

UCLA

UCLA Previously Published Works

Title

Scalable gradients enable Hamiltonian Monte Carlo sampling for phylodynamic inference under episodic birth-death-sampling models.

Permalink

<https://escholarship.org/uc/item/0zn3k1nt>

Journal

PLoS Computational Biology, 20(3)

Authors

Shao, Yucai

Magee, Andrew

Suchard, Marc

et al.

Publication Date

2024-03-01

DOI

10.1371/journal.pcbi.1011640

Copyright Information

This work is made available under the terms of a Creative Commons Attribution License, available at <https://creativecommons.org/licenses/by/4.0/>

Peer reviewed

RESEARCH ARTICLE

Scalable gradients enable Hamiltonian Monte Carlo sampling for phylodynamic inference under episodic birth-death-sampling models

Yucui Shao¹, Andrew F. Magee², Tetyana I. Vasylyeva^{3,4}, Marc A. Suchard^{1,2,5*}

1 Department of Biostatistics, University of California, Los Angeles, California, United States of America, **2** Department of Biomathematics, University of California, Los Angeles, California, United States of America, **3** Department of Medicine, University of California San Diego, La Jolla, California, United States of America, **4** Department of Population Health and Disease Prevention, University of California Irvine, Irvine, California, United States of America, **5** Department of Human Genetics, University of California, Los Angeles, California, United States of America

* msuchard@ucla.edu

OPEN ACCESS

Citation: Shao Y, Magee AF, Vasylyeva TI, Suchard MA (2024) Scalable gradients enable Hamiltonian Monte Carlo sampling for phylodynamic inference under episodic birth-death-sampling models. *PLoS Comput Biol* 20(3): e1011640. <https://doi.org/10.1371/journal.pcbi.1011640>

Editor: Joëlle Barido-Sottani, Ecole Normale Supérieure, FRANCE

Received: October 30, 2023

Accepted: March 10, 2024

Published: March 29, 2024

Copyright: © 2024 Shao et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: BEAST XML files for the analyses in this paper are available at https://github.com/suchard-group/hmc_ebds. The approximate gradients have been implemented in the hmc-clock branch of BEAST (<https://github.com/beast-dev/beast-mcmc/tree/hmc-clock>) and the v4.0.0 release of BEAGLE (<https://github.com/beagle-dev/beagle-lib/releases/tag/v4.0.0>).

Funding: This work was supported through National Institutes of Health grants U19 AI135995 (MAS), R01 AI153044 (YC, AFM and MAS) and

Abstract

Birth-death models play a key role in phylodynamic analysis for their interpretation in terms of key epidemiological parameters. In particular, models with piecewise-constant rates varying at different epochs in time, to which we refer as episodic birth-death-sampling (EBDS) models, are valuable for their reflection of changing transmission dynamics over time. A challenge, however, that persists with current time-varying model inference procedures is their lack of computational efficiency. This limitation hinders the full utilization of these models in large-scale phylodynamic analyses, especially when dealing with high-dimensional parameter vectors that exhibit strong correlations. We present here a linear-time algorithm to compute the gradient of the birth-death model sampling density with respect to all time-varying parameters, and we implement this algorithm within a gradient-based Hamiltonian Monte Carlo (HMC) sampler to alleviate the computational burden of conducting inference under a wide variety of structures of, as well as priors for, EBDS processes. We assess this approach using three different real world data examples, including the HIV epidemic in Odesa, Ukraine, seasonal influenza A/H3N2 virus dynamics in New York state, America, and Ebola outbreak in West Africa. HMC sampling exhibits a substantial efficiency boost, delivering a 10- to 200-fold increase in minimum effective sample size per unit-time, in comparison to a Metropolis-Hastings-based approach. Additionally, we show the robustness of our implementation in both allowing for flexible prior choices and in modeling the transmission dynamics of various pathogens by accurately capturing the changing trend of viral effective reproductive number.

Author summary

Epidemic control and forecasting relies on accurate quantification of transmission and recovery dynamics. This quantification is achievable through the analysis of phylogenetic

R01 AI162611 (AFM and MAS). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. We also gratefully acknowledge support from Advanced Micro Devices, Inc. with the donation of parallel computing resources used for this research.

Competing interests: The authors have declared that no competing interests exist.

relationships among pathogen strains obtained from infected individuals. As a key analytical tool for such inference, we concentrate on the study of the episodic birth-death-sampling (EBDS) models. These models define a probability distribution on time-calibrated phylogenies that enable estimation of time-varying rates of pathogens' transmission, recovery, and sampling. Advances in sequencing technology, however, have led to an increasing amount of genetic data collected from these pathogens. Consequently, the traditional computational methods for analyzing these large-scale data under EBDS models have become inadequate due to their high computational load. We aimed to break this computational bottleneck by developing a new approach, based on the Hamiltonian Monte Carlo sampling, that considerably accelerates inference of all rate parameters compared to the traditional random-walk Metropolis-Hasting method. Our method greatly improves the ability to explore the complex distributions that arise when we want to understand disease dynamics at a more granular temporal resolution. This advancement delivers value to public health as it helps rapid data-driven decision-making during outbreaks and enhances our understanding of the spread of infectious diseases.

1 Introduction

Phylogenetic models represent a suite of powerful tools designed to unravel the interactions between epidemiological and evolutionary processes. These models offer significant insights into the pathogens' dynamics and play a crucial role in advancing outbreak response efforts for diseases such as SARS-CoV-2, Ebola, Dengue, and HIV [1–5]. In this paper, our primary emphasis is directed toward the inference of epidemiological dynamics, rather than estimation of the underlying phylogeny through sequence analysis. Specifically, we start with a sample of molecular sequences, which can be used to reconstruct the evolutionary relationships between organisms, often viral pathogens, and yield inference on dynamics of the larger pathogen population over time while relegating the phylogeny the status of a nuisance parameter. To provide this link, a vital component of phylogenetic analysis is the use of birth-death models, which belong to an important subclass of continuous-time Markov chains (CTMCs). We use birth-death models to define the probability distribution on time-calibrated phylogenies to reflect the fluctuations of the population size [6]. In this context, birth-death models posit three major types of events: birth, which refers to the creation of new lineages through pathogen transmission between hosts; death, which represents host death/recovery or other removal from the studied population, and sampling, which means the collection of a sequence derived from the pathogen in a single infected host and included in the data set under analysis [7].

The past few decades have delivered a wide range of birth-death models. These span from a simple, constant-over-time formulation [8] to models that allow both birth and death rates to vary over time [9, 10]. Further extensions incorporate additional processes, both statistical and biological, such as the collection of samples in continuous time [11], migration [12], or the dependency of rates of birth and death on key biological traits [13–15]. One powerful variant, the episodic birth-death-sampling (EBDS) model [9, 16–18] permits birth, death, and sampling rates to change in discrete epochs throughout time to capture more complicated population dynamics. Recent inference based on EBDS models has found its way already into many applications, especially on the understanding of the spread of infectious disease [4, 19, 20].

With increasingly rich and complex molecular sequence datasets across fields, improving the scalability of inference under EBDS models remains challenging both in terms of the number of sequences and the number of epochs. The most commonly employed inference methods based on Markov chain Monte Carlo (MCMC) [21, 22] use random-walk transition kernels generally to propose new parameter values in a blind fashion. Consequently, they lead to many birth-death model likelihood evaluations and slow exploration across the state space, especially for high-dimensional problems. The potentially complex correlation structure between epoch parameters can further exacerbate inference. This is where gradient-based sampling methods, such as Hamiltonian Monte Carlo (HMC) [23, 24], are expected to shine. HMC has recently become very popular as a MCMC algorithm that overcomes many of the limitations of random-walk Metropolis-Hasting (MH) methods. Instead of making random proposals, HMC exploits the gradient of the log posterior with respect to (wrt) its model parameters to propose new states that are likely to be accepted and are far from the current state. Since HMC can make large moves in the state space while still maintaining a high acceptance rate, it can lead to faster convergence and better mixing than MH approaches, if one can efficiently evaluate not only the log posterior (up to a constant) but also its gradient. Successful implementation of HMC transition kernels has proved fruitful in terms of boosting sampling performance in other phylogenetic inference frameworks, including for different clock models (which describe how rates of molecular evolution vary among different organisms over time [25, 26]), divergence times (the internal-node heights of phylogenies [27]) and non-parametric coalescent models (which fall into another category of phylodynamic models assuming effective population size as a piecewise-constant form of time [28]).

In this paper, we incorporate gradient-based sampling methods into phylodynamic analysis based on EBDS models, thereby enabling scalable inference within this framework. First, we refactor the EBDS (log) likelihood to show explicitly that the computational complexity scales linearly both in terms of the number of sequences and the number of epochs. With this refactoring in hand, we deliver a novel linear-time algorithm to evaluate the gradient of this (log-) likelihood wrt all epoch parameters simultaneously. Then we design and deploy an efficient HMC sampler that enables us to fit a large class of EBDS models in a Bayesian framework and provide an open-source implementation in the popular Bayesian Evolutionary Analysis by Sampling Trees (BEAST) software [29].

Current approaches to Bayesian inference for EBDS epoch parameters employ a variety of prior assumptions to model the dependence structure between parameters across epochs. Some priors assume that birth, death and sampling rates across epochs are independent and identically distributed (iid) [4, 9, 17]. To smooth rate variation over time, temporally-auto-correlated priors such as Ornstein-Uhlenbeck smoothing prior [18], Gaussian Markov random fields (GMRF) priors [30, 31] and the horseshoe Markov random field for EBDS models [32] have been considered. Conveniently, both our linear-time gradients and our HMC approach generalize across all of these choices of prior without the need to construct model-specific sampling techniques and allow us to introduce the Bayesian bridge shrinkage prior to yield parsimonious time-varying rate patterns.

Across three real-world infectious disease examples that vary in the number of sequences, model dimension, and prior specification, we demonstrate the performance gain achieved by our implementation of an HMC transition kernel compared to random walk transition kernels. Moreover, for each of these datasets we infer key epidemiological parameters and demonstrate the utility of our scalable approach for providing reasonable estimates of pathogen transmission dynamics over time.

2 Methods

2.1 Setup

In an infectious disease setting, suppose an infected individual initiates an epidemic at time (measured backwards from the present day) $t_{or} > 0$, called the time of the origin. Then, each currently and newly infected individual disseminates the pathogen to others at a time-varying birth rate $\lambda(t)$ and transitions into a noninfectious state at a time-varying death rate $\mu(t)$. At any given time, we may sample an infected individual with time-varying sampling rate $\psi(t)$, at which point we add the time of sampling and a molecular sequence of their infectious agent into our time-stamped molecular sequence alignment \mathbf{Y} . Further, we may posit K fixed time-points at which we randomly sample all infected individuals with associated vector of probabilities $\boldsymbol{\rho} = (\rho_1, \dots, \rho_K)$, adding the time and molecular sequence to \mathbf{Y} . Note that this means that several individuals can be sampled at the same time point. The choice of the time-points is dependent on the dataset at hand and will be discussed later in this section. Every sampled infected individual may be treated and then become noninfectious with time-varying probability $r(t)$ which we assume equal to one everywhere for complete sampling.

The model defined above provides a forward in time portrayal of the epidemiological process. Considering the N sampled and time-stamped sequences in \mathbf{Y} as tree tips, there exists a (possibly unknown) phylogeny \mathcal{T} that depicts the evolutionary relationships among these sequences. Specifically, \mathcal{T} is a rooted, bifurcating tree with N tip nodes that correspond to the sampled sequences or their hosts from the population and $N - 1$ internal nodes that represent transmission events between hosts. We define the height of the nodes as the length of time between the time of the corresponding transmission/sampling events and the time of the most recent sampled sequence, which we refer to the present time, 0. Each node of \mathcal{T} is then associated with a node-height ≥ 0 relative to the present, such that the difference between the parent node-height and its child node-height is a branch length measured in the units of real time (e.g., years). We call the earliest internal node in \mathcal{T} the root and its node-height corresponds to the time of the most recent common ancestor (TMRCA). Therefore, we can further define the node heights of internal nodes to be bifurcation times and that of leave nodes to be sampling times. Accordingly, for a vector of bifurcation times, we have $\mathbf{v} = (v_1, v_2, \dots, v_{N-1})$ where $v_1 < \dots < v_{N-1}$. And we let $\mathbf{u} = (u_1, u_2, \dots, u_N)$ be a vector of serial sampling times where $u_1 < \dots < u_N$.

For an episodic model, we make the assumption that all the rate parameters are piece-wise constant across K different epochs with cut points $\mathbf{t} = (t_0, \dots, t_K)$, with $t_0 = 0 < t_1 < \dots < t_{K-1} < t_K$. We also require $t_{or} \leq t_K$. Under this assumption, we can rewrite the time dependent birth rate $\lambda(t)$ in terms of some unknown epoch-specific birth rate $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_K)$, where $\lambda(t) = \lambda_k$ for $t_{k-1} < t \leq t_k$. Similar parametrization applies to other parameters, so that we can express $\mu(t)$ in terms of $\boldsymbol{\mu} = (\mu_1, \dots, \mu_K)$, $\psi(t)$ in terms of $\boldsymbol{\psi} = (\psi_1, \dots, \psi_K)$ and $r(t)$ in terms of $\mathbf{r} = (r_1, \dots, r_K)$. Without loss of generality, we let intensive sampling events happen at every time point in \mathbf{t} . Then we define $\boldsymbol{\rho} = (\rho_1, \dots, \rho_K)$, where $\rho(t) = \rho_k$ for $t = t_{k-1}$. We can remove these intensive sampling events at the epoch switching times from our model simply by setting $\boldsymbol{\rho} = \mathbf{0}$.

After reparametrizing the rates of the EBDS model, we can arrive at some key epidemiological quantities. For example, if we assume there are no intensive sampling events, we can specify the effective reproductive number as $R_e(t) = \frac{\lambda(t)}{\mu(t) + \psi(t)r(t)}$. Other parameters that are important include the total rate of becoming noninfectious, which is defined as $\delta(t) = \mu(t) + \psi(t)r(t)$, and the sampling proportion, defined as $\zeta(t) = \frac{\psi(t)r(t)}{\mu(t) + \psi(t)r(t)}$. If we also assume removal of lineages

upon sampling, these formulas can be further simplified by letting $r(t)$ be constant and always equal to 1.

2.2 Probability density of a sampled phylogeny

Recall we break time into intervals with cut points $t = (t_0, \dots, t_K)$ defined by epochs. Within each epoch, we define a series of subintervals such that a new subinterval start at every bifurcation time v , sampling time u and epoch switching time t (see Fig 1). We delineate the subinterval by indices j , which begins at s_j and terminates at s_{j+1} (where $s_j < s_{j+1}$). If $t_{or} = t_K$, then the grids $s = (s_1, \dots, s_{2N+K-1})$ can be obtained by joining the time points in v , u and t according to their ascending order when none of these times coincide with each other. If $t_{or} < t_K$, we have $s_{2N+K-1} = t_{or}$ instead of t_K .

Consequently, each subinterval, inclusive on the left, is partitioned in such a way that it precludes the occurrence of an epoch switching, birth or sampling event within its boundaries. Within the k -th epoch, the first subinterval starts at $s_j = t_{k-1}$ and the last subinterval ends at $s_{m_k+1} = t_k$. Note that for the last epoch K , the last subinterval ends at t_{or} . We assign $L(j)$ to account for the number of lineages in \mathcal{T} that are extant in subinterval time $(s_j, s_{j+1}]$.

Our likelihood derivation falls into the common framework with Stadler *et al.* (2013), Gavryushkina *et al.* (2014) and Magee and Höhna (2021) [9, 17, 33]. However, instead of writing the likelihood in terms of the times of node and epochs, we write it in terms of the subintervals j . This representation highlights the fact that the likelihood can be computed in one pass,

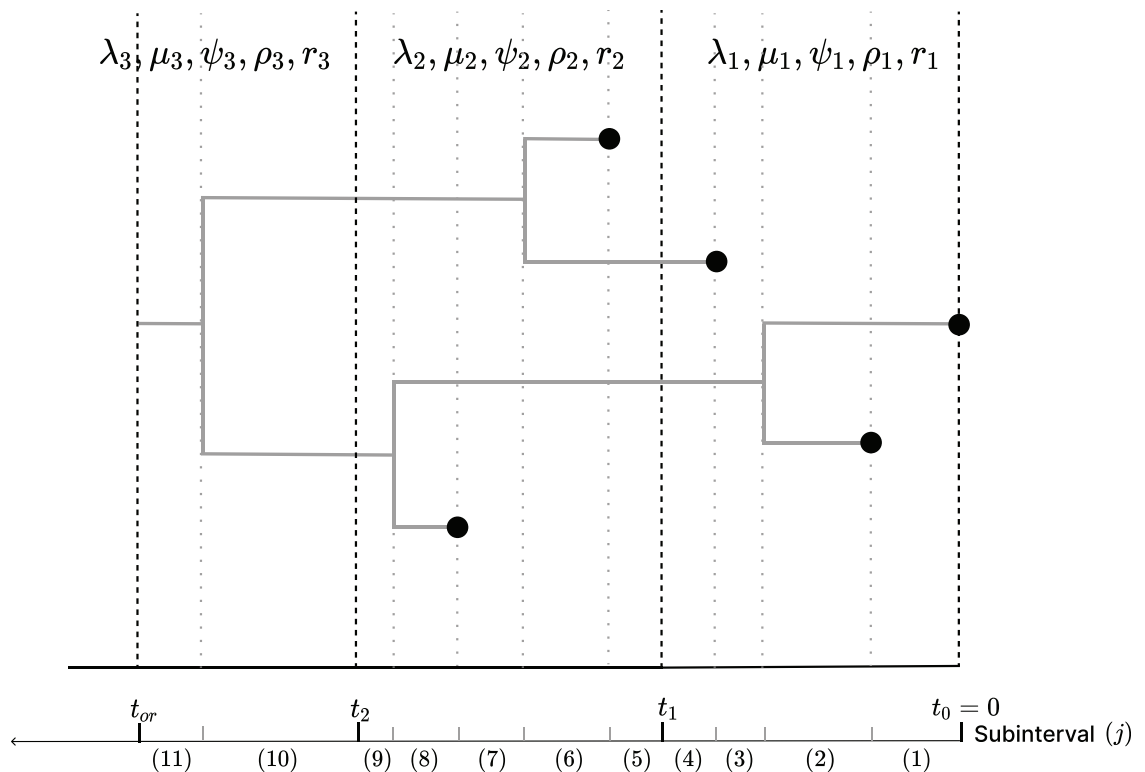


Fig 1. A phylogeny arising from an EBDs model. This sampled phylogeny has three epochs (with epoch switching time t_1, t_2) and thus three sets of model parameters including rates and probabilities. For every epoch, each branch is further divided into subinterval that starts at s_j and ends at time s_{j+1} so that no epoch switching, birth or sampling event occurs within it. Each subinterval within each epoch k is represented by a phylogeny segment index, j .

<https://doi.org/10.1371/journal.pcbi.1011640.g001>

starting at the present and ending at the origin. The interval-based representation of the log-likelihood is as follows:

$$\log \mathbb{P}[\mathcal{T} \mid \lambda, \mu, \psi, \rho, r, t] = N_1 \log \rho_1 + \sum_{k=1}^K \sum_{j=1}^{m_k} \left(\log I_k(E_j) + L(j) \log \left(\frac{q_k(s_{j+1})}{q_k(s_j)} \right) \right), \quad (1)$$

$\sum_{k=1}^K m_k \leq 2N + K - 2^*$

where m_k is the total number of subintervals in epoch k . (*: equality holds when no events happens at the exact same time except for the current).

The indicator function $I_k(E_j)$ is labelled by the index k . This implies that the function is concerned with events occurring within the time frame $(t_{k-1}, t_k]$. We have E_j represent the event that takes place at the termination of subinterval j within epoch k . In most phylodynamic studies, ancestral sampling scenarios are not taken into account; therefore, our model is based on the assumption of a strictly bifurcating phylogenetic tree and does not involve considerations of ancestral sampling cases, which is distinctive from the work of Gavryushkina *et al.* (2014) [17]. Nonetheless, incorporating ancestral sampling into our framework is relatively straightforward. This can be achieved by setting the treatment probability to be less than 1 and adding the term $\psi_k(1 - r_k)$ to our indicator function to account for events involving ancestral samples. Consequently, this indicator function takes the following form:

$$I_k(E_j) = \begin{cases} 1, & E_j = \text{a epoch switching event happens on } s_{j+1} \\ \lambda_k, & E_j = \text{a birth event happens at } s_{j+1} \\ \psi_k((1 - r_k)p_k(s_{j+1}) + r_k), & E_j = \text{a tip sampling event happens at } s_{j+1} \\ \rho_k^{N_k}((1 - r_k)p_{k-1}(s_{j+1}) + r_k)^{N_k} \cdot (1 - \rho_k)^{L(j) - N_k}, & E_j = \text{an intensive sampling event happens at } s_{j+1} = t_{k-1}. \end{cases} \quad (2)$$

Note that $p_k(t)$ is the probability that an infected individual at time t has no sampled descendants when the process is stopped (i.e., at time t_0), and $q_k(t)$ is the probability density of an individual at time t giving rise to an edge between t and t_{k-1} (not t_k since we define time to flow backwards which is the reverse of the generative process) for $t_{k-1} < t < t_k$ in epoch k . We have $p_0(t_0 = 0) = 1$.

The intensive sampling probability at time t_{k-1} is ρ_k and the corresponding number of leaves sampled at that time is N_k . The index here is intentionally misaligned to reconcile the fact that we model the epoch as left inclusive in time.

The definitions of the underlying functions, $q_k(t)$ and $p_k(t)$, follow the work from Stadler *et al.* (2013) and the detailed formulas are included in S1 Text [9]. Note that our Eq 1 does not condition the tree likelihood upon any particular properties, such as the presence of at least one sampled individual. Without loss of generality, additional conditioning schemes can be integrated by adding a factor to the log-likelihood; relevant discussions on this subject are available in Table S3 from the study by MacPherson *et al.* (2022) [6].

As stated previously, our representation of the likelihood differs from the more standard nodewise representation (see for example [9, 17, 33, 34]). Our representation makes it explicit that the likelihood computation can be accomplished in $\mathcal{O}(N + K)$ time (see S1 Text for computational details). We also demonstrate this behavior empirically in S1 Text. On the other hand, as we show in S1 Text section 6, the conventional nodewise representation leads to ambiguities in the cost and a wide potentially range of computational complexities depending on implementation decisions. In S1 Text section 7 we show empirically that formulations based on the nodewise representation include both implementations which are of the same

computational order as ours (namely BEAST2 [35] and RevBayes [36]) and which scale worse in the number of epochs (TreePar [9]).

2.3 Inference

In a Bayesian inference procedure, as introduced in Section 2.1, we use a multiple sequence alignment with the sampling times, the time-stamped sequences, \mathbf{Y} , as the input data. Based on \mathbf{Y} , we can form the posterior distribution over the product space of trees and EBDS model parameters as follows. First, a phylogeny \mathcal{T} is generated from the EBDS process defined in Section 2. Then we specify a molecular clock model that controls the rate at which evolution occurs on each branch of \mathcal{T} . Under a molecular character-based CTMC substitution model, the columns in the sequence alignment evolve independently along the branches of the tree. Adoption of different substitution models is contingent upon the distinct attributes of the dataset under investigation (see Section 2.6.1). For the sake of notational convenience, we refer to the vector encompassing both substitution and clock model parameters as $\boldsymbol{\omega}$. We denote by $\mathbb{P}(\mathbf{Y} | \boldsymbol{\omega}, \mathcal{T})$ the probability of the time-stamped sequences under the CTMC substitution model, known as the phylogenetic likelihood. Subsequently, we can factorize the posterior in the following manner:

$$\begin{aligned} \mathbb{P}[\mathcal{T}, \boldsymbol{\lambda}, \boldsymbol{\mu}, \boldsymbol{\psi}, \boldsymbol{\rho}, \boldsymbol{r}, \boldsymbol{t}, \boldsymbol{\omega} | \mathbf{Y}] &\propto \mathbb{P}(\mathbf{Y} | \boldsymbol{\omega}, \mathcal{T}) \mathbb{P}[\mathcal{T} | \boldsymbol{\lambda}, \boldsymbol{\mu}, \boldsymbol{\psi}, \boldsymbol{\rho}, \boldsymbol{r}, \boldsymbol{t}] \\ &\quad \times \mathbb{P}[\boldsymbol{\lambda}, \boldsymbol{\mu}, \boldsymbol{\psi}, \boldsymbol{\rho}, \boldsymbol{r}, \boldsymbol{t}, \boldsymbol{\omega}] \\ &\propto \mathbb{P}(\mathbf{Y} | \boldsymbol{\omega}, \mathcal{T}) \mathbb{P}(\boldsymbol{\omega}) \mathbb{P}[\mathcal{T} | \boldsymbol{\lambda}, \boldsymbol{\mu}, \boldsymbol{\psi}, \boldsymbol{\rho}, \boldsymbol{r}, \boldsymbol{t}] \\ &\quad \times \mathbb{P}(\boldsymbol{\lambda}) \mathbb{P}(\boldsymbol{\mu}) \mathbb{P}(\boldsymbol{\psi}) \mathbb{P}(\boldsymbol{\rho}) \mathbb{P}(\boldsymbol{r}) \mathbb{P}(\boldsymbol{t}). \end{aligned} \tag{3}$$

In phylodynamic analyses, it is sometimes advantageous to streamline the model by maintaining the death rate as constant. We can also presume the intensive sampling probability to be 0 and treatment probability to uniformly be 1 across all epochs. In handling time-varying parameters, we choose either iid priors or Markov random field models based on dataset-dependent assumptions pertaining to the patterns of change expected in rate parameters. In this paper, we specifically consider the GMRF and the Bayesian bridge Markov random field model, the latter of which we describe below.

With increasing complexity of the existing EBDS models, we seek to integrate Bayesian regularization methods to help manage the potentially vast quantity of model parameters. Specifically, we consider Markov random field priors which specify distributions on the incremental difference between the log-transformed rate parameters. By assigning a normal distribution to the incremental changes, we arrive at the GMRF priors that induce a smoothing effect on the change of rate parameters across contiguous epochs. This approach naturally leads to adjacent epochs exhibiting similar rate values. However, a strong data signal indicative of a rate change can still manifest in the resulting trajectory. By placing a heavy-tailed Bayesian bridge prior [37] on these, we achieve a more generalized extension of the GMRF model. The key distinction resides in the specification of the standard deviation arising from the normal priors on the increments. In this resulting Bayesian bridge Markov random field framework, each epoch's increment is assigned an additional variable to account for variation, thereby affording greater flexibility to the model.

Supposing we have varied birth rates, we define the birth rate on the log scale $\lambda_k^* = \log(\lambda_k)$. Then we have the prior on increments, $\mathbb{P}(\lambda_k^* - \lambda_{k-1}^* | \tau) \propto \exp\left\{-\left|\frac{\lambda_k^* - \lambda_{k-1}^*}{\tau}\right|^\alpha\right\}$ for $k > 1$, where τ is the global scale parameter that controls the overall degree of parameter variation. As α

diminishes, the function $\mathbb{P}(\lambda_k^* - \lambda_{k-1}^*)$ accrues an increased density close to zero. For the purpose of our study, we establish $\alpha = 0.25$ to address a potent prior assumption that $\lambda_k^* - \lambda_{k-1}^*$ is approximately 0 without inducing any problems related to mixing issues. In other words, we do not anticipate substantial fluctuations in the birth rates across consecutive epochs (but allow for rapid rate shift, for example during the exponential growth phase.) Another important parameter is the local scale, denoted as v_k , which is specific to an individual increment $\lambda_k^* - \lambda_{k-1}^*$. Its density regulates the magnitude of the spike and the tail behavior of the above marginal $\lambda_k^* - \lambda_{k-1}^* \mid \tau$.

Note that the GMRF model can be perceived as a specific instance of the Bayesian bridge MRF, where all the local scale parameters are equalized to 1 and α is fixed at 2. In this case, the increment differences adhere to a normal distribution whose variance is solely governed by a single global scale parameter.

To complete our model, a normal prior is assigned to λ_1^* in adherence with the method outlined in Magee *et al.* (2020) [32]. We obtain the mean parameter of the prior using an empirical Bayes method. This provides a crude estimate of the log rate parameter, coupled with a standard deviation that is sufficiently large to encompass all possible values (see S1 Text). We apply a Gamma(1,1) prior to $\phi = \tau^{-\alpha}$. This selection is grounded on a combination of theoretical considerations and empirical validation and allows for an efficient Gibbs sampler for τ .

To regularize the tail behavior, we leverage the shrunken-shoulder version of the Bayesian bridge prior and limit the bridge to have light tails past the slab width, ξ [37, 38]. An efficient update of Markov random field models global and local scale parameters (for Bayesian bridge priors) follows Nishimura and Suchard (2023) [38]. In this framework, the prior on the increment space represented as a scale mixture of normal distributions:

$$\mathbb{P}(\lambda_k^* - \lambda_{k-1}^* \mid v_k, \tau, \xi) = \text{N}\left(0, \left(\frac{1}{\xi^2} + \frac{1}{v_k^2 \tau^2}\right)^{-1}\right), \quad (4)$$

where v_k is called the local scale parameter and τ is the global scale parameter. (Note that v_k has an exponentially tilted stable distribution with characteristic exponent $\alpha/2$.) This mixture representation aids in clarifying the local adaptivity of the Bayesian bridge prior as considerable changes in rates can be accommodated by an increase in v_k without necessitating a rise in τ . The inclusion of the slab width helps to bound the variance of increments to ξ^2 . We set $\xi = 2$, which creates a reasonable upper limit on the variations in birth rate between consecutive epochs.

In our study, we primarily focus on sampling $\mathbb{P}[T \mid \lambda, \mu, \psi, t]$. With increasing numbers of epochs, the parameter space of the EBDS model expands quickly, exhibiting substantial correlation between adjacent epochs. To improve the sampling efficiency, we utilize HMC method to concurrently sample the time varying model parameters and ensure a high acceptance rate.

2.4 Hamiltonian Monte Carlo sampling

Hamiltonian Monte Carlo is a widely-used Markov chain Monte Carlo method to sample from a target distribution effectively. Given a target parameter θ with a posterior probability density $\pi(\theta)$, HMC iteratively generates samples from the target distribution by simulating the dynamics of a physical system whose equilibrium distribution is equal to $\pi(\theta)$. In particular, HMC introduces an auxiliary momentum parameter \mathbf{d} , which is typically chosen to follow a multivariate normal distribution with zero mean and covariance matrix \mathbf{M} , i.e., $\mathbf{d} \sim \mathcal{N}(0, \mathbf{M})$. \mathbf{M} is also known as the mass matrix, which serves as a hyperparameter. The Hamiltonian

function of the system is defined as:

$$H = U(\boldsymbol{\theta}) + K(\mathbf{d}), \tag{5}$$

where $U(\boldsymbol{\theta}) = -\log(\pi(\boldsymbol{\theta}))$ is the potential energy, and $K(\mathbf{d}) = \mathbf{d}^{(T)} \mathbf{M}^{(-1)} \mathbf{d}/2$ is the kinetic energy of the system.

Starting from the current state $(\boldsymbol{\theta}_0, \mathbf{d}_0)$, HMC updates the state according to the following differential equations:

$$\begin{aligned} \frac{d\mathbf{d}}{dt} &= -\nabla U(\boldsymbol{\theta}) = \nabla \log \pi(\boldsymbol{\theta}) \\ \frac{d\boldsymbol{\theta}}{dt} &= +\nabla K(\mathbf{d}) = \mathbf{M}^{-1} \mathbf{d}. \end{aligned} \tag{6}$$

The simple and effective ‘‘leapfrog’’ method [24] approximates the solution to (6) numerically:

$$\begin{aligned} \mathbf{d}_{t+\epsilon/2} &= \mathbf{d}_t + \frac{\epsilon}{2} \nabla \log \pi(\boldsymbol{\theta}_t) \\ \boldsymbol{\theta}_{t+\epsilon} &= \boldsymbol{\theta}_t + \epsilon \mathbf{M}^{-1} \mathbf{d}_{t+\epsilon/2} \\ \mathbf{d}_{t+\epsilon} &= \mathbf{d}_{t+\epsilon/2} + \frac{\epsilon}{2} \nabla \log \pi(\boldsymbol{\theta}_{t+\epsilon}), \end{aligned} \tag{7}$$

where ϵ is the size of each leapfrog step, and n steps are required to simulate the Hamiltonian dynamics from time $t = 0$ to $t = n\epsilon$. In practice, the ‘‘leapfrog’’ method has been shown to be stable and accurate for a wide range of step sizes [24]. In this work, we fix $n = 15$ and auto-adapt ϵ within BEAST to achieve an acceptance rate of approximately 70%.

The default choice of the mass matrix is the identity matrix. However, using a different \mathbf{M} , such as a log-posterior Hessian approximation can largely enhance the efficiency of HMC sampling. In this work, \mathbf{M} is adaptively tuned to estimate the expected (diagonal) Hessian averaged over the prior distribution. This design choice alleviates some computational burden, following the work of Ji *et al.* (2020) [25].

2.5 Gradient

HMC sampling of the model parameters requires the gradient of the log-likelihood derived from (1) wrt the EBDS model rate parameters. The gradient is the collection of derivatives wrt model parameters:

$$\nabla_{\boldsymbol{\theta}} \mathbb{P}[\mathcal{T} \mid \boldsymbol{\lambda}, \boldsymbol{\mu}, \boldsymbol{\psi}, \boldsymbol{\rho}, \mathbf{r}, \mathbf{t}] = \left(\frac{\partial \mathbb{P}}{\partial \theta_1}, \dots, \frac{\partial \mathbb{P}}{\partial \theta_k}, \dots, \frac{\partial \mathbb{P}}{\partial \theta_K} \right)^T, \tag{8}$$

where $\theta_k \in \{\lambda_k, \psi_k, \mu_k, \rho_k\}$ is a unified parameter to reduce notation clutter.

Given the piece-wise constant nature of the model, the likelihood assumes a consistent form across all epochs. Therefore, we can examine the gradient of the log-likelihood at each epoch separately. We denote the log-likelihood at epoch k and phylogeny segment j as:

$$\log \mathbb{P}_k(j) = \log I_k(E_j) + L(j) \log \left(\frac{q_k(s_{j+1})}{q_k(s_j)} \right). \tag{9}$$

We can further get individual terms in (8) by accumulating contributions from each epoch

and the corresponding phylogeny segments:

$$\frac{\partial \log \mathbb{P}}{\partial \theta_k} = \sum_{a=k}^K \sum_{j=1}^{m_k} \frac{\partial \log \mathbb{P}_a(j)}{\partial \theta_k}, \quad \theta_k \in \{\lambda_k, \psi_k, \mu_k, \rho_k\}. \tag{10}$$

By examining the interdependency between epochs, we discern that a given epoch k exerts influence on the gradient of parameters pertaining to that and all preceding epochs. Consequently, it becomes necessary to consider $\frac{\partial \log \mathbb{P}_k(j)}{\partial \theta_k}$ and $\frac{\partial \log \mathbb{P}_k(j)}{\partial \theta_{k-i}}$ respectively, where i is a positive integer ranging between 1 and $k - 1$.

First, we consider the gradient contribution at epoch k wrt the current model parameters $\frac{\partial \log \mathbb{P}_k(j)}{\partial \theta_k}$, where $\theta_k \in \{\lambda_k, \psi_k, \mu_k, \rho_k\}$.

Then we have the following cases:

$$\frac{\partial \log \mathbb{P}_k(j)}{\partial \theta_k} = \begin{cases} \text{If } E_j \text{ is a birth event happens at subinterval end } s_{j+1} : & \mathbb{1}_{\theta_k=\lambda_k} \frac{1}{\theta_k} + L(j) \cdot \frac{\partial Q_k(s_{j+1}, s_j)}{\partial \theta_k}, & (11) \\ \text{If } E_j \text{ is a serial sampling event happens at subinterval end } s_{j+1} : & \mathbb{1}_{\theta_k=\psi_k} \frac{1}{\theta_k} + \frac{1 - r_k}{(1 - r_k)p_k(s_j) + r_k} \cdot \frac{\partial p_k(s_j)}{\partial \theta_k} + L(j) \cdot \frac{\partial Q_k(s_{j+1}, s_j)}{\partial \theta_k}, & (12) \\ \text{If } E_j \text{ is an intensive sampling event happens at subinterval end } s_{j+1} = t_{k-1} : & \mathbb{1}_{\theta_k=\rho_k} \left(\frac{N_k}{\theta_k} + \frac{L(j) - N_k}{(1 - \theta_k)} \right) + \frac{1 - r_k}{(1 - r_k)p_{k-1}(s_j) + r_k} \cdot \frac{\partial p_{k-1}(s_j)}{\partial \theta_k}, & (13) \\ \text{If } E_j \text{ is a epoch switching event happens at subinterval end } s_{j+1} : & L(j) \cdot \frac{\partial Q_k(s_{j+1}, s_j)}{\partial \theta_k}. & (14) \end{cases}$$

Note that $\mathbb{1}$ is the indicator function. We leave the explicit expression of the shared terms in (11)–(14) to [S1 Text](#).

Second, we consider the gradient at epoch k wrt the previous model parameters $\frac{\partial \log \mathbb{P}_k(j)}{\partial \theta_{k-i}}$, where $\theta_{k-i} \in \{\lambda_{k-i}, \psi_{k-i}, \mu_{k-i}, \rho_{k-i}\}$:

$$\frac{\partial \log \mathbb{P}_k(j)}{\partial \theta_{k-i}} = \begin{cases} \text{If } E_j \text{ is a birth event or epoch switching event happens at subinterval end } s_{j+1} : & L(j) \cdot \frac{\partial Q_k(s_{j+1}, s_j)}{\partial \theta_{k-i}}, & (15) \\ \text{If } E_j \text{ is a serial sampling event happens at subinterval end } s_{j+1} : & \frac{1 - r_k}{(1 - r_k)p_k(s_j) + r_k} \cdot \frac{\partial p_k(s_j)}{\partial \theta_{k-i}} + L(j) \cdot \frac{\partial Q_k(s_{j+1}, s_j)}{\partial \theta_{k-i}}. & (16) \end{cases}$$

We also leave the explicit expression of the shared terms in (15)–(16) in [S1 Text](#).

Third, we discuss the gradient at epoch k wrt the treatment probability \mathbf{r} . In (1), the treatment probabilities at different epochs only affect the current epoch. Therefore, we only need to

consider $\frac{\partial \log \mathbb{P}_k(j)}{\partial r_k}$ as follows:

$$\frac{\partial \log \mathbb{P}_k(j)}{\partial r_k} = \begin{cases} \text{If } E_j \text{ is a serial sampling event happens at subinterval end } s_{j+1}: \\ \frac{1 - p_k(s_j)}{(1 - r_k)p_k(s_j) + r_k}, & (17) \\ \text{If } E_j \text{ is a intensive sampling event happens at subinterval end } s_{j+1} = t_{k-1}: \\ \frac{1 - p_{k-1}(s_j)}{(1 - r_k)p_{k-1}(s_j) + r_k}. & (18) \end{cases}$$

The total gradient wrt \mathbf{r} can be obtained similar to (10).

To determine the computation complexity of gradient evaluation, we can assume the gradient calculation for $\frac{\partial \log \mathbb{P}_k(j)}{\partial \theta_k}$ takes constant time. The model has K epochs, where each epoch has $\frac{(2N+K-2)}{K}$ phylogeny segments in average. According to (10), the total computation complexity is $\mathcal{O}(K \cdot \frac{(2N+K-2)}{2}) \sim \mathcal{O}(NK)$, since $K \ll N$. We demonstrate this result through a series of timing experiments presented in S1 Text where we also compare the efficiency of gradients calculations with the automatic differentiation algorithm implemented in the VBSKY [39] package based on JAX library [40]. Figure E in S1 Text shows our analytical gradients implemented in BEAST significantly outpace the VBSKY method.

2.6 Analysis

2.6.1 Examples. We evaluate the relative effectiveness of MH-MCMC and HMC transition kernels under the EBDS model using three phylodynamic examples. The first example comprises 274 sequences of the Pol locus of HIV-1 subtype A sampled in Odesa, Ukraine from 2000 to 2020 that Vasylyeva *et al.* (2020) previously analyzed to assess the population-level impact of the transmission reduction intervention project (TRIP) on HIV transmission [4, 41]. Following this previous analysis, we establish a cutoff point of 50 years for the EBDS model. Within this period of time, we let the birth, death and sampling rates vary across 10 epochs mirroring the grid points specified by Vasylyeva *et al.* (2020) [4]. Note that for better comparability to the original work [4], we place iid lognormal priors on the rate parameters. Both the previous and our analysis assume an HKY nucleotide substitution [42] model with discrete-gamma-distributed rate variation among sites (HKY+G) [43], and an uncorrelated lognormal relaxed molecular clock model [44] (UCLD), with a CTMC rate-reference prior [45] on the clock-model mean, truncated between 1×10^{-3} – 3×10^{-3} , and a normal prior (with mean = 5×10^{-4} and standard deviation = 5×10^{-4}) on the standard deviation. We use a normal distribution prior (with mean = 35, standard deviation = 5) on the time to the most recent common ancestor, in accordance with the previous study.

Second, we examine the transmission dynamics of 637 human influenza A/H3N2 HA genes across 12 epidemic seasons sampled from New York state following the study of Parag *et al.* (2020) [46, 47]. We set an EBDS model cutoff value of 13 years and infer time-varying birth and sampling rates across 78 epochs, each representing 2 months in time, and a constant-over-time death rate. Preceding studies focused on the evolutionary dynamics of influenza A/H1N1 virus mostly utilize the coalescent models. These studies predominantly rely on Gaussian process smoothing [48, 49]. Following the same path, we seek to use GMRF prior distributions for the birth and sampling rates. Our approach accommodates the considerable variability in the effective reproductive number across different flu seasons from 1993 to 2005.

We adopt the same substitution and clock models from Rambaut *et al.* (2008) [46]. Specifically, to account for potential differences in the rate of substitution between the first and second codon positions compared to the third, we employ the SRD06 substitution model [50] and apply an HKY nucleotide substitution model with discrete-gamma distributed rate heterogeneity for both codon-position partitions (1st + 2nd, and 3rd). We further assume a UCLD clock model and employ the default priors from BEAST on the substitution and clock model parameters.

Lastly, to demonstrate the potential our linear-time algorithms afford phylodynamic analyses on larger data sets, we examine 1610 full Ebola virus (EBOV) genomes sampled between 17 March 2014 and 24 October 2015 from West Africa [2] to explore the factors contributing to the spread of Ebola during the 2014–2016 epidemic. We set a EBDS model cutoff value of 2 years and infer time-varying birth and sampling rates for 24 epochs, each corresponding to a month in time, and a constant death rate. For choosing the priors on the rate parameters, we incorporate information from previous studies on the transmission dynamics of Ebola virus disease in West Africa [51, 52]. The number of confirmed cases first persisted at a relatively low level and started to soar in the mid-Summer of 2014, followed by a consistent peak and a dramatic decrease after the initiation of some key intervention events. Considering the potential fast shifts projected to the effective reproductive number, we apply the Bayesian bridge MRF model as the prior for the incremental differences in the birth and sampling rates. Based on Dudas *et al.* (2017), we assume a HKY+G substitution model independently across four partitions (codon positions 1, 2, 3 and non-coding intergenic regions) and a log-normally-distributed relaxed molecular clock model with a CTMC reference prior on the clock model mean, and leave all other priors on substitution and clock model parameters at their BEAST defaults [2].

2.6.2 Implementation. We conduct all analyses using extensions to BEAST 1.10 [29] and the high-performance BEAGLE 4.0 library [53] for efficient computation on central processing units (CPUs). We take the timing measurements using a Macbook Pro equipped with an M1 Pro chip that features 8 CPU cores and 32GB of RAM. For all experiments involving BEAST, we utilized the Azul Zulu Builds of OpenJDK version 18 on the ARM architecture.

To validate our implementation of the EBDS model within the BEAST framework, we employ continuously-integrated unit-testing. These tests cover both the EBDS likelihood and its gradients. For the likelihood assessment, we perform a comparative analysis of the log-density evaluation between our implementation and that of existing software. We also evaluate our analytic gradient against a numerically computed central-difference approximation all within BEAST, setting a criterion that the maximum absolute relative difference between the gradient dimensions does not exceed 1×10^{-3} .

To compare the performance of the two transition kernels in estimating the EBDS model parameters, we conduct efficiency comparison analyses that focused solely on the estimation of the birth-death model's rate parameters. Specifically, we fix the phylogeny to the maximum clade credibility (MCC) tree, a tree with the maximum product of the posterior clade probabilities summarized from the Bayesian joint phylogeny inference. We analyze all data sets using BEAST with logging performed every 1000 iterations. We run our algorithm on the HIV example for 300 million iterations when using MH-MCMC transition kernel and 30 million iterations for HMC transition kernel. Also, to obtain convergent results for the influenza example, we run analyses using MH-MCMC and HMC transition kernels for 300 million and 50 million iterations, respectively. For the Ebola example, we run analyses using MH-MCMC and HMC transition kernels for 100 million and 30 million iterations, respectively.

For a more comprehensive evaluation of HMC and MH-MCMC samplers, we extend our efficiency comparison beyond the fixed phylogeny assumption by conducting a joint analysis

where we simultaneously infer the phylogenetic tree from sequence data. We limit this comparative analysis to the HIV dataset as a proof of concept, acknowledging that tree inference can be impractically time-consuming for the larger two examples using only univariate transition kernels on the EBDS model rates. We run both sets of transition kernels for a total of 550 million iterations for this joint analysis comparison.

We adopt a Metropolis-within-Gibbs approach [54] and develop a random-scan Gibbs cycle to systematically cycle through the sampling of the phylogenetic tree topology, rate parameters, and subsequently other model components. For all analyses, we discard 10% of the MCMC chain samples as burn-in.

We calculate the effective sample size (ESS) for each rate parameter of interest using the coda package [55] in CRAN [56]. ESS quantifies the degree to which auto-correlation within MCMC iterations contributes to uncertainty in estimates [57]. We average ESS per compute-hour for each parameter across 10 independent runs to reduce Monte Carlo error in each estimate, aiming for a maximal Monte Carlo error of 10%. We report the relative increase in ESS per hour of the HMC sampler compared with the MH-MCMC sampler over all rate parameters.

We also conduct phylodynamic analysis for each of the three examples under a joint phylogeny inference scheme to mitigate potential bias from the fixed phylogeny, following the model specifications discussed in Section 2.6.1. Under these settings, we simulate MCMC chains for all examples of 500 million iterations using HMC transition kernel with logging performed every 1000 iterations.

3 Results

3.1 Performance improvements

Fig 2 shows the binned ESS per hour estimates of the EBDS model rates (λ , μ , ψ) that the MH-MCMC and HMC samples generate for all three viral examples. Table 1 summarizes the performance improvements by reporting the relative increase in the minimum ESS per hour comparing both samplers across all rate parameters.

The HIV example assumes that time-varying rates are *a priori* independent across epochs and for inference on a fixed phylogeny, HMC demonstrates an approximate 245-fold acceleration relative to MH-MCMC. Likewise, the influenza example imposes a GMRF across epochs and returns an approximate 79.4-fold speed-up. On the other hand, the EBOV example enforces heavier shrinkage, and hence higher *a priori* correlation between epochs, and yields a smaller yet computationally impactful (approximately 12.7-fold) performance increase.

These efficiency improvements also extend to more comprehensive analyses in which we simultaneously learn the phylogenetic tree structure and nucleotide substitution process. In the HIV example, while minimum ESS per hour estimates for the EBDS model rates decrease by approximately an order-of-magnitude when jointly inferring the phylogeny (Fig 2), the relative speed-up that HMC affords actually grows slightly to a 277-fold increase (Table 1). This suggests that HMC can handle well the additional parameter correlation that joint analyses often induce. We observe, however, impractically long convergence and mixing times using the univariate transition kernels for the two larger examples under joint analyses, and so can not directly compare their efficiency with the joint analyses that HMC succeeds in enabling for these examples.

S1 Text provides additional performance metrics comparing HMC vs MH-MCMC, including run-times to reach min ESS > 200 across all EBDS model rates.

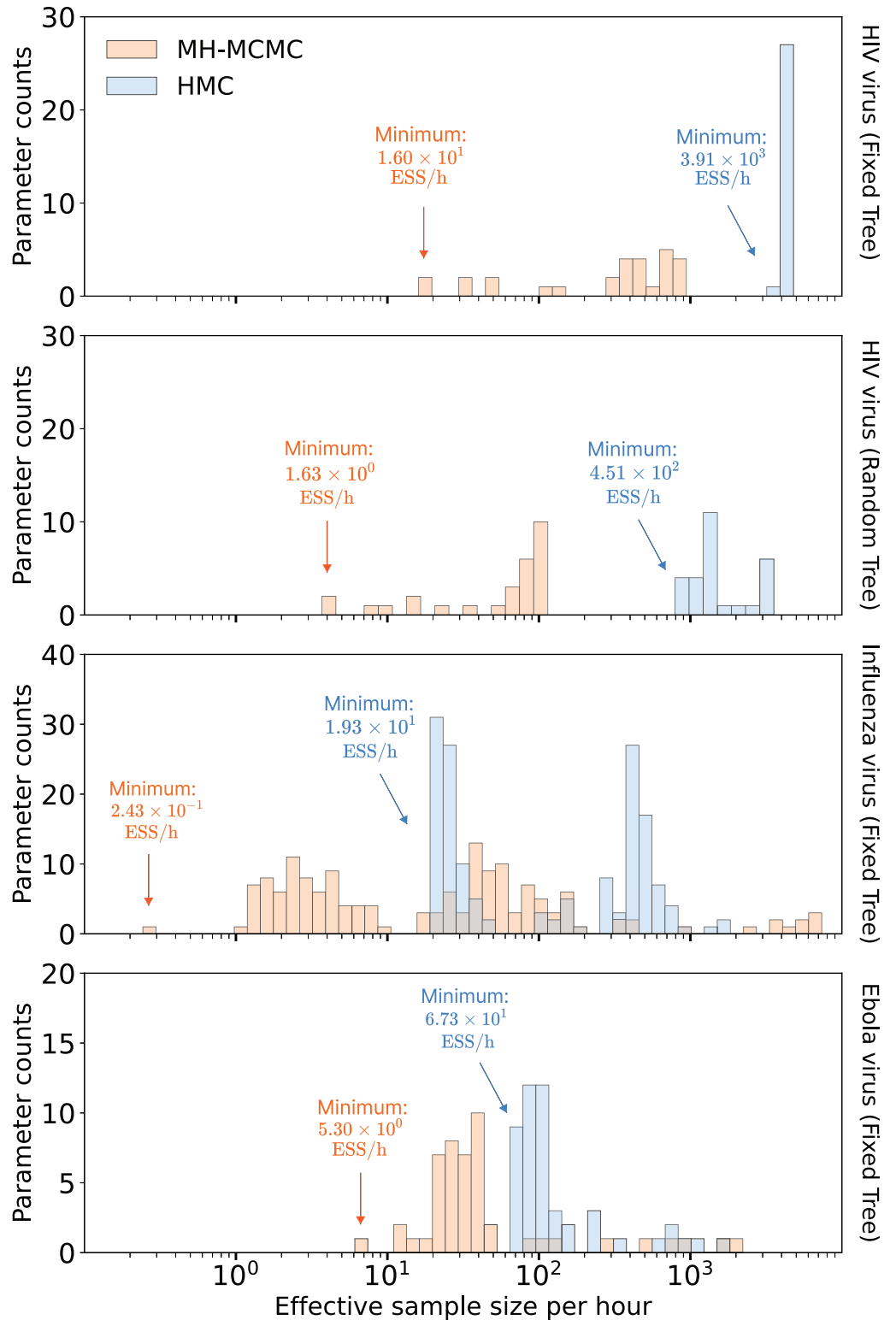


Fig 2. Efficiency Comparison between random walk Metropolis-Hastings (MH-MCMC) and Hamiltonian Monte Carlo (HMC) samplers. Bars correspond to the estimated effective sample size per hour averaged across 10 independent runs for all rate parameters. The height of each bar indicates the number of parameters that achieve the given ESS per hour value.

<https://doi.org/10.1371/journal.pcbi.1011640.g002>

Table 1. Relative speedup in terms of effective sample size per hour (ESS/h) of HMC over MH-MCMC for HIV example from both fixed and random phylogeny analyses, and for Influenza and Ebola examples from fixed phylogeny analyses.

Example	Minimum ESS/h		HMC
	MH-MCMC	HMC	Speedup
HIV (Fixed tree, 10 epochs)	1.60×10^1	3.91×10^3	2.45×10^2 times
HIV (Random tree, 10 epochs)	1.63×10^0	4.51×10^2	2.77×10^2 times
Influenza (Fixed tree, 78 epochs)	2.43×10^{-1}	1.93×10^1	7.94×10^1 times
Ebola (Fixed tree, 24 epochs)	5.30×10^0	6.73×10^1	1.27×10^1 times

<https://doi.org/10.1371/journal.pcbi.1011640.t001>

3.2 HIV dynamics in Odesa, Ukraine

In the context of conducting phylodynamic analyses using EBDS models, we are primarily interested in the value and trend of effective reproductive number over time $R_e(t)$ that is the average number of secondary cases per infectious case in a population made up of both susceptible and non-susceptible hosts. If $R_e > 1$, the number of cases is growing, such as at the start of an epidemic; if $R_e = 1$, the disease is endemic; and if $R_e < 1$, there is an expected decrease in transmission [58]. Under the EBDS model, given the absence of intensive sampling events, if an individual becomes infected at time t , we can use the rate parameters at time t to obtain an estimated $R_e(t) = \frac{\lambda(t)}{\mu(t) + r(t)\psi(t)}$. Furthermore, in all our analyses for infectious disease phylodynamics, we maintain $r(t) = 1$ as constant. This assertion carries the assumption that upon diagnosis and sequencing, an individual ceases to be a source of infection. This could be due to treatment, death, or geographical relocation, rendering them incapable of onward transmission.

To assess the effects of TRIP for reducing the transmission of HIV in Odesa, we fit the EBDS model with varying birth, death and sampling rates and plot the resulting $R_e(t)$ trend estimate in Fig 3. We apply iid lognormal priors on the rate parameters to stay consistent with the methods in previous study [4].

Estimates of $R_e(t)$ appear mostly to accord with previous findings that identify a drop in infection rate subsequent to the implementation of the TRIP intervention. Focusing on the period from 2013 to early 2016, when TRIP was enacted, our posterior mean estimate of R_e is 2.64 (95% CI: 1.18–5.43); while post-intervention, the posterior mean reduces to 0.152 (95% CI: 0.03–0.32). This latter value, falling below the critical threshold of 1, signifies the potential deceleration of HIV transmission.

3.3 Seasonal influenza in New York state

While influenza viruses circulate throughout the year, peak influenza outbreaks in the United States typically occur between December and February. Rambaut *et al.* (2008) employed a non-parametric coalescent model to elucidate the cyclical patterns of variation in the population size, uncovering a notable increase in genetic diversity at the beginning of each winter flu season [46]. Subsequently, Parag *et al.* (2020) demonstrated that incorporating sampling intensity into the otherwise sampling-naïve non-parametric coalescent process improves the precision of these inferred cycles [47]. With a GMRF smoothing prior on increments, our model also offers the potential for accurately inferring seasonal behaviour and achieving the precision of parameter estimations.

Fig 4 presents posterior estimates of the effective reproductive number $R_e(t)$ for the alignment of 637 A/H3N2 HA sequences from New York state. As expected, the trajectory is highly cyclic, and all peaks lie near the midpoint of the influenza seasons with estimated R_e larger

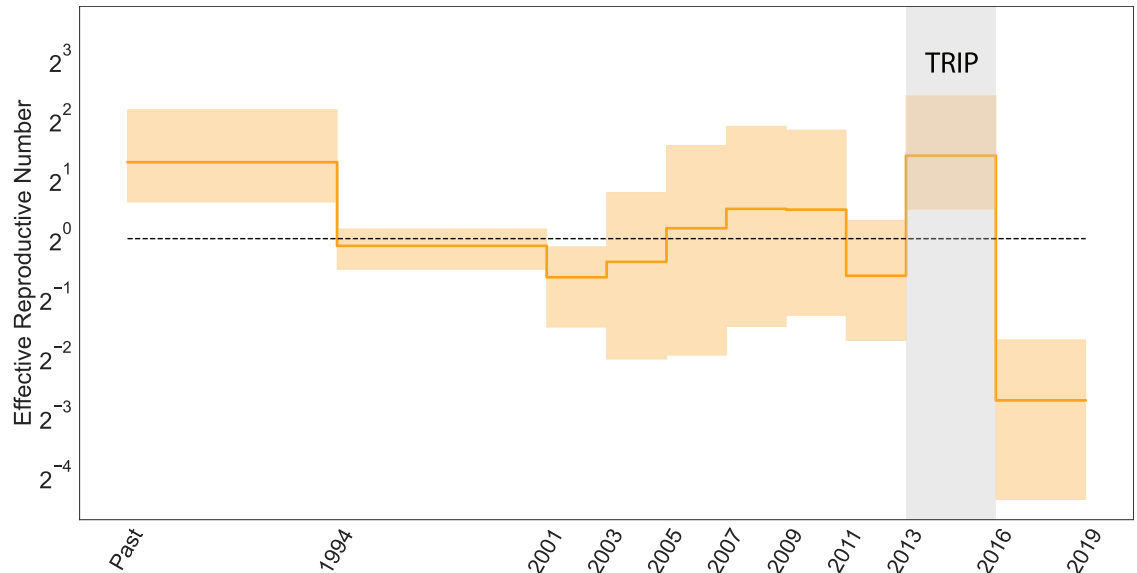


Fig 3. Posterior median (solid line) and 95% credible intervals (CI) indicated by the orange shaded areas of the effective reproductive number estimates (R_e) through time for HIV epidemic in Odesa, where the black dotted line represents the epidemiological threshold of $R_e(t) = 1$. The gray shaded region in the figure represents the duration of the transmission reduction intervention project (TRIP) on HIV transmission.

<https://doi.org/10.1371/journal.pcbi.1011640.g003>

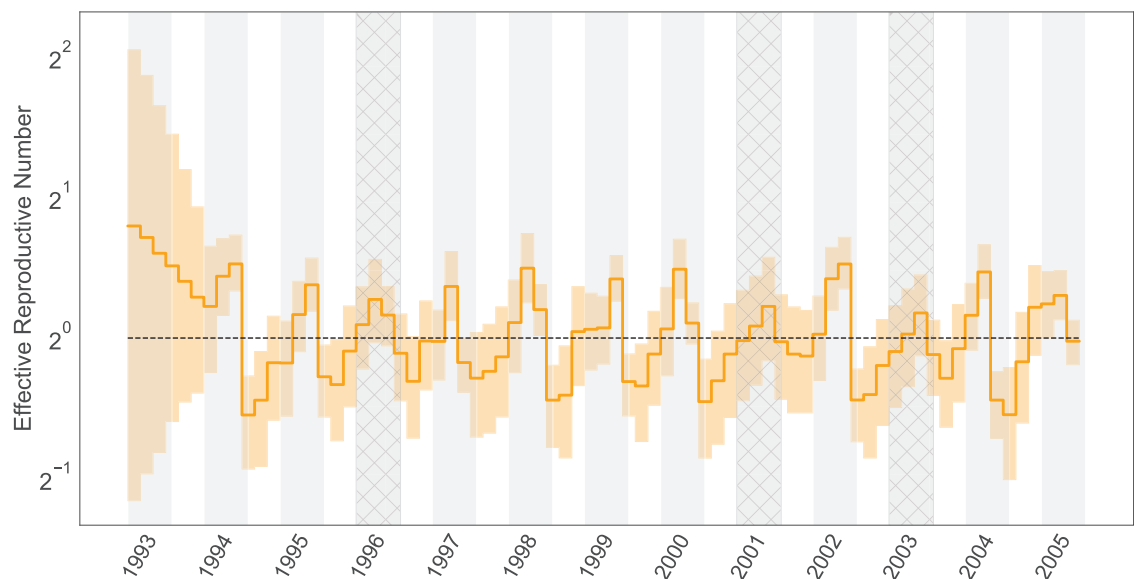


Fig 4. Median (solid orange line) and 95% credible intervals indicated by the shaded orange areas for the effective reproductive number estimates (R_e) through time. Gray shading in the graph represents the rough duration of influenza monitored in New York state for each season, spanning from epidemiological week 40 to week 20 of the following year. Seasons where A/H3N2 was not the dominant influenza virus subtype are cross-hatched.

<https://doi.org/10.1371/journal.pcbi.1011640.g004>

than 1. For the 2000/2001 and 2002/2003 seasons, where almost all infections were attributed to other sub-types of influenza viruses as indicated by the surveillance data and previous work [47, 59], we observe the 95% CI of the estimated peak cover values from 0.68 to 1.3 and from 0.48 to 1.4, respectively. This suggests that their true R_e values might have fallen below 1. Similar to the results given by the non-parametric coalescent with sampling analysis [47], we capture a minor peak in the 1995/1996 season, where the inferred R_e is slightly above one. This again echoes with the fact that the influenza case composition during the 1995/1996 season was characterized by a mix of A/H1N1 and A/H3N2 infections [60]. This diversity in infection types led to a less significant elevation in the effective reproductive number for that specific year.

3.4 Ebola epidemic in West Africa

Using EBDS model assisted by the HMC sampler, we are able to analyze the 2014 Ebola epidemic in West Africa using the full 1610-sequence alignment and metadata of sampling times taken from the work by Dudas *et al.* (2017) [2]. Previously, researchers have applied birth-death models extensively for the phylodynamic analysis of the Ebola outbreak. Stadler *et al.* (2014) adopted a series of birth death models to capture the early trend of the infection of Ebola virus in Sierra-Leone [61]. They used 72 Ebola samples from late May to mid June 2014 with three epochs, and estimated the corresponding effective reproductive number in each period. Zhukova *et al.* (2022) applied the multi-type birth death models to the 1610 sequence data [62]. However, their analysis was based on the maximum likelihood estimation. To demonstrate the scalability of our method, we also take the 1610 sequence data and fit the EBDS model with 24 epochs for a finer time resolution to provide more precise estimation of the effective reproductive number. Here, we employ a Bayesian bridge MRF prior on rate increments to avoid spurious rate variations while capturing significant rate shifts.

Our inference results give an estimated posterior mean effective reproductive number at the beginning of the epidemic before December 2013 as 1.65 (95% CI: 0.41–3.05). Dudas *et al.* (2017) show that after the international border closure of Sierra Leone on 11 June 2014, followed by Liberia on 27 July 2014, and Guinea on 9 August 2014, the relative contribution of international border to overall viral migration is significantly lower [2]. The change-point probability is the highest from August to September. As shown in Fig 5, this finding stands clearly compatible with our EBDS inference that demonstrates a drop of R_e from 1.3 (posterior mean, 95% CI: 1.01–1.59) to 0.79 (95% CI: 0.62–0.91) after September 2014 when the international travel restrictions are in place across the three countries.

4 Discussion

Birth-death models serve as fundamental tools for modeling the temporal progression of epidemics. In extending the work of Stadler *et al.* (2013) and Gavryushkina *et al.* (2014) [9, 17], we have provided a systematic representation of the EBDS model for phylodynamics that promotes scalability. Our general re-formalization of the EBDS likelihood identifies that its computation is simply $\mathcal{O}(N + K)$, foreshadowing an $\mathcal{O}(NK)$ algorithm to deliver its gradient wrt time-varying birth, death or sampling rates across K epochs. This optimal scaling enables HMC sampling to more efficiently explore the high-dimensional joint distribution of rates as we increase the number of sequences and the number of model epochs to learn these processes at a finer time-resolution. HMC also emits an agnostic approach to incorporate a variety of prior assumptions about these time-varying trends, without the need to hand-craft specialized transitions kernels for specific priors. Moreover, as suggested by Ji *et al.* (2020) [25], we take

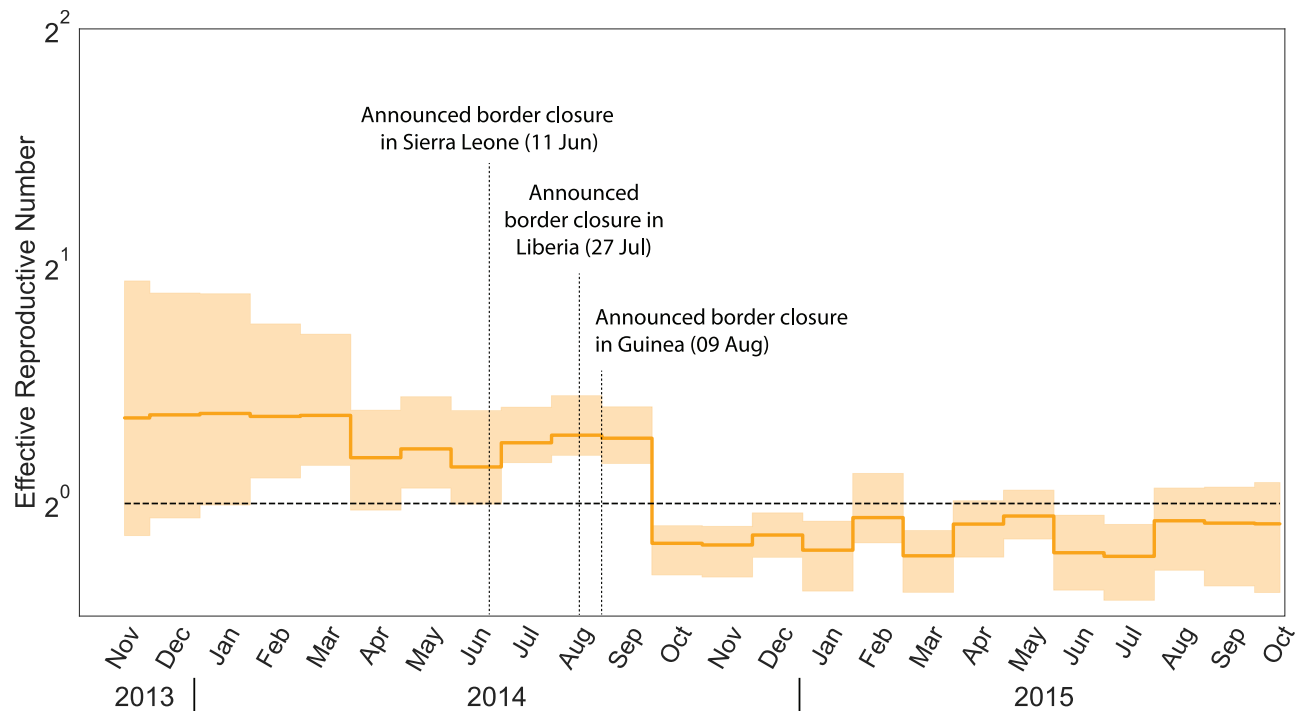


Fig 5. Median (solid line) and 95% credible intervals indicated by the shaded areas of the effective reproductive number estimates (R_e) through time for Ebola outbreak in west Africa. The black dotted line represents the epidemiological threshold of $R_e = 1$.

<https://doi.org/10.1371/journal.pcbi.1011640.g005>

measures to enhance the efficiency of our HMC sampler by preconditioning the mass matrix based on the Hessian of the log-prior.

Through three viral epidemic examples, we show that our HMC-assisted approach considerably accelerates Bayesian inference across three very different choices of prior models. Our preconditioned HMC sampler achieves roughly 10- to 200-fold increase over the widely used MH-MCMC sampler in terms of the minimum ESS per unit-time. The enhanced efficiency gains are particularly beneficial given the increasing use of phylodynamic inference techniques, often in resource-limited settings, in conducting real-time evaluations of outbreak patterns.

For applying our model in phylodynamic analyses of disease epidemics, we first examine our EBDS model on the effects of TRIP for reducing the transmission of HIV in Ukraine, and our inference results support a decreased rate of transmission following the TRIP intervention. Applied to seasonal Influenza in New York city, our model is able to accurately capture the complex pattern of variation in R_e during each influenza season. Applied to the Ebola outbreak in West Africa, our model supports the effect of international travel restrictions characterized as a noticeable decrease in R_e following the border closure of the three countries in West Africa.

In the EBDS model, Stadler and colleagues [9] have indicated that the three rate parameters, λ , μ , and ψ , cannot be simultaneously identified. This issue of unidentifiability in complex birth-death processes has also been recently discussed by Louca and Pennell (2020) [63]. In our own empirical analysis, problems related to unidentifiability seldom manifest when we restrict ourselves to estimating no more than two time-varying rate parameters. Instead, the primary challenge appears to be the multimodal nature of the posterior distribution. Legried and Terhorst (2022) have demonstrated that, under certain conditions, piecewise constant

birth-death models can be reliably inferred and differentiated [64]. Furthermore, Kopperud *et al.* (2023) showed that rapidly changing speciation or extinction rates can be accurately estimated [65]. This lends credence to the identifiability of patterns we observed in our phylodynamic analysis of pandemics such as the seasonal influenza and the Ebola outbreaks.

Current methods to estimate the expected Hessian averaged over the posterior distribution improves upon the previous work [66] by avoiding excessive computational burden. However, it relies on numerical approximations to compute the Hessian, leaving room for potential performance enhancements. To further optimize the methodology, we can advance beyond analytical solutions solely for gradients and extend them to encompass the analytical Hessian. This would smooth the path of updating the adaptive mass matrix, offering opportunities for better outcomes in terms of both efficiency and accuracy.

In many scenarios, the examination of EBDS models is contingent upon having some preliminary understanding of how to identify the epoch switching time and the length of duration of each epoch. However, it is possible that information available through epidemiological surveillance is insufficient. Moreover, the choice of epoch duration can be related to the uncertainty in the timing of the rate shifts [32]. In this study, our strategy aims to increase the number of epochs and leverage regularizing priors, striving to achieve a refined grid of timelines. Nevertheless, constraints persist on the maximum epochs feasible with our HMC algorithm, particularly when confronted with computational limitations or models exhibiting multimodality challenges. One possible solution entails simultaneously inferring epoch duration, epoch switching times, and rate parameters via the reversible-jump MCMC method [34]. However, this method requires one to integrate across models with differing dimension, which demands substantial effort and might be impractical for large datasets.

Considering these cases, if the piece-wise constant model assumptions can be lifted so that we can obtain a smoothly differentiable likelihood function, it would inherently help the gradient's derivation concerning node ages and epoch switching times. This advancement would, in turn, improve our current implementation, empowering us to infer, rather than presuppose, epoch switching times, with better scalability prospects. It would also enhance the sampling efficiency from joint phylogeny posterior distributions, by enabling us to take advantage of recent work by Ji *et al.* (2021), yielding a pronounced improvement in the analytical capacity of our models [27].

In anticipation of future advancements that will improve upon standard HMC methods and broaden the applicability of the current EBDS model, we present a comprehensive framework in this manuscript. This framework facilitates phylodynamic analysis on large-scale sequence data and employs regularization techniques to yield a finely-resolved grid that effectively aids in our understanding of the impact of the pandemics.

Supporting information

S1 Text. Appendices with additional likelihood/gradient equations and real data analysis details.

(PDF)

Author Contributions

Conceptualization: Yucai Shao, Andrew F. Magee, Marc A. Suchard.

Data curation: Yucai Shao, Andrew F. Magee, Tetyana I. Vasylyeva.

Formal analysis: Yucai Shao, Andrew F. Magee, Marc A. Suchard.

Funding acquisition: Marc A. Suchard.

Investigation: Yucai Shao.

Methodology: Yucai Shao, Andrew F. Magee, Marc A. Suchard.

Project administration: Yucai Shao, Andrew F. Magee, Marc A. Suchard.

Resources: Tetyana I. Vasylyeva.

Software: Yucai Shao, Andrew F. Magee, Marc A. Suchard.

Supervision: Andrew F. Magee, Marc A. Suchard.

Validation: Yucai Shao, Andrew F. Magee, Marc A. Suchard.

Visualization: Yucai Shao, Andrew F. Magee, Marc A. Suchard.

Writing – original draft: Yucai Shao, Marc A. Suchard.

Writing – review & editing: Yucai Shao, Andrew F. Magee, Tetyana I. Vasylyeva, Marc A. Suchard.

References

1. Nunes MR, Palacios G, Faria NR, Sousa EC Jr, Pantoja JA, Rodrigues SG, et al. Air travel is associated with intracontinental spread of dengue virus serotypes 1–3 in Brazil. *PLoS Neglected Tropical Diseases*. 2014; 8(4):e2769. <https://doi.org/10.1371/journal.pntd.0002769> PMID: 24743730
2. Dudas G, Carvalho LM, Bedford T, Tatem AJ, Baele G, Faria NR, et al. Virus genomes reveal factors that spread and sustained the Ebola epidemic. *Nature*. 2017; 544:309–15. <https://doi.org/10.1038/nature22040> PMID: 28405027
3. Lau MS, Grenfell BT, Worby CJ, Gibson GJ. Model diagnostics and refinement for phylodynamic models. *PLoS Computational Biology*. 2019; 15:e1006955. <https://doi.org/10.1371/journal.pcbi.1006955> PMID: 30951528
4. Vasylyeva TI, Zarebski A, Smyrnov P, Williams LD, Korobchuk A, Liulchuk M, et al. Phylodynamics helps to evaluate the impact of an HIV prevention intervention. *Viruses*. 2020; 12:469. <https://doi.org/10.3390/v12040469> PMID: 32326127
5. Attwood SW, Hill SC, Aanensen DM, Connor TR, Pybus OG. Phylogenetic and phylodynamic approaches to understanding and combating the early SARS-CoV-2 pandemic. *Nature Reviews Genetics*. 2022; 23(9):547–62. <https://doi.org/10.1038/s41576-022-00483-8> PMID: 35459859
6. MacPherson A, Louca S, McLaughlin A, Joy JB, Pennell MW. Unifying phylogenetic birth–death models in epidemiology and macroevolution. *Systematic Biology*. 2022; 71:172–89. <https://doi.org/10.1093/sysbio/syab049>
7. Crawford FW. General birth-death processes: probabilities, inference, and applications. UCLA; 2012.
8. Yang Z, Rannala B. Bayesian phylogenetic inference using DNA sequences: a Markov Chain Monte Carlo method. *Molecular biology and evolution*. 1997; 14:717–24. <https://doi.org/10.1093/oxfordjournals.molbev.a025811> PMID: 9214744
9. Stadler T, Kühnert D, Bonhoeffer S, Drummond AJ. Birth–death skyline plot reveals temporal changes of epidemic spread in HIV and hepatitis C virus (HCV). *Proceedings of the National Academy of Sciences*. 2013; 110:228–33. <https://doi.org/10.1073/pnas.1207965110> PMID: 23248286
10. Höhna S. Likelihood inference of non-constant diversification rates with incomplete taxon sampling. *PLoS one*. 2014; 9:e84184. <https://doi.org/10.1371/journal.pone.0084184> PMID: 24400082
11. Stadler T. Sampling-through-time in birth–death trees. *Journal of Theoretical Biology*. 2010; 267:396–404. <https://doi.org/10.1016/j.jtbi.2010.09.010> PMID: 20851708
12. Barido-Sottani J, Vaughan TG, Stadler T. A multitype birth–death model for Bayesian inference of lineage-specific birth and death rates. *Systematic Biology*. 2020; 69:973–86. <https://doi.org/10.1093/sysbio/syaa016> PMID: 32105322
13. Maddison WP, Midford PE, Otto SP. Estimating a binary character’s effect on speciation and extinction. *Systematic biology*. 2007; 56:701–10. <https://doi.org/10.1080/10635150701607033> PMID: 17849325
14. FitzJohn RG. Quantitative traits and diversification. *Systematic biology*. 2010; 59:619–33. <https://doi.org/10.1093/sysbio/syq053> PMID: 20884813

15. FitzJohn RG. Diversitree: comparative phylogenetic analyses of diversification in R. *Methods in Ecology and Evolution*. 2012; 3:1084–92. <https://doi.org/10.1111/j.2041-210X.2012.00234.x>
16. Lambert A, Stadler T. Birth–death models and coalescent point processes: The shape and probability of reconstructed phylogenies. *Theoretical Population Biology*. 2013; 90:113–28. <https://doi.org/10.1016/j.tpb.2013.10.002> PMID: 24157567
17. Gavryushkina A, Welch D, Stadler T, Drummond AJ. Bayesian inference of sampled ancestor trees for epidemiology and fossil calibration. *PLoS Computational Biology*. 2014; 10:e1003919. <https://doi.org/10.1371/journal.pcbi.1003919> PMID: 25474353
18. Du Plessis L. Understanding the spread and adaptation of infectious diseases using genomic sequencing data. ETH Zurich; 2016.
19. Novitsky V, Kühnert D, Moyo S, Widenfelt E, Okui L, Essex M. Phylogenetic analysis of HIV sub-epidemics in Mochudi, Botswana. *Epidemics*. 2015; 13:44–55. <https://doi.org/10.1016/j.epidem.2015.07.002> PMID: 26616041
20. Minosse C, Salichos L, Taibi C, Luzzitelli I, Nardozi D, Capobianchi MR, et al. Phylogenetic and Phylogenetic Analyses of HCV Strains Circulating among Patients Using Injectable Drugs in Central Italy. *Microorganisms*. 2021; 9:1432. <https://doi.org/10.3390/microorganisms9071432> PMID: 34361868
21. Hastings WK. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*. 1970; 57:97–109. <https://doi.org/10.1093/biomet/57.1.97>
22. Morlon H, Parsons TL, Plotkin JB. Reconciling molecular phylogenies with the fossil record. *Proceedings of the National Academy of Sciences*. 2011; 108:16327–32. <https://doi.org/10.1073/pnas.1102543108> PMID: 21930899
23. Duane S, Kennedy AD, Pendleton BJ, Roweth D. Hybrid Monte Carlo. *Physics Letters B*. 1987; 195:216–22. [https://doi.org/10.1016/0370-2693\(87\)91197-X](https://doi.org/10.1016/0370-2693(87)91197-X)
24. Neal RM, et al. MCMC using Hamiltonian dynamics. *Handbook of Markov Chain Monte Carlo*. 2011; 2:2.
25. Ji X, Zhang Z, Holbrook A, Nishimura A, Baele G, Rambaut A, et al. Gradients do grow on trees: a linear-time $O(N)$ -dimensional gradient for statistical phylogenetics. *Molecular Biology and Evolution*. 2020; 37:3047–60. <https://doi.org/10.1093/molbev/msaa130> PMID: 32458974
26. Fisher AA, Ji X, Nishimura A, Lemey P, Suchard MA. Shrinkage-based random local clocks with scalable inference. *arXiv preprint arXiv:210507119*. 2021.
27. Ji X, Fisher AA, Su S, Thorne JL, Potter B, Lemey P, et al. Scalable Bayesian divergence time estimation with ratio transformations. *arXiv preprint arXiv:211013298*. 2021.
28. Baele G, Gill MS, Lemey P, Suchard MA. Hamiltonian Monte Carlo sampling to estimate past population dynamics using the skygrid coalescent model in a Bayesian phylogenetics framework. *Wellcome Open Research*. 2020; 5:53. <https://doi.org/10.12688/wellcomeopenres.15770.1> PMID: 32923688
29. Suchard MA, Lemey P, Baele G, Ayres DL, Drummond AJ, Rambaut A. Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10. *Virus Evolution*. 2018; 4:vey016. <https://doi.org/10.1093/ve/vey016> PMID: 29942656
30. Condamine FL, Rolland J, Höhna S, Sperling FA, Sanmartín I. Testing the role of the Red Queen and Court Jester as drivers of the macroevolution of Apollo butterflies. *Systematic Biology*. 2018; 67:940–64. <https://doi.org/10.1093/sysbio/syy009> PMID: 29438538
31. Silvestro D, Tejedor MF, Serrano-Serrano ML, Loiseau O, Rossier V, Rolland J, et al. Early arrival and climatically-linked geographic expansion of New World monkeys from tiny African ancestors. *Systematic Biology*. 2019; 68:78–92. <https://doi.org/10.1093/sysbio/syy046> PMID: 29931325
32. Magee AF, Höhna S, Vasylyeva TI, Leaché AD, Minin VN. Locally adaptive Bayesian birth-death model successfully detects slow and rapid rate shifts. *PLoS Computational Biology*. 2020; 16:e1007999. <https://doi.org/10.1371/journal.pcbi.1007999> PMID: 33112848
33. Magee AF, Höhna S. Impact of K-Pg mass extinction event on crocodylomorpha inferred from phylogeny of extinct and extant taxa. *bioRxiv*. 2021:2021–01.
34. Wu CH. Bayesian approaches to model uncertainty in phylogenetics [Ph.D. thesis]. University of Auckland; 2014.
35. Bouckaert R, Vaughan TG, Barido-Sottani J, Duchêne S, Fourment M, Gavryushkina A, et al. BEAST 2.5: An advanced software platform for Bayesian evolutionary analysis. *PLoS computational biology*. 2019; 15:e1006650. <https://doi.org/10.1371/journal.pcbi.1006650> PMID: 30958812
36. Höhna S, Landis MJ, Heath TA, Boussau B, Lartillot N, Moore BR, et al. RevBayes: Bayesian phylogenetic inference using graphical models and an interactive model-specification language. *Systematic biology*. 2016; 65:726–36. <https://doi.org/10.1093/sysbio/syw021> PMID: 27235697
37. Piironen J, Vehtari A. Sparsity information and regularization in the horseshoe and other shrinkage priors. 2017.

38. Nishimura A, Suchard MA. Shrinkage with shrunken shoulders: Gibbs sampling shrinkage model posteriors with guaranteed convergence rates. *Bayesian Analysis*. 2023; 18:367–90. <https://doi.org/10.1214/22-BA1308>
39. Ki C, Terhorst J. Variational phylodynamic inference using pandemic-scale data. *Molecular Biology and Evolution*. 2022; 39:msac154. <https://doi.org/10.1093/molbev/msac154> PMID: 35816422
40. Bradbury J, Frostig R, Hawkins P, Johnson MJ, Leary C, Maclaurin D, et al. JAX: Composable Transformations of Python+ NumPy Programs (v0.2.5). Software available from <https://github.com/google/jax>. 2018.
41. Nikolopoulos GK, Pavlitina E, Muth SQ, Schneider J, Psychogiou M, Williams LD, et al. A network intervention that locates and intervenes with recently HIV-infected persons: The Transmission Reduction Intervention Project (TRIP). *Scientific reports*. 2016; 6:38100. <https://doi.org/10.1038/srep38100> PMID: 27917890
42. Hasegawa M, Kishino H, Yano Ta. Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. *Journal of Molecular Evolution*. 1985; 22:160–74. <https://doi.org/10.1007/BF02101694> PMID: 3934395
43. Yang Z. Maximum likelihood phylogenetic estimation from DNA sequences with variable rates over sites: approximate methods. *Journal of Molecular Evolution*. 1994; 39:306–14. <https://doi.org/10.1007/BF00160154> PMID: 7932792
44. Drummond AJ, Ho SYW, Phillips MJ, Rambaut A. Relaxed phylogenetics and dating with confidence. *PLoS Biology*. 2006; 4:e88. <https://doi.org/10.1371/journal.pbio.0040088> PMID: 16683862
45. Ferreira MA, Suchard MA. Bayesian analysis of elapsed times in continuous-time Markov chains. *Canadian Journal of Statistics*. 2008; 36:355–68. <https://doi.org/10.1002/cjs.5550360302>
46. Rambaut A, Pybus OG, Nelson MI, Viboud C, Taubenberger JK, Holmes EC. The genomic and epidemiological dynamics of human influenza A virus. *Nature*. 2008; 453:615–9. <https://doi.org/10.1038/nature06945> PMID: 18418375
47. Parag KV, du Plessis L, Pybus OG. Jointly inferring the dynamics of population size and sampling intensity from molecular sequences. *Molecular Biology and Evolution*. 2020; 37:2414–29. <https://doi.org/10.1093/molbev/msaa016> PMID: 32003829
48. Karcher MD, Carvalho LM, Suchard MA, Dudas G, Minin VN. Estimating effective population size changes from preferentially sampled genetic sequences. *PLoS Computational Biology*. 2020; 16: e1007774. <https://doi.org/10.1371/journal.pcbi.1007774> PMID: 33044955
49. Bhattacharjee U, Chakrabarti AK, Kanungo S, Dutta S. Evolutionary dynamics of influenza A/H1N1 virus circulating in India from 2011 to 2021. *Infection, Genetics and Evolution*. 2023; 110:105424. <https://doi.org/10.1016/j.meegid.2023.105424> PMID: 36913995
50. Shapiro B, Rambaut A, Drummond AJ. Choosing appropriate substitution models for the phylogenetic analysis of protein-coding sequences. *Molecular Biology and Evolution*. 2006; 23:7–9. <https://doi.org/10.1093/molbev/msj021> PMID: 16177232
51. Fang LQ, Yang Y, Jiang JF, Yao HW, Kargbo D, Li XL, et al. Transmission dynamics of Ebola virus disease and intervention effectiveness in Sierra Leone. *Proceedings of the National Academy of Sciences*. 2016; 113:4488–93. <https://doi.org/10.1073/pnas.1518587113> PMID: 27035948
52. Nyenswah TG, Kateh F, Bawo L, Massaquoi M, Gbanyan M, Fallah M, et al. Ebola and its control in Liberia, 2014–2015. *Emerging Infectious Diseases*. 2016; 22:169. <https://doi.org/10.3201/eid2202.151456> PMID: 26811980
53. Ayres DL, Cummings MP, Baele G, Darling AE, Lewis PO, Swofford DL, et al. BEAGLE 3: improved performance, scaling, and usability for a high-performance computing library for statistical phylogenetics. *Systematic Biology*. 2019; 68:1052–61. <https://doi.org/10.1093/sysbio/syz020> PMID: 31034053
54. Tierney L. Markov chains for exploring posterior distributions. *the Annals of Statistics*. 1994:1701–28.
55. Plummer M, Best N, Cowles K, Vines K. CODA: Convergence Diagnosis and Output Analysis for MCMC. *R News*. 2006; 6:7–11. Available from: <https://journal.r-project.org/archive/>.
56. R Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria; 2021. Available from: <https://www.R-project.org/>.
57. Ripley BD. Stochastic simulation. John Wiley & Sons; 2009.
58. Nishiura H, Chowell G. The effective reproduction number as a prelude to statistical estimation of time-dependent epidemic trends. *Mathematical and Statistical Estimation Approaches in Epidemiology*. 2009:103–21. https://doi.org/10.1007/978-90-481-2313-1_5
59. Centers for Disease Control and Prevention. Key Facts About Influenza (Flu);. Accessed: 2023-05-31. <https://www.cdc.gov/flu/about/keyfacts.html>.
60. Ferguson NM, Galvani AP, Bush RM. Ecological and immunological determinants of influenza evolution. *Nature*. 2003; 422:428–33. <https://doi.org/10.1038/nature01509> PMID: 12660783

61. Stadler T, Kühnert D, Rasmussen DA, du Plessis L. Insights into the early epidemic spread of Ebola in Sierra Leone provided by viral sequence data. *PLoS Currents*. 2014; 6. <https://doi.org/10.1371/currents.outbreaks.02bc6d927ecee7bbd33532ec8ba6a25f> PMID: 25642370
62. Zhukova A, Hecht F, Maday Y, Gascuel O. Fast and Accurate Maximum-Likelihood Estimation of Multi-Type Birth-Death Epidemiological Models from Phylogenetic Trees. *medRxiv*. 2022:2022–08.
63. Louca S, Pennell MW. Extant timetrees are consistent with a myriad of diversification histories. *Nature*. 2020; 580:502–5. <https://doi.org/10.1038/s41586-020-2176-1> PMID: 32322065
64. Legried B, Terhorst J. A class of identifiable phylogenetic birth–death models. *Proceedings of the National Academy of Sciences*. 2022; 119:e2119513119. <https://doi.org/10.1073/pnas.2119513119> PMID: 35994663
65. Kopperud BT, Magee AF, Höhna S. Rapidly changing speciation and extinction rates can be inferred in spite of nonidentifiability. *Proceedings of the National Academy of Sciences*. 2023; 120:e2208851120. <https://doi.org/10.1073/pnas.2208851120> PMID: 36757894
66. Girolami M, Calderhead B. Riemann manifold Langevin and Hamiltonian Monte Carlo methods. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*. 2011; 73:123–214. <https://doi.org/10.1111/j.1467-9868.2010.00765.x>