

Revisions MS: MBE-25-0816

**Title: The phylodynamic threshold of measurably evolving populations**

**Associate Editor**

**Editors' comments to the author:**

Dear authors,

I have received comments from two expert reviewers, and have carefully read the paper myself. While the first reviewer had few remarks, the second highlights major issues, which I had independently noted in my own reading. The central problem we both identified was the limited—and somewhat ill-suited—choice of priors to be explored. Since this choice is the starting point of a central aspect of the work, revising it implies an extensive redo of the study. For this reason, I cannot recommend acceptance.

I would have liked to see a much larger set of priors than simply two sets of values shown in table 1. Moreover, this seems like the perfect context to explore hyperpriors, i.e., priors on the parameters of the exponential and gamma distributions. Here again, a set of hyperpriors could be explored. Reducing the number of replicates in each configuration of priors would be a reasonable comprise to save on computation.

I wish to emphasize that I consider the central idea of the study to have merit. If you feel able to redo a new study based on the reviewers' comments (notably reviewer 2) and mine, and address all points raised carefully, I would encourage you to resubmit as a new manuscript.

**Response:**

We would like to thank the Editor and the two reviewers for their helpful feedback. We have attempted to address all comments and concerns in our revised manuscript, which entailed a novel and extensive set of simulation experiments.

Our revisions include analyses of all data with five sampling windows, and nine configurations for which the prior on the mean evolutionary rate has different degrees of uncertainty and concentrated on the value used to generate the data, and a bias of one order of magnitude above or below the truth. We also conducted a set of analyses using hyper-prior distributions for the parameters that govern the relaxed molecular clock with an underlying lognormal distribution. The sum of our novel analyses consisted of a total of 5,000 new simulation replicates for the conditions stated above. Please see our detailed actions in our point-by-point response.

**Reviewer(s)' comments to author:**

**Reviewer: 1**

Comments to the Author The paper "The phylodynamic threshold of measurably evolving populations" by Ariane Weber et al represents a useful and systematic assessment of the so-called phylogenetic threshold and its sensitivity to data, sampling and model priors.

I have a couple of comments, most of them minor.

Major:

Lines 136-140: Here the authors state that *M. tuberculosis* was considered to evolve too slowly for molecular clock calibration, citing their own paper from 2016. On the contrary, a study from 2015 clearly demonstrated it to be measurably evolving (Eldholm 2015 Nat Commun) at the scale of decades. This was later confirmed, not shown, by the papers cited (Kuhnert, Menardo, Merker). The authors should correct this

framing and cite all key papers on the matter.

**Response:** We have edited this section and included all references mentioned by the reviewer:

*The causative agent of tuberculosis, the bacterium *Mycobacterium tuberculosis*, was commonly considered to evolve too slowly for calibrating the molecular clock using samples collected over a few years Duchene et al. (2016). A range of studies have shown, however, that for *M. tuberculosis* a genome sampling window of a few decades might be sufficient for reliable clock calibration (Eldholm et al., 2015, Kühnert et al., 2018, Menardo et al., 2019, Merker et al., 2022).*

Minor:

Line 36: Rate variation among branches - but also among sites?

**Response:** We added a reference for the context of evolutionary rate variation among sites (Ho, 2014). The molecular clock as described in this section only models rate variation among lineages. Models that allow for clock rate variation among sites are known as heterotachy models and we would like to retain the distinction here. This section now reads as follows:

*The development of molecular clock models as statistical processes relaxes this and other assumptions by allowing for rate variation among branches (and sometimes sites see Ho 2014) in phylogenetic trees (reviewed by Guindon (2020), Ho and Duchêne (2014)).*

Line 54: I would say that most (all?), not "some" bacteria are measurably evolving. This applies even to very diverse species, as long as one zooms in on particular clades etc.

**Response:** We agree that most bacteria are measurably evolving, but we contend that not all viruses and bacteria are measurably evolving for a range of reasons. To avoid confusion we rephrased this section as follows:

*Rapidly evolving organisms, notably many viruses and bacteria, have been found to accrue an appreciable number of mutations over the sampling timescale.*

Figure 4: The shades of blue in panel b aren't easily separated by eye. I suggest trying some more diverging hues/colours.

**Response:** We edited the colours to improve the contrast.

## Reviewer: 2

Comments to the Author Major Comments:

This manuscript begins with a nice and illuminating discussion of measurably evolving populations, the phylodynamic threshold, and temporal signal. This paper aims to tackle an important set of questions, and I appreciate the authors' contribution. Unfortunately, I think that the study design could have been much better and the exposition could have been much clearer. I also believe that the end results are not as compelling and insightful as one might hope for. I believe that there are many aspects of the paper that should be redone or expanded.

**Response:**

The Reviewer raises an important point that we attempted to address comprehensively. To this end, we conducted novel simulation experiments using nine configurations for the prior on the mean evolutionary rate that include those where this parameter is dramatically over- and underestimated and with varying degrees of uncertainty. Please see our detailed actions below.

A major focus of this paper is clarifying the impact of prior distributions on the phylogenetic threshold. I was therefore surprised by the very limited range of prior distributions that were used to analyze the simulated data. These analyses form the heart of this paper, and a much wider range of priors (not simply or two additional sets of priors) should be examined. I understand that these analyses are time consuming, but I think fewer replicates of each simulation scenario, with a wider range of analyses (corresponding to a wider range of priors) for each replicate would be a good trade-off.

**Response:**

We conducted analyses with a range of priors for the mean evolutionary rate,  $M$  (the mean rate of the relaxed molecular clock with underlying lognormal distribution, where branch rates  $r_i$  are given by  $r_i \sim \text{Lognormal}(\mu, \sigma)$ ).

- First, we considered the situation where the prior on  $M$  is centred around the value used to generate the data ( $1.5 \times 10^{-5}$  subs/site/year), but with three degrees of uncertainty:

$M \sim \text{Gamma}(\text{shape} = 1.5, \text{rate} = 1 \times 10^5)$ , with mean of  $1.5 \times 10^{-5}$  and 95% quantile width of three times the mean.

$M \sim \text{Gamma}(\text{shape} = 0.3, \text{rate} = 2 \times 10^4)$ , with mean of  $1.5 \times 10^{-5}$  and 95% quantile width of six times the mean.

$M \sim \text{Gamma}(\text{shape} = 15, \text{rate} = 1 \times 10^6)$ , with mean of  $1.5 \times 10^{-5}$  and 95% quantile width of about the same as the mean.

- Second, we used three prior configurations for the mean  $\mu$  was one order of magnitude higher than the truth:

$M \sim \text{Gamma}(\text{shape} = 1.5, \text{rate} = 1 \times 10^4)$ , with a mean of  $1.5 \times 10^{-4}$ .

$M \sim \text{Gamma}(\text{shape} = 0.3, \text{rate} = 2 \times 10^3)$ , with a mean of  $1.5 \times 10^{-4}$ .

$M \sim \text{Gamma}(\text{shape} = 15, \text{rate} = 1 \times 10^5)$ , with a mean of  $1.5 \times 10^{-4}$ .

- Third, we considered the situation where the mean  $\mu$  was an order of magnitude lower than the truth:

$M \sim \text{Gamma}(\text{shape} = 1.5, \text{rate} = 1 \times 10^6)$ , with a mean of  $1.5 \times 10^{-6}$ .

$M \sim \text{Gamma}(\text{shape} = 0.3, \text{rate} = 2 \times 10^5)$ , with a mean of  $1.5 \times 10^{-6}$ .

$M \sim \text{Gamma}(\text{shape} = 15, \text{rate} = 1 \times 10^7)$ , with a mean of  $1.5 \times 10^{-6}$ .

In fig 1 we show the corresponding prior distributions.

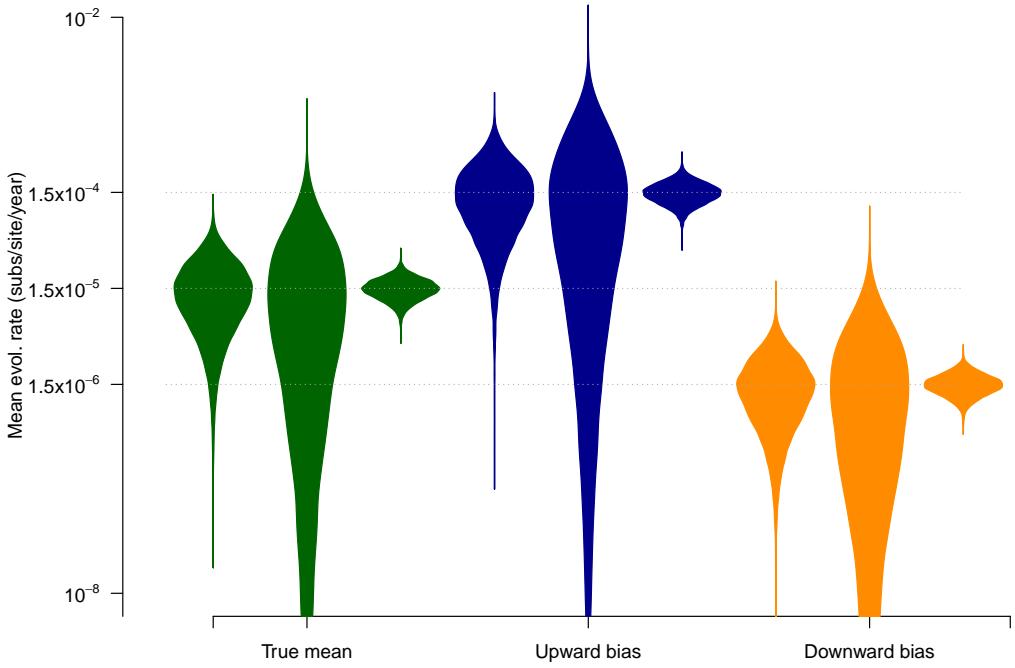


Figure 1: Violin plots of prior distributions used in revisions. The ‘true mean’ in green are Gamma distributions for which the mean is  $1.5 \times 10^{-5}$  and with three degrees of uncertainty. ‘Upward bias’ is for a three prior distributions with mean  $1.5 \times 10^{-4}$ , and ‘downward bias’ correspond to priors with mean  $1.5 \times 10^{-6}$ .

We also conducted analyses where we included a hyper prior distributions for the shape and rate of the Gamma distribution that we use as a prior for the mean evolutionary rate as follows:

$$\mu \sim \text{Gamma}(\text{shape}, \text{rate})$$

$$\text{shape} \sim \text{Lognormal}(M = 1, \sigma = 5)$$

$$\text{rate} \sim \text{Lognormal}(M = 1, \sigma = 5)$$

We conducted 100 simulation replicates in each case, for a total of 5,000 new analyses.

Our results are consistent with our previous findings, but much more comprehensive. We find that a prior that is highly informative (low uncertainty) tends to bias rate estimates. Such bias is negligible when the sampling window is 100 times the phylodynamic threshold, but in some cases, such as when the prior is positively biased and not overly informative, a sampling window of 0.5 of the phylodynamic threshold may be sufficient. In practice, our results suggest that a prior that is not overly informative is advisable, for example with a 95% quantile width that is six times the mean.

Another major issue I have is that the choice of prior distributions used for comparison seem arbitrary and not very well motivated. For example, consider the “reasonable prior” and “misleading prior” detailed

on Page 20. The population size parameter priors are both relatively vague, and the "misleading" prior comfortably includes the true value within its 95% quantile interval. On the other hand, for the molecular clock mean rate, the "misleading" prior has a 95% quantile interval that is 1/10-th as wide as that of the "reasonable" prior, and it does not include the true value. What if the "misleading" priors were centered far away from the true values but were as informative/uninformative (meaning, had the same variance) as the "reasonable" priors? Or what if we kept the same "misleading" downward biased prior for the mean rate and paired it with a prior for the population size that is upward biased (as the "misleading" prior is) AND much more informative than the "reasonable" prior? It seems that the authors' motivation for the "misleading" prior was choosing the mean rate to be much lower and the mean population size to be much higher, but it is puzzling why they seem to have ignored the variance of these priors: the "misleading" mean rate prior has much smaller variance than the "reasonable" mean rate prior, while the the "misleading" population size prior has much larger variance than the "reasonable" population size prior. As it stands, the results of the simulation study show: 1) that we have more temporal signal when we increase the sampling window, and 2) the scenario in which the data have the lowest amount of signal and the model employs a relatively strong prior results in the posterior being dominated by the prior. Neither of these findings are new or add much insight. What would be insightful is if the authors could provide some guidance (by exploring a wide range of priors) on how vague the prior must be in the case of each sampling window to ensure that the prior does not override the temporal signal in the data.

**Response:**

The Reviewer raises an entirely valid point about the uncertainty in the prior distributions. In our revised study we only concentrate on the prior on the evolutionary rate. For the population size of the coalescent,  $\theta$ , we use a conservative prior of the form  $\theta \sim \text{Exponential}(\text{mean} = 5,000)$ . This is the 'reasonable prior' in our previous version of the manuscript, because its mean is the value used to generate the data. For the prior on the evolutionary rate we considered a range of Gamma distributions that place more weight on values that would over- or underestimate the evolutionary rate and with three different degrees of uncertainty. Please see our response above and fig 1.

In the discussion, the authors write: "Maximally uninformative priors, such as Jeffrey's prior, offer an attractive approach, but such priors can ignore useful expert knowledge, they can be particularly difficult to sample, and are not necessarily proper probability distributions". I do not think that this is a very compelling point. It is easy to specify a suitably vague prior distribution that is proper and will not unnecessarily complicate MCMC sampling. On the other hand, if a researcher does not want their expert knowledge to be ignored, then they should not be alarmed if the prior has a relatively high impact on the posterior.

**Response:**

This is a valid point and we deleted this sentence.

I found Figures 2 and 3 (and similar figures in the supplementary material) to be confusing. The blue/gold color combination makes it difficult to determine where intervals are overlapping. Also, the solid circles for the mean estimates are very difficult (in many instances, impossible) to see.

**Response:**

We revised our previous figures. At present we only show violin plots for the prior distribution of the evolutionary rate and 95% credible interval of the the posterior is shown with black bars and no points.

Also, apologies if I have misunderstood, but I am confused by how the yellow bars (95% quantile ranges for the prior) in each plot vary. From the text, it seems that the same prior distribution is being used for all analyses corresponding to a given plot, so why do the yellow bars in a given plots have varying widths?

**Response:** Our figures are different and do not show individual priors for the analyses. However, we would like to clarify this choice in the previous version of our manuscript.

The prior for the evolutionary rate is indeed expected to be identical over simulation replicates. However, this is not necessarily the case for prior on the tree height because this parameter is a function of the prior on the population size, which is identical for all simulations, and the number and distribution of sampling times. In our simulations, the number of samples are fixed, but the sampling times are drawn from a uniform distribution between the present and a pre-defined time (e.g. the phylodynamic threshold multiplied by 0.5). As such, the prior on the tree height is not necessarily identical between simulation replicates. This point has been described in some detail in Boskova et al. (2014).

Finally, it seems that the simulation replicates in each plot have been ordered by posterior mean (or something like that). If so, I suggest making this clear. Otherwise, the "increasing" trend in blue bars from left to right might confuse readers.

**Response:**

Yes, we sorted simulation replicates by their posterior mean to facilitate visualisation. We have now explained this in the figure captions.

I found the description of "uniformly distributed" samples in the "Temporal sampling bias" section to be misleading/confusing. The samples do not appear to be uniformly distributed between the present and the roots of the trees in Figure 4. Please clarify.

**Response:**

The sampling times are uniformly distributed, but note that the branch lengths are in  $\log_{10}$  scale. We have clarified this in the figure caption.

Some Minor Suggestions:

In the abstract, change "ties two key concepts" to "ties together two key concepts." Also, it may be better to avoid the word "dogma" and choose a different word.

**Response:** We changed the wording in this sentence as follows:

*Our current understanding suggests that populations meeting these criteria are suitable for molecular clock calibration via sampling times.*

Throughout the text, single quotations are used when double quotations are appropriate.

**Response:**

We are using the conventions from the Australian Government Style Manual (see <https://www.stylemanual.gov.au/grammar-punctuation-and-conventions/punctuation/quotation-marks>), which recommends single quotes in the instances we could identify in our manuscript. To the best of our knowledge in this journal it is acceptable to use American or British English.

Throughout the text, when referring to the gamma distribution, I suggest to write "Gamma" rather than using the Greek letter.

**Response:**

We made the corresponding change, except for the HKY+ $\Gamma_4$  substitution model.

Caption for Figure 3: insert "for the mean evolutionary rate" at the end, to clarify which parameter the Gamma(shape = 1.5, rate =  $10^6$ ) prior distribution is being applied to.

**Response:**

We replaced all figures and ensured that the captions are complete.

## References

- V. Boskova, S. Bonhoeffer, and T. Stadler. Inference of epidemiological dynamics based on simulated phylogenies using birth-death and coalescent models. *PLoS Computational Biology*, 10(11):e1003913, 2014.
- S. Duchene, K. E. Holt, F.-X. Weill, S. Le Hello, J. Hawkey, D. J. Edwards, M. Fourment, and E. C. Holmes. Genome-scale rates of evolutionary change in bacteria. *Microbial genomics*, 2(11):e000094, 2016.
- V. Eldholm, J. Monteserin, A. Rieux, B. Lopez, B. Sobkowiak, V. Ritacco, and F. Balloux. Four decades of transmission of a multidrug-resistant mycobacterium tuberculosis outbreak strain. *Nature communications*, 6(1):7119, 2015.
- S. Guindon. Rates and rocks: strengths and weaknesses of molecular dating methods. *Frontiers in Genetics*, 11:526, 2020.
- S. Y. Ho. The changing face of the molecular evolutionary clock. *Trends in Ecology & Evolution*, 29(9):496–503, 2014.
- S. Y. Ho and S. Duchêne. Molecular-clock methods for estimating evolutionary rates and timescales. *Molecular Ecology*, 23(24):5947–5965, 2014.
- D. Kühnert, M. Coscolla, D. Brites, D. Stucki, J. Metcalfe, L. Fenner, S. Gagneux, and T. Stadler. Tuberculosis outbreak investigation using phylodynamic analysis. *Epidemics*, 25:47–53, 2018.
- F. Menardo, S. Duchêne, D. Brites, and S. Gagneux. The molecular clock of mycobacterium tuberculosis. *PLoS pathogens*, 15(9):e1008067, 2019.
- M. Merker, J.-P. Rasigade, M. Barbier, H. Cox, S. Feuerriegel, T. A. Kohl, E. Shitikov, K. Klaos, C. Gaudin, R. Antoine, et al. Transcontinental spread and evolution of mycobacterium tuberculosis w148 european/russian clade toward extensively drug resistant tuberculosis. *Nature Communications*, 13(1):5105, 2022.