Report on "Coupling adaptive molecular evolution to phylodynamics using fitness-dependent birth-death models"

Seth D. Temple[*] , Department of Statistics, University of Washington, Seattle, WA 98195, USA

1 Introduction

Selection is one of the main evolutionary forces amongst genetic drift, mutation, and migration that modify the genomes of organisms. It plays a major role in the evolutionary potential of a (large) population to adapt over time. As a result, evolutionary biologists, population geneticists, and epidemiologists have long endeavored to model its characteristics such as direction (positive or negative selection), magnitude (selection coefficient), and adversarial behavior (balancing or frequency-dependent selection) (Eyre-Walker & Keightley, 2007). For multicellular organisms a standard way to study selection is to take an independence assumption and decouple it from the mutational process. That is, while both selection and mutation may be understood in light of a phylogenetic tree of shared ancestry which drives genomic sequence evolution, these two processes do not interact. These mutations of appreciable fitness may result in more branching events and/or longer or shorter branch lengths, affecting the tree and the tip sequences. (Rasmussen & Stadler, 2019) advance that this naïve independence may be appropriate for nearly neutral (mutation) theory (Kimura, 1983) or macroevolutionary tree processes of mammals and other multicellular organisms where speciation and extinction happen slowly, but not for viruses and bacteria.

This incompatibility with the naïve independence is supported by the viral life cycle. During reproduction, a new offspring virus may develop mutations from imprecise replication machinery. These mutations occur at rate 1e-4 to 1e-6 for viruses of genome size 1e4 to 1e6, such that ~1 new mutation is expected in each offspring (Sanjuán et al., 2010). The offspring virus then must survive a host immune system and infect a host cell to replicate its genetic material in the form of a next offspring. While ~1 new mutation may not provide a large fitness advantage to viability or fertility, viruses proliferate rapidly and have a compact sequence dense with genes such that adaptive mutations may develop on the order of days/weeks in pandemic times. This discussion of viral life history posits a justification for modeling mutation and selection jointly in how they drive sequence evolution. (Huge caveat: I am not a virologist.)

Birth-death processes, similar to or mathematically consistent with stochastic epidemic models (MacPherson et al., 2021), have been developed to estimate genotype or character effects on viral sequence evolution. For instance, (Stadler & Bonhoeffer, 2013) propose a multi-type birth-death (MTBD) model to analyze viral sequence data ascertained from people infected with human immunodeficiency virus (HIV). They study character types, heterosexual, intravenous drug users, and men who have sex with men, as they relate to pathogen spread. The MTBD approach formulates a series of ordinary differential equations (ODEs) resulting from continuous time Markov chains. These ODEs are then numerically or analytically solved with respect to key parameters of interest, most often via the maximum likelihood method. In terms of ODEs on tree structures, a peeling (Cannings et al., 1978) or pruning (Felsenstein, 1981) approach computes on subtrees and passes information upwards towards the root. Finally, these parameter MLEs are interpreted according to the underlying model and communicated to other parties (government, public health, policy officials). The main ODE for MTBD

$$\frac{d}{dt}D_{n,i}(t) = -\left(\lambda_i + \sum_{j=1}^{M}\gamma_{i,j} + d_i\right)D_{n,i}(t) + 2\lambda_i E_i(t)D_{n,i}(t) + \sum_{j=1}^{M}\gamma_{i,j}D_{n,j}(t)$$

studies how a character type i affects birth, substitution, and death rates λ, γ, d, where other notation is put forth in Supplementary Material and mathematical motivation is given in Appendix A1. Serious limitations of MTBD are that computation scales exponentially with character types and that if character types are multisite genotypes then site-specific and interaction effects are not directly modeled. Given these challenges, (Rasmussen & Stadler, 2019) introduce the marginal fitness birth-death (MFBD) model to study site-specific fitness effects in a tractable manner.

2 Methods

The MFBD model extends the MTBD model by making marginal independence assumptions to express character/genotype type birth rates λ_i (and/or death rates d_i) as a function of marginal sites k. As a running example, consider 3 biallelic sites such that 8 genotypes may exist: abc, Abc, ABc, ABC, aBc, aBC, abC, AbC. The capital A,B,C mutations may induce some relative fitness effect over lowercase a,b,c. Let equation $\lambda_n = f_g \lambda_0$ describe the lineage-specific birth rates according to its genotype fitness f_g. Since f_g is intractable as the # of sites grows, MFBD formulates marginal fitness $f_{n,k,i}$ for site k and state i and places these in the context of MTBD. In doing so, MFBD does the following:

1. Approximate site probabilities $\omega_{n,k,i}$ with a Bayes formula
2. Approximate genotype probabilities $\omega_{n,g}$ by site probabilities
3. Express marginal fitness as an average weighted by genotype probabilities

I explain and interpret these three steps as they relate to sequence evolution models. For a subtree $\mathcal{T}_n$ descendant from node n with tip sequences $\mathcal{S}_n$, step 1 amounts to

$$\omega_{n,k,i} = Pr(i|\mathcal{T}_n, \mathcal{S}_n) = \frac{Pr(\mathcal{T}_n, \mathcal{S}_n|i)q(i)}{\sum_{j=1}^{J} Pr(\mathcal{T}_n, \mathcal{S}_n|j)q(j)} = \frac{Pr(\mathcal{T}_n, \mathcal{S}_n|i)}{\sum_{j=1}^{J} Pr(\mathcal{T}_n, \mathcal{S}_n|j)} = \frac{D_{n,k,i}}{\sum_{j=1}^{J} D_{n,k,j}},$$

a Bayes formula with discrete uniform prior type i at site k and $D_{n,k,i}$ estimable via numerical ODE. While uniform priors are commonplace in the absence of additional information, here the uniform prior exchanges mathematical convenience for a very strong assumption. A better prior may be that a lineage is in state i with probability $q(i)$ equal to the population allele frequency. As well, this $\omega_{n,k,i}$ depends mostly on the phylogenetic tree, a mathematical object that is difficult to infer and know for certain. Next, step 2 assumes independence over sites to estimate genotype probabilities as

$$\omega_{n,g} = \prod_{k=1}^{K} \omega_{n,k,g_k} / \left(\sum_g \prod_{k=1}^{K} \omega_{n,k,g_k} \right),$$

where g_k is the type of genotype g at site k and the denominator normalizes the probability space to 1. Given our running example, independence means $P(ABC) = P(A)P(B)P(C)$. This assumption does appear discordant with the modeling task and motivation as in applied examples on Ebola and influenza where epistatic interaction effects are suspected. Lastly, step 3 supposes discrete random variables for genotype and site-specific fitnesses and computes them via the method of moments; namely, $f_{n,k,i} = \sum_{g_k=i} f_g \omega_{n,g}$. In the 3 sites genotype example, and the MTBD HIV example, these f_g could be directly modeled. For many more sites, multiplicative fitness helps scale:

$$f_{n,k,i} = \sigma_{k,i} \prod_{l \neq k} \sum_{j=1}^{J} \sigma_{l,j} \omega_{n,l,j},$$

where fitness $\sigma_{k,i}$ at site k of type i is modified by marginal fitnesses at other sites. This choice is also a strong assumption to say that dominance, recessive, or additive models do not better fit marginal fitness. And, again, this equation is discordant with step 2 that assumes site

independence. (Rasmussen & Stadler, 2019) calculate these formulas in chronological order for each subtree as they evaluate the pruning procedure for MFBD. For simulation and applied analyses, they situate MFBD and MTBD in the Bayesian phylogenetic software `BEAST2` to assess statistical uncertainty.

3 Results

To give evidence to their methods, (Rasmussen & Stadler, 2019) simulate 100 phylogenetic trees with the (Gillespie, 1977) algorithm applied to birth-death processes with site-specific fitness effects (2, 5, and 10 sites). Due to runtime limitations, I replicated this simulation study with 4 phylogenies (Figure A1) generated from their Matlab code and their `Lumiere` package in Java. Figure 1 shows site-specific fitness effects for 2 sites across 4 phylogenetic simulations, and Figure A2 shows birth rates for the 4 simulations. These plots suggest coverage in these 4 cases, but the authors report less than 80% coverage for 95% credible intervals. They also show posterior medians not centered about the true population average fitness 1.0, which is concordant with their positive correlation between true and posterior median fitness, but noteworthy poor estimation. One explanation for this observation could be that uncertainty in phylogenetic tree space is not modeled by their MFBD procedure.

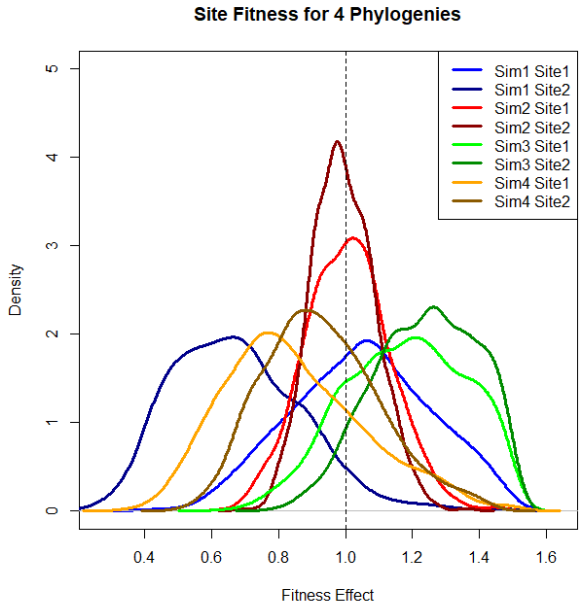

Figure 1 Posterior distributions for 2 site-specific fitness values across 4 simulated phylogenies. Vertical black dashed line is the mean site-specific fitness that the true fitness effects are drawn from. These results compare to Figures 5-6 in (Rasmussen & Stadler, 2019).

Another simulation study the authors conducted was to demonstrate that MFBD is a good approximation for MTBD in the few genotypes scenario (e.g., 4 genotypes). They conclude that MFBD is a good approximation except when site-specific or epistatic fitness effects grow (Figures 2-4; (Rasmussen & Stadler, 2019)). On the other hand, the main point of MFBD is to study strong fitness and epistasis, so I have remaining doubts about independence assumptions across sites.

Informed by another study on epistasis in the Ebola genotypes of a 2013-16 outbreak, the authors run MFBD to infer genotype fitnesses for populations in West Africa as opposed to

fitness via cell infectivity experiments (Urbanowicz et al., 2016). There data involves 1610 viral sequences. I compiled their `Lumiere` package and performed the Ebola analysis with geographic effects for Guinea, Sierra Leone, and Liberia; however, I did not arrive at posterior credible intervals similar to the authors. Providing details on the Bayesian analysis such as thinning, burn-in, and # of iterations could promote more reproducible science. My results diverge largely from those of the authors. I did find a genotype with a relative fitness effect that did not overlap 1.0 (blue), and some relative fitness effects with large credible intervals (red). An explanation for some wider credible intervals could be sampling frequency biases in ascertainment (column 2 in Table 1; (Rasmussen & Stadler, 2019)) which is common as resources and capabilities get diverted in early stages of outbreaks. I also find more uncertainty in a geographic location (Figure A6), which is another sampling bias related to resource availability in geographic space.

Posterior Fitness for Ebola 2013-16			
Fitness Effect	*Quantile*		
	0.025	Median	0.975
1	**0.764**	**0.811**	**0.862**
2	0.98	0.996	1.014
3	0.774	0.89	1.005
4	0.925	0.987	1.059
5	0.967	1.042	1.123
6	**0.977**	**1.086**	**1.199**
7	**0.894**	**1.014**	**1.133**
8	0.907	0.976	1.059
9	0.933	0.985	1.041

Table 1 Posterior fitness effects for 9 genotypes in Ebola outbreak data. This table ought to be compared to (Table 1; Rasmussen & Stadler, 2019). Blue row is only credible interval not overlapping 1.0. Red rows have the widest credible intervals.

In another case study, the authors perform a Bayesian analysis with MFBD for 2150 influenza sequences in the United States. This study involves more than 50 sites. They relate lineage fitness to results from a deep mutational scanning experiment with an unmotivated linear equation in terms of log2 relative amino acid preferences (and linear and exponent scaling terms). It is unclear how this linear model relates back to MFBD as presented here. I ran their code for 5000 iterations which took 7 hours. Diagnostic plots suggest that the MCMC chain is nowhere near convergence and that Metropolis-Hastings acceptance probabilities are low (Appendix A10. My effort here points out that statistical inference for many parameters is challenging in the Bayesian paradigm and that multivariate parameter sets may incur slow convergence. In the end, the flu and Ebola case studies conclude that there is a rank correlation between cell infectivity and population fitness studies, but that population fitnesses are much more muted than amino acid preferences.

4 Discussion

The marginal fitness birth-death model extends the multi-type birth-death model to directly link mutational sites and their fitness effects with the genotypes of tip sequences in a phylogenetic tree. As far as I know, this is the first such effort in multi-type sequence evolution modeling to

jointly model mutational types as the output of sequence evolution and the phylogenetic process. Regarding the tree, a higher birth rate λ_i for type i would result in more branching events. The resulting sequences that share those branching events would share a lot of the same genetic sequence. This perspective unifies the mutational process along sequences and the phylogenetic process in time for the cases in the microbial world where these two processes occur contemporaneously. While the authors focus on MFBD to estimate marginal fitness, I am not sure this effort aligns with laboratory experiments suggesting few, multiple mutation genotypes surpassing neutrality, nor do I think Bayesian parameter uncertainty is the most important uncertainty to study here. I may suggest collaboration with laboratory scientists to identify genotypes with multiple mutations of interest (assuming other genotypes are of a neutral type), computing MLEs of a smaller parameter set with deterministic ODEs, and a parametric bootstrap of the Gillespie type to characterize uncertainty in tree space. Another reason for my suggestion above is to develop simpler models with less multivariate effects to consider. Even in simulation studies of mine and (Rasmussen & Stadler, 2019), posterior medians for most every parameter do not estimate the true parameter value well and credible intervals are wide, indicating that the data provides little information to study a multivariate parameter set of substantial covariance.

Appendix

A1 Mathematical Details

Birth-death (BD), stochastic epidemic (SE), or susceptible-infected-removed (SIR) compartmental models usually involve many parameters. I provide as supplementary material a notational glossary for these parameters. Where appropriate, I define the parameter or function in this appendix section.

A1.1 Multitype birth-death model (MTBD)

These models fit ordinary differential equations with respect to a tree for for time-indexed functions $D_{n,i}(t)$ and $E_i(t)$. The $D_{n,i}(t)$ function refers to the probability density of a subtree. It depends on a lineage n being in character state i between present-day backward in time to t. The birth, substitution, and death rates λ, γ, d parameterize the model as follows:

$$\frac{d}{dt}D_{n,i}(t) = (1) + (2) + (3)$$

- (1) := $-\left(\lambda_i + \sum_{j=1}^{J}\gamma_{i,j} + d_i\right)D_{n,i}(t)$
  - There are no unobserved births, substitutions from character state j of J possibilities, nor deaths
- (2) := $2 \times \lambda_i E_i(t) D_{n,i}(t)$
  - There is an unobserved birth event because one or the other lineage when extinct with probability density $E_i(t)$
- (3) := $\sum_{j=1}^{J}\gamma_{i,j}D_{n,i}(t)$
  - The lineage n actually converted from character state j among J states to the observed state i sometime between time 0 and t

I note that the probability density of an extinct lineage $E_i(t)$ must be defined and solved with ODEs. Its formulation is similar to $D_{n,i}(t)$; namely,

$$\frac{d}{dt}E_i(t) = (1e) + (2e) + (3e) + (4e)$$

- (1e) := $-\left(\lambda_i + \sum_{j=1}^{J}\gamma_{i,j} + d_i\right)\mathrm{E}_i(t)$
- (2e) := $\lambda_i E_i(t)^2$
  - There is a birth event and both lineages go extinct
- (3e) := $\sum_{j=1}^{J}\gamma_{i,j}\mathrm{E}_i(t)$
- (4e) := $(1 - s_i)d_i$
  - A lineage of type i dies and it is not sampled upon dying with probability $(1 - s_i)$ where s_i is the probability of being sampled upon dying

The main differences in this formulation is (4e) and replacing $D_{n,i}(t)$ with $E_i(t)$. These calculations are done for subtrees and the density is passed up the subtrees progressively towards the phylogenetic root (Felsenstein, 1981). The full likelihood is then

$$D_n = \sum_{j=1}^{J}\frac{D_{n,j}(t_{root})}{1 - E_i(t_{root})} \times q_i$$

where q_i is prior density that root is in state i and the denominator normalizes that the root did not go extinct. Incorporating the sequence data conditional on the phylogenetic tree amounts to a product term over sites $k = 1, \ldots, K$ with is equation (23) in the original paper.

A1.2 Marginal fitness birth-death (MFBD)

The changes to the MFBD from the MTBD and their motivations are described in the report body. Formally, the MFBD ODE for $D_{n,k,i}(t)$ is

$$\frac{d}{dt}D_{n,i,k}(t) = -\left(\lambda_{k,i} + \sum_{j=1}^{J}\gamma_{i,j} + d_{k,i}\right)D_{n,k,i}(t) + 2\lambda_{k,i}E_u(t)D_{n,k,i}(t) + \sum_{j=1}^{J}\gamma_{i,j}\,D_{n,k,i}(t)$$

where $\lambda_{k,i}, d_{k,i}$ depend on (approximated) marginal fitness values $f_{n,k,i}$. The authors of the original manuscript only implement MFBD for $\lambda_{k,i}$ depending on $f_{n,k,i}$ and death rate d fixed. The term $E_u(t)$ is a computational simplification to consider binned marginal fitness u on a one-dimensional fitness landscape instead of multivariate sequence genotypes. In my running example, the states $j = 1, \ldots, J$ are actually $J = 2$ for biallelic single nucleotide (or allele) polymorphisms. So, this extension just changes the character-specific λ_i to be character-specific and site-specific.

A2 Software and Replication Efforts

This paper provides some code to replicate their results. They present these resources as a software package `Lumiere` on GitHub: https://github.com/davidrasm/Lumiere. This software fits within the `BEAST` computational paradigm in phylogenetic analysis. `Lumiere` was developed and tested on `BEAST2` versions 2.4.0 to 2.5.2, which were the current versions in 2019, but these versions are no longer current. The first author reports a plan to incorporate `Lumiere` in as a `BEAST2` package available for installation via `BEAUti` such that it can easily be used in the `BEAST2` graphical user interface (GUI). This plan never came to fruition.

As a result, users must download the necessary dependencies on their local devices and compile `Lumiere` with a Java integrated development environment (IDE) like Eclipse. I successfully followed the first author's instructions and compiled `Lumiere` on my local machine. I executed some of the examples provided to get a scope of the computational runtime. These examples include the applied analyses of Ebola from 2013-2016 and Influenza from 2009-2012, and some simulated phylogenies for 4 genotypes and 2, 5, and 10 mutational sites. I note that `Lumiere`, and `BEAST` in general, conduct Markov Chain Monte Carlo (MCMC) to explore the posterior space of key parameters. Bayesian analysis is standardly taught in graduate programs for Statistics and Biostatistics. Moreover, I have summarized the main ideas of Bayesian analysis elsewhere (Temple et al., 2022) as well as other authors (Gelman et al., 1995), and therefore exclude details in this appendix.

MCMC is known to be a computationally intensive algorithm. The computation grows more so for large parameter sets where the parameters exhibit strong correlations. For example, (Temple et al., 2022) required 8+ hours one Thanksgiving in 2021 to perform 50,000 iterations with 10-20 parameters for a Gibbs sampler in a generalized linear model. While recent developments leverage online learning to build in new samples to an MCMC analysis for phylogenetics, these efforts still report days to weeks' worth of computation (Suchard et al., 2018). I report here some rough estimates of runtimes for some of the `Lumiere` examples: 7 hours for ~5000 iterations in flu example, 6 hours for ~70,000 iterations in Ebola example, 1.5 hours for 100,000

iterations in 4 genotypes example, 1.5 hours for 100,000 iterations in 2 sites example, 4.5 hours for 100,000 iterations in 5 sites example. These examples involve 1 MCMC chain, whereas standard Bayesian analysis involve multiple chains to give some evidence that convergence results hold when varying random seeds. Additionally, these examples are for a single simulated phylogeny. In (Rasmussen & Stadler, 2019) 100 simulated phylogenies are studied. Figures A1-3 apply to a simulated 2 site-specific fitness effects over 4 phylogenies example. These demonstrate that MFBD is a reasonable approximation for MTBD but that Bayesian phylogenetic analysis may be poorly calibrated and highly uncertain. Figures A4-5 suggest with some phylogenetic tree visualization that the examples provided at the GitHub page for `Lumiere` match those in the original paper. I find this concerning because my replication results in Table 1 do not match the Ebola analysis table in the original paper. Figures A6-10 show additional replication results for the Ebola and flu analyses, especially highlighting challenges and standard practices in Bayesian analysis.

Finally, I provide some criticism with respect to reproducible science. I found `Lumiere` to be difficult to use and understand. Various functions do not run without editing them, which is challenging given that Matlab is no longer taught in most academic majors. The first author suggests that MCMC experiments be performed on a cluster by compiling a Java JAR applet, but the documentation on such compilation is limited. Compiling such an applet would likely require an experienced Java developer. As a practice, I strive to give well-documented and reproducible software for downstream users in the scientific community. This practice involves: (1) documenting inputs and outputs of functions and data types, (2) a minimal sketch of the workflow or application programming interface (API), and (3) providing examples [or a single example] on how run the code.

A3 Figures

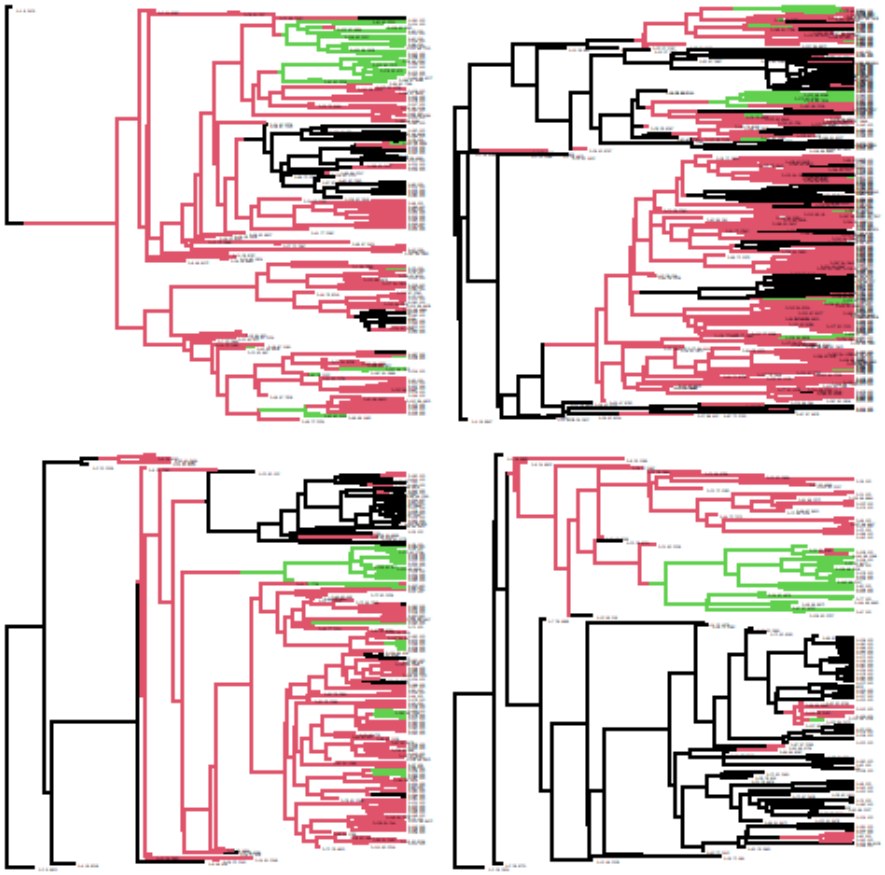

Figure A1 Visualization of 4 phylogenetic trees from `Lumiere` Matlab scripts and using R package `phytools`. Green and pink lineages denote site-specific mutations, while black is the reference. These four phylogenies present considerable uncertainty in phylogenetic tree space.

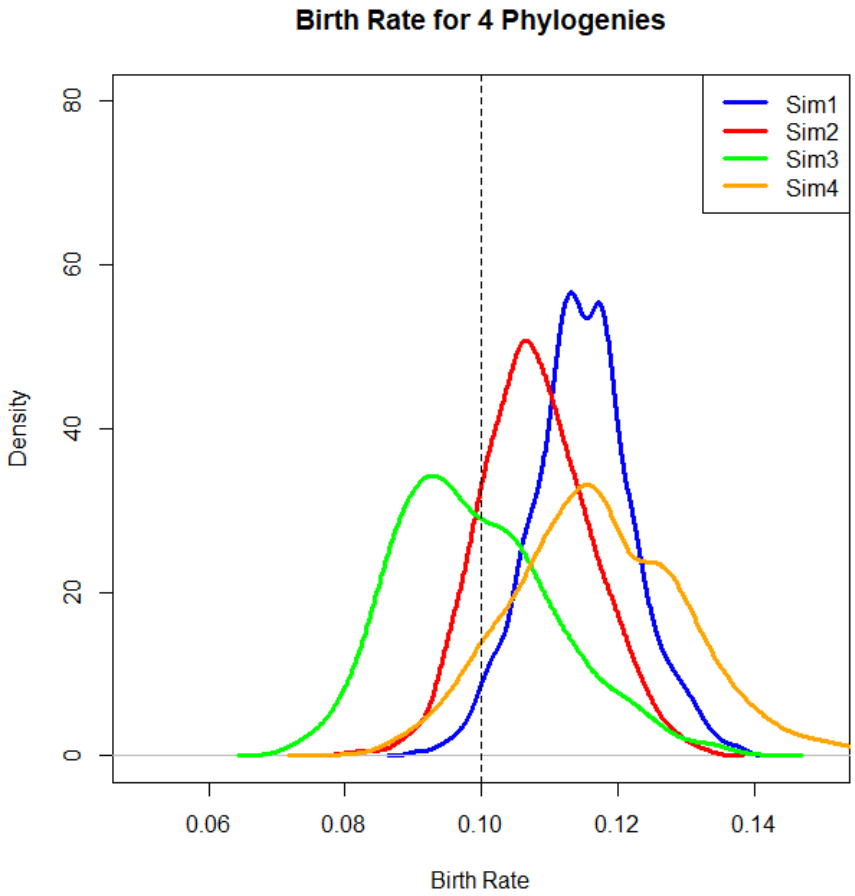

Figure A2 Posterior distributions for birth rate λ_0 across 4 simulation phylogenies and associated Bayesian analyses. Credible intervals cover the true birth rate (vertical dashed black line), but the posterior median is often a poor estimate.

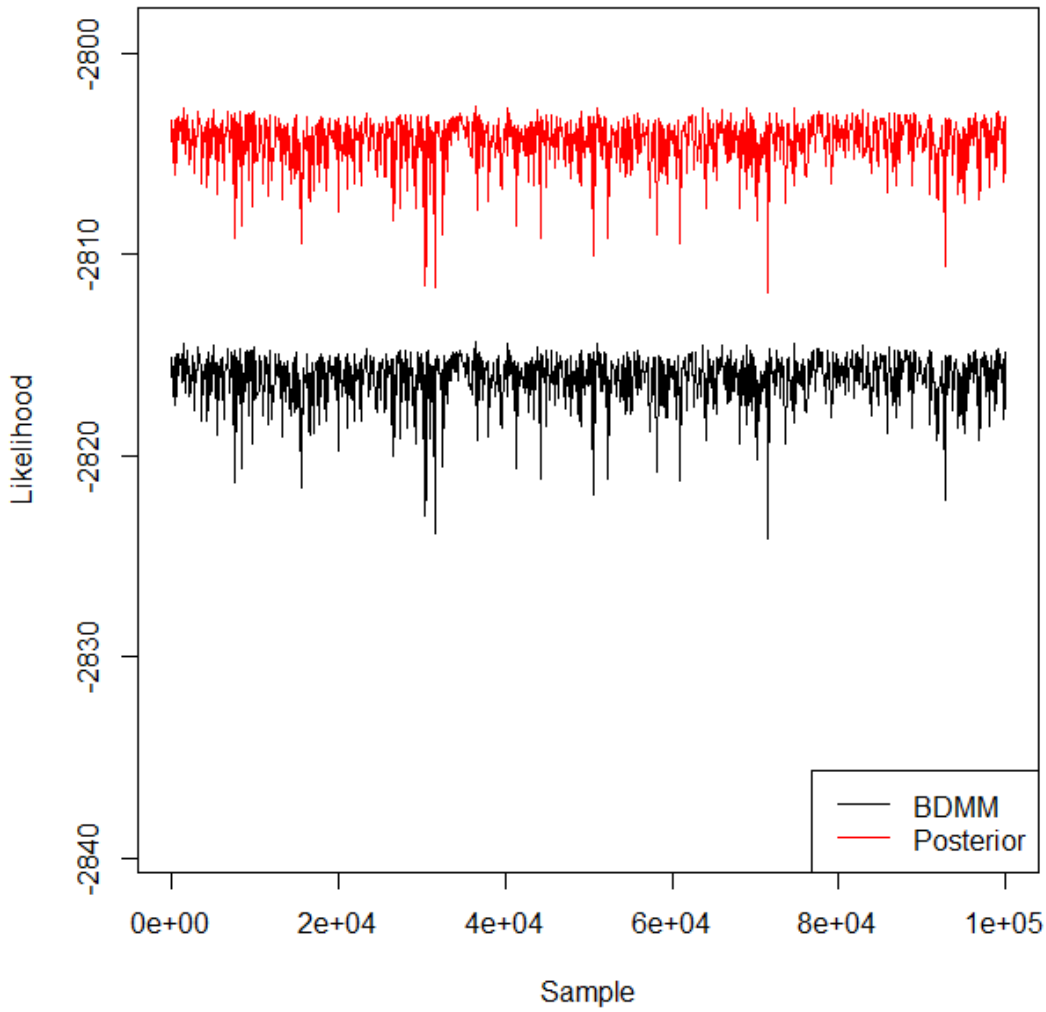

Figure A3 For a simulated phylogeny with 2 site-specific fitness effects, the MTBD (=BDMM) and MFBD (=Posterior) functions track each other well, albeit they do not match exactly. This is because the MFBD is a computationally tractable approximation for MTBD.

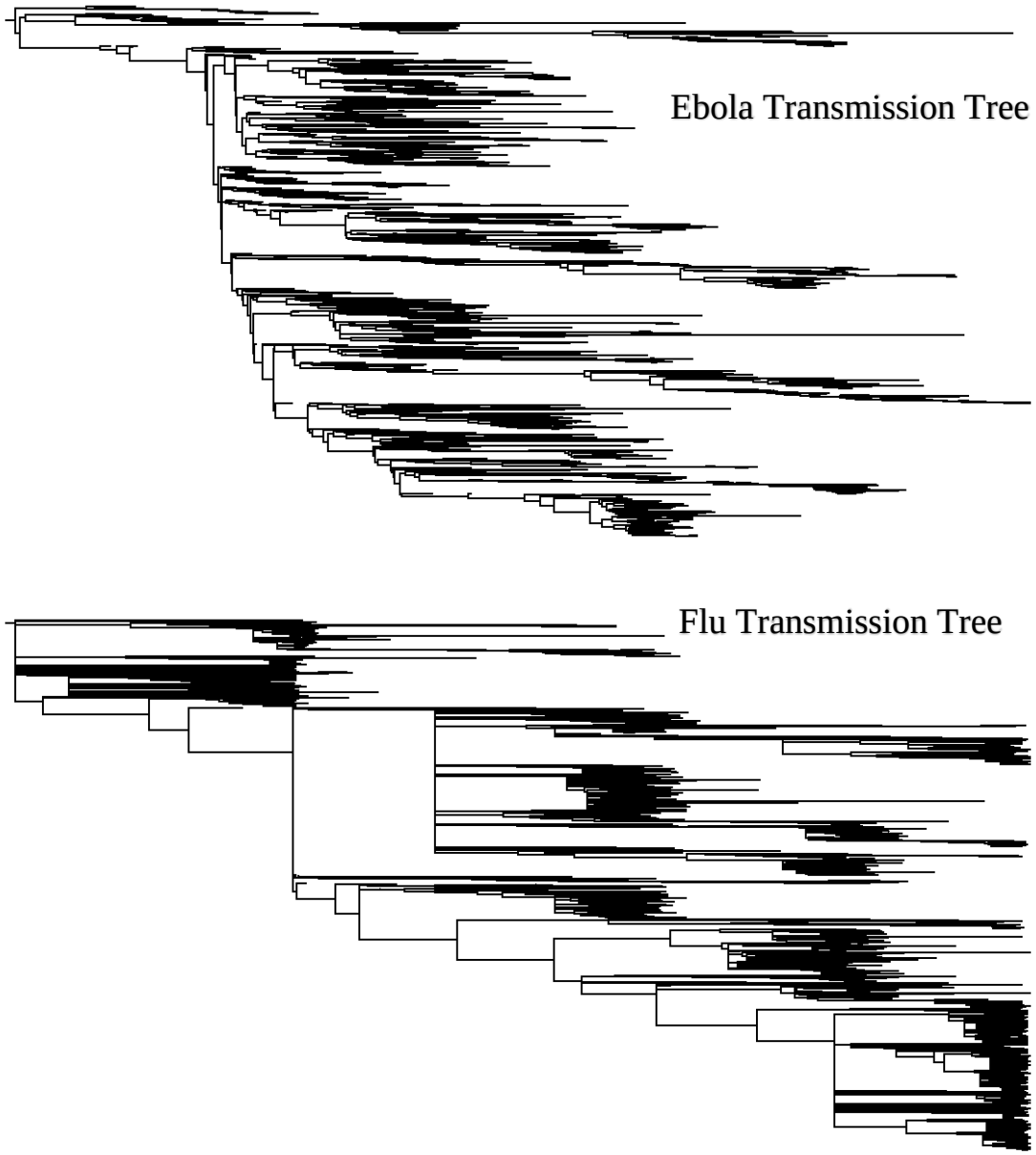

Fig A4 Transmission trees from JavaScript web app https://icytree.org/ for example Ebola (top) and flu (bottom) analyses. These trees mirror Figures 7 and 9 in the original paper, giving evidence that these files from GitHub may be the phylogenetic trees used in the paper analyses. Annotations that may be relevant to create the coloring in Figures 7 and 9 were removed to create Newick-style tree files for IcyTree.

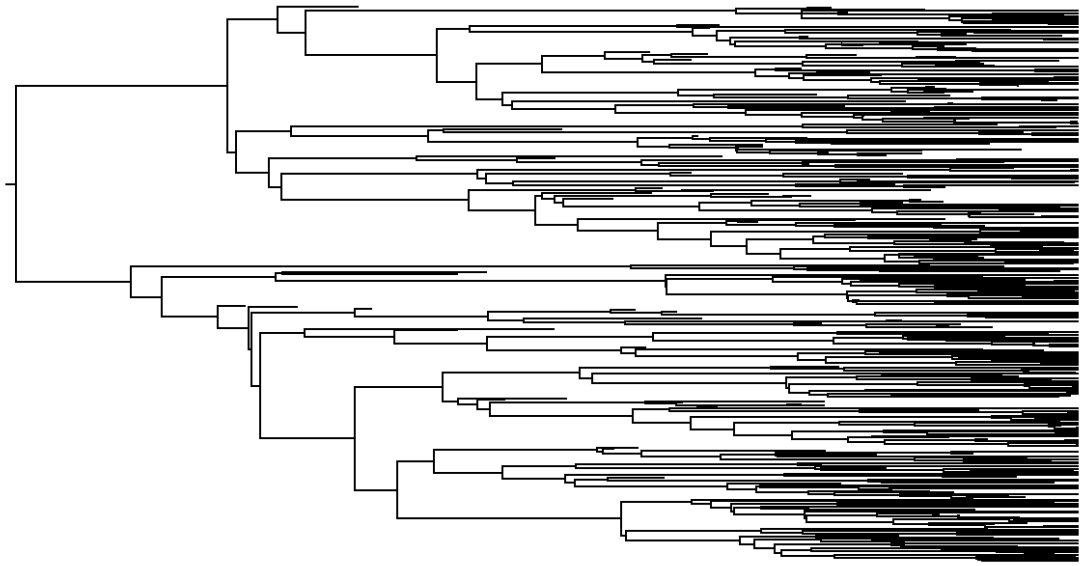

Fig A5 Transmission tree from JavaScript web app https://icytree.org/ for 4 genotypes example analysis. This trees mirrors Figure 2 in the original paper, giving evidence that this file from GitHub may be the phylogenetic tree used in the paper analysis. Annotations that may be relevant to create the coloring in Figure 2 were removed to create Newick-style tree files for IcyTree.

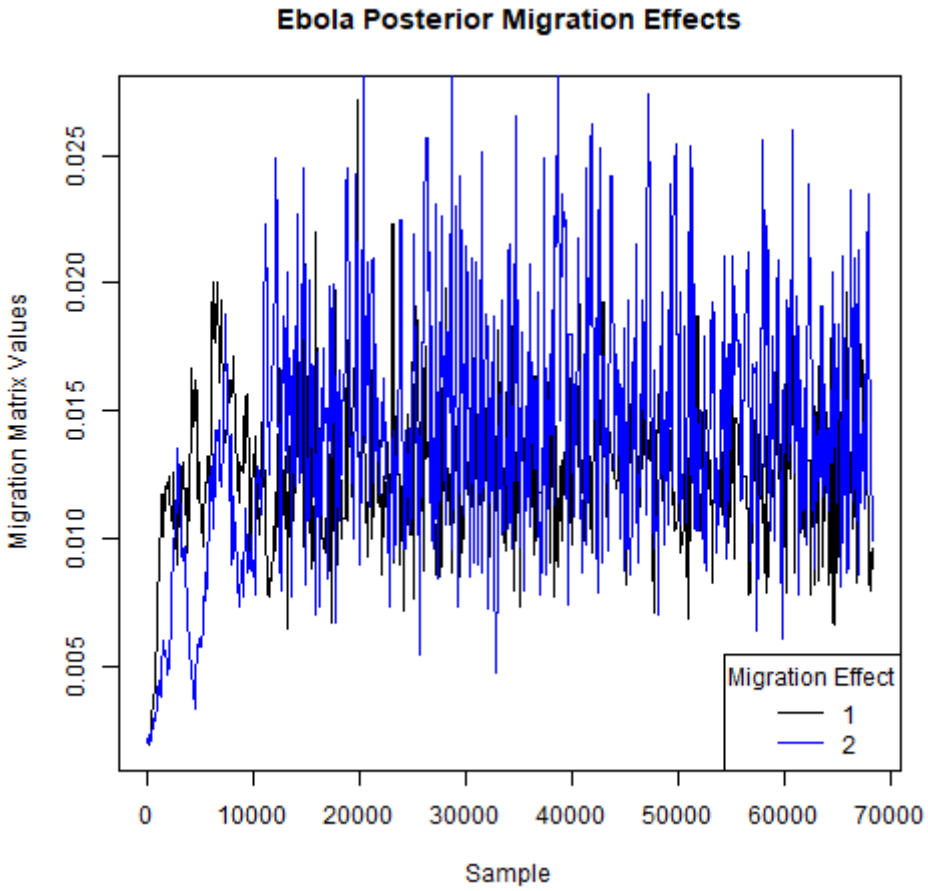

Figure A6 A single MCMC chain for "migration effects" reported in Lumiere code. This diagnostic plot suggests that the MCMC chain converges in ~15,000 iterations, at least marginally for these model effects. Moreover, it seems as if these two effects should have similar posterior medians and credible intervals. It is unclear in the code and in the original paper what these "migration effects" are. They discuss a model with geographic effects for Guinea, Sierra Leone, and Liberia, but they never describe how they model these geographic effects. They say that these are relative transmission rates by location, so the term "migration" appears misleading. They say that the relative transmission rates are ~1.0, so I assume these small "migration effects" are some sort of exponents.

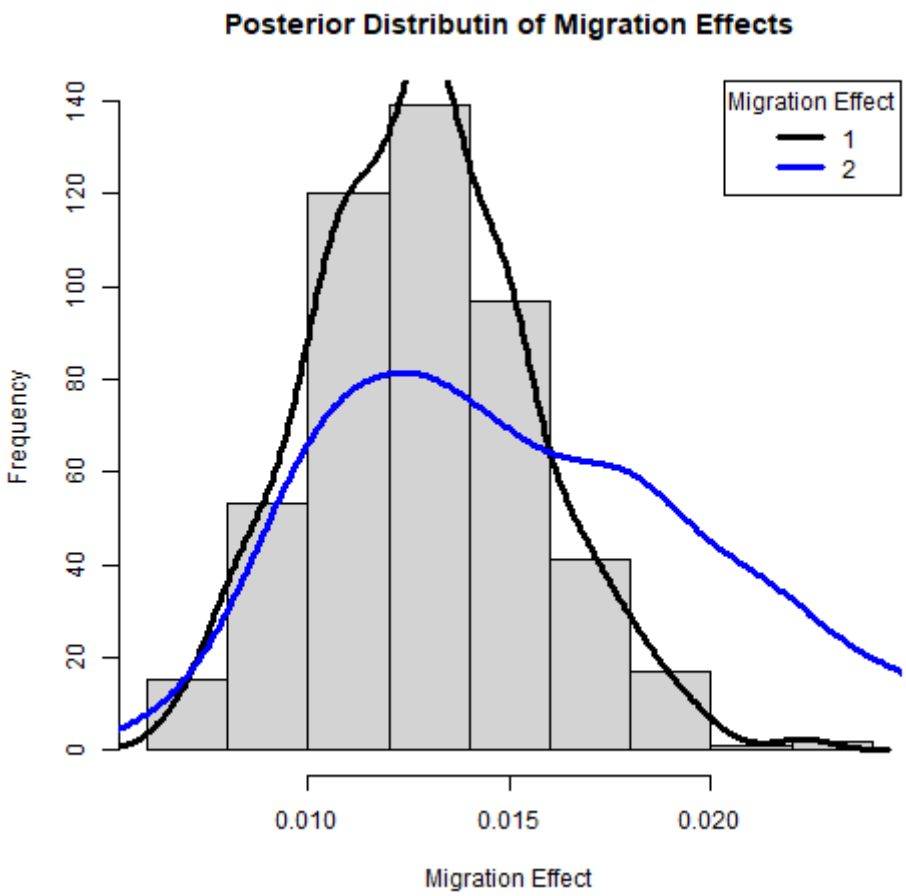

Figure A7 Histogram of MCMC draws from Ebola model + geographic effects. These draws are from after a 10,000 iterations burn-in. Criticisms of the lack of specificity in the code and paper are found in the previous figure. It appears that "migration effects" 1 and 2 have similar posterior medians, but that "migration effect" 2 has a wider credible interval. Assuming these effects correspond to localities, this could indicate a sampling bias for some countries over others.

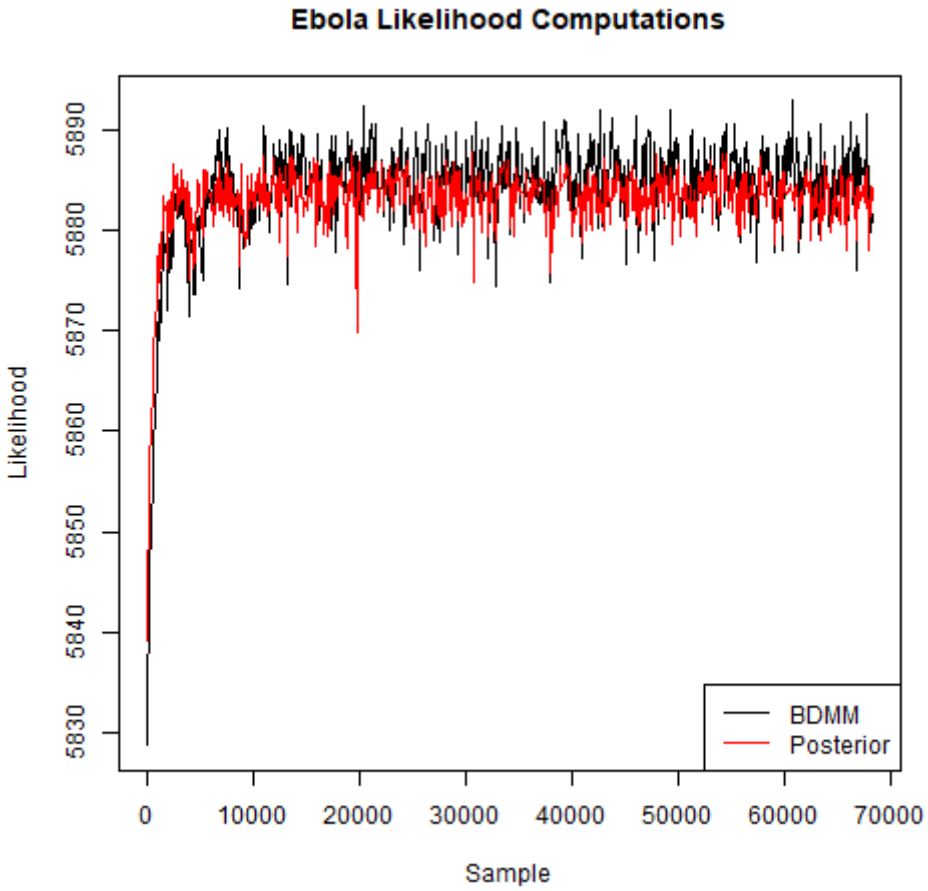

Figure A8 Posterior and "BDMM" "likelihood" values for the Ebola model + geographic effects analysis. The term value "likelihood" is assumed to be a negative log likelihood given the observed domain. In the `Lumiere` code, the first author never specifies what "posterior" is versus "BDMM", as in the original paper they use MTBD for multi-type birth-death and MFBD for marginal fitness birth-death models. I assume "BDMM" = MTBD and that "Posterior" = MFBD. In any event, these negative log likelihood computations appear similar for the Ebola analysis of 9 non-rare genotypes and the single MCMC chain appears to converge after ~15,000 iterations.

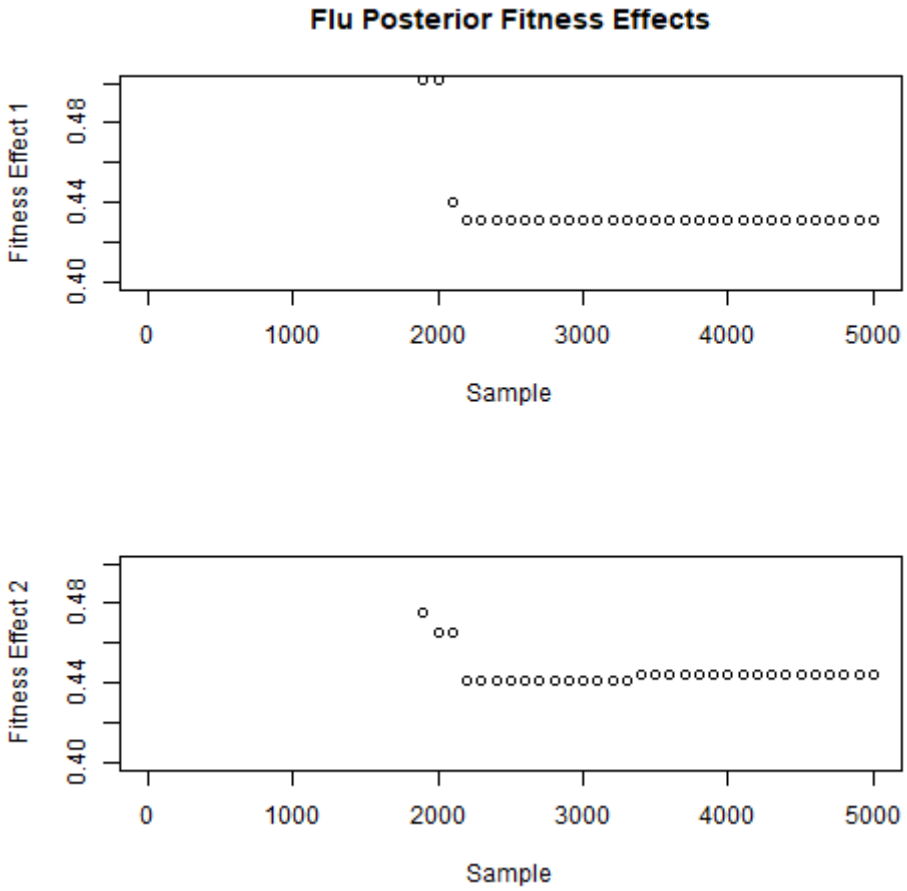

Fig A9 Evaluation of MCMC iterations for 1 chain of applied example flu analysis. 2 of 52 relative fitness effects are compared for 5000 iterations. Within 2000 iterations the fitness effects approach a suggestive convergence. However, the fitness effects appear constant over many iterations, indicating that proposals from the Metropolis Hastings algorithm have an undesirable low acceptance probability. Many iterations with considerable thinning may be necessary to fully characterize the posterior uncertainty.

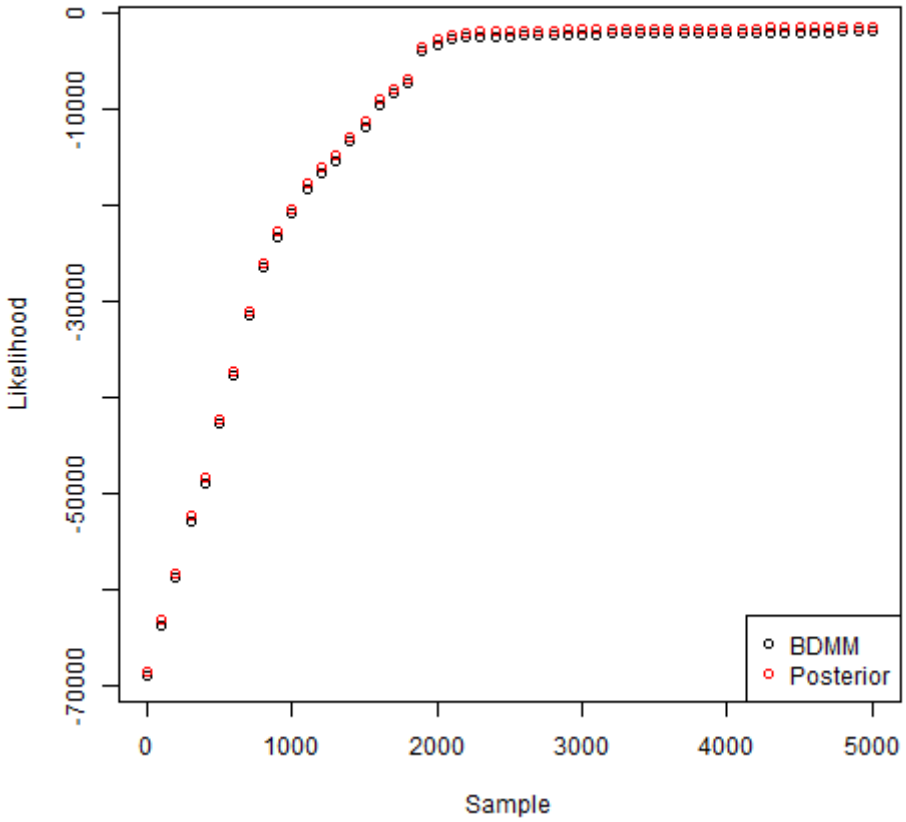

Figure A10 Comparison of likelihood computations for applied example flu analysis. 5000 iterations of 1 MCMC chain are run due to computational limitations. This analysis involves 52 relative fitness effects for the US states and territories. BDMM stands for the MTBD model. Posterior stands for the MFBD model. Posterior values (red) are jittered up by 500 to be visible. Posterior and BDMM values are nearly identical.

References

Cannings, C., Thompson, E. A., & Skolnick, M. H. (1978). Probability functions on complex pedigrees. *Advances in Applied Probability*, *10*(1), 26–61.

Eyre-Walker, A., & Keightley, P. D. (2007). The distribution of fitness effects of new mutations. *Nature Reviews. Genetics*, *8*(8), 610–618.

Felsenstein, J. (1981). Evolutionary trees from DNA sequences: a maximum likelihood approach. *Journal of Molecular Evolution*, *17*(6), 368–376.

Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, D. B. (1995). Bayesian Data Analysis. *Bayesian Data Analysis*. https://doi.org/10.1201/9780429258411

Gillespie, D. T. (1977). Exact stochastic simulation of coupled chemical reactions. *The Journal of Physical Chemistry*, *81*(25), 2340–2361.

Kimura, M. (1983). *The Neutral Theory of Molecular Evolution*. Cambridge University Press.

MacPherson, A., Louca, S., McLaughlin, A., Joy, J. B., & Pennell, M. W. (2021). Unifying Phylogenetic Birth–Death Models in Epidemiology and Macroevolution. *Systematic Biology*. https://doi.org/10.1093/sysbio/syab049

Rasmussen, D. A., & Stadler, T. (2019). Coupling adaptive molecular evolution to phylodynamics using fitness-dependent birth-death models. *ELife*, *8*. https://doi.org/10.7554/eLife.45562

Sanjuán, R., Nebot, M. R., Chirico, N., Mansky, L. M., & Belshaw, R. (2010). Viral mutation rates. *Journal of Virology*, *84*(19), 9733–9748.

Stadler, T., & Bonhoeffer, S. (2013). Uncovering epidemiological dynamics in heterogeneous host populations using phylogenetic methods. *Philosophical*

Transactions of the Royal Society of London. Series B, Biological Sciences, *368*(1614), 20120198.

Suchard, M. A., Lemey, P., Baele, G., & Ayres, D. L. (1093). Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10 Virus Evolution 4 (1): vey016. *DOI: Http://Doi. Org/10.1093*.

Temple, S. D., Manore, C. A., & Kaufeld, K. A. (2022). Bayesian time-varying occupancy model for West Nile virus in Ontario, Canada. *Stochastic Environmental Research and Risk Assessment: Research Journal*. https://doi.org/10.1007/s00477-022-02257-4

Urbanowicz, R. A., Patrick McClure, C., Sakuntabhai, A., Sall, A. A., Kobinger, G., Müller, M. A., Holmes, E. C., Rey, F. A., Simon-Loriere, E., & Ball, J. K. (2016). Human Adaptation of Ebola Virus during the West African Outbreak. *Cell*, *167*(4), 1079-1087.e5.

