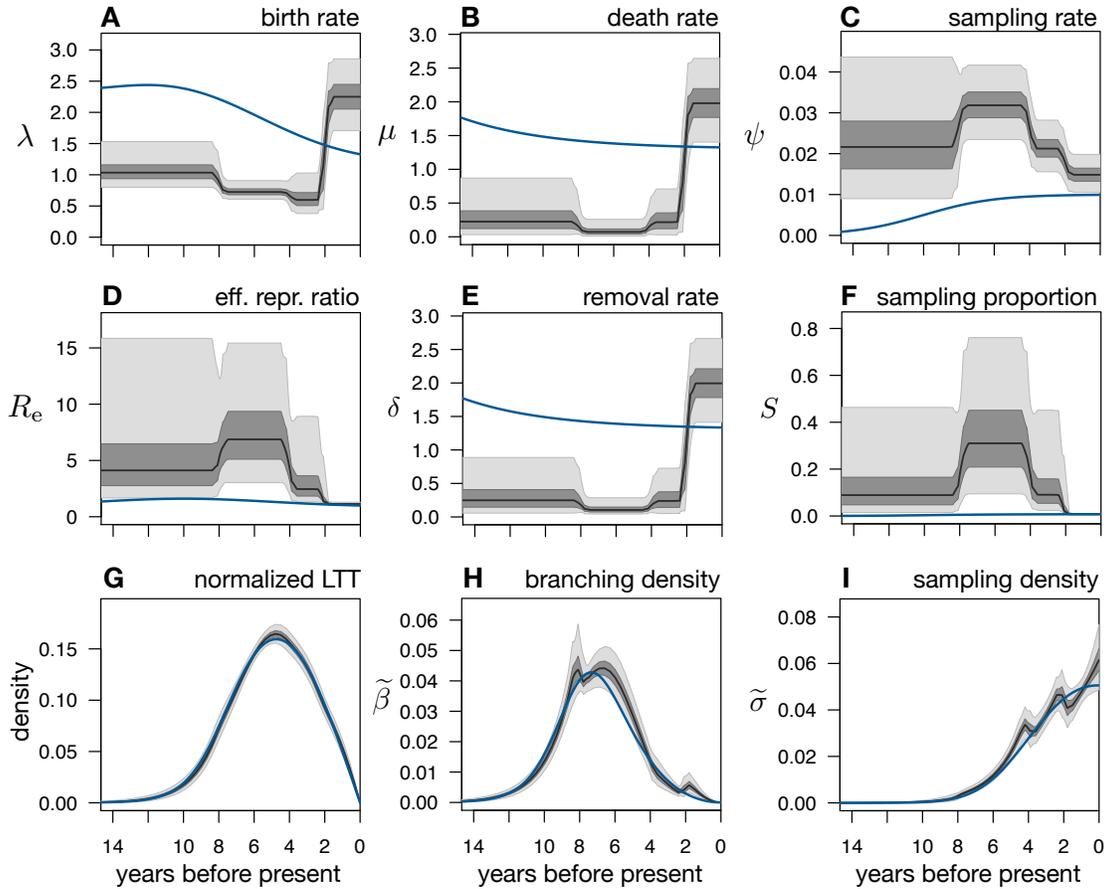


Figure S5: Reconstructing an epidemic's dynamics in a Bayesian framework (BEAST2 run U1). (A–F) Posterior distributions of the birth rate (λ), death rate (μ), sampling rate (ψ), effective reproduction ratio (R_e), removal rate ($\delta = \mu + \psi$) and sampling proportion ($S = \psi / (\mu + \psi)$), as inferred from 590 sequences simulated under a hypothetical birth-death-sampling scenario (blue curves) using BEAST2 (tree in Supplemental Fig. S3A). Black curves show posterior median, dark and light shades represent equal-tailed 50%- and 95%-credible intervals of the posterior. All rates are in yr^{-1} . The present-day sampling proportion was fixed to its true value during fitting, to account for previously reported identifiability issues [2]. Model adequacy was confirmed using predictive posterior simulations with multiple test statistics. Note the poor agreement between the predicted and true profiles. (G–I) Distributions of the deterministic lineages-through-time curves (normalized to unit area under the curve), branching densities ($\tilde{\lambda}$) and sampling densities ($\tilde{\psi}$), corresponding to the same posterior scenarios as in A–F, compared to their true profiled (blue curves). The relatively good agreement between the inferred and true profiles shows that BEAST2 closely reconstructed the epidemiological history's congruence class but not the epidemiological history itself. See Supplemental Fig. S10 for the corresponding posterior distributions of the molecular evolution parameters.

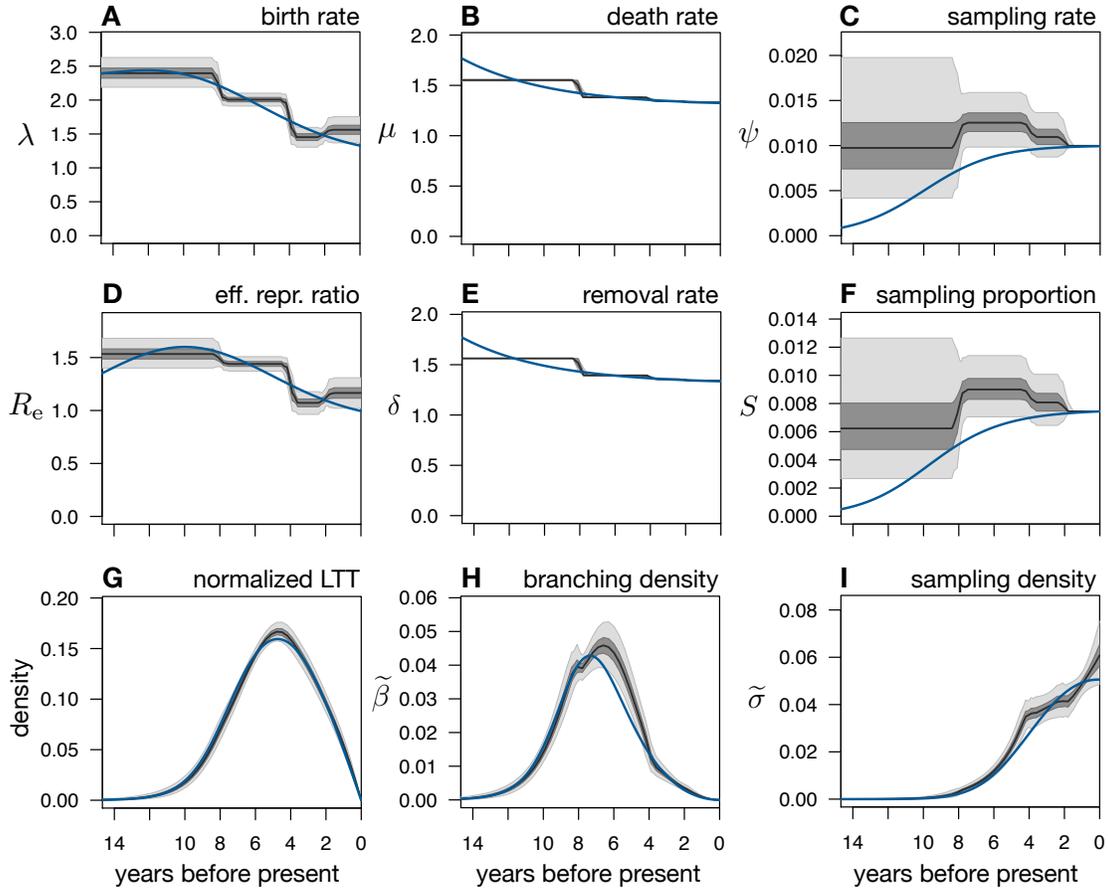


Figure S6: Reconstructing an epidemic's dynamics in a Bayesian framework (BEAST2 run F1). (A–F) Posterior distributions of the birth rate (λ), death rate (μ), sampling rate (ψ), effective reproduction ratio (R_e), removal rate ($\delta = \mu + \psi$) and sampling proportion ($S = \psi / (\mu + \psi)$), as inferred from 590 sequences simulated under a hypothetical birth-death-sampling scenario (blue curves) using BEAST2 (tree in Supplemental Fig. S3A). Black curves show posterior median, dark and light shades represent equal-tailed 50%- and 95%-credible intervals of the posterior. All rates are in yr^{-1} . The present-day sampling proportion as well as the removal rate (all intervals) were fixed to their true values during fitting. Model adequacy was confirmed using predictive posterior simulations with multiple test statistics. Note the much better agreement with the true profiles, compared to the situation where the removal rate was not fixed (Supplemental Fig. S5). (G–I) Distributions of the deterministic lineages-through-time curves (normalized to unit area under the curve), branching densities ($\tilde{\beta}$) and sampling densities ($\tilde{\sigma}$), corresponding to the same posterior scenarios as in A–F, compared to their true profiled (blue curves). See Supplemental Fig. S11 for the corresponding posterior distributions of the molecular evolution parameters.

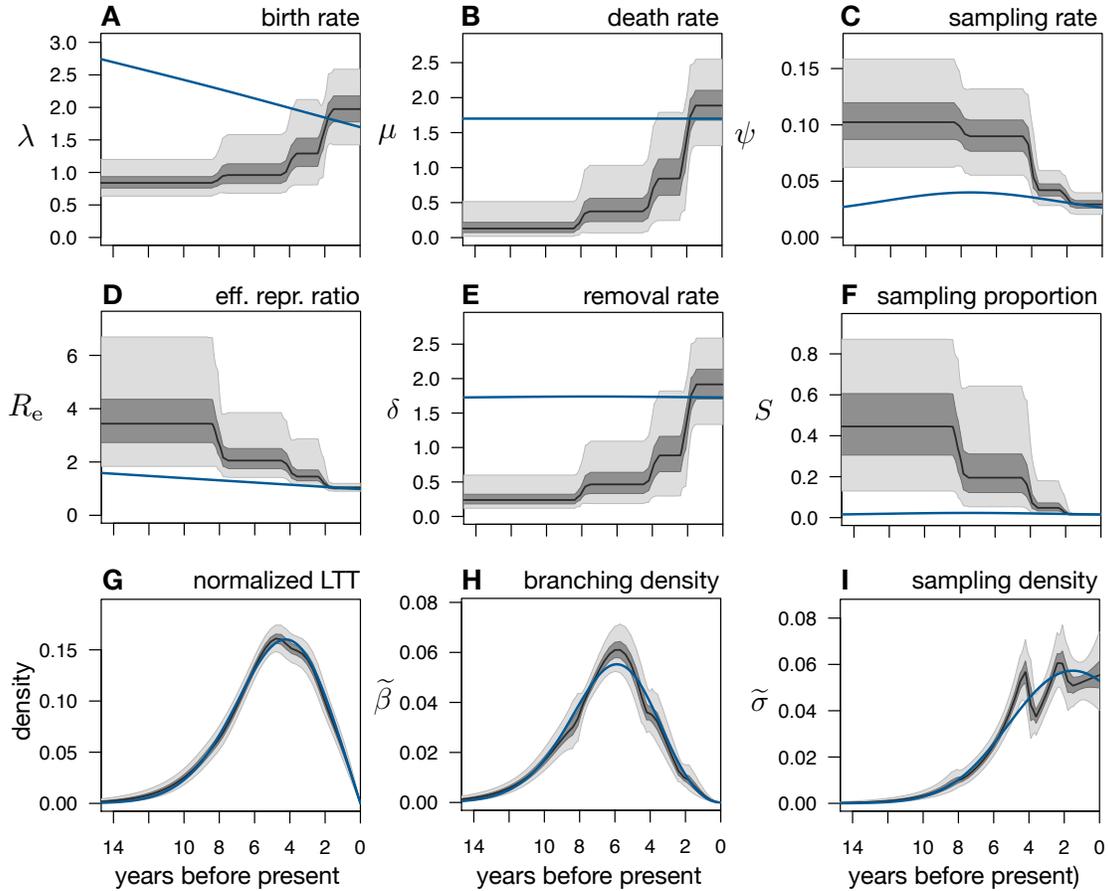


Figure S7: Reconstructing an epidemic's dynamics in a Bayesian framework (BEAST2 run U2). (A–F) Posterior distributions of the birth rate (λ), death rate (μ), sampling rate (ψ), effective reproduction ratio (R_e), removal rate ($\delta = \mu + \psi$) and sampling proportion ($S = \psi / (\mu + \psi)$), as inferred from 540 sequences simulated under a hypothetical birth-death-sampling scenario (blue curves) using BEAST2 (tree in Supplemental Fig. S3B). Black curves show posterior median, dark and light shades represent equal-tailed 50%- and 95%-credible intervals of the posterior. All rates are in yr^{-1} . The present-day sampling proportion was fixed to its true value during fitting, to account for previously reported identifiability issues [2]. Model adequacy was confirmed using predictive posterior simulations with multiple test statistics. Note the poor agreement between the predicted and true profiles. (G–I) Distributions of the deterministic lineages-through-time curves, branching densities ($\tilde{\lambda}$) and sampling densities ($\tilde{\psi}$), corresponding to the same posterior scenarios as in A–F, compared to their true profiled (blue curves). The relatively good agreement between the inferred and true profiles shows that BEAST2 closely reconstructed the epidemiological history's congruence class but not the epidemiological history itself. See Supplemental Fig. S12 for the corresponding posterior distributions of the molecular evolution parameters.

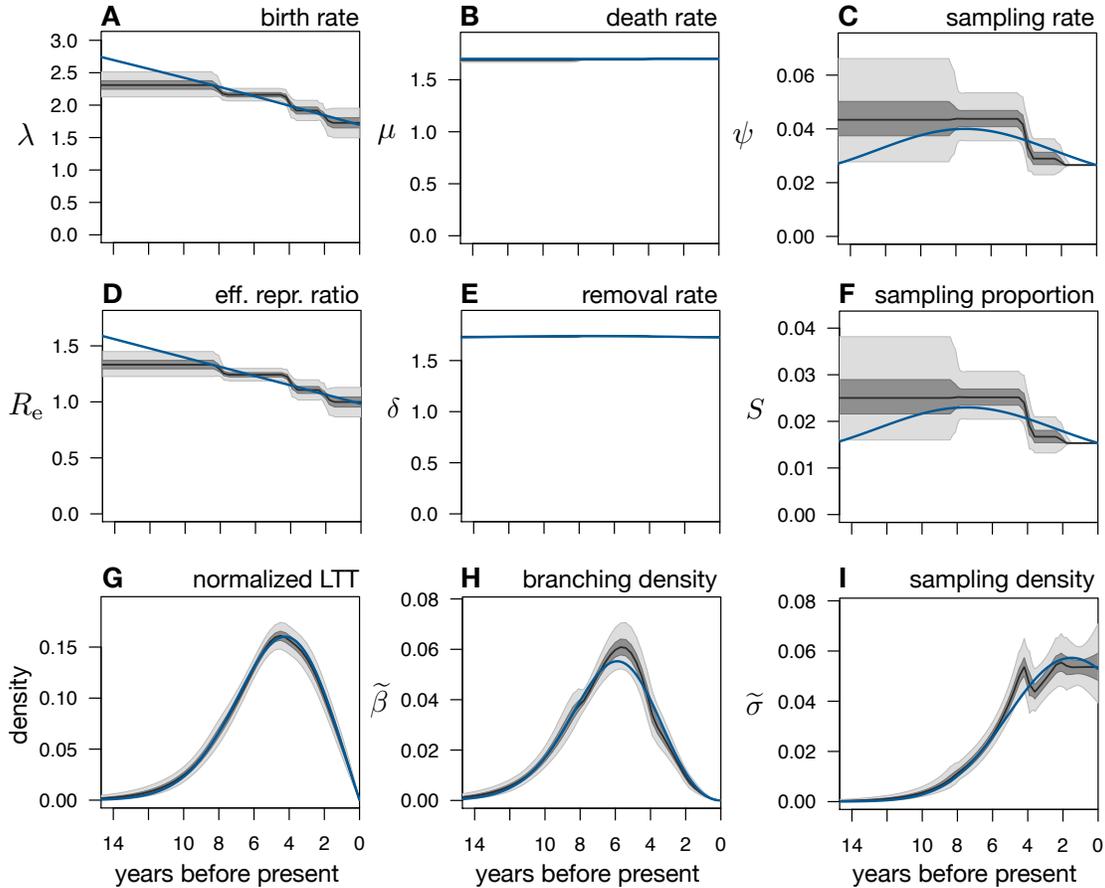


Figure S8: Reconstructing an epidemic's dynamics in a Bayesian framework (BEAST2 run F2). (A–F) Posterior distributions of the birth rate (λ), death rate (μ), sampling rate (ψ), effective reproduction ratio (R_e), removal rate ($\delta = \mu + \psi$) and sampling proportion ($S = \psi / (\mu + \psi)$), as inferred from 540 sequences simulated under a hypothetical birth-death-sampling scenario (blue curves) using BEAST2 (tree in Supplemental Fig. S3B). Black curves show posterior median, dark and light shades represent equal-tailed 50%- and 95%-credible intervals of the posterior. All rates are in yr^{-1} . The present-day sampling proportion and the removal rate (all time intervals) were fixed to their true value during fitting. Model adequacy was confirmed using predictive posterior simulations with multiple test statistics. Note the good agreement between the predicted and true profiles, made possible by fixing one of the model's parameters. (G–I) Distributions of the deterministic lineages-through-time curves (normalized to unit area under the curve), branching densities ($\tilde{\beta}$) and sampling densities ($\tilde{\sigma}$), corresponding to the same posterior scenarios as in A–F, compared to their true profiled (blue curves). See Supplemental Fig. S14 for the corresponding posterior distributions of the molecular evolution parameters.

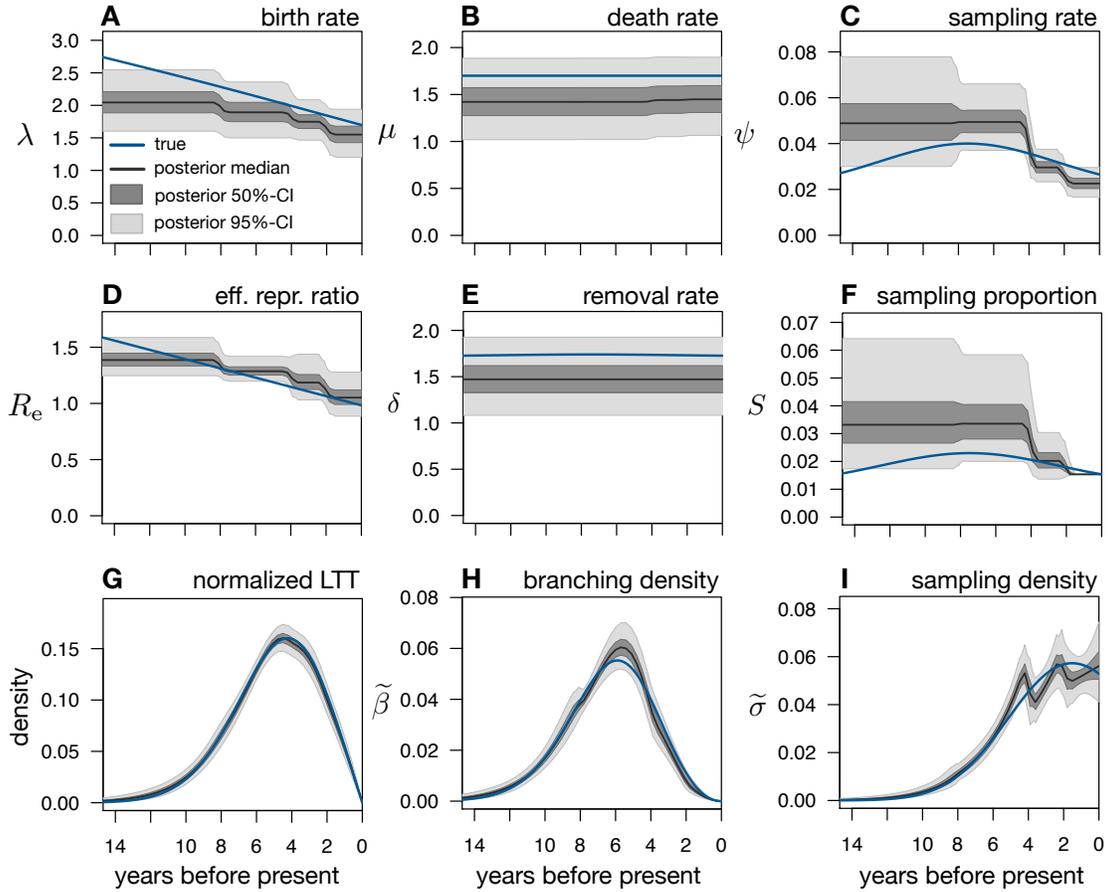


Figure S9: Reconstructing an epidemic's dynamics in a Bayesian framework (BEAST2 run C2). (A–F) Posterior distributions of the birth rate (λ), death rate (μ), sampling rate (ψ), effective reproduction ratio (R_e), removal rate ($\delta = \mu + \psi$) and sampling proportion ($S = \psi / (\mu + \psi)$), as inferred from 540 sequences simulated under a hypothetical birth-death-sampling scenario (blue curves) using BEAST2 (tree in Supplemental Fig. S3B). Black curves show posterior median, dark and light shades represent equal-tailed 50%- and 95%-credible intervals of the posterior. All rates are in yr^{-1} . The removal rate was constrained to be constant over time. The present-day sampling proportion was fixed to its true value during fitting, to account for previously reported identifiability issues [2]. Model adequacy was confirmed using predictive posterior simulations with multiple test statistics. (G–I) Distributions of the deterministic lineages-through-time curves (normalized to unit area under the curve), branching densities ($\tilde{\lambda}$) and sampling densities ($\tilde{\psi}$), corresponding to the same posterior scenarios as in A–F, compared to their true profiled (blue curves). See Supplemental Fig. S13 for the corresponding posterior distributions of the molecular evolution parameters.

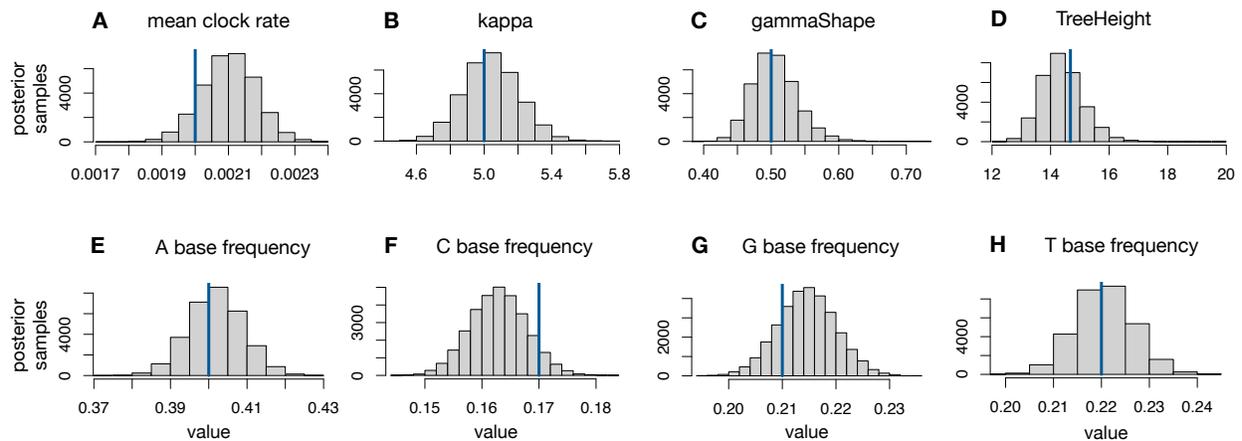


Figure S10: Posterior distributions of molecular evolution parameters (BEAST2 run U1). Posterior distributions of (A) the mean molecular clock rate, (B) the transition/transversion ratio κ , (C) the shape parameter of the discretized gamma distribution of substitution rates, (D) the height of the tree or root age, and (E–H) the stationary frequencies of nucleotide bases A,C,G,T, as inferred from 590 sequences simulated under a hypothetical birth-death-sampling scenario using BEAST2 (epidemiological parameters in Supplemental Fig. S5, tree in Supplemental Fig. S3A). Histogram bars show frequencies. Blue vertical lines show the true values used in the simulation.

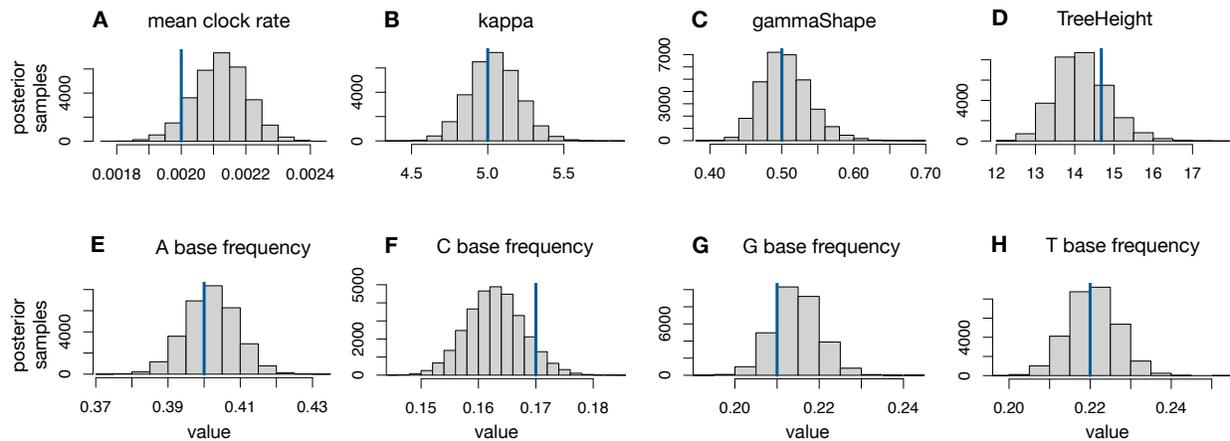


Figure S11: Posterior distributions of molecular evolution parameters (BEAST2 run F1). Posterior distributions of (A) the mean molecular clock rate, (B) the transition/transversion ratio κ , (C) the shape parameter of the discretized gamma distribution of substitution rates, (D) the height of the tree or root age, and (E–H) the stationary frequencies of nucleotide bases A,C,G,T, as inferred from 590 sequences simulated under a hypothetical birth-death-sampling scenario using BEAST2 (epidemiological parameters in Supplemental Fig. S6, tree in Supplemental Fig. S3A). Histogram bars show frequencies. Blue vertical lines show the true values used in the simulation.

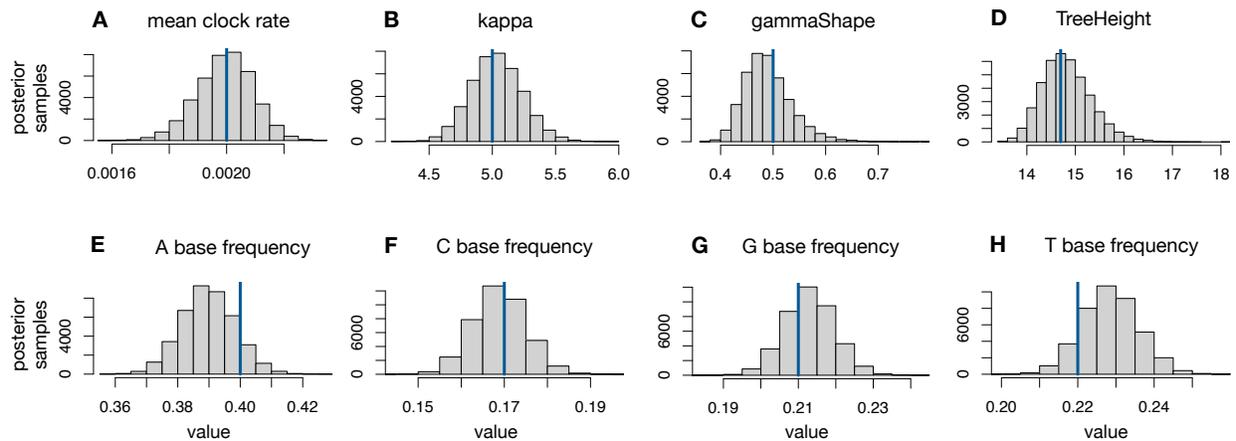


Figure S12: Posterior distributions of molecular evolution parameters (BEAST2 run U2). Posterior distributions of (A) the mean molecular clock rate, (B) the transition/transversion ratio κ , (C) the shape parameter of the discretized gamma distribution of substitution rates, (D) the height of the tree or root age, and (E–H) the stationary frequencies of nucleotide bases A,C,G,T, as inferred from 540 sequences simulated under a hypothetical birth-death-sampling scenario using BEAST2 (epidemiological parameters in Supplemental Fig. S7, tree in Supplemental Fig. S3B). Histogram bars show frequencies. Blue vertical lines show the true values used in the simulation.

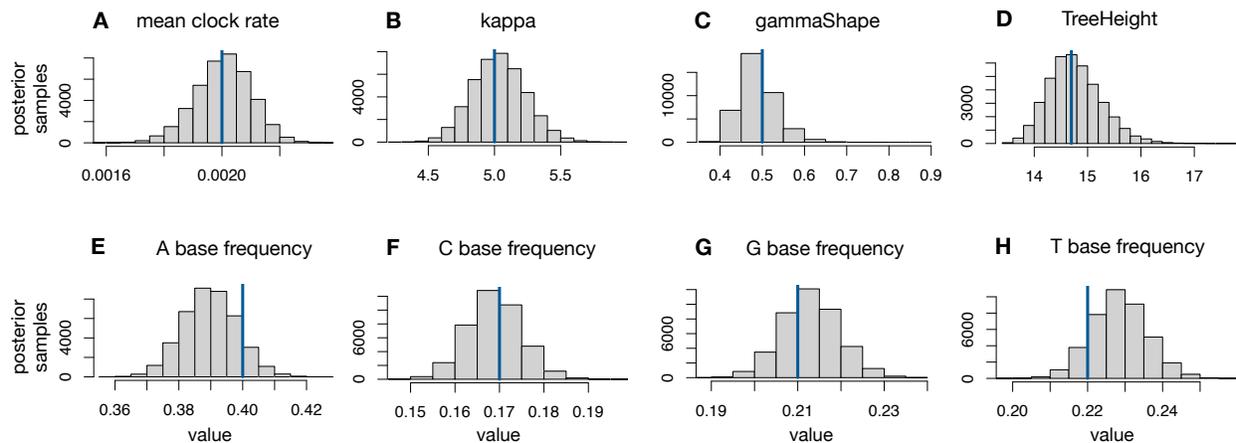


Figure S13: Posterior distributions of molecular evolution parameters (BEAST2 run C2). Posterior distributions of (A) the mean molecular clock rate, (B) the transition/transversion ratio κ , (C) the shape parameter of the discretized gamma distribution of substitution rates, (D) the height of the tree or root age, and (E–H) the stationary frequencies of nucleotide bases A,C,G,T, as inferred from 540 sequences simulated under a hypothetical birth-death-sampling scenario using BEAST2 (epidemiological parameters in Supplemental Fig. S9, tree in Supplemental Fig. S3B). Histogram bars show frequencies. Blue vertical lines show the true values used in the simulation.

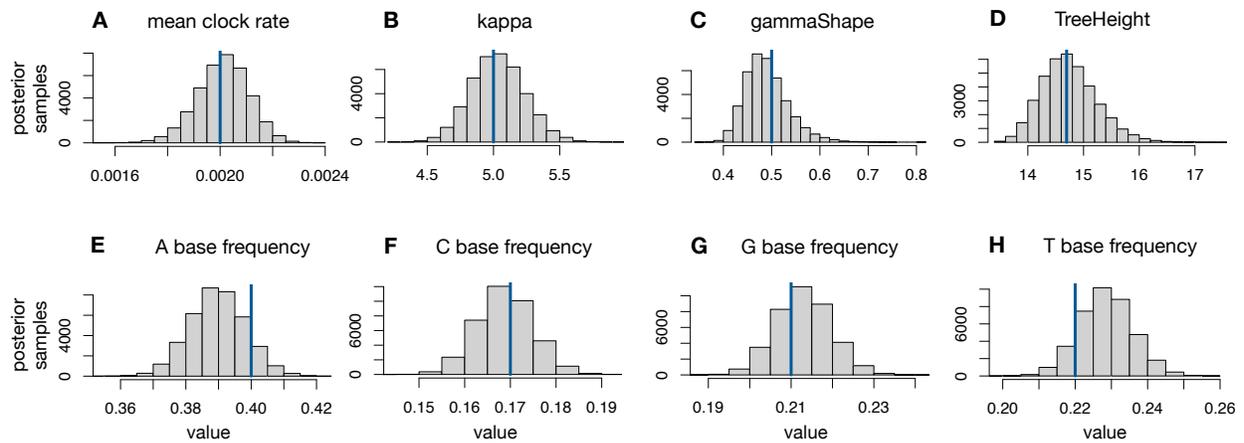


Figure S14: Posterior distributions of molecular evolution parameters (BEAST2 run F2). Posterior distributions of (A) the mean molecular clock rate, (B) the transition/transversion ratio κ , (C) the shape parameter of the discretized gamma distribution of substitution rates, (D) the height of the tree or root age, and (E–H) the stationary frequencies of nucleotide bases A,C,G,T, as inferred from 540 sequences simulated under a hypothetical birth-death-sampling scenario using BEAST2 (epidemiological parameters in Supplemental Fig. S8, tree in Supplemental Fig. S3B). Histogram bars show frequencies. Blue vertical lines show the true values used in the simulation.

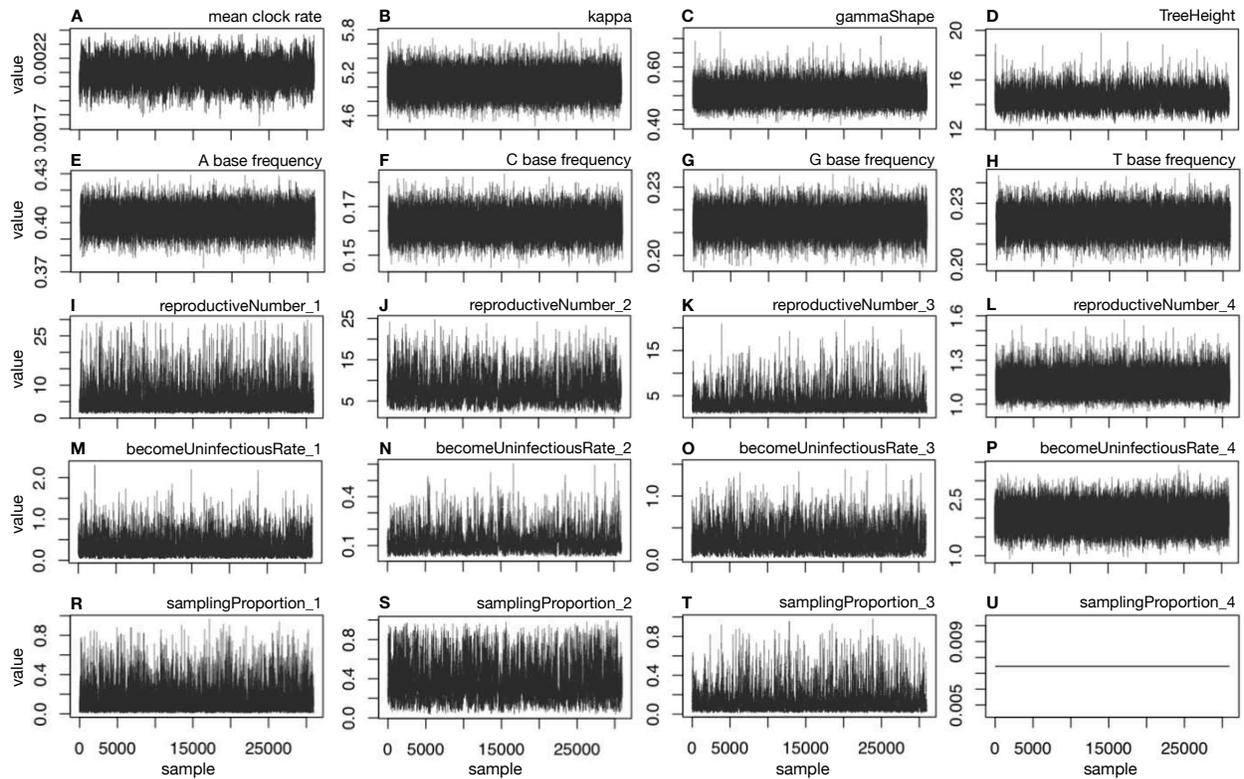


Figure S15: MCMC traces of molecular evolution and epidemiological parameters (BEAST2 run U1). MCMC trace plots of molecular evolution and epidemiological (birth-death-sampling skyline model) parameters generated by BEAST2 (2 independent MCMC chains), based on the sequences simulated under the 1st hypothetical birth-death-sampling scenario (epidemiological parameters in Supplemental Fig. S5, tree in Supplemental Fig. S3A). Samples are shown after burn-in removal and after thinning of each MCMC chain, and after concatenating the two MCMC chains.

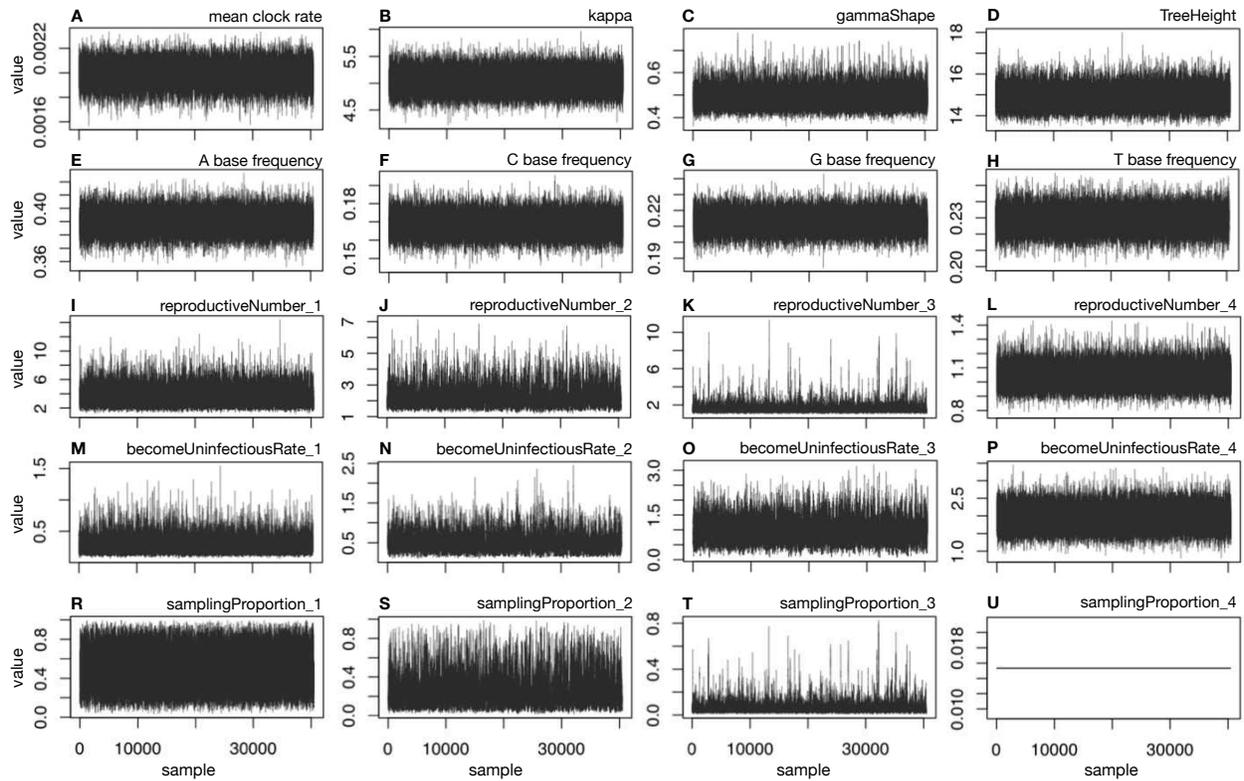


Figure S16: MCMC traces of molecular evolution and epidemiological parameters (BEAST2 run U2). MCMC trace plots of molecular evolution and epidemiological (birth-death-sampling skyline model) parameters generated by BEAST2 (2 independent MCMC chains), based on the sequences simulated under the 2nd hypothetical birth-death-sampling scenario (epidemiological parameters in Supplemental Fig. S7, tree in Supplemental Fig. S3B). Samples are shown after burn-in removal and after thinning of each MCMC chain, and after concatenating the two MCMC chains.

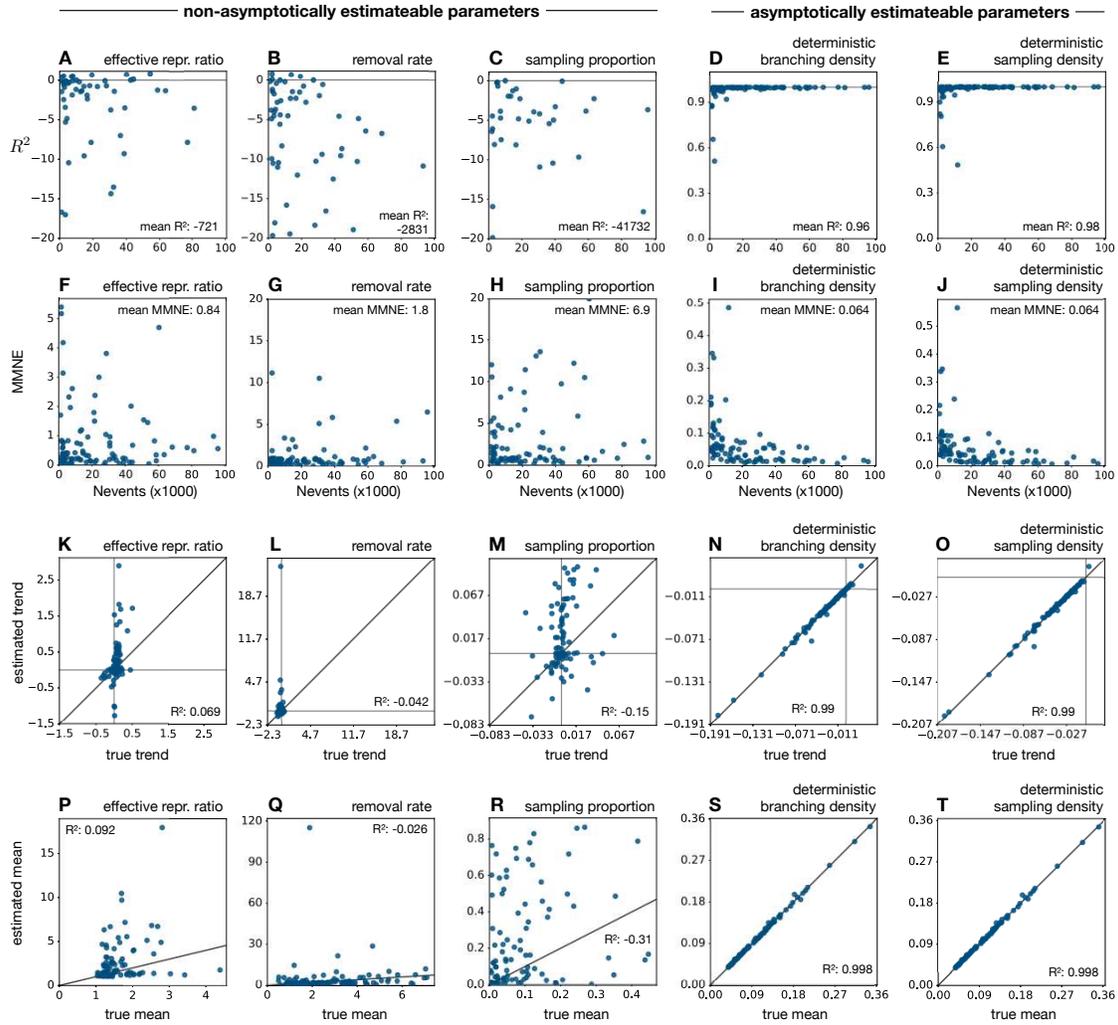


Figure S17: Estimation accuracies using 92 simulated trees (exponential scenarios). Figures summarize the results of fitting BDS models with piecewise-linear profiles for λ , μ and ψ to simulated trees, via maximum likelihood and choosing the grid size based on AIC. In each figure, each point corresponds to a different tree and its best fitted model. Trees were simulated under random epidemiological scenarios in which each λ , μ and ψ had simple exponential forms over time. Each tree comprised between 500 and 50,000 tips and was adequately described by the fitted model (according to Kolmogorov-Smirnov tests for the distribution of node ages, tip ages and edge lengths, $P > 0.05$ in all cases). Note that BDS skyline models were also fitted to the same trees, but almost always achieved a worse AIC than the piecewise linear models. (A) Coefficient of determination (R^2 , vertical axis) for the estimated effective reproduction ratio R_e compared to the true R_e . Cases with R^2 below -20 are not shown. The average R^2 across all 92 trees is written in the figure. The horizontal axis denotes the total number of events in the tree (thousands of tips + branching points). (B-E) Similar to (A) and for the same trees and models, but showing the R^2 of the fitted removal rate (δ), sampling proportion (S), deterministic branching density ($\tilde{\beta}$) and deterministic sampling density ($\tilde{\sigma}$). Note the much higher R^2 for $\tilde{\beta}$ and $\tilde{\sigma}$, consistent with the fact that these are asymptotically estimateable from phylogenetic data. (F-J) Similar to A-E, but showing the mean modulus of the normalized error (MMNE) of the fitted parameters on the vertical axis instead of their R^2 . MMNEs above 20 are not shown. The average MMNE across all 92 trees is written in the figure. Again, note the much lower MMNE for $\tilde{\beta}$ (I) and $\tilde{\sigma}$ (J). (K) Estimated linear trend (vertical axis) of R_e , compared to its true linear trend (horizontal axis), for the same fitted models as in A. Units are 1/year. The diagonal is shown for reference; the R^2 of the diagonal is written in the figure (a value of 1 would mean a perfect agreement between estimated and true trends). Linear trends were calculated from the fitted and true R_e profiles using linear regression with intercept. (L-O) Similar to K, but for other model parameters. (P-T) Similar to K-O, but comparing the true versus estimated mean parameter values. The diagonal is shown for reference; the R^2 of the diagonal is written in the figure. For concrete simulation examples see Supplemental Fig. S18. Methods details in Supplement S.3.

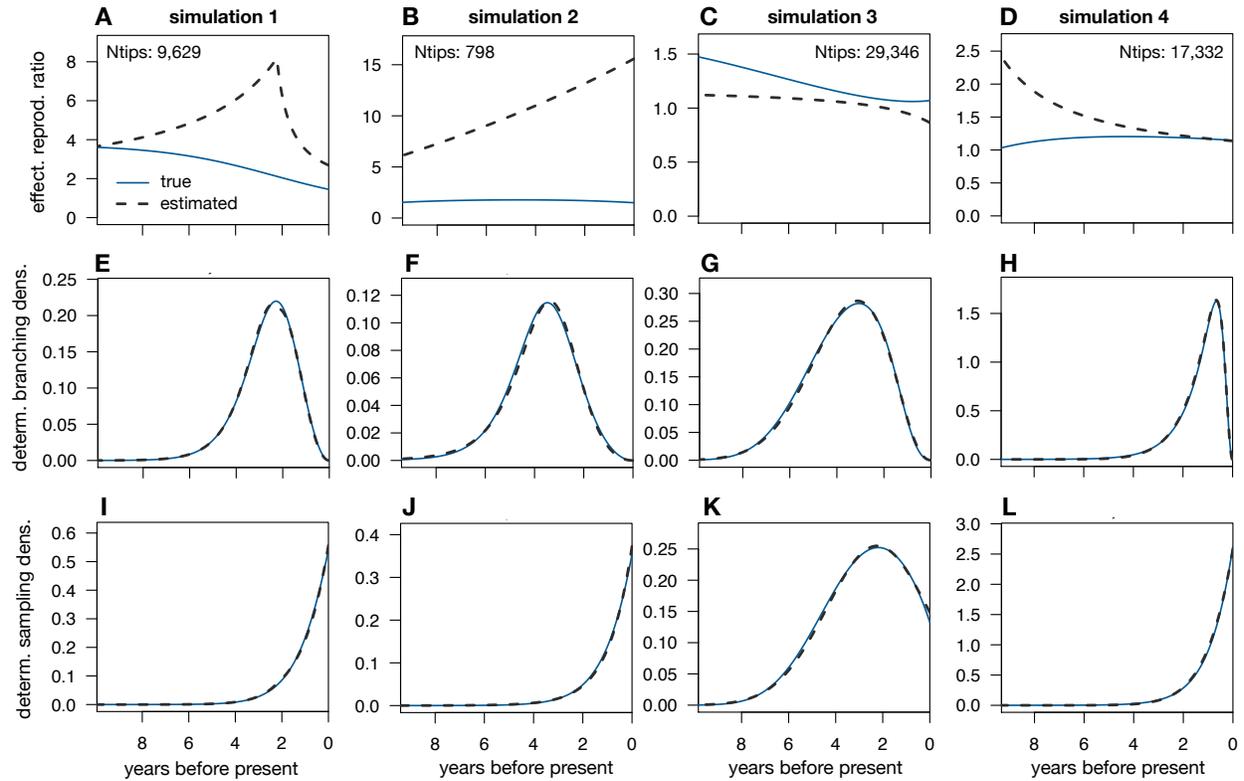


Figure S18: Examples of R_e fitted to simulated trees (4 exponential scenarios). (A) Estimated effective reproduction ratio R_e over time (dashed curve), estimated by fitting a BDS model with piecewise-linear profiles for λ , μ and ψ to a tree generated under a hypothetical epidemiological scenario as in Supplemental Fig. S17. The true R_e is shown for comparison (continuous curve). The number of tips in the tree is written in the figure. (B–D) As in A, but for different simulations. (E–H) Estimated and true deterministic branching densities ($\tilde{\beta}$), corresponding to the model fits in A–D. (I–L) Estimated and true deterministic sampling densities ($\tilde{\sigma}$), corresponding to the model fits in A–D. Note the good agreement between the estimated and true $\tilde{\beta}$ and $\tilde{\sigma}$, and the general bad agreement between the estimated and true R_e , showing that the fitted models accurately inferred the true scenario’s congruence class but not the true scenario itself. The fitted models adequately described the trees (i.e., could not be rejected), based on Kolmogorov-Smirnov tests for the distribution of node ages, tip ages and edge lengths, ($P > 0.05$ in all cases). Methods details in Supplement S.3.

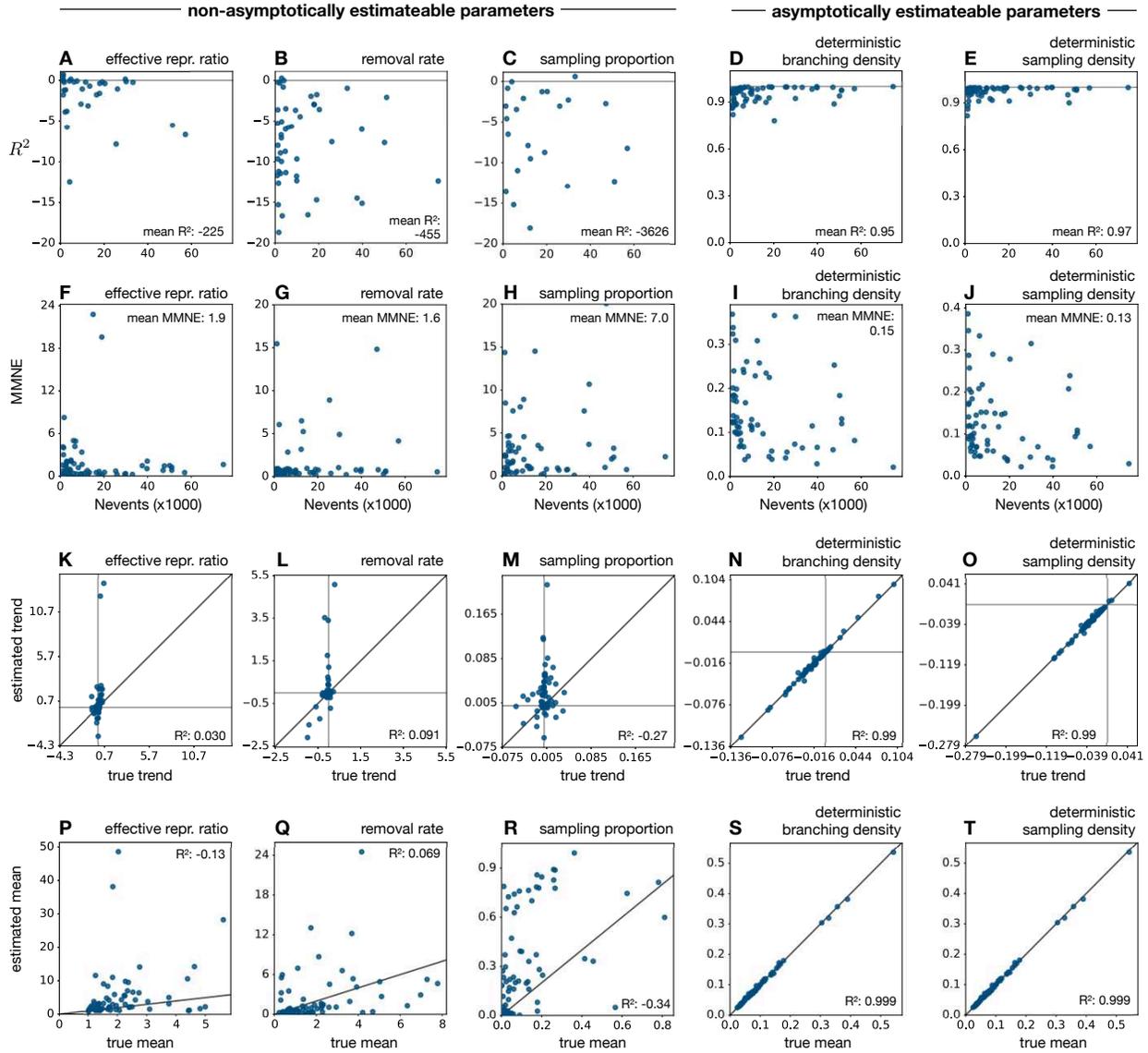


Figure S19: Estimation accuracies using 66 simulated trees (Ornstein-Uhlenbeck scenarios). Panels are as in Supplemental Fig. S17, but for trees simulated under epidemiological scenarios with Ornstein-Uhlenbeck-type profiles of λ , μ and ψ . For concrete simulation examples see Supplemental Fig. S20. Methods details in Supplement S.3.

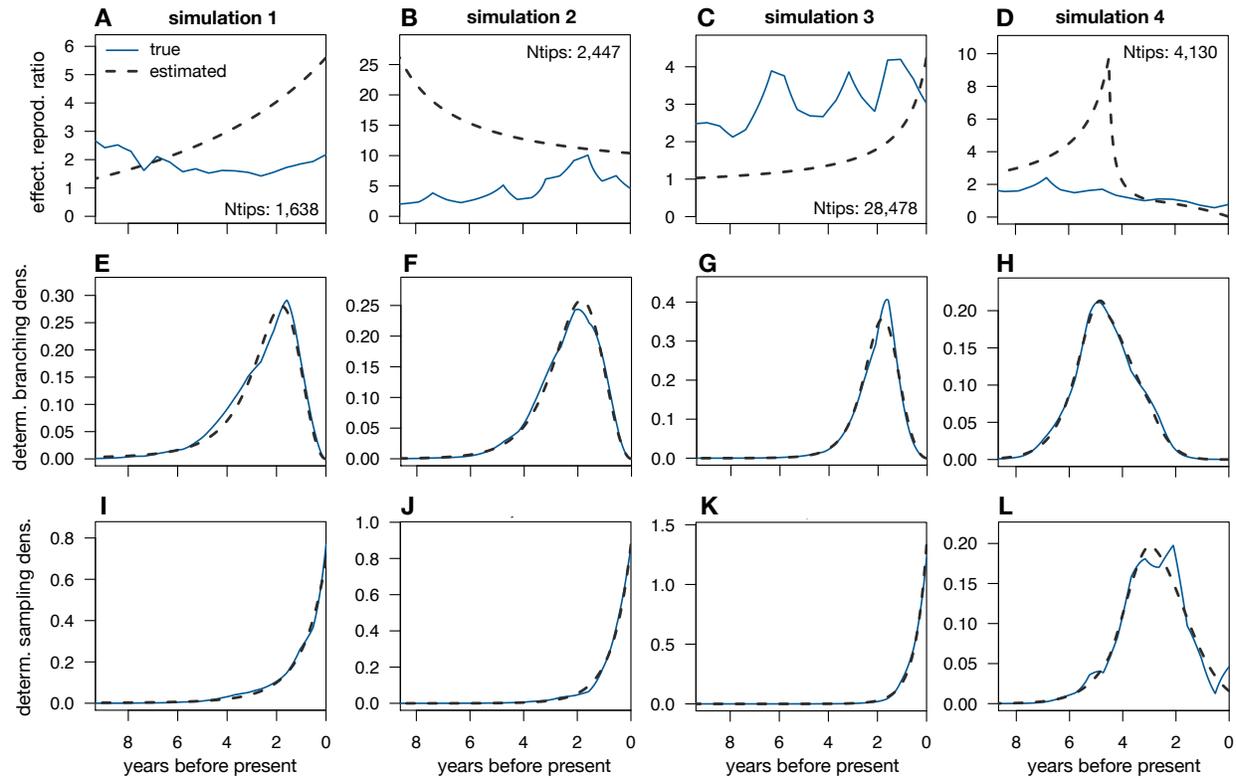


Figure S20: Examples of R_e fitted to simulated trees (4 Ornstein-Uhlenbeck scenarios). (A) Estimated effective reproduction ratio R_e over time (dashed curve), estimated by fitting a BDS model with piecewise-linear profiles for λ , μ and ψ to a tree generated under a hypothetical epidemiological scenario as in Supplemental Fig. S19. The true R_e is shown for comparison (continuous curve). The number of tips in the tree is written in the figure. (B–D) As in A, but for different simulations. (E–H) Estimated and true deterministic branching densities ($\tilde{\beta}$), corresponding to the model fits in A–D. (I–L) Estimated and true deterministic sampling densities ($\tilde{\sigma}$), corresponding to the model fits in A–D. Note the good agreement between the estimated and true $\tilde{\beta}$ and $\tilde{\sigma}$, and the general bad agreement between the estimated and true R_e (including wrong trends and major spurious features), showing that the fitted models accurately inferred the true scenario’s congruence class but not the true scenario itself. The fitted models adequately described the trees (i.e., could not be rejected), based on Kolmogorov-Smirnov tests for the distribution of node ages, tip ages and edge lengths, ($P > 0.05$ in all cases). Methods details in Supplement S.3.

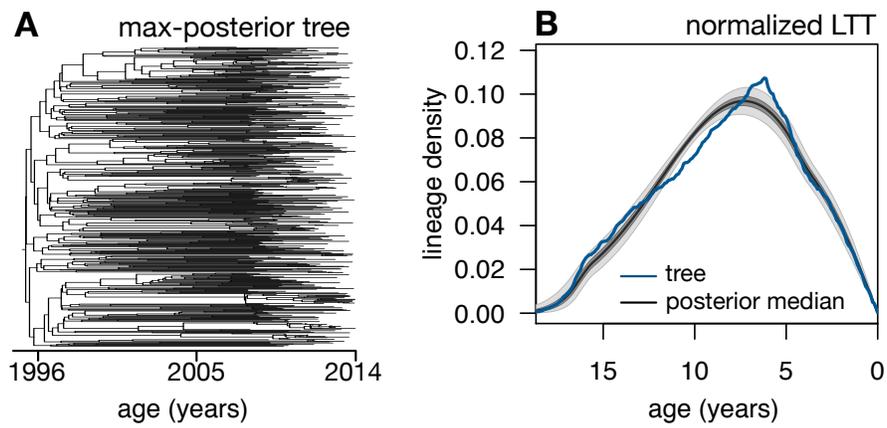


Figure S21: HIV tree and LTT. (A) Maximum-posterior-probability HIV timetree sampled with BEAST2. (B) Lineages-through-time (LTT) curve of the tree normalized to have unit area under-the-curve (blue curve), compared to the normalized deterministic LTTs of the posterior skyline models (black curve shows median, dark and light shades show 50% and 95% equal-tailed credible intervals).

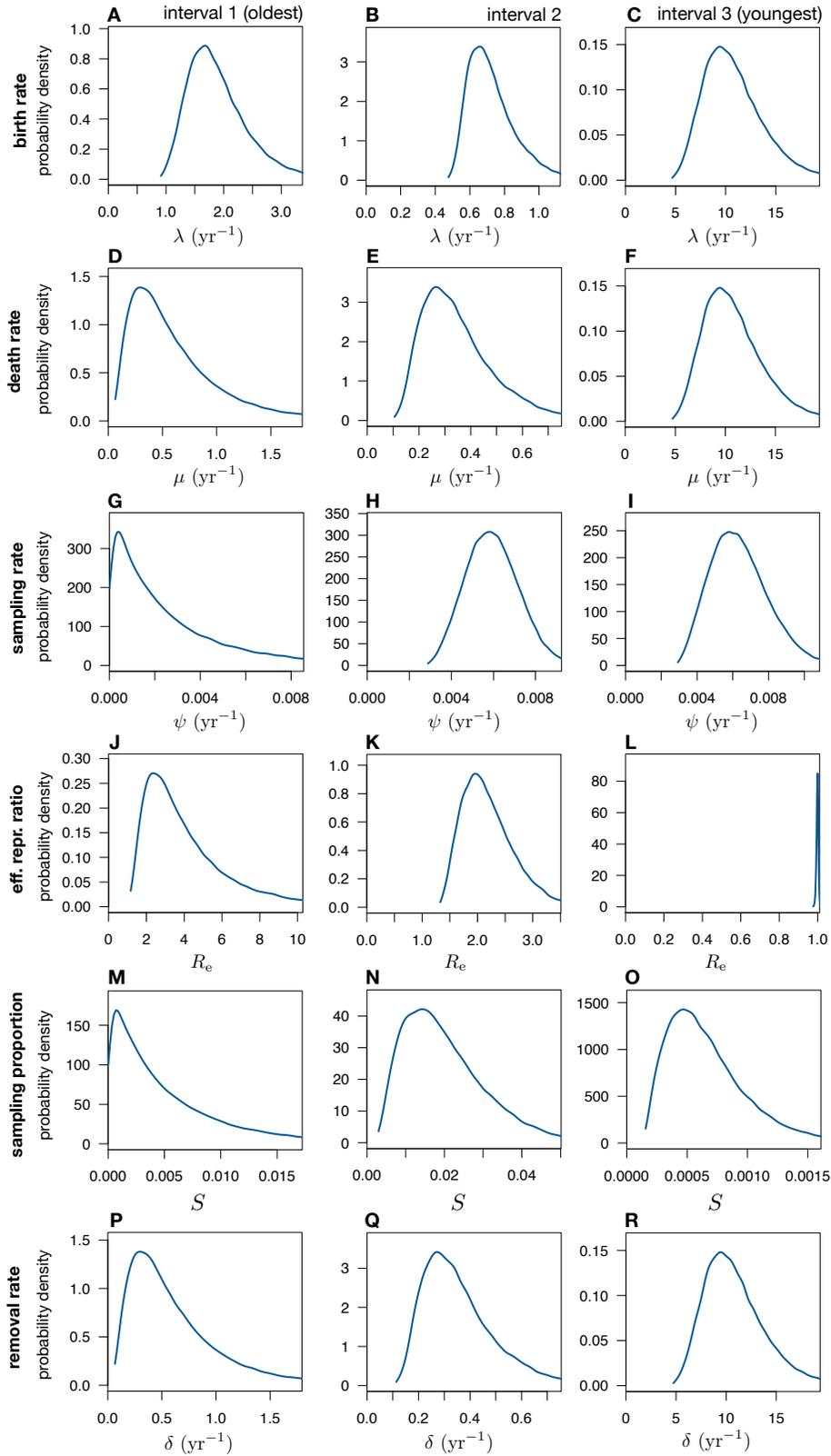


Figure S22: HIV sampled posteriors - probability densities. Probability densities of various epidemiological parameters according to the posterior model distribution sampled by BEAST2, in each of the 3 time intervals. Time intervals cover the periods 0–4, 4–16 and ≥ 16 years before present.

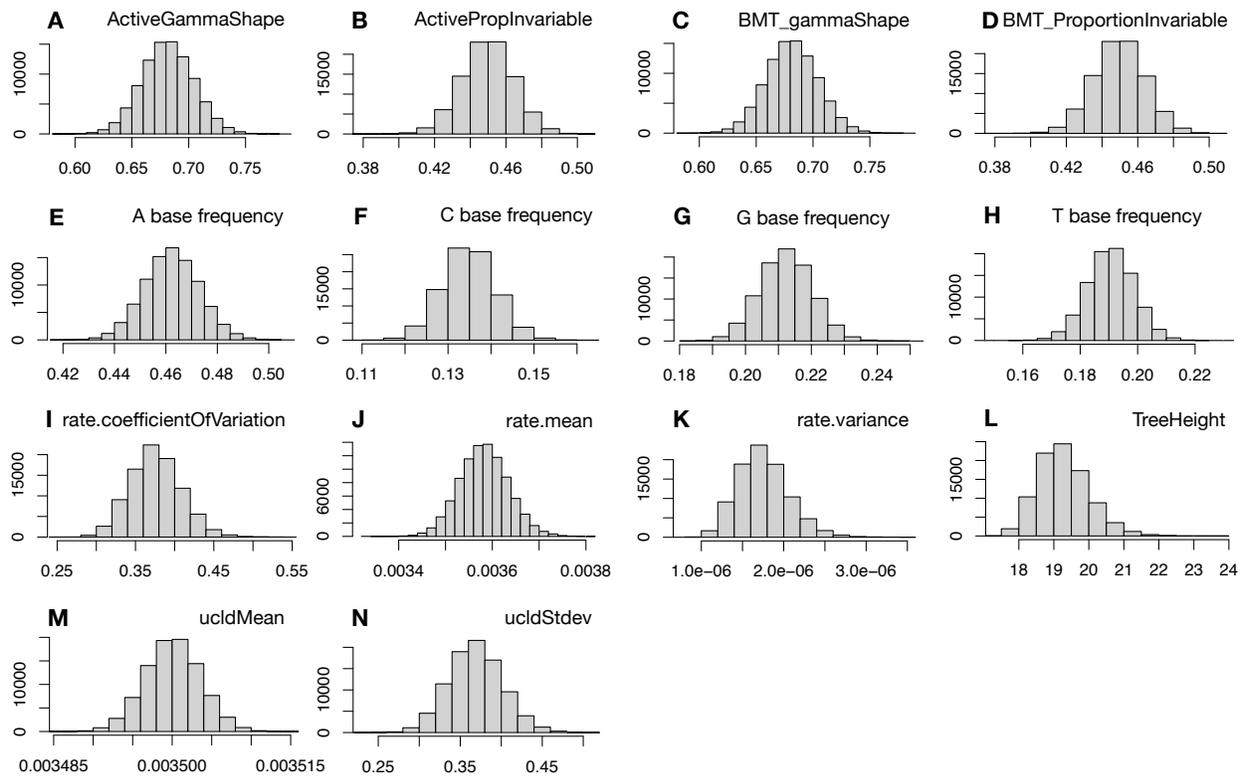


Figure S23: Posterior distributions of molecular evolution parameters (HIV). Histograms of various molecular evolution parameters drawn from the posterior distribution using BEAST2, for the HIV dataset. Vertical bars depict frequencies.

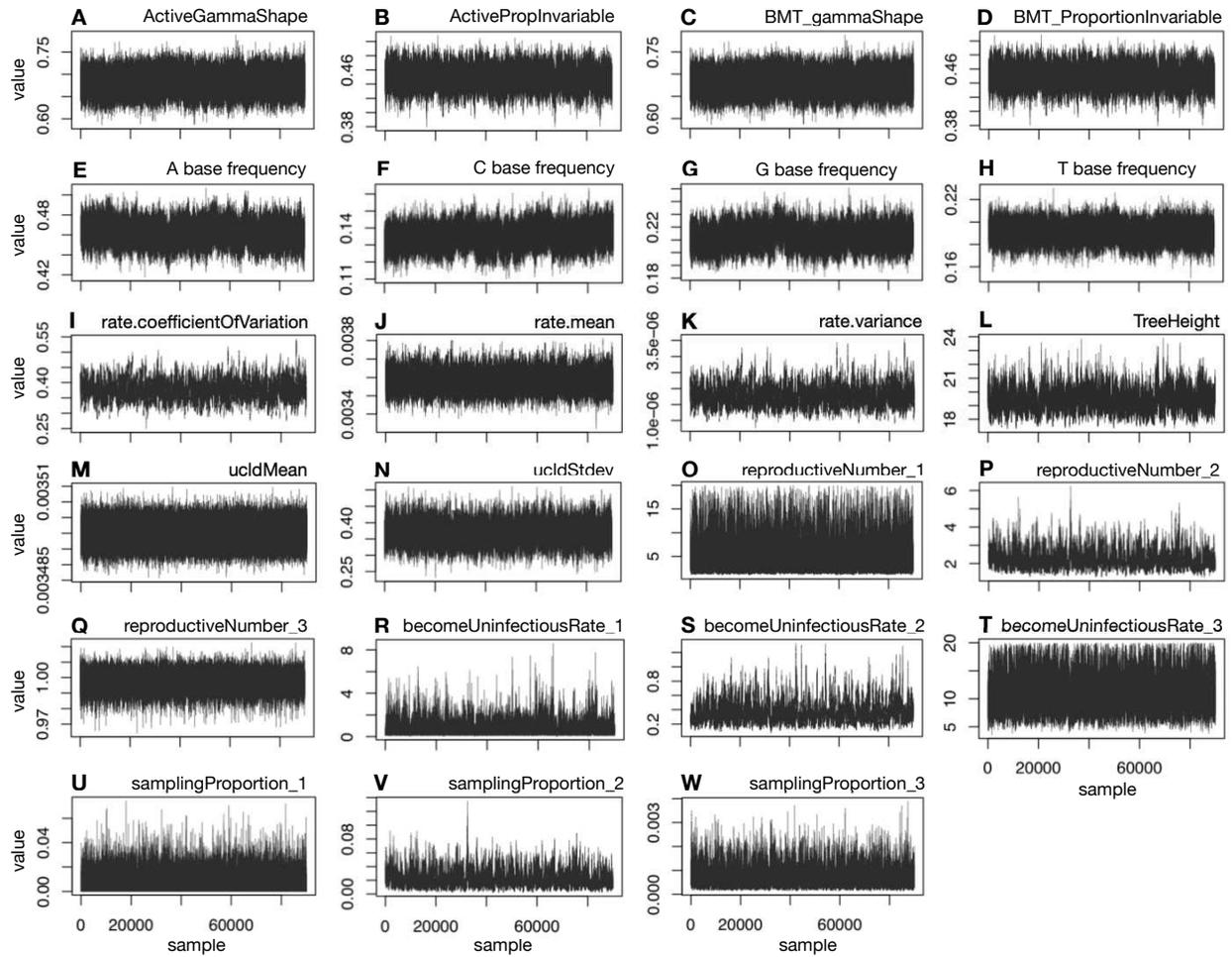


Figure S24: MCMC traces of molecular evolution and epidemiological parameters (HIV). MCMC trace plots of molecular evolution and epidemiological (birth-death-sampling skyline model) parameters generated by BEAST2, for the HIV dataset (2 independent MCMC chains). Samples are shown after burn-in removal and after thinning of each MCMC chain, and after concatenating the two MCMC chains.

Table S1: Priors and other settings specified for the BEAST2 analysis of the simulated HIV timetrees. * denotes that the prior parameters are in linear space (default is log).

Tree model	BD skyline serial
Substitution model	HKY+ I + G
Gamma categories	4
Clock model	Strict molecular clock
Clock rate	LN($\mu=0.003$, $s=1$)*
Reproductive number	LN($\mu=0$, $s=1.25$); 4 intervals
Become uninfected rate	LN($\mu=2$, $s=1.25$)*; 1 or 4 intervals
Sampling proportion	BETA($\alpha=2$, $\beta=2$); 4 intervals
Rate change times	2, 4, 8 years before present
Proportion invariant	BETA($\alpha=2$, $\beta=2$)
Origin	LN($\mu=20$, $s=1.25$)*
Frequency parameter	UNIF(0-1), estimated
Gamma shape	EXP(1, offset=0)
Kappa	LN($\mu=1$, $s=1.25$)
Number of chains	2
Chain length	100–200 million
Sampling frequency	10 000

Table S2: Priors and other settings specified for the BEAST2 analysis of the empirical HIV-1 data from Northern Alberta. * denotes that the prior parameters are in linear space (default is log).

Tree model	BD skyline serial
Substitution model	bModelTest, Dirichlet (1, 1, 1, 1)
Clock model	Uncorrelated log normal relaxed
Clock mean	LN(mu=0.0035, s=0.001)*
Clock std.dev.	Gamma(alpha=0.5396, beta=0.3819)
Reproductive number	LN(mu=0, s=1.25); 3 intervals
Become uninfected rate	LN(mu=0, s=1.25); 3 intervals
Sampling proportion	BETA(alpha=1, beta=100); 3 intervals
Rate change times	4, 16 years before most recent sample (2014)
Proportion invariant	BETA(alpha=1, beta=4)
Origin	LN(mu=25, s=1)*
Frequency parameter	UNIF(0-1), estimated
Gamma shape	EXP(1, offset=0)
Kappa	LN(mu=1, s=1.25)
Number of chains	2
Chain length	500 million
Sampling frequency	10 000

References

- [1] MacPherson, A., Louca, S., McLaughlin, A., Joy, J. B. & Pennell, M. W. A general birth-death-sampling model for epidemiology and macroevolution. *bioRxiv* 2020.10.10.334383 (2020).
- [2] Stadler, T., Kühnert, D., Bonhoeffer, S. & Drummond, A. J. Birth–death skyline plot reveals temporal changes of epidemic spread in HIV and hepatitis C virus (HCV). *Proceedings of the National Academy of Sciences* **110**, 228–233 (2013).
- [3] Louca, S. & Pennell, M. W. A general and efficient algorithm for the likelihood of diversification and discrete-trait evolutionary models. *Systematic Biology* **69**, 545–556 (2020).
- [4] Louca, S. & Doebeli, M. Efficient comparative phylogenetics on large trees. *Bioinformatics* **34**, 1053–1055 (2018).
- [5] Akaike, H. Likelihood of a model and information criteria. *Journal of Econometrics* **16**, 3–14 (1981).
- [6] Schwarz, G. Estimating the dimension of a model. *Annals of Statistics* **6**, 461–464 (1978).
- [7] Engl, H. W., Hanke, M. & Neubauer, A. *Regularization of inverse problems*, vol. 375 (Springer Science & Business Media, 1996).
- [8] Oraintara, S., Karl, W. C., Castanon, D. A. & Nguyen, T. Q. A method for choosing the regularization parameter in generalized tikhonov regularized linear inverse problems. In *Image Processing, 2000. Proceedings. 2000 International Conference on*, vol. 1, 93–96 (IEEE, 2000).
- [9] Calvetti, D., Morigi, S., Reichel, L. & Sgallari, F. Tikhonov regularization and the L-curve for large discrete ill-posed problems. *Journal of Computational and Applied Mathematics* **123**, 423–446 (2000).
- [10] Uhlenbeck, G. E. & Ornstein, L. S. On the theory of the Brownian motion. *Physical Review* **36**, 823–841 (1930).
- [11] Gupta, A., Manceau, M., Vaughan, T., Khammash, M. & Stadler, T. The probability distribution of the reconstructed phylogenetic tree with occurrence data. *Journal of Theoretical Biology* **488**, 110115 (2020).
- [12] Featherstone, L. A., Di Giallonardo, F., Holmes, E. C., Vaughan, T. G. & Duchêne, S. Infectious disease phylodynamics with occurrence data. *bioRxiv* (2020).
- [13] Manceau, M., Gupta, A., Vaughan, T. & Stadler, T. The probability distribution of the ancestral population size conditioned on the reconstructed phylogenetic tree with occurrence data. *Journal of Theoretical Biology* **509**, 110400 (2021).