viruses

Article

Robust Phylodynamic Analysis of Genetic Sequencing Data from Structured Populations

Jérémie Scire [1,2], Joëlle Barido-Sottani [1,2,3], Denise Kühnert [4], Timothy G. Vaughan [1,2] and Tanja Stadler [1,2,*]

[1] Department of Biosystems Science and Engineering, ETH Zürich, 4058 Basel, Switzerland; jeremie.scire@bsse.ethz.ch (J.S.); joelle.barido-sottani@m4x.org (J.B.-S.); timothy.vaughan@bsse.ethz.ch (T.G.V.)
[2] Swiss Institute of Bioinformatics, 1015 Lausanne, Switzerland
[3] Institut de Biologie de l'ENS, École normale supérieure, CNRS, INSERM, Université PSL, 75005 Paris, France
[4] Transmission, Infection, Diversification and Evolution Group, Max Planck Institute for the Science of Human History, 07745 Jena, Germany; kuehnert@shh.mpg.de
* Correspondence: tanja.stadler@bsse.ethz.ch

Abstract: The multi-type birth–death model with sampling is a phylodynamic model which enables the quantification of past population dynamics in structured populations based on phylogenetic trees. The BEAST 2 package *bdmm* implements an algorithm for numerically computing the probability density of a phylogenetic tree given the population dynamic parameters under this model. In the initial release of *bdmm*, analyses were computationally limited to trees consisting of up to approximately 250 genetic samples. We implemented important algorithmic changes to *bdmm* which dramatically increased the number of genetic samples that could be analyzed and which improved the numerical robustness and efficiency of the calculations. Including more samples led to the improved precision of parameter estimates, particularly for structured models with a high number of inferred parameters. Furthermore, we report on several model extensions to *bdmm*, inspired by properties common to empirical datasets. We applied this improved algorithm to two partly overlapping datasets of the Influenza A virus HA sequences sampled around the world—one with 500 samples and the other with only 175—for comparison. We report and compare the global migration patterns and seasonal dynamics inferred from each dataset. In this way, we show the information that is gained by analyzing the bigger dataset, which became possible with the presented algorithmic changes to *bdmm*. In summary, *bdmm* allows for the robust, faster, and more general phylodynamic inference of larger datasets.

Keywords: phylogenetics; Bayesian inference; phylodynamics; population structure

check for updates

Citation: Scire, J.; Barido-Sottani, J.; Kühnert, D.; Vaughan, T.G.; Stadler, T. Robust Phylodynamic Analysis of Genetic Sequencing Data from Structured Populations. *Viruses* **2022**, *14*, 1648. https://doi.org/10.3390/v14081648

Academic Editor: Viktor Müller

Received: 9 June 2022
Accepted: 22 July 2022
Published: 27 July 2022

1. Introduction

Genetic sequencing data taken from a measurably evolving population contain fingerprints of past population dynamics [1]. In particular, the phylogeny spanning the sampled genetic data contains information about the mixing pattern of different populations and thus contains information beyond what is encoded in classic occurrence data; see, e.g., Hey and Machado [2], Stadler and Bonhoeffer [3]. Phylodynamic methods [4,5] aim at quantifying past population dynamic parameters, such as migration rates, from genetic sequencing data. Such tools have been widely used to study the spread of infectious diseases in structured populations; see, e.g., Dudas et al. [6], Faria et al. [7] as examples of analyses of recent epidemic outbreaks. The Bayesian phylodynamic inference framework BEAST2 [8] is one of the software frameworks within which such analyses can be carried out. With BEAST2, tree topologies, parameters from phylodynamic, molecular clock, and substitution models can be jointly inferred via Markov-Chain Monte-Carlo (MCMC) sampling. Both the host population and the pathogen population may be structured (e.g., the host population may be geographically structured), and the pathogen population may consist of a drug-sensitive and a drug-resistant subpopulation. Understanding how these subpopulations interact with one another—whether they are separated by geographic distance, lifestyles of the hosts, or other barriers—is a key determinant in understanding how an epidemic spreads. In macroevolution, different species may also be structured into different "subpopulations", e.g., due to their geographic distribution or to trait variations; see, e.g., Hodges [9]. Phylodynamic tools aim at quantifying the rates at which species migrate or the rates at which traits are gained or lost, as well as the rates of speciation and extinction within the "subpopulations"; see, e.g., Goldberg et al. [10], Mayrose et al. [11], Goldberg et al. [12].

The phylodynamic analysis of structured populations can be performed using two classes of models, namely coalescent-based and birth–death-based approaches. Both have their unique advantages and disadvantages [13,14]. Here, we report major improvements on a multi-type birth–death-based approach.

A multi-type birth–death model is a linear birth–death model accounting for structured populations. Under this model, the probability density of a phylogenetic tree can be calculated by numerically integrating a system of differential equations. The use of this model within a phylodynamic setting and the associated computational approach were initially proposed for the analysis of species phylogenies [15] and later for the analysis of pathogen phylogenies [3,13]. The BEAST2 package *bdmm* generalizes the assumptions of these two initial approaches [16]. It further allows for co-inferring phylogenetic trees together with the model parameters and thus explicitly takes phylogenetic uncertainty into account. Datasets containing more than 250 genetic sequences could not be analyzed using the original *bdmm* package [16] due to numerical instability. This limitation was a strong impediment to obtaining reliable results, particularly for the analysis of structured populations as the quantification of parameters which characterize each subpopulation requires a significant amount of samples from each of them. The instability was due to numerical underflow in the probability density calculations, which meant that probability values extremely close to zero could not be accurately calculated and stored. We were able to solve the numerical instability issue of *bdmm*, thereby lifting the hard limit on the number of samples that could be analyzed. In addition, the practical usefulness of the *bdmm* package had previously been restricted by the amount of computation time required to carry out the analyses. Here, we report on significant improvements in computation efficiency. As a result, *bdmm* can now handle datasets containing several hundred genetic samples. Finally, we made the multi-type birth–death model more general in several ways: homochronous sampling events can now occur at multiple time points (not only the present), individuals are no longer necessarily removed upon sampling, and the migration rate specification has been made more flexible by allowing for piecewise-constant changes through time.

Overall, these model generalizations and implementation improvements enable more reliable and ambitious empirical data analyses. Below, we use the new release of *bdmm* to quantify the Influenza A virus spread around the globe as a sample application and compare the results obtained with those from the reduced dataset analyzed in Kühnert et al. [16].

2. Methods

2.1. Description of the Extended Multi-Type Birth–Death Model

First, we formally define the multi-type birth–death model on d types [16], including the generalizations introduced in this work. The process starts at time 0 with one individual; this is also called the origin of the process and the origin of the resulting tree. This individual is of type $i \in \{1 \ldots d\}$, with a probability of h_i (where $\sum_{i=1}^{d} h_i = 1$). The process ends after T time units (in the present). The time interval $(0, T)$ is partitioned into n intervals through $0 < t_1 < \ldots < t_{n-1} < T$, and we define $t_0 := 0$ and $t_n := T$. Each individual at time t, $t_{k-1} \leq t < t_k$, $k \in \{1 \ldots n\}$ of type $i \in \{1 \ldots d\}$, gives rise to an additional individual of type $j \in \{1 \ldots d\}$, with a birth rate of $\lambda_{ij,k}$, migrates to type j with a rate of $m_{ij,k}$ (with $m_{ii,k} = 0$), dies with a rate of $\mu_{i,k}$, and is sampled with a rate of $\psi_{i,k}$. At time t_k, each individual of type i is sampled with a probability of $\rho_{i,k}$. Upon sampling (either with a rate of $\psi_{i,k}$ or a probability of $\rho_{i,k}$), the individual is removed from the infectious pool with a probability of $r_{i,k}$. We summarize all birth rates $\lambda_{ij,k}$ in $\boldsymbol{\lambda}$, migration rates $m_{ij,k}$ in $\boldsymbol{m}$, death rates $\mu_{i,k}$ in $\boldsymbol{\mu}$, sampling rates $\psi_{i,k}$ in $\boldsymbol{\psi}$, sampling probabilities $\rho_{i,k}$ in $\boldsymbol{\rho}$, and removal probabilities $r_{i,k}$ in $\boldsymbol{r}$, $i, j \in \{1, \dots, d\}, k \in \{1, \dots, n\}$. The model described in Kühnert et al. [16] is a special case of the above and assumes that migration rates are constant through time (i.e., do not depend on k), removal probabilities are constant through time and across types (i.e., do not depend on k and i), and that $\rho_{i,k} = 0$ for $k < n$ and $i \in \{1 \dots d\}$.

This process gives rise to complete trees on sampled and non-sampled individuals, with types being assigned to all branches at all times (Figure 1, left). Following each branching event, one offspring is assigned to be the "left" offspring and another to be the "right" offspring, with each assignment having a probability of $\frac{1}{2}$. In the figure, we draw the branch with the assignment "left" on the left and the branch with the assignment "right" on the right. Such trees are called oriented trees, and considering oriented trees will facilitate the calculation of the probability densities of trees. Pruning all lineages without sampled descendants leads to the *sampled phylogeny* (Figure 1, middle and right). The orientation of a branch in the sampled phylogeny is the orientation of the corresponding branch descending from the common branching event in the complete tree. When the sampled phylogeny is annotated with the types along each branch, we refer to it as a *branch-typed tree* (Figure 1, middle). On the other hand, if we discard these annotations but keep the types of the sampled individuals, we call the resulting object a *sample-typed (or tip-typed) tree* (Figure 1, right).

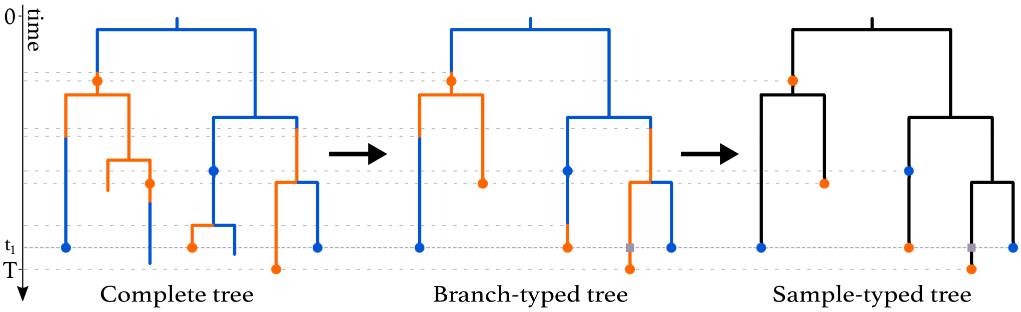

Figure 1. Complete tree (**left**) and sampled trees (**middle** and **right**) obtained from a multi-type birth–death process with two types. The orange and blue dots on the trees represent sampled individuals and are colored according to the type these individuals belong to. A ρ-sampling event happens at time t_1. The grey squares represent degree-2 nodes added to branches crossing this event. ρ-sampling also happens in the present (time T). As seen in the complete tree, the three individuals who were first sampled were not removed from the population upon sampling, whereas the three individuals sampled at time t_1 were.

Here, we give an overview of the computation of the probability density of the sampled tree (i.e., the sample-typed or branch-typed tree) given the multi-type birth–death parameters $\boldsymbol{\lambda}$, $\boldsymbol{m}$, $\boldsymbol{\mu}$, $\boldsymbol{\psi}$, $\boldsymbol{\rho}$, $\boldsymbol{r}$, $\boldsymbol{h}$, T. This probability density is obtained by integrating probability densities g from the leaf nodes (or "tips"), backwards along all the edges (or "branches") to the origin of the tree. Our notation here is based on previous work [16,17], and the probabilities $p_{i,k}(t)$ and $g^{e}_{i,k}(t)$ relate to E and D in Stadler and Bonhoeffer [3], Maddison et al. [15], respectively.

Every branching event in the sampled tree gives rise to a node of degree 3 (i.e., 3 branches are attached). Every sampling event gives rise to a node of degree 2 (called "sampled ancestor") or 1 (called "tip", as defined above). A sampling event at time $t = t_k$, $k \in \{1, \dots, n\}$, is referred to as a ρ-sampled node. All other nodes corresponding to samples are referred to as ψ-sampled nodes.

Furthermore, degree-2 nodes are placed at time t_k on all lineages crossing time t_k, $k = 1, \dots, n-1$, as shown at time t_1 in Figure 1. In a branch-typed tree, a node of degree 2 also occurs on a lineage at a time point when a type change occurs. Such type changes may be the result of either migrations or birth events in which one of the descendant subtrees is unsampled (Figure 1, middle).

We highlight that in *bdmm*, we assume that the most recent sampling event happens at time T. This is equivalent to assuming that the sampling effort was terminated directly after the last sample was collected and overcomes the necessity for users to specify the time between the last sample and the termination of the sampling effort at time T.

The derivation of the probability density of a sampled tree under the extended multi-type birth–death model is developed in Appendix A. This probability density, also called a "phylodynamic likelihood", can be used to estimate the multi-type birth–death parameters $\boldsymbol{\lambda}$, $\boldsymbol{\mu}$, $\boldsymbol{\psi}$, $\boldsymbol{m}$, T, when used in a Bayesian phylodynamic framework such as BEAST 2 by Bouckaert et al. [8]. Note that unlike other parameters of this model, $\boldsymbol{h}$ is typically not estimated via MCMC sampling. h_i values can be set according to different rationales: the root type can be fixed to a particular type k ($h_k = 1$, $h_i = 0$ for $i \neq k$), or all types can be equally likely ($h_i = \frac{1}{n}$), or they can be set to the equilibrium distribution (derived by Stadler and Bonhoeffer [3]) given that the process was already in equilibrium at the time of origin.

2.2. Implementation Improvements

The computation of the probability densities of sampled trees under the multi-type birth–death model requires the numerical solving of Ordinary Differential Equations (ODEs) along each tree branch. We were able to significantly improve the robustness of the original *bdmm* implementation, which suffered from instabilities caused by the underflow of these numerical calculations. Compared to the original implementation, we prevented the underflow by implementing an extended precision floating point representation (EPFP) for storing intermediary calculation results. In addition to this improvement in stability, we improved the efficiency of the probability density calculations by (1) using an adaptive step-size integrator for numerical integration, (2) performing preliminary calculations and storing the results for use during the main calculation step, and (3) distributing calculations among threads running in parallel. Details can be found in Appendix B.

The latest release with our updates, *bdmm* 1.0, is freely available as an open access package of BEAST 2. The source code can be accessed at https://github.com/denisekuehnert/bdmm (accessed on July 26th 2022).

3. Results

3.1. Evaluation of Numerical Improvements

We compared the robustness and efficiency of the improved *bdmm* package against its original version. We considered tree sizes varying between 10 and 1000 samples. For each tree size, we simulated 50 branch-typed and 50 sample-typed trees under the multi-type birth–death model using randomly drawn parameter values from the distributions shown in Table A1. The distributions from which the parameters were drawn were selected to reflect a wide range of scenarios. For each simulated tree, we measured the time taken to perform the calculation of the probability density, given the parameter values under which the tree was simulated, using the updated and the original *bdmm* implementation. We report here the wall-clock time taken to perform this calculation 5000 times (Figure 2). All computations were performed on a MacBook Pro with a dual-core 2.3 GHz Intel Core i5 processor. The new implementation of *bdmm* is on average nine times faster than the original (Figure 2A). The robustness of the updated implementation was demonstrated by reporting how often the implementations returned $-\infty$ for the probability density in log space. We call these calculations "failures", the most likely cause of error being the underflow. Our new implementation showed no calculation failure for trees containing a thousand samples, while in the original implementation calculations often failed for trees with more than 250 samples (Figure 2B).

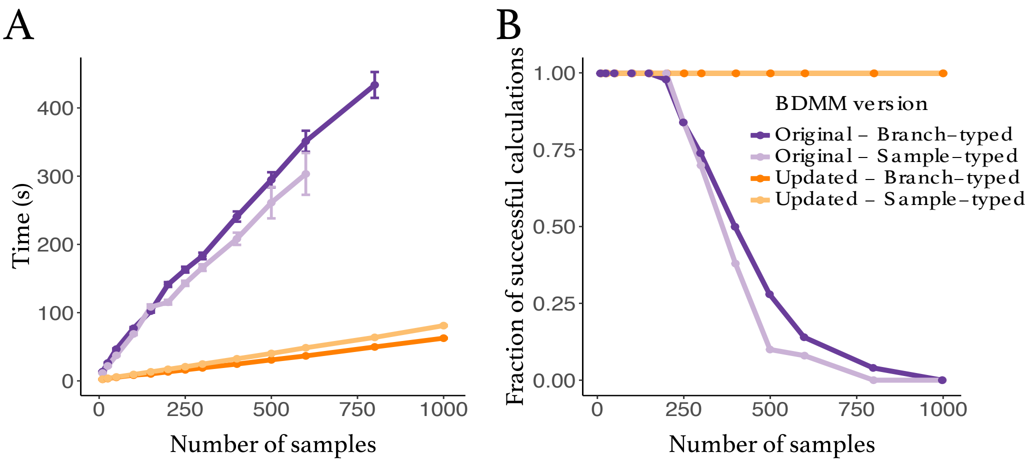

Figure 2. Comparison between the original and the updated implementations of the multi-type birth–death model. (**A**) Speed comparison. Only successful calculations were taken into account, i.e., calculations where the log probability density was different from $-\infty$. (**B**) Success in calculating probability density values plotted against tree size. The values presented in this panel correspond to the same set of calculations as the one in panel (**A**).

3.2. Validation Against Original Implementation

To ensure that no errors were introduced into the updated *bdmm*, we validated the improved implementation against the original *bdmm* version by comparing the results of likelihood computation on a handful of randomly simulated trees. We used simulated trees with 10 or 100 tips, well below the limit of reliability of the original *bdmm* version (approximately 250 tips). Details of this procedure can be found in Appendix C. Figure 3 shows one of such simulated trees along with tree likelihood values (or the probability density of the sampled tree given the multi-type birth–death parameters) computed with each *bdmm* version. Likelihood computation results are identical for all trees and parameters tested for both implementations (difference in log-likelihoods $< 1 \times 10^{-6}$). Figure A3 shows that the same results were obtained with other trees or when varying other parameters. Therefore, we conclude that the results of the full validation, along with error and bias assessment performed by Kühnert et al. [16] on the original *bdmm* version, hold true for the improved *bdmm* implementation we present in this article.

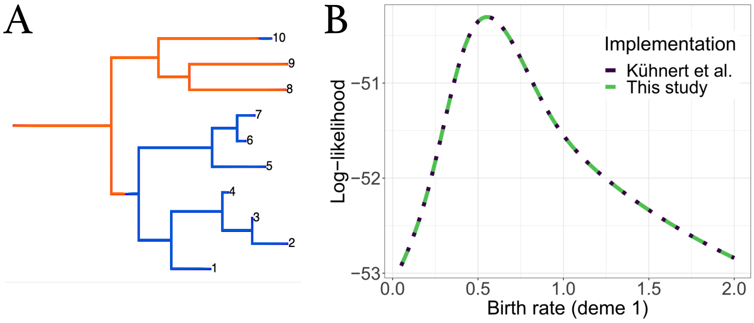

Figure 3. Comparison of computation results between the original *bdmm* and improved *bdmm* versions. (**A**) Randomly simulated tree with 10 tips and 2 demes, used for comparison. (**B**) Log-likelihood values obtained with each *bdmm* version as a function of λ_1 (birth rate of orange deme).

3.3. Influenza A virus (H3N2) Analysis

As an example of a biological question that can be investigated with the use of *bdmm*, we analyzed 500 H3N2 influenza virus HA sequences sampled around the globe from 2000 to 2006; we aimed to recover the seasonal dynamics of the global epidemics. The dataset is a random subset of the data analyzed by Vaughan et al. [18], taken from three different regions around the globe: New York (North, $n = 167$), New Zealand (South, $n = 215$), and

Hong Kong (Tropics, $n = 118$). The dataset of 980 samples assembled by Vaughan et al. [18] was built with the aim of gathering samples from three locations with relatively similar population sizes, each representative of the northern, southern, or equatorial regions.

As a comparison, we performed an identical analysis of the H3N2 influenza dataset with 175 sequences sampled between 2003 and 2006 that was used in [16]. This dataset of 175 sequences was also a subset of the data by Vaughan et al. [18], and it also grouped samples from three locations denoted as North (for the northern hemisphere), South (for the southern hemisphere), and Tropic (for tropical regions). Note that the latter dataset came from more geographically spread samples, and thus we did not expect results from both analyses to be perfectly comparable. As we were dealing with pathogen sequence data, we adopted the epidemiological parametrization of the multi-type birth–death model, as detailed in Kühnert et al. [16]. The epidemiological parametrization substitutes birth, death, and sampling rates with effective reproduction numbers within types, rate at which hosts become noninfectious, and sampling proportions. To study the seasonal dynamics of the global epidemic, we allowed the effective reproduction number R_e to vary through time. To do so, we subdivided time into six-month intervals (starting April 1st and October 1st), and we constrained the effective reproduction number values corresponding to the same season across different years to be equal for each particular location, assuming that the R_e values were the same in the summer seasons and the same in the winter seasons. The testing of this hypothesis' validity by estimating the R_e values that varied in each six-month period was not performed as we expected little information from the data for the additional parameters. For the same reason, the migration rate was not varied through time. Further details on the data analysis configuration can be found in Appendix D.

The analysis of the larger dataset (500 samples) allowed for the reconstruction of the phylogenetic tree encompassing a longer time period, and therefore gave a more long-term and detailed view of the evolution of the global epidemic (see Figure 4 for the Maximum Clade Credibility trees).

As can be expected for the tropical location, in both analyses, the effective reproduction numbers for H3N2 influenza A were inferred to be close to one throughout the year (Figure 5A). Conversely, strong seasonal variations can be observed in the Northern and Southern hemisphere locations, where the effective reproduction number was close to one in winter and was much lower in summer. Inferences from the small and large datasets are mostly in agreement. A subtle difference appears: in the larger dataset, the effective reproduction numbers in the winter seasons and in the tropical location are closer to one, with less variation across estimates. This seems to indicate that the variations between estimates observed with the smaller dataset, including samples from 2003 to 2006 (for instance, R_e in winter in the North compared to R_e in winter in the South), are due to stochastic fluctuations, which are averaged out when considering a longer period of transmission dynamics in the larger dataset covering the years 2000–2006.

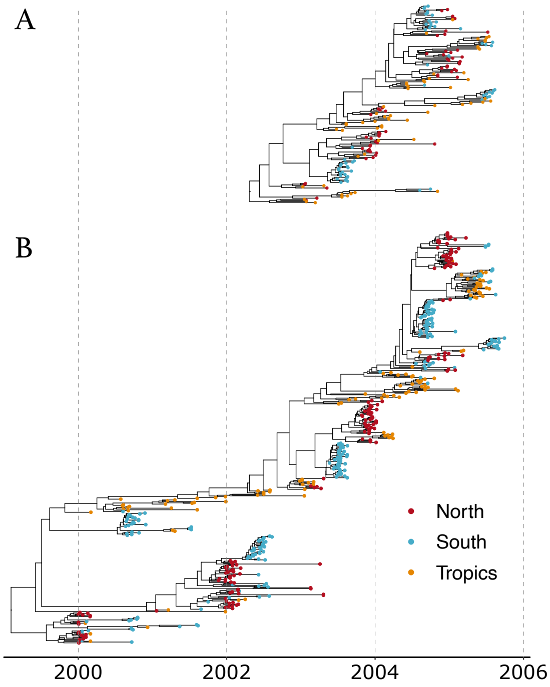

Figure 4. Maximum Clade Credibility (MCC) trees from analyses of (**A**) 175 samples and (**B**) 500 samples.

The precise inference of migration rates is more difficult, as reflected by the significant uncertainty we obtained on the estimates (Figure 5B). However, we still observed that the uncertainty was generally reduced for the inference performed with the larger dataset, as expected. A significant difference between the migration rates inferred from the Southern to Tropical locations arose between the two analyses. With the larger dataset, the estimated rate was much lower than that with the smaller one, and it was more in range with the other migration rate estimates. Detailed results of all the parameter estimates for both analyses are available in Table A4. Most notably, the estimates of the root locations for both datasets are very similar. In both cases, the tropical location is most likely to be the location of the root; however, neither of the two other locations can be entirely excluded.

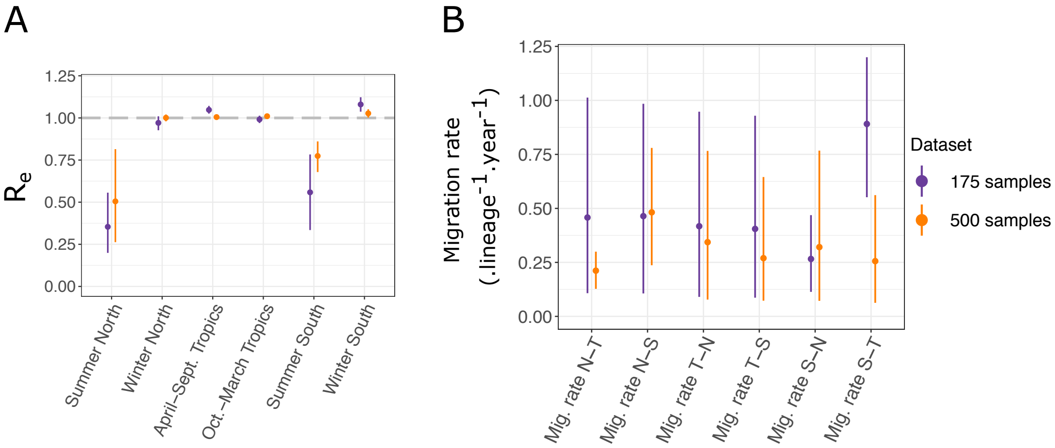

Figure 5. (**A**) Seasonal effective reproduction numbers for each sample location, for both datasets. (**B**) Migration rates inferred for each dataset. N, S, and T refer respectively to North, South, and Tropics. For instance, "*Mig. rate N-T*" represents the migration rate from the Northern location to the Tropical one.

3.4. Properties of bdmm

3.4.1. Identifiability of Parameters

In birth–death models with sampling through time or in the present, only two of the three parameters of birth rate, death rate, and sampling rate/sampling probability can be jointly estimated [19,20]. Thus, independent prior data need to be employed to quantify all three parameters. In recent work, the question of identifiability in time-dependent birth–death sampling models has been thoroughly investigated [20,21]. The issue of identifiability in state-dependent birth–death sampling models remains, to our knowledge, largely unanswered. The interactions between migration rates, rates of birth-among-demes, and other multi-type birth–death parameters is not well-known. It is likely that different parameter combinations of the multi-type birth–death model can yield the same likelihood value. Informative prior information on some of the birth–death parameters mitigates parameter non-identifiability issues.

3.4.2. Computational Costs

Despite the implementation improvements presented in this manuscript, phylodynamics analyses performed in *bdmm* are still limited in practice by the number of genetic sequences they can handle. This limitation, unlike the previously existing limitation caused by underflow, is not a hard boundary but rather a soft boundary imposed by the practical constraints of computational analyses. Limitations with regard to the complexity of the analyses that could be carried out with the improved version of *bdmm* derive from the time required to carry out computations and from the complexity of the probability space that must be explored. For instance, each MCMC chain for the 500-sample Influenza A analysis required about 15 days to compute. Keeping the same analysis setup and increasing the number of genetic samples will have a linear effect on the time required to compute the phylodynamic likelihood with *bdmm*. With our updated *bdmm* implementation, the core bottleneck is the complexity of exploring tree space, which increases exponentially with additional samples. Due to this complexity, only trees with up to around 1000 samples can be successfully estimated with BEAST2.

4. Discussion

The multi-type birth–death model with its updated implementation in the *bdmm* package for BEAST 2 provides a flexible method for taking into account the effect of population structure when performing a phylodynamic genetic sequence analysis. Compared to the original implementation, we now prevent the underflow of numerical calculations and speed up calculations by almost an order of magnitude. The size limit of around 250 samples for datasets that could be handled by *bdmm* is thus lifted, and significantly larger datasets can now be analyzed. Now, the bottleneck lies in the search for tree space with MCMC rather than with *bdmm*. We demonstrate this improvement by analyzing two datasets of Influenza A virus H3N2 genetic data from around the globe. One dataset has 500 samples and could not have been analyzed with the original version of *bdmm*, while the other one contains 175 samples and is the original sample dataset analyzed in [16]. Overall, we observed that analyzing a dataset with more samples, as expected, gives a more long-term picture of the global transmission patterns and reduces the general uncertainty concerning parameter estimates.

With the addition of the so-called ρ-sampling events in the past, intense sampling efforts limited to short periods of time (leading to many samples being taken nearly simultaneously) can be easily modeled as instantaneous sampling events across the entire population (or subpopulation) rather than as non-instantaneous sampling over small sampling intervals. This simplifies the modeling of intense pathogen sequencing efforts in very short time windows. By allowing the removal probability r (the probability for an individual to be removed from the infectious population upon sampling) to be type-dependent and to vary across time intervals, as well as allowing migration rates between types to vary across time intervals, we further increase the generality and flexibility of the multi-type birth–death model. A sample *bdmm* analysis with a ρ-sampling event in the past was added to the software package to guide users who may want to set up such an analysis with their own data.

We focused on an epidemiological application of *bdmm*, where we co-infer the phylogenetic trees to take into account the phylogenetic uncertainty. However, the *bdmm* modeling assumptions are equally applicable to the analysis of macroevolutionary data, in which context *bdmm* allows for the joint inference of trees with fossil samples under structured models [22]. When using a multi-type birth–death model in the macroevolutionary framework, ρ-sampling can be used to model fossil samples originating from the same rock layer. In the context of the exploration of trait-dependent speciation, structured birth–death models such as the binary-state speciation and extinction model (BiSSE) [23,24] have been shown to possibly produce spurious associations between character state and speciation rate when applied to empirical phylogenies [25]. When used in this fashion, users of *bdmm* should assess the propensity of their dataset analysis for Type I errors through neutral trait simulations, as suggested by Rabosky and Goldberg [25].

In summary, the new release of *bdmm* overcomes several constraints when analyzing sequencing data in BEAST2. As it stands, the main constraint now is given by the efficiency of the BEAST2 MCMC tree space sampler rather than *bdmm* itself. We expect the new release of *bdmm* to become a standard tool for the phylodynamic analysis of sequencing data and phylogenetic trees from structured populations.

Supplementary Materials: The following are available at https://www.mdpi.com/article/10.3390/v14081648/s1. The XML files for the replication of the BEAST2 analyses of the two datasets with 175 and 500 samples of the H3N2 influenza virus HA sequences are available as supplementary files. To run these XML files using BEAST2, the *bdmm* and *feast* packages must be installed.

Author Contributions: Conceptualization, J.S., J.B.-S., D.K., T.G.V., and T.S.; software, J.S., D.K., and T.G.V.; writing—original draft preparation, J.S., T.G.V., and T.S.; writing—review and editing, J.S., J.B.-S., D.K., T.G.V., and T.S.; supervision, T.S. All authors have read and agreed to the published version of the manuscript.

Funding: T.S. was supported in part by the European Research Council under the Seventh Framework Programme and under Horizon 2020 of the European Commission (PhyPD: grant agreement number 335529, PhyCogy: grant agreement number 101001077).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are either included in Supplementary Materials 4 or openly available at https://github.com/jscire/bdmm_paper_code (accessed on July 26th 2022).

Acknowledgments: We thank Nicola Müller, David Rasmussen, Jūlija Pečerska for very valuable discussions and Fábio Kuriki Mendes for helpful comments on the manuscript.

Conflicts of Interest: The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

Appendix A. Derivation of the Probability Density of a Sampled Tree

Appendix A.1. Probability of Having No Sampled Descendants

In order to compute the probability density of a sampled tree, we need to calculate the probability of an individual with type $i \in \{1 \dots d\}$ at time $t_{k-1} \leq t \leq t_k$ to have no sampled descendant, which we denote by $p_{i,k}(t)$. In the interval $t_{k-1} \leq t < t_k$, this probability satisfies the ordinary differential equation (ODE)

$$\begin{aligned}-\frac{d}{dt}p_{i,k}(t) = &-\left(\sum_{j=1}^{d}(\lambda_{ij,k}+m_{ij,k})+\mu_{i,k}+\psi_{i,k}\right)p_{i,k}(t)\\ &+\sum_{j=1}^{d}\lambda_{ij,k}p_{i,k}(t)p_{j,k}(t)+\sum_{j=1}^{d}m_{ij,k}p_{j,k}(t)+\mu_{i,k}.\end{aligned} \tag{A1}$$

The terms on the right-hand side in the first line correspond to the probability of no event happening, and the terms in the second line correspond to either a branching, migration, or death event happening. For a detailed derivation via Master equations, the reader is referred to Maddison et al. [15] or Stadler and Bonhoeffer [3].

For $t = t_k$, $k \in \{0 \dots n\}$, we have

$$p_{i,k}(t_k) = (1-\rho_{i,k}) \times \begin{cases}1 & \text{if } k = n,\\ p_{i,k+1}(t_k) & \text{otherwise.}\end{cases} \tag{A2}$$

Appendix A.2. Probability Density of a Sample-Typed Subtree

We use $g^e_{i,k}(t)$ to denote the probability density that an individual represented by branch e at time t (with $t_{k-1} \le t \le t_k$, $k \in \{1 \dots n\}$) in state $i \in \{1, \dots, d\}$ has evolved between t and T, as observed in the tree. Branch e is connected to two nodes, and we denote the more recent node with n_e, occurring at time $t_{k-1} < t_e \le t_k$. Node n_e represents either a sampling event (Figure A1A), a branching event (Figure A1B), or a degree-2 node at $t_e = t_k$ with or without sampling (Figure A1C,D). One can show that $g^e_{i,k}(t)$ for $t < t_e$ satisfies the ODE

$$\begin{aligned}-\frac{d}{dt}g^e_{i,k}(t) = &-\left(\sum_{j=1}^{d}(\lambda_{ij,k}+m_{ij,k})+\mu_{i,k}+\psi_{i,k}\right)g^e_{i,k}(t)\\ &+\sum_{j=1}^{d}m_{ij,k}g^e_{j,k}(t)+\sum_{j=1}^{d}\lambda_{ij,k}p_{j,k}(t)g^e_{i,k}(t)+\sum_{j=1}^{d}\lambda_{ij,k}p_{i,k}(t)g^e_{j,k}(t),\end{aligned} \tag{A3}$$

with the derivation being analogous to that of $p_{i,k}(t)$.

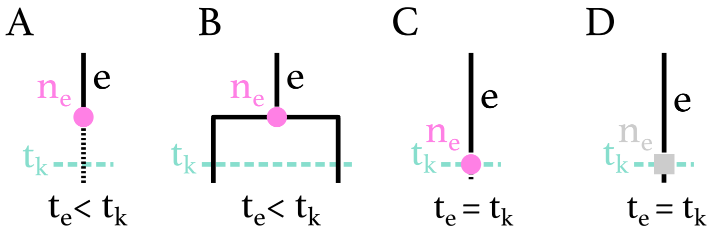

Figure A1. Possible configurations for node n_e on branch e at time $t_{k-1} < t_e \le t_k$. (**A**) n_e is a ψ-sampled node at time $t_e < t_k$, with or without sampled descendants. (**B**) n_e is a branching event at time $t_e < t_k$. (**C**) n_e is a ρ-sampling node at time $t_e = t_k$, with or without sampled descendants. (**D**) n_e is a degree-2 node at time $t_e = t_k$ without sampling.

We denote the branch (or two branches) descending from n_e at time t_e with e_1 (respectively e_1 and e_2). The initial conditions for the differential equations at node n_e, i.e., the values of the probability densities at the most recent end of the branch e, are as follows:

$$g^e_{i,k}(t_e) = \begin{cases} \psi_{i,k}(r_{i,k} + (1-r_{i,k})p_{i,k}(t_e)) & \text{if } n_e \text{ is a } \psi\text{-sampled tip of type } i, \\ \psi_{i,k}(1-r_{i,k})g^{e_1}_{i,k}(t_e) & \text{if } n_e \text{ is a } \psi\text{-sampled ancestor of type } i, \\ \rho_{i,k}(r_{i,k} + (1-r_{i,k})p_{i,k+1}(t_e)) & \text{if } n_e \text{ is a } \rho\text{-sampled tip of type } i, \\ \rho_{i,k}(1-r_{i,k})g^{e_1}_{i,k+1}(t_e) & \text{if } n_e \text{ is a } \rho\text{-sampled ancestor of type } i, \\ 0 & \text{if } n_e \text{ is a sample of type } j \neq i, \\ (1-\rho_{i,k})g^{e_1}_{i,k+1}(t_e) & \text{if } n_e \text{ is not a sample and } t_e = t_k, \\ \frac{1}{2}\sum_{j=1}^{d}\lambda_{ij,k}\left[g^{e_1}_{i,k}(t_e)g^{e_2}_{j,k}(t_e) + g^{e_1}_{j,k}(t_e)g^{e_2}_{i,k}(t_e)\right] & \text{if } n_e \text{ has two descendant branches.} \end{cases} \tag{A4}$$

The $\frac{1}{2}$ in the last equation is needed since we are computing the probability density of an oriented tree.

Appendix A.3. Probability Density of a Branch-Typed Subtree

When inferring branch-typed trees, we condition on the type of a branch at all times. Thus, we do not integrate over migration events or unobserved branching events that change the type of the tree lineage. We define $\eta_{i,j} = 1$ for $i \neq j$ and $\eta_{i,j} = 2$ for $i = j$. Equation (A3) is replaced by

$$-\frac{d}{dt}g^e_{i,k}(t) = -\left(\sum_{j=1}^{d}(\lambda_{ij,k} + m_{ij,k}) + \mu_{i,k} + \psi_{i,k}\right)g^e_{i,k}(t) + \sum_{j=1}^{d}\eta_{i,j}\lambda_{ij,k}p_{j,k}(t)g^e_{i,k}(t), \tag{A5}$$

with the initial conditions

$$g^e_{i,k}(t_e) = \begin{cases} \psi_{i,k}\,(r_{i,k} + (1-r_{i,k})\,p_{i,k}(t_e)) & \text{if } n_e \text{ is a } \psi\text{-sampled tip of type } i, \\ \psi_{i,k}\,(1-r_{i,k})\,g^{e_1}_{i,k}(t_e) & \text{if } n_e \text{ is a } \psi\text{-sampled ancestor of type } i, \\ \rho_{i,k}\,(r_{i,k} + (1-r_{i,k})\,p_{i,k+1}(t_e)) & \text{if } n_e \text{ is a } \rho\text{-sampled tip of type } i, \\ \rho_{i,k}\,(1-r_{i,k})\,g^{e_1}_{i,k+1}(t_e) & \text{if } n_e \text{ is a } \rho\text{-sampled ancestor of type } i, \\ 0 & \text{if } n_e \text{ is a sampled tip of type } j \neq i, \\ (1-\rho_{i,k})\,g^{e_1}_{i,k+1}(t_e) & \text{if } n_e \text{ is not a sample and } t_e = t_k, \\ m_{ij,k}\,g^{e_1}_{j,k}(t_e) + \lambda_{ij,k}\,g^{e_1}_{j,k}(t_e)\,p_{i,k}(t_e) & \text{if } n_e \text{ has one descendant branch with type } j \neq i, \\ \frac{1}{2}\eta_{i,j}\lambda_{ij,k}\,g^{e_1}_{i,k}(t_e)\,g^{e_2}_{j,k}(t_e) & \text{if } n_e \text{ has two descendant branches.} \end{cases} \tag{A6}$$

Here, the $\frac{1}{2}$ in the last equation is also needed since we are computing the probability density of an oriented tree. In branch-typed trees, a branch is always of one single type. The type of branch e being i implies that $g^e_{j,k}(t) = 0$ for $j \neq i$. Indeed, Equation (A6) states that $g^e_{j,k}(t_e) = 0$ for $i \neq j$. Furthermore, Equation (A5) specifies $g^e_{j,k}(t) = 0$ for all $t < t_e$.

Appendix A.4. Probability Density of a Sampled Tree

The probability density of a sampled tree, with the lineage at time $t = 0$ being of type i and the branch being labelled with e, is the product of the probability density that the individual evolved as observed in the tree ($g^e_{i,1}(0)$ and the probability h_i that the individual at the start of the process is in type i.

Hence, the probability density of an oriented sampled tree $\mathcal{T}$ under the multi-type birth–death model is

$$f(\mathcal{T}|\boldsymbol{\lambda}, \boldsymbol{\mu}, \boldsymbol{\psi}, \boldsymbol{m}, \boldsymbol{h}, T) = \sum_{i=1}^{d} h_i\, g^e_{i,1}(0).$$

In BEAST 2, we infer labeled sampled trees; thus, we need to calculate the probability of a labeled sampled tree. In a labeled sampled tree, each sample has a unique label, and orientations at branching events are ignored. In order to obtain the probability density of a labeled sampled tree, we need to transform the oriented tree probability density by multiplying by $2^M/N!$, where M is the number of branching events in the tree, and N is the number of samples [19]. The probability density of a labeled tree $\mathcal{T}$ under the multi-type birth–death model is thus

$$\tilde{f}(\mathcal{T}|\boldsymbol{\lambda}, \boldsymbol{\mu}, \boldsymbol{\psi}, \boldsymbol{m}, \boldsymbol{h}, T) = \frac{2^M}{N!} \sum_{i=1}^{d} h_i \, g_{i,0}^e(0). \tag{A7}$$

Appendix B. Improved Implementation of the Tree Probability Density Evaluation

The core component of the *bdmm* package is the evaluation of the tree probability density given by Equation (A7). This involves numerical integration to solve the system of ODEs defined by Equations (A1) and (A2), and by either Equations (A3) and (A4) (for sample-typed trees) or Equations (A5) and (A6) (for branch-typed trees).

In what follows, we discuss the numerical stability of our calculations.

Appendix B.1. Extended Numerical Representation

The traditional floating point representation of a real number x is the closest number $\hat{x}$ to x, where

$$\hat{x} = m \times 2^q \tag{A8}$$

and where m (the *mantissa*) and q (the *exponent*) are signed binary integers of a specified number of bits. Note that the mantissa is understood to represent a fixed-precision number between 1 and 2. In the standard 64-bit double-precision floating point (DPFP) used by the original *bdmm* implementation [16], m is restricted to 53 bits and q is restricted to 11 bits. Ignoring the mantissa, this implies that the smallest, representable, non-zero absolute value is on the order of $2^{-2^{10}} \simeq 10^{-320}$. We calculate a probability density $\tilde{f}$ over a tree space (together with intermediate values $g_{i,k}^e(t)$ as part of the ODE integration), and these values can fall well below the lower limit imposed by the standard 64-bit DPFP representation, resulting in underflow errors.

To work around this limitation, our new implementation of *bdmm* employs a 96-bit extended precision floating point representation (EPFP), in which the mantissa is represented using a standard 64-bit double precision floating point number, and the exponent is represented using a 32-bit signed integer value. This dramatically extends the range of possible values; in particular, the smallest non-zero absolute value is $2^{-2^{31}} \simeq 10^{-646456993}$.

A naive approach to employing the EPFP representation would be to use numbers of this kind at each and every step in the probability density calculation. This would ensure that all intermediate calculation results were accurately stored and would allow for accurate calculations at the next step.

However, this approach has two main drawbacks. Firstly, existing numerical integration libraries almost exclusively use the DFPF representation. While integration algorithms could certainly be implemented for other representations such as the EPFP, this would be extremely time-consuming. Secondly, since DPFP calculations are implemented at a hardware level on modern processors, calculations using this representation are performed very efficiently. Abandoning this primitive data type thus makes basic calculation steps considerably less efficient.

For these reasons, our approach involves a mixture of using both representations. The numerical integration of the ODEs along each branch of the sampled phylogeny (Equations (A3) and (A5)) is performed using the DPFP representation, while the combination of these integration results—accomplished to provide the boundary condition for the next integration (Equations (A4) and (A6))—is achieved using the EPFP representation.

In order to avoid underflow at all times, the EPFP numbers are scaled before they are converted to the DPFP representation used by the integrator. The scaling procedure amounts to an integration by substitution. We scale $g_{i,k}^e(t_k)$ by the linear substitution function $G_{i,k}^e(t_k) = \delta \times g_{i,k}^e(t_k)$, with δ being a scale factor. Since the original Equations (A3) and (A5) are both linear in g_i^e, this scaling does not affect the results of their integration once the inverse scaling at time t_{k-1} has been completed. The method employed to choose an appropriate scale factor is described in detail below. The goal of scaling is to make sure all values stored as DPFP actually fit the window of values that this representation can express. After each numerical integration step, the results are converted back to EPFP and rescaled accordingly. The differential equation for the probability $p_{i,k}(t)$ of having no sampled descendants is not scaled at all as its integration does not cause any underflows in practice.

Appendix B.2. Choice of a Scale Factor

The scale factor δ is shared among all subpopulations. Therefore, it has to be carefully chosen so that all initial conditions fit into the window of values that can be represented in DPFP. To choose δ, we use three rules: A, B, and C. We apply rule A if possible, otherwise we apply rule B, and otherwise, rule C. These rules are described below and illustrated in Figure A2, with a brief summary in the legend.

Let $(\hat{x}_i)$ represent a set of floating point representations of real positive numbers (x_i), as defined in Equation (A8). $(\hat{x}'_i)$ are the scaled values

$$\hat{x}'_i = \hat{x}_i \times 2^{\delta}. \tag{A9}$$

q_{min} and q_{max} denote the minimum and maximum exponents over all $\hat{x}_i$. We define $q_{low} = -1022$ and $q_{high} = 1023$, respectively the lowest and highest acceptable exponents for a scaled value $\hat{x}'_i$. We define $q_{gap} = 2040$, the largest accepted difference between exponents across pairs of scaled values $\hat{x}'_i$. The values q_{low}, q_{high}, and q_{gap} were chosen to accommodate 64-bit DPFP representation, with some wiggle room to eliminate the need to account for edge cases.

Rule A is applied if:

$$q_{max} - q_{min} < q_{high}. \tag{A10}$$

In this case, the scale factor δ is simply $-q_{min}$. With this rule, the minimal exponent of scaled values becomes zero. Thus, scaled values are located between 1 and the maximum value in DPFP representation.

Rule B is applied if:

$$q_{high} \leq q_{max} - q_{min} < q_{gap}. \tag{A11}$$

The scale factor is $\delta = q_{low} - q_{min} + (q_{gap} - (q_{max} - q_{min}))/2$. With this rule, the exponents of scaled values are approximately centered around zero. Sets of values $\hat{x}'_i$ with greater variation between the smallest and the biggest elements can be handled as compared with rule A.

In the rare case that neither rule A nor rule B can be applied, rule C is applied. All $(\hat{x}_i)$ values with exponents q_i such that

$$q_i \leq q_{max} - q_{gap} \tag{A12}$$

are set to zero. These are the smallest values among $\hat{x}_i$ values. Then, rule B is applied to all non-zero $\hat{x}_i$ values.

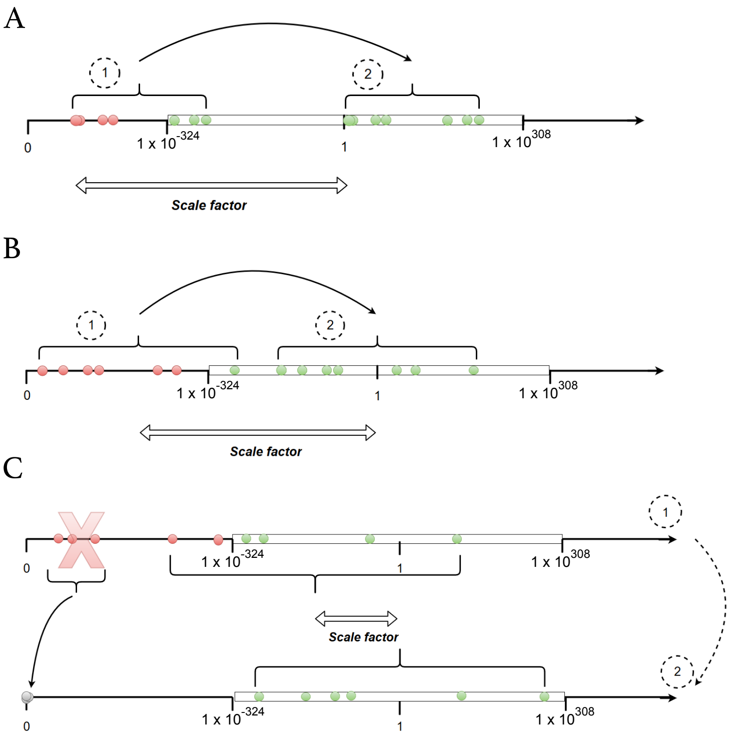

Figure A2. Scale factor choice. (**A**) Simplest case. The scale factor is the inverse of the smallest non-scaled value. (**B**) If A is not applicable, the scale factor is chosen such that the initial conditions are centered inside the range of acceptable values. The mid-point (on a log scale) of this interval is approximately 1. (**C**) Last case, if all scaled values cannot fit at once inside the range of accepted values, the lowest non-scaled values are dropped and set to zero so that the problem is simplified to case 1 (panel **A**) or 2 (panel **B**). In all panels, the white rectangle represents values that can be represented using DPFP. Dots represent the values of initial conditions for the differential equations of the multi-type birth–death model, before (1) and after (2) scaling. Red dots represent values that are initially outside the window of values that can be represented using DPFP.

Appendix B.3. Performance Improvements

We use various techniques to increase the efficiency of the numerical calculations performed by *bdmm*.

Prior to the integration of the coupled differential equations $p_{i,k}$ and $g^e_{i,k}$ backwards in time, we calculate the values of $p_{i,k}$ for every sampling event in the tree by numerically integrating the ODEs for $p_{i,k}$. We store the values obtained and use them when calculating the initial conditions for $g^e_{i,k}$ for every tip.

We also implement a parallelization scheme. An initial recursive tree traversal step is necessary to reach the tips of the tree before launching the numerical integration of the system of ODEs on $g^e_{i,k}$ and $p_{i,k}$ along the tree branches. During this traversal, when a node (whose left and right child subtrees are of significant size compared to the total tree size) is reached, a new computation thread is spawned and assigned to the traversal of one of the two subtrees. The initial thread continues onward with the traversal of the other one. This split between two threads is only executed when both subtrees represent more than a user-defined fraction of the total tree length, which is a tenth by default. This is performed in order to prevent an excessive number of threads from being created, since thread creation itself carries a computational overhead.

Finally, we replace the fixed-step size Runge–Kutta integrator used as a default integrator in the original implementation by the fifth-order Dormand–Prince integrator for ODEs [26]. This integrator uses a step-size control, which improves the efficiency and accuracy of numerical integration steps. We use the existing implementation of these integrators from the *Apache Commons Math* Java library [27].

Appendix C. Details on Likelihood Comparisons

To compare the results of computations performed with each *bdmm* version, we randomly simulated sampled trees with 2 demes with fixed multi-type birth–death parameter values using MASTER [28]. We then computed tree likelihood values, varying one of the multi-type birth–death parameters (either the birth or death rate of deme 1, corresponding to the tree origin location). Parameter values used for simulation and likelihood computation are listed in Table A2. In Figures 3 and A3, we show the results for three ten-tip trees, for two of them we vary λ_1, and for the third we vary μ_1. In Figure A3, we also show the results with a hundred-tip tree, varying λ_1. For all simulated trees, the likelihood results are identical between the original and the new *bdmm* versions.

Appendix D. Additional Details on Influenza Data Analysis

As we deal with pathogen sequence data, we adopt the epidemiological parametrization of the multi-type birth–death model, as detailed in Kühnert et al. [16]. The epidemiological parametrization substitutes birth, death, and sampling rates with effective reproduction numbers within types, rate at which hosts become noninfectious, and sampling proportions.

For $i \in \{1, \ldots, d\}$ and $k \in \{1, \ldots, n\}$,

$$R_{i,k} = \frac{\lambda_{ii,k}}{\mu_{i,k} + r_{i,k}\,\psi_{i,k}}.$$

The rate of becoming noninfectious $\delta_{i,k}$ represents the inverse of the mean duration of the infection:

$$\delta_{i,k} = \mu_{i,k} + r_{i,k}\,\psi_{i,k}.$$

Based on our data, $\rho_{i,k} = 0$ for all i, k, as there is no singular point in time when a population-wide sampling effort had been carried out and had led to multiple simultaneous samples. Thus, the probability of an individual being sampled, or the sampling proportion, is calculated as follows:

$$s_{i,k} = \frac{\psi_{i,k}}{\mu_{i,k} + \psi_{i,k}}.$$

We assume that $r = 1$ for all i, k in our analyses, i.e., individuals become noninfectious upon sampling. Furthermore, we assume that δ is constant across locations i and time intervals k. Sampling proportion and migration are assumed not to change through time.

To study the seasonal dynamics of the global epidemic, we allow the effective reproduction number R to vary through time. To do so, we subdivide time into six-month intervals (starting on April 1st and October 1st) for the time period during which we have samples. We set all values corresponding to the same season across different years to be equal for a particular location. Therefore, we infer two different values of R for each location X, one which corresponds to the April–September period: R_{X_1}, and another one for the October–March period: R_{X_2}. The location-specific sampling proportions s_i are assumed to be constant for the time interval in which we have samples and null before the first sample.

Migration rates between types are inferred as products of a unique migration factor σ_m and relative rates $M_{i,j}$:

$$\forall\, i, j \in \{1 \ldots d\}, m_{i,j} = \sigma_m\, M_{i,j}.$$

This setting allows us to use rather informative priors of around 1 for the relative rates, and only requires a less informative prior for the unique migration factor. Table A3 lists the prior distributions assumed for the birth–death parameters.

Following [18], we use a GTR+Γ substitution model [29] and a strict molecular clock with the same priors as in [18].

The BEAST 2 analysis infers the birth–death model parameters, the substitution model parameters, and the clock model parameters together with the phylogenetic tree. In our analysis, the inferred phylogenetic trees are sample-typed trees. Thus, we do not attempt to reconstruct the history of migrations; rather, we marginalize over all the possible migration histories. This is realized in order to allow the MCMC chain to converge in a shorter time as compared to an analysis inferring branch-typed trees.

For each data analysis, we ran 10 parallel MCMC chains for 9×10^7 steps, each using the new implementation of *bdmm* as a package of BEAST 2.6 [8,30]. The XML files providing all analyses specifications are available as Supplementary Material.The computations were run on the Euler computation cluster at ETH Zürich. Each MCMC chain in the 175-sample analysis required 150 computation hours on average (1500 h in total for ten chains). In the 500-sample analysis, each MCMC required 360 computation hours on average (3600 h for ten chains). We used Tracer v1.6 [31] to check for convergence. The effective sample size was above 200 for all parameters. We combined the chains run in parallel into one using LogCombiner. We obtained the MCC tree using TreeAnnotator. Both LogCombiner and TreeAnnotator are available as part of BEAST 2.6. Finally, we plotted the numerical results with the R package *ggplot2* [32] and the MCC tree with *ggtree* [33].

Appendix E. Tables

Table A1. Distributions from which parameters were sampled for the simulation of trees. All parameters were constant through time.

Parameter	Distribution
λ_i	Unif(1, 3)
μ_i	$\lambda_i \times$ Unif(0, 1)
$m_{i,j}$	Unif(0, 0.5)
ψ_i	Unif(0.05, 0.5)
r_i	Unif(0, 1)

Table A2. Fixed parameter values for tree simulation and likelihood computations.

Parameter	Value
λ_1	0.4
λ_2	0.3
μ_1	0.27
μ_2	0.17
ψ_1	0.03
ψ_2	0.03
$m_{1,2}$	0.03
$m_{2,1}$	0.03
Initial root state	1

Table A3. Prior distributions for parameters of the multi-type birth–death model in the seasonal influenza analysis.

Parameter	Prior Distribution
R	LogNormal(0, 1.0)
δ	LogNormal(4.5, 0.15)
σ_m	LogNormal(0, 2.0)
$M_{i,j}$	LogNormal(0, 0.5)
s	Exp(0.001) truncated on $[0, 1]$
r	Beta(10.0, 1.5)
T	LogNormal(2.0, 1.0)

Table A4. Inferred parameter values for Influenza A virus analysis under the multi-type birth–death model. For each parameter, the lower and upper bounds for the 95% Highest Posterior Density interval (hpd_low, hpd_high) are given along with the median (m). N, S, and T refer respectively to the North, South, and Tropics. For effective reproduction numbers R, the first subscript is the location while the second one refers to the period of the year. s, w, as, om respectively refer to *summer*, *winter*, *april–september*, and *october–march*. Thus, for instance, R_{S_w} refers to the effective reproduction number of samples from the southern hemisphere during the winter season. The tree height t is given in number of years, M is given in migrations/lineage/year, and δ is given in years^{-1}. The remaining parameters are unitless.

	175 Samples			**500 Samples**		
	m	**hpd_low**	**hpd_high**	**m**	**hpd_low**	**hpd_high**
t	3.342	3.048	3.643	6.645	6.313	7.031
δ	90.464	75.836	105.207	101.582	87.197	116.768
R_{N_s}	0.354	0.199	0.556	0.505	0.263	0.815
R_{N_w}	0.971	0.927	1.01	1.001	0.98	1.02
$R_{T_{as}}$	1.048	1.026	1.071	1.005	0.992	1.017
$R_{T_{om}}$	0.991	0.969	1.013	1.01	0.998	1.022
R_{S_s}	0.558	0.335	0.783	0.774	0.679	0.861
R_{S_w}	1.08	1.037	1.123	1.027	1.002	1.051
σ_m	0.475	0.196	0.869	0.304	0.147	0.524
$M_{N,T}$	0.871	0.245	1.923	1.064	0.3	2.422
$M_{N,S}$	0.894	0.253	1.965	0.838	0.264	1.744
$M_{T,N}$	2.183	0.877	4.174	1.568	0.635	2.973
$M_{T,S}$	0.561	0.205	1.08	0.694	0.332	1.248
$M_{S,N}$	1.001	0.273	2.261	0.887	0.275	1.93
$M_{S,T}$	1.005	0.279	2.173	1.13	0.368	2.468
h_N	0.292	0	0.777	0.296	0	0.779
h_T	0.3	0	0.784	0.292	0	0.777
h_S	0.283	0	0.769	0.289	0	0.767
P(root in N)	0.024	0	0.535	0.023	0	0.501
P(root in T)	0.849	0.167	1	0.863	0.176	1
P(root in S)	0.041	0	0.731	0.044	0	0.67

Appendix F. Additional Figures

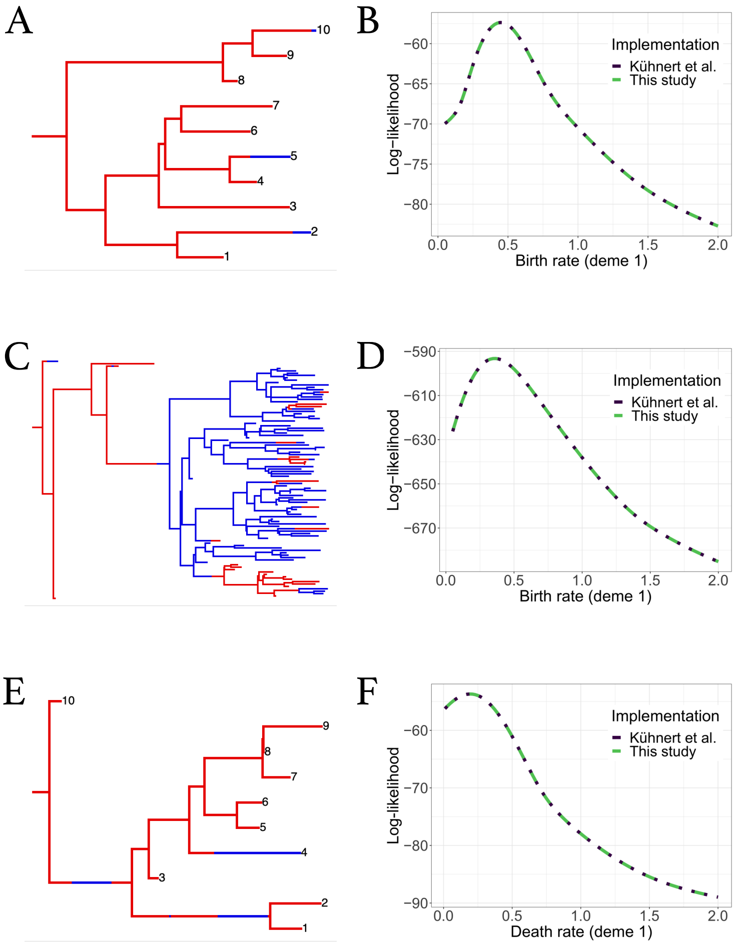

Figure A3. Comparisons of likelihood computation results between the original and improved *bdmm* versions for additional trees. (**A**,**B**) Randomly simulated ten-tip tree and log-likelihood computation results against λ_1 (birth rate of red deme). (**C**,**D**) Randomly simulated hundred-tip tree and log-likelihood computation results against λ_1 (birth rate of red deme). (**E**,**F**) Randomly simulated ten-tip tree and log-likelihood computation results against μ_1 (death rate of red deme).

References

1. Felsenstein, J. Estimating effective population size from samples of sequences: Inefficiency of pairwise and segregating sites as compared to phylogenetic estimates. *Genet. Res.* **1992**, *59*, 139?147. https://doi.org/10.1017/S0016672300030354.
2. Hey, J.; Machado, C.A. The study of structured populations ? new hope for a difficult and divided science. *Nat. Rev. Genet.* **2003**, *4*, 535–543. https://doi.org/10.1038/nrg1112.
3. Stadler, T.; Bonhoeffer, S. Uncovering epidemiological dynamics in heterogeneous host populations using phylogenetic methods. *Philos. Trans. R. Soc. London. Ser. B Biol. Sci.* **2013**, *368*, 20120198. https://doi.org/10.1098/rstb.2012.0198.
4. Grenfell, B.T.; Pybus, O.G.; Gog, J.R.; Wood, J.L.; Daly, J.M.; Mumford, J.A.; Holmes, E.C. Unifying the epidemiological and evolutionary dynamics of pathogens. *Science* **2004**, *303*, 327–332.
5. Kühnert, D.; Wu, C.H.; Drummond, A.J. Phylogenetic and epidemic modeling of rapidly evolving infectious diseases. *Infect. Genet. Evol.* **2011**, *11*, 1825–1841. https://doi.org/10.1016/j.meegid.2011.08.005.
6. Dudas, G.; Carvalho, L.M.; Bedford, T.; Tatem, A.J.; Baele, G.; Faria, N.R.; Park, D.J.; Ladner, J.T.; Arias, A.; Asogun, D.; et al. Virus genomes reveal factors that spread and sustained the Ebola epidemic. *Nature* **2017**, *544*, 309.
7. Faria, N.R.; Kraemer, M.U.; Hill, S.; De Jesus, J.G.; Aguiar, R.; Iani, F.C.; Xavier, J.; Quick, J.; Du Plessis, L.; Dellicour, S.; et al. Genomic and epidemiological monitoring of yellow fever virus transmission potential. *Science* **2018**, *361*, 894–899.
8. Bouckaert, R.; Heled, J.; Kühnert, D.; Vaughan, T.; Wu, C.H.; Xie, D.; Suchard, M.A.; Rambaut, A.; Drummond, A.J. BEAST 2: A software platform for Bayesian evolutionary analysis. *PLoS Comput. Biol.* **2014**, *10*, e1003537. https://doi.org/10.1371/journal.pcbi.1003537.
9. Hodges, S.A. Floral nectar spurs and diversification. *Int. J. Plant Sci.* **1997**, *158*, S81–S88.

10. Goldberg, E.E.; Kohn, J.R.; Lande, R.; Robertson, K.A.; Smith, S.A.; Igić, B. Species selection maintains self-incompatibility. *Science* **2010**, *330*, 493–495.
11. Mayrose, I.; Zhan, S.H.; Rothfels, C.J.; Magnuson-Ford, K.; Barker, M.S.; Rieseberg, L.H.; Otto, S.P. Recently formed polyploid plants diversify at lower rates. *Science* **2011**, *333*, 1257–1257.
12. Goldberg, E.E.; Lancaster, L.T.; Ree, R.H. Phylogenetic inference of reciprocal effects between geographic range evolution and diversification. *Syst. Biol.* **2011**, *60*, 451–465.
13. Volz, E.M.; Frost, S.D. Sampling through time and phylodynamic inference with coalescent and birth–death models. *J. R. Soc. Interface* **2014**, *11*, 20140945.
14. Boskova, V.; Bonhoeffer, S.; Stadler, T. Inference of epidemiological dynamics based on simulated phylogenies using birth-death and coalescent models. *PLoS Comput. Biol.* **2014**, *10*, e1003913.
15. Maddison, W.P.; Midford, P.E.; Otto, S.P. Estimating a binary character's effect on speciation and extinction. *Syst. Biol.* **2007**, *56*, 701–710. https://doi.org/10.1080/10635150701607033.
16. Kühnert, D.; Stadler, T.; Vaughan, T.G.; Drummond, A.J. Phylodynamics with Migration: A Computational Framework to Quantify Population Structure from Genomic Data. *Mol. Biol. Evol.* **2016**, *33*, 2102–2116. https://doi.org/10.1093/molbev/msw064.
17. 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). *Pnas* **2013**, *110*, 228–233, https://doi.org/10.1073/pnas.1207965110/-/DCSupplemental.www.pnas.org/cgi/doi/10.1073/pnas.1207965110.
18. Vaughan, T.G.; Kühnert, D.; Popinga, A.; Welch, D.; Drummond, A.J. Efficient Bayesian inference under the structured coalescent. *Bioinformatics* **2014**, *30*, 2272–2279. https://doi.org/10.1093/bioinformatics/btu201.
19. Gavryushkina, A.; Welch, D.; Stadler, T.; Drummond, A.J. Bayesian inference of sampled ancestor trees for epidemiology and fossil calibration. *PLoS Comput. Biol.* **2014**, *10*, e1003919.
20. Louca, S.; Pennell, M.W. Extant timetrees are consistent with a myriad of diversification histories. *Nature* **2020**, *580*, 502–505.
21. Louca, S.; McLaughlin, A.; MacPherson, A.; Joy, J.B.; Pennell, M.W. Fundamental identifiability limits in molecular epidemiology. *Mol. Biol. Evol.* **2021**, *38*, 4010–4024.
22. MacPherson, A.; Louca, S.; McLaughlin, A.; Joy, J.B.; Pennell, M.W. Unifying Phylogenetic Birth–Death Models in Epidemiology and Macroevolution. *Syst. Biol.* **2022**, *71*, 172–189.
23. Maddison, W.P. Confounding asymmetries in evolutionary diversification and character change. *Evolution* **2006**, *60*, 1743–1746.
24. FitzJohn, R.G. Diversitree: Comparative phylogenetic analyses of diversification in R. *Methods Ecol. Evol.* **2012**, *3*, 1084–1092.
25. Rabosky, D.L.; Goldberg, E.E. Model inadequacy and mistaken inferences of trait-dependent speciation. *Syst. Biol.* **2015**, *64*, 340–355.
26. Dormand, J.R.; Prince, P.J. A family of embedded Runge-Kutta formulae. *J. Comput. Appl. Math.* **1980**, *6*, 19–26.
27. Math, C. The Apache Commons Mathematics Library, 2016. Available online: https://commons.apache.org/proper/commons-math/ (accessed on July 26th 2022).
28. Vaughan, T.G.; Drummond, A.J. A stochastic simulator of birth–death master equations with application to phylodynamics. *Mol. Biol. Evol.* **2013**, *30*, 1480–1493.
29. Lanave, C.; Preparata, G.; Sacone, C.; Serio, G. A new method for calculating evolutionary substitution rates. *J. Mol. Evol.* **1984**, *20*, 86–93.
30. Bouckaert, R.; Vaughan, T.G.; Barido-Sottani, J.; Duchêne, S.; Fourment, M.; Gavryushkina, A.; Heled, J.; Jones, G.; Kühnert, D.; De Maio, N.; et al. BEAST 2.5: An advanced software platform for Bayesian evolutionary analysis. *PLoS Comput. Biol.* **2019**, *15*, 1–28. https://doi.org/10.1371/journal.pcbi.1006650.
31. Rambaut, A.; Drummond, A.J.; Xie, D.; Baele, G.; Suchard, M.A. Posterior Summarization in Bayesian Phylogenetics Using Tracer 1.7. *Syst. Biol.* **2018**, *67*, 901–904. https://doi.org/10.1093/sysbio/syy032.
32. Wickham, H. *ggplot2: Elegant Graphics for Data Analysis*; Springer: New York, NY, USA, 2016.
33. Yu, G.; Smith, D.K.; Zhu, H.; Guan, Y.; Lam, T.T.Y. ggtree: An r package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods Ecol. Evol.* **2017**, *8*, 28–36. https://doi.org/10.1111/2041-210X.12628.