# Impact of the tree prior on estimating clock rates during epidemic outbreaks

Simon Möller[a,1], Louis du Plessis[b,1], and Tanja Stadler[a,c,2]

[a]Department of Biosystems Science and Engineering, ETH Zürich, 4058 Basel, Switzerland; [b]Department of Zoology, University of Oxford, Oxford OX1 3SY, United Kingdom; and [c]Swiss Institute of Bioinformatics, 1015 Lausanne, Switzerland

**Bayesian phylogenetics aims at estimating phylogenetic trees together with evolutionary and population dynamic parameters based on genetic sequences. It has been noted that the clock rate, one of the evolutionary parameters, decreases with an increase in the sampling period of sequences. In particular, clock rates of epidemic outbreaks are often estimated to be higher compared with the long-term clock rate. Purifying selection has been suggested as a biological factor that contributes to this phenomenon, since it purges slightly deleterious mutations from a population over time. However, other factors such as methodological biases may also play a role and make a biological interpretation of results difficult. In this paper, we identify methodological biases originating from the choice of tree prior, that is, the model specifying epidemiological dynamics. With a simulation study we demonstrate that a misspecification of the tree prior can upwardly bias the inferred clock rate and that the interplay of the different models involved in the inference can be complex and nonintuitive. We also show that the choice of tree prior can influence the inference of clock rate on real-world Ebola virus (EBOV) datasets. While commonly used tree priors result in very high clock-rate estimates for sequences from the initial phase of the epidemic in Sierra Leone, tree priors allowing for population structure lead to estimates agreeing with the long-term rate for EBOV.**

molecular clock | Bayesian phylodynamics | tree inference | phylogenetics | Ebola

**B**ayesian inference is a powerful tool for the study of phylogenetics and phylodynamics. It allows seamless integration of complicated models with various parameters along with varying degrees of uncertainty. Rather than point estimates, we can compute marginal posterior distributions of our parameters of interest, incorporating the overall uncertainty in the parameters, provided the model fits the data. While the Bayesian phylogenetic framework as a whole is conceptually straightforward, carrying out an analysis can be very complex and thus dedicated software tools have been developed (1–4).

Sequence data alone allow us to infer phylogenetic trees in which branch lengths correspond to the expected number of substitutions along that branch. For sequence data collected serially through time, the dates of the sequences inform us about the branch lengths and the substitution rate in calendar time units, provided that evolution happens on the timescale that was sampled (5). For fast-evolving pathogens, like RNA viruses, serially sampled data collected only within a few months may be sufficient to obtain an estimate of the calendar timescale. For studies of recent macroevolutionary history, ancient DNA (aDNA) (6) samples may inform us about the calendar timescale. Macroevolutionary processes into deep time lack the availability of serially sampled sequences. Instead, contemporary sequences together with fossil samples are used to time calibrate the phylogenetic trees (7).

As a model parameter, the clock rate—namely the rate of nucleotide changes in units of calendar time—is well defined and together with the branch length (in units of calendar time) determines the expected amount of nucleotide change along a branch. These changes may be mutations or substitutions. Many mod-

els and analyses acknowledge the fact that there are multiple different clock rates varying between branches and sites. Comparisons of clock rates therefore need to be done very carefully. Nevertheless, such comparisons have shown for many different empirical datasets stemming from viral outbreaks that the clock rate decreases as the sampling period is increased (8). This was also observed during the 2013–2016 Ebola virus disease (EVD) epidemic in West Africa, where an early study inferred an elevated substitution rate (9) (albeit with a large degree of uncertainty). Further data collection made it evident that substitutions were occurring at a similar rate as suggested by long-term observations (10).

The phenomenon that the clock-rate estimate depends on the timescale used for calibration is not limited to viruses and was first observed more than 10 years ago on avian and primate mitochondrial DNA (11). That paper (11) suggests that the most likely cause is incomplete purifying selection. On shorter timescales, slightly deleterious mutations are still observed in the data and artificially inflate the clock rate. Over a longer time frame these mutations are purged due to purifying selection (see ref. 12 for an illustrative example). It was quickly shown, though, that purifying selection alone cannot explain the observed decline (13). Multiple other factors, such as calibration errors, model misspecification, and sequencing errors, can all contribute to inflated clock-rate estimates (see ref. 12 for a review). The debate about which of these factors contribute and to which degree is still very much ongoing, in particular with

## Significance

**Genetic sequencing data of pathogens allow one to quantify the evolutionary rate together with epidemiological dynamics using Bayesian phylodynamic methods. Such tools are particularly useful for obtaining a timely understanding of newly emerging epidemic outbreaks. During the West African Ebola virus disease epidemic, an unusually high evolutionary rate was initially estimated, promoting discussions regarding the potential danger of the strain quickly evolving into an even more dangerous virus. We show here that such high evolutionary rates are not necessarily real but can stem from methodological biases in the analyses. While most analyses of epidemic outbreak data are performed such that these biases may be present, we suggest a solution to overcome these biases in the future.**

[1]S.M. and L.d.P. contributed equally to this work.

[2]To whom correspondence should be addressed. Email: tanja.stadler@bsse.ethz.ch.

regard to the question of how big a role purifying selection plays (14–16). To understand the complex interplay, simulation studies and analyses of empirical datasets are both important.

For a time-dated phylogenetic analysis in a Bayesian framework we need to specify at least an evolutionary model consisting of the clock model and the substitution model, together with a population dynamic model specifying the tree prior. These components interact in a way that is sometimes counterintuitive. Some efforts have been made to make it easier for researchers to select the most appropriate clock and substitution models (17–22), but fewer efforts have been made to choose the appropriate tree prior. Even if we are interested only in the clock rate and integrate out the uncertainty in tree space, the tree prior can still have an appreciable impact on the posterior distribution of the evolutionary parameters. Even though the models for the clock and the tree are independent components of the analysis, the tree length (i.e., the sum of all branch lengths) and clock rate are highly negatively correlated, as their product needs to explain the overall diversity that is observed in the data. While we put an explicit prior directly on the clock rate, this is not true for the tree length. Rather, the tree length obtains a prior indirectly from the specified tree prior. This indirect influence has not been studied in detail, except for some analytical results for a coalescent (23) and a Yule model (24) with contemporary tips. Results for serially sampled tips or for birth–death processes are to our knowledge not available.

New models for tree priors are regularly investigated using simulation studies in which the model itself, or simpler models, is used to generate phylogenies and sequence alignments (25, 26). While this is a valuable contribution to show that the model can recover true values under ideal circumstances, it offers no information about the robustness of inferences to violations of the underlying model assumptions.

All of the currently available tree priors are huge simplifications to the full range of dynamics seen in a real epidemic. It is thus likely that the true tree will be very poorly supported under the tree prior. If the data are informative enough, the prior will not contribute significantly to the posterior and the true tree can be recovered, given that it has a nonzero probability den-

sity under the tree prior, regardless of how atypical the tree is under the tree prior. However, in data-limited scenarios (such as the start of an epidemic), using a tree prior that provides a poor description of the epidemiological process could result in highly biased estimates of model parameters.

In this paper, we identify some nontrivial conceptual issues arising from the choice of tree priors when estimating clock rates. In particular, we perform a simulation study using a fixed empirical—rather than a simulated—tree and simulate sequence evolution on that tree. We obtained the empirical tree from an analysis of sequences from Guinea during the 2013–2016 EVD epidemic (27, 28). Using an empirical tree (rather than a simulated tree using the tree prior) allows us to assess the robustness of the inference of the (known) clock rate from the simulated sequences, when the tree prior potentially poorly models the underlying tree. By simulating under known substitution and clock models we ensure that any biases observed in the inference must be due to the tree prior and not due to other complicating factors, such as incomplete purifying selection, that play a role in the real world.

We show that for short and moderate sequence lengths the simulated data are not informative enough to lead to unbiased estimates of the clock rate or the tree length, when using commonly employed coalescent and birth–death model tree priors with nonstructured populations. We then analyze the Guinea sequencing data, as well as a dataset sampled during the first month of the epidemic in Sierra Leone (9), using classic tree priors ignoring population structure as well as a tree prior accounting for population structure. We show that tree priors assuming a population without structure lead to the Sierra Leone clock rate being inflated compared with the long-term estimates for Ebola virus (EBOV). A tree model that accounts for structure within the population leads not only to a better fit to the data, but also to Sierra Leone clock-rate estimates that are in good agreement with estimates we obtained for Guinea, as well as the long-term estimates for EBOV.

## Results

**Simulation Study.** The empirical tree used to simulate sequences for the simulation study is shown in Fig. 1A. The tree is obviously
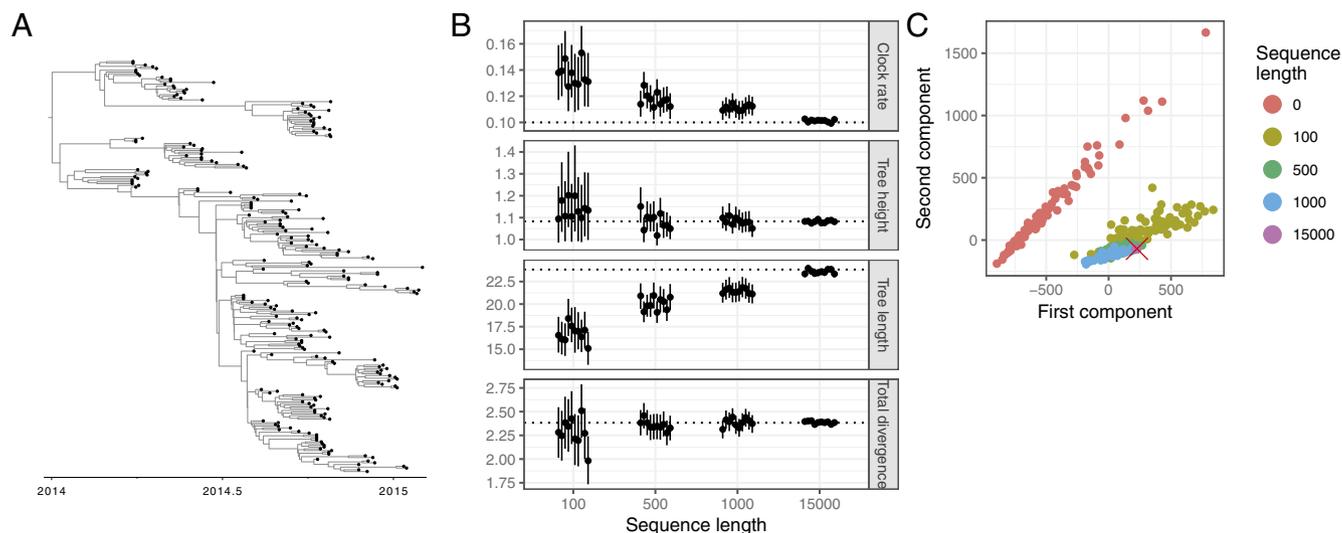


**Fig. 1.** Results of the simulation study. (A) The tree that was used in the simulation study [this tree is the maximum clade credibility (MCC) tree of an analysis under a birth–death skyline model on a dataset consisting of the coding regions of 236 EBOV genomes sampled from patients in Guinea]. (B) The median values and 95% HPD intervals for key parameters estimated from simulated sequences. The dashed lines indicate the true values used in simulations. Clock rate is reported in substitutions per site per year, tree height and tree length in years, and total divergence (product of clock rate and tree length) in substitutions per site. (C) The distribution of topologies of posterior tree samples for analyses of simulated datasets of different sequence lengths, where we projected the Euclidean distances between real-valued representations of the topologies onto a 2D space. The red cross marks the true tree.

less balanced than a typical constant-size coalescent tree (compare with *SI Appendix*, Fig. S2). The median estimate and 95% highest posterior density (HPD) intervals for clock rate, tree height, tree length, and total divergence for each replicate of the simulation study under a constant size coalescent are shown in Fig. 1*B* and *SI Appendix*, Fig. S3*A*. The dashed lines in each panel indicate the true values. The HPD intervals for tree height and total divergence include the true value in all but 1 replicate each, for all sequence lengths, and become smaller with increasing sequence length. The estimates for clock rate and tree length include the true value in none of the 30 replicates for tree length up to sequence length 1,000. The estimates are biased upward and downward, respectively. The bias and the variance decrease as the sequence length increases (*SI Appendix*, Fig. S3*D*), but the true value is covered only by the HPD intervals of 7 and 8 of 10 replicates for sequence length 15,000, for tree length and clock rate, respectively. Without any sequence data (i.e., sequence length of 0), the median inferred value for the clock rate is unbiased and all HPD intervals contain the true value, as expected, since the prior is centered around the true rate (*SI Appendix*, Fig. S4*A*).

Fig. 1*C* shows the posterior distribution of tree topologies, for the first replicate (of 10), projected onto 2D Euclidean space (29), after down-sampling to 101 trees per sequence length and discarding 10% as burn-in. The points representing topologies obtained with sequence data form a cluster around the true topology (marked with a red cross), while topologies originating from the analysis without sequence data are clearly separated.

To illustrate that the above biases result from the empirical Guinea tree being very different from a typical coalescent tree (i.e., the tree prior) we repeated the simulation study on 10 trees simulated under the constant-size coalescent. As expected, about 95% of the HPDs contain the true values. Furthermore, the observed biases are very small and parameter estimates become unbiased with very small HPD intervals for sequences of length 500 or longer (*SI Appendix*, Figs. S5 and S6).

We assessed the robustness of our findings by repeating the simulation study with different clock rates (*SI Appendix*, Figs. S3 and S4) and with a less informative clock rate prior (*SI Appendix*,

Figs. S7 and S8). We further changed the constant population-size assumption of the coalescent tree to exponential growth (*SI Appendix*, Fig. S9) and also repeated the experiment with a birth–death tree prior (*SI Appendix*, Fig. S10). Finally, it has been noted that not exploring the correct topological space can result in biases to branch-length estimates (30). To test whether this hypothesis is responsible for the observed biases we fixed the tree topology (but not the branch lengths) to that of the empirical tree (*SI Appendix*, Fig. S11). Although the magnitudes of the biases change between analyses, the same pattern remains visible in all of our sensitivity analyses.

*SI Appendix*, Fig. S12 shows an analysis of the simulated alignments in a maximum-likelihood framework using the tools RAxML (31) and least-squares dating (32). In this case the clock rate is slightly underestimated for sequences of length 100, whereas it was severely overestimated in a Bayesian framework. For sequences of length 500 and more, the true value is within 1 SD of the inferred mean.

**Empirical Ebola Study.** Fig. 2 shows the results for the two EBOV datasets. For the Guinea dataset the birth–death model leads to the highest clock rate with a median of roughly $1.3 \times 10^{-3}$ substitutions per site per year. Under all of the other models the inferred rate is slightly below $1.2 \times 10^{-3}$ substitutions per site per year. The HPD intervals are largely overlapping. These estimates are in good agreement with the long-term rate estimated over the course of the epidemic [$\sim 1.2 \times 10^{-3}$ substitutions per site per year (10)]. For the tree height the large HPD intervals for the constant coalescent stand out. This model also has the highest median of around 1.1 years, while the lowest estimate of slightly below 1 year is obtained under the birth–death model. The inferred tree length shows the opposite trend to the clock rate, with the birth–death model leading to a median estimate of 17 years whereas the other models result in an estimate between 18 and 19 years. For total divergence there are no noticeable differences between any of the models.

For Sierra Leone all of the unstructured models lead to a median estimate of the clock rate of about $2 \times 10^{-3}$ substitutions per site per year. In contrast, the structured coalescent
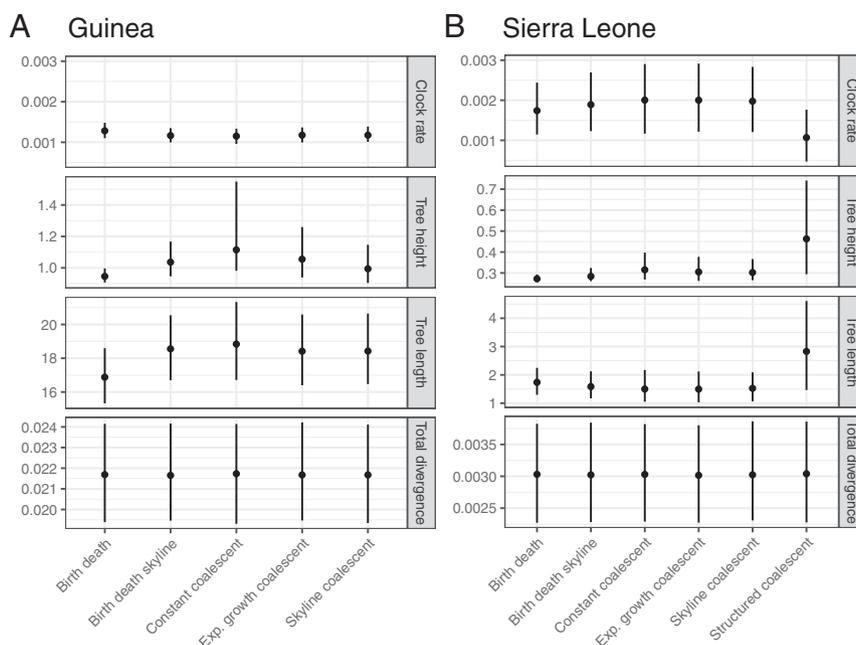
**Fig. 2.** Median and 95% HPD intervals for key parameters inferred from the Guinea (*A*) and Sierra Leone (*B*) datasets under different tree priors. For units refer to the Fig. 1 legend.

model results in a median rate of $1.3 \times 10^{-3}$ substitutions per site per year which agrees with the long-term rate estimated over the course of the epidemic [$\sim 1.2 \times 10^{-3}$ substitutions per site per year (10)]. The opposite trend is observed for the tree length. While the unstructured models result in estimates of around 1.5 years, the structured coalescent leads to a median value of 2.3 years. Similarly, for the tree height, the medians of the unstructured models are around 0.3 years but the birth–death models result in narrower HPD intervals than the coalescent models. Using the structured coalescent, where we assigned demes based on genetic similarity (see *SI Appendix, Supporting Methods*, for details), we obtain a much larger HPD interval around a median of 0.46 years. Again, we find no noticeable difference for total divergence.

*SI Appendix*, Fig. S13 shows the results of the model comparison. For both datasets the structured coalescent is clearly the best-fitting model among those examined. We assessed the robustness of this finding by running path sampling with a varying number of steps (*SI Appendix*, Table S2). While the structured coalescent always presents the best fit, the ranking of the other models varies.

The assignment to demes in the structured model analysis is not necessarily clear a priori and it could be argued that our results are applicable only to the particular assignment we used. To investigate this dependence we assigned sequences in the Sierra Leone dataset randomly to demes. Random deme assignment does not appear to affect estimates of the clock rate, tree height, tree length, or total divergence (*SI Appendix*, Fig. S14). However, we advise caution when interpreting these results, as the analyses mix poorly for the migration model-specific parameters, although all other parameters have effective sample sizes above 200.

## Discussion and Conclusion

Most common tree priors are relatively simple and do not take into account the interplay between the phylogeny, population structure, selection, and other factors that affect the population dynamics. Thus, empirical trees are often less balanced, with different distributions of branching times compared with trees under the prior. The simulation study shows that when simulating along a tree that is based on empirical data, it can be surprisingly difficult to recover the true clock rate, even when very simple clock and substitution models are used. The biases are observed despite the fact that the correct substitution model and an unbiased prior on the clock rate are used and persist even when conditioning on the true topology and allowing only branch lengths to vary. The problem arises from a misspecification of the tree prior (Fig. 1 and *SI Appendix*, Fig. S4) which will be difficult to detect in empirical datasets where the truth is unknown. Thus, the biases may be overcome by using more appropriate tree priors; however, the biases we observe will not vanish when using more complex clock-rate priors [such as the continuous-time Markov chain reference prior (33)], as our simulated data do not contain rate variation. Instead, it would be more difficult to disentangle biases due to the tree prior from those due to the clock-rate prior when using more complex clock-rate priors. The tree prior implies an indirect prior for the tree length, which may cause a bias in the clock-rate estimate. Biases on the clock rate disappear when sampling from the prior without sequence data (*SI Appendix*, Fig. S4).

These results may initially seem counterintuitive. To explain them, we briefly review the Bayesian phylogenetic framework. Let the clock rate and substitution rate be denoted by $\mu$, the tree by $\mathcal{T}$, the parameters of the tree prior by $\theta$, and the data (sequence alignment) by $\mathcal{D}$. In a simple form, the posterior is given by $P(\mu, \theta, \mathcal{T} | \mathcal{D}) \propto P(\mathcal{D} | \mu, \mathcal{T}) f(\mathcal{T} | \theta) f(\mu) f(\theta)$, if we assume that the clock rate and tree priors are independent of each other ($f(\mu, \theta) = f(\mu) f(\theta)$), as is the case in all of the models

we analyzed. Without any sequence data, we have $P(\mathcal{D} | \mu, \mathcal{T}) = 1$ and therefore, when sampling from the posterior, $\mathcal{T}$ and $\mu$ can change independently; thus we do not observe biases in our analyses without sequence data. Adding only a small amount of sequence data can make a large difference in the inference as it links the tree and the clock rate via $P(\mathcal{D} | \mu, \mathcal{T})$. If the tree prior causes the tree length to be underestimated, the model will compensate for this by increasing the clock rate to explain the overall diversity in the data; this is seen in our simulation analyses with sequences.

The prior distribution on the tree space, $f(\mathcal{T} | \theta)$, is a distribution over topologies and branch lengths. This indirectly gives rise to the prior on the tree length (depicted in *SI Appendix*, Fig. S4). Upon adding sequence data the topologies that before had a high prior support become very unlikely (Fig. 1C) and under this constrained topological space, the indirect prior on the tree length is altered as well (see *SI Appendix*, Fig. S2 for an illustration). The constrained topological space causes a downward bias in tree length for sequence lengths 100, 500, and 1,000 as seen in Fig. 1B, which in turn causes the clock rate to be overestimated. Our simulation study indicates that in data-limited scenarios, none of the commonly used unstructured tree priors provide unbiased estimates for data simulated along an empirical tree (Fig 1B and *SI Appendix*, Figs. S9 and S10). Despite very different priors for the dynamics in the underlying unstructured population, too much weight is given to certain trees, causing a bias in the tree length. The dependence that even a small amount of sequence data introduces between the tree prior and the clock-rate prior, along with the negative correlation between the tree length and the clock rate, in turn results in biases to the clock rate. Negative correlations among parameters that are independent in the priors can be indicative of a model that is overparameterized (34). However, this is not necessarily the case, and the fact that we recover unbiased estimates as the sequence length is increased shows that this is unlikely to be a factor here. Since trees sampled from real epidemics are likely to be highly atypical under the currently available tree priors, the chosen tree prior may result in biased estimates, especially during the early phase of an epidemic.

The influence of sequence data on the topological space and on the tree length can be illustrated with a toy example (Fig. 3). Consider a tree with two contemporary samples and one past sample. For a small population size a coalescent tree prior would give high probability to the tree topology in which the two contemporary tips form a cherry. When sequence data are added, it may become obvious, though, that the cherry should be formed between one of the contemporary tips and the sample from the past. This effectively puts a lower bound on the tree length. We point out that the tree prior restricts not only the topological space but also the branch lengths and thus restricts the tree length. This restriction also leads to biases, as observed when constraining our simulation analysis to the true tree topology, but allowing branch lengths to vary.
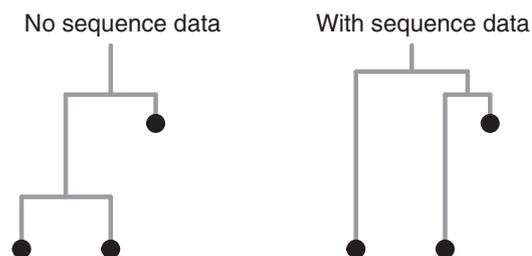


**Fig. 3.** A toy example of how the sequence data can influence the branch length via changing the topology.

EVOLUTION

The usual approach to assess whether the data are informative on a parameter of interest (e.g., the clock rate) is to look for a departure of the posterior distribution from the prior. However, without enough data to overcome methodological biases from model misspecification, simply showing a departure from the prior is not sufficient. In our simulation study the clock rate and tree length both initially show very clear departures from their priors for shorter sequence lengths. However, it is only once more sequencing data are added that estimates become unbiased. Thus, in the absence of independent estimates, departures from the prior should not be taken as evidence that the data are informative enough to produce unbiased estimates.

We note that the tree height can be estimated much more reliably than the tree length (Fig. 1*B*). Like the total divergence, it is a global parameter and a small amount of data are already informative about it. As there are many trees of the same height but with different lengths (e.g., Fig. 3), inferring the length correctly is a much harder problem and thus more susceptible to biases from the tree prior.

We showed that a maximum-likelihood analysis which does not use a tree prior did not suffer from an upward bias in clock-rate estimates, offering a potential way to check for clock-rate biases in Bayesian analyses. In contrast, maximum-likelihood estimates were underestimated for the shortest sequences (although much shorter sequences were needed to obtain unbiased estimates than within a Bayesian framework). These results are reminiscent of ref. 35, where the authors show that branch lengths tend to be underestimated in a Bayesian framework, whereas maximum-likelihood estimates tend to be inflated.

The analysis of the two empirical datasets also confirms that the tree prior can influence the inferred clock rate. For Sierra Leone, where 81 sequences were sampled over 3 months, we see that the choice of tree prior can heavily influence the estimated clock rate. If a structured model had been used in the original analysis (9), then the difference between the short- and long-term estimates would have disappeared. For the Guinea dataset (236 sequences spanning 10 months) we get similar clock-rate estimates across tree priors. In fact, the median clock rates of the Sierra Leone dataset under the structured coalescent and the Guinea dataset under any tree prior are estimated to be in the range $1.15 - 1.3 \times 10^{-3}$, which is in good agreement with the long-term rate of $1.2 \times 10^{-3}$ substitutions per site per year estimated over the course of the epidemic (10).

We did not use the structured coalescent in its usual form, as a model of migrations between discrete locations. Instead, we assigned sequences to different demes in the structured model based on genetic distance between sequences, as well as in a random manner. In this sense, the structured model allows distinct lineages to coalesce at different rates, introducing a greater degree of flexibility in the tree prior. Since clock-rate estimates are not affected by the particular chosen deme assignment, we suggest that the reduction in biases does not stem from the introduction of realistic population structure, but because the structured model assigns a higher probability to unbalanced trees with higher tree lengths than any of the unstructured models. It may be fruitful to apply a structured model that does not rely on classified tips, thereby avoiding arbitrary deme assignments (e.g., refs. 36 and 37). Analogous to our findings regarding clock rates in epidemiological studies, simulation studies in the context of aDNA have shown that complex population structure in the past can lead to biased estimates of clock rate if the data are analyzed under a model that is too simple (38).

Model comparison suggests that structured models are strongly preferred, indicating that the data appear to demand a model that allows for more variability in the tree distribution than provided by unstructured models. However, correct estima-tion of the marginal likelihood is a difficult and computationally demanding task and the results should therefore be taken with a grain of salt. Furthermore, marginal likelihoods and Bayes factors say nothing about the absolute goodness of fit (39). This can be assessed only with even more computationally demanding methods like posterior predictive simulation (40). Regardless of its shortcomings, our results show the importance of carefully choosing a tree prior and that this choice can strongly influence the clock-rate estimates.

In this paper we highlight some conceptual problems in the inference of the clock rate when using Bayesian phylogenetic tools. The interaction between tree prior and clock-rate estimation can be complex and nonintuitive. We used a simulation study to demonstrate that deviation of the posterior clock-rate distribution from the prior does not necessarily imply a signal in the data and can be a mere artifact of the chosen tree prior. The study indicated that even under the simplest substitution and clock models it may not be possible to recover the true parameter values if the tree prior is a poor description of the evolutionary process. The reanalysis of an Ebola dataset from Sierra Leone showed that the high mutation rate that was reported originally could be due to biases introduced by model misspecification and that the inferred rate under a more flexible tree prior comes very close to the long-term estimate. Overall this emphasizes the need to choose the tree prior carefully—even if the parameter of interest is the clock rate—and demands further investigations into how overall model fit in Bayesian phylogenetic analyses can be assessed.

## Materials and Methods

**Simulation Study.** We simulated sequence evolution along a fixed empirical tree using simple clock and substitution models and subsequently analyzed the resulting alignment using the same clock and substitution models under a simple unstructured coalescent tree prior. The empirical tree was obtained from an analysis of the coding regions of 236 Ebola virus genomes sampled from patients in Guinea over a period of 10 months (previously described in refs. 27 and 28). Details of the analysis are described in *SI Appendix, Supporting Methods*.

We simulated sequences of length 100 base pairs (bp), 500 bp, 1,000 bp, and 15,000 bp using a Jukes–Cantor (JC69) (41) substitution model and a fixed clock rate of 0.1 substitutions per site per year. For each sequence length we performed 10 independent simulations. The parameters in the setup of this simulation are not meant to be biologically relevant but rather an illustrative example using the simplest possible substitution model. The simulated alignments were subsequently analyzed in Beast2 (2), using the same model of molecular evolution used to simulate the sequences (i.e., strict clock and JC69 model). We chose a normal distribution with standard deviation of 0.02 around the true value of 0.1 as a prior for the clock rate and a lognormal distribution with $M = 0$ and $S = 0.5$ for the population size of a constant-size coalescent tree prior. In sensitivity analyses we varied the clock rate used in simulations, the precision of the clock-rate prior, and the tree prior (using different coalescent and birth–death tree priors). Model specifics and further details can be found in *SI Appendix, Supporting Methods*.

**Empirical EBOV Study.** We investigated the dependence of clock-rate inference on the chosen tree prior on two EBOV datasets. The first dataset is the one used to generate the tree for the simulation study and we refer to it as the Guinea dataset. The second dataset contains whole-genome data of 81 sequences sampled over 3 months, with all but 3 earlier sequences from Guinea sampled during the first month of the epidemic in Sierra Leone (9). Estimates of the clock rate from this dataset are about twice as high during the outbreak as between outbreaks, albeit with wide confidence intervals (9). We refer to this dataset as the Sierra Leone dataset.

For the analyses in Beast2 we used a strict clock and the Hasegawa, Kishino and Yano (HKY) substitution model (42) without site heterogeneity for all models. We used six different tree priors: constant-rate birth–death, birth–death skyline (26), constant population-size coalescent, exponential growth coalescent, coalescent skyline (25), and structured coalescent (43). We subsequently used path sampling (44) to assess the relative goodness of fit of the different models. Model specifics are listed in *SI Appendix, Supporting Methods*.

    

1. Drummond AJ, Rambaut A (2007) BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol Biol* 7:214.
2. Bouckaert R, et al. (2014) BEAST 2: A software platform for Bayesian evolutionary analysis. *PLoS Comput Biol* 10:e1003537.
3. Huelsenbeck JP, Ronquist F (2001) MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* 17:754–755.
4. Höhna S, et al. (2016) RevBayes: Bayesian phylogenetic inference using graphical models and an interactive model-specification language. *Syst Biol* 65:726–736.
5. Drummond AJ, Pybus OG, Rambaut A, Forsberg R, Rodrigo AG (2003) Measurably evolving populations. *Trends Ecol Evol* 18:481–488.
6. Lambert DM, et al. (2002) Rates of evolution in ancient DNA from adelie penguins. *Science* 295:2270–2273.
7. Donoghue PCJ, Yang Z (2016) The evolution of methods for establishing evolutionary timescales. *Philos Trans R Soc Lond B Biol Sci* 371:20160020.
8. Duchêne S, Holmes EC, Ho SYW (2014) Analyses of evolutionary dynamics in viruses are hindered by a time-dependent bias in rate estimates. *Proc R Soc B Biol Sci* 281:20140732.
9. Gire SK, et al. (2014) Genomic surveillance elucidates Ebola virus origin and transmission during the 2014 outbreak. *Science* 345:1369–1372.
10. Holmes EC, Dudas G, Rambaut A, Andersen KG (2016) The evolution of Ebola virus: Insights from the 2013–2016 epidemic. *Nature* 538:193–200.
11. Ho SYW, Phillips MJ, Cooper A, Drummond AJ (2005) Time dependency of molecular rate estimates and systematic overestimation of recent divergence times. *Mol Biol Evol* 22:1561–1568.
12. Ho SYW, et al. (2011) Time-dependent rates of molecular evolution. *Mol Ecol* 20:3087–3101.
13. Woodhams M (2006) Can deleterious mutations explain the time dependency of molecular rate estimates? *Mol Biol Evol* 23:2271–2273.
14. Emerson BC, Hickerson MJ (2015) Lack of support for the time-dependent molecular evolution hypothesis. *Mol Ecol* 24:702–709.
15. Ho SYW, Duchêne S, Molak M, Shapiro B (2015) Time-dependent estimates of molecular evolutionary rates: Evidence and causes. *Mol Ecol* 24:6007–6012.
16. Emerson BC, Alvarado-Serrano DF, Hickerson MJ (2015) Model misspecification confounds the estimation of rates and exaggerates their time dependency. *Mol Ecol* 24:6013–6020.
17. Ho SYW, Duchêne S (2014) Molecular-clock methods for estimating evolutionary rates and timescales. *Mol Ecol* 23:5947–5965.
18. Duchêne DA, Duchêne S, Holmes EC, Ho SY (2015) Evaluating the adequacy of molecular clock models using posterior predictive simulations. *Mol Biol Evol* 32:2986–2995.
19. Duchêne S, Giallonardo FD, Holmes EC (2015) Substitution model adequacy and assessing the reliability of estimates of virus evolutionary rates and time scales. *Mol Biol Evol* 33:255–267.
20. Bouckaert RR, Drummond AJ (2017) bModelTest: Bayesian phylogenetic site model averaging and model comparison. *BMC Evol Biol* 17:42.
21. Kelchner SA, Thomas MA (2007) Model use in phylogenetics: Nine key questions. *Trends Ecol Evol* 22:87–94.
22. Ripplinger J, Sullivan J (2010) Assessment of substitution model adequacy using frequentist and Bayesian methods. *Mol Biol Evol* 27:2790–2803.
23. Eriksson A, Mehlig B, Rafajlovic M, Sagitov S (2010) The total branch length of sample genealogies in populations of variable size. *Genetics* 186:601–611.
24. Stadler T, Steel M (2012) Distribution of branch lengths and phylogenetic diversity under homogeneous speciation models. *J Theor Biol* 297:33–40.
25. Drummond AJ, Rambaut A, Shapiro B, Pybus OG (2005) Bayesian coalescent inference of past population dynamics from molecular sequences. *Mol Biol Evol* 22:1185–1192.
26. Stadler T, Kuhnert D, Bonhoeffer S, Drummond AJ (2013) Birth-death skyline plot reveals temporal changes of epidemic spread in HIV and hepatitis C virus (HCV). *Proc Natl Acad Sci USA* 110:228–233.
27. Carroll MW, et al. (2015) Temporal and spatial analysis of the 2014–2015 Ebola virus outbreak in West Africa. *Nature* 524:97–101.
28. Simon-Loriere E, et al. (2015) Distinct lineages of Ebola virus in Guinea during the 2014 West African epidemic. *Nature* 524:102–104.
29. Kendall M, Colijn C (2016) Mapping phylogenetic trees to reveal distinct patterns of evolution. *Mol Biol Evol* 33:2735–2743.
30. Wang Y, Yang Z (2014) Priors in Bayesian phylogenetics. *Bayesian Phylogenetics: Methods, Algorithms, and Applications*, eds Chen MH, Kuo L, Lewis PO (CRC Press, Boca Raton, FL), pp 5–24.
31. Stamatakis A (2014) RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30:1312–1313.
32. To TH, Jung M, Lycett S, Gascuel O (2015) Fast dating using least-squares criteria and algorithms. *Syst Biol* 65:82–97.
33. Ferreira MAR, Suchard MA (2008) Bayesian analysis of elapsed times in continuous-time Markov chains. *Can J Stat* 36:355–368.
34. Rannala B (2002) Identifiability of parameters in MCMC Bayesian inference of phylogeny. *Syst Biol* 51:754–760.
35. Schwartz RS, Mueller RL (2010) Branch length estimation and divergence dating: Estimates of error in Bayesian and maximum likelihood frameworks. *BMC Evol Biol* 10:5.
36. Stadler T, Bonhoeffer S (2013) Uncovering epidemiological dynamics in heterogeneous host populations using phylogenetic methods. *Philos Trans R Soc Lond B Biol Sci* 368:20120198.
37. Barido-Sottani J, Stadler T (2017) Accurate detection of HIV transmission clusters from phylogenetic trees using a multi-state birth-death model. bioRxiv:215491.
38. Navascués M, Emerson BC (2009) Elevated substitution rate estimates from ancient DNA: Model violation and bias of Bayesian methods. *Mol Ecol* 18:4390–4397.
39. Gatesy J (2007) A tenth crucial question regarding model use in phylogenetics. *Trends Ecol Evol* 22:509–510.
40. Brown JM (2014) Detection of implausible phylogenetic inferences using posterior predictive assessment of model fit. *Syst Biol* 63:334–348.
41. Jukes TH, Cantor CR (1969) Evolution of protein molecules. *Mamm Protein Metab* 3:21–132.
42. Hasegawa M, Kishino H, Yano Ta (1985) Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. *J Mol Evol* 22:160–174.
43. Vaughan TG, Kühnert D, Popinga A, Welch D, Drummond AJ (2014) Efficient Bayesian inference under the structured coalescent. *Bioinformatics* 30:2272–2279.
44. Baele G, et al. (2012) Improving the accuracy of demographic and molecular clock model comparison while accommodating phylogenetic uncertainty. *Mol Biol Evol* 29:2157–2167.

EVOLUTION