

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 compromise 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.

[Action: Explore wider set of priors for the simulations](#)

Response: Pending

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.

[Action: cite those papers and tone down MTB stuff.](#)

Response: Pending

Minor:

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

Response: Pending

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: Pending

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

Response: Pending

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: Pending

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: Pending

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 "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.

Action: Pending.

Response: Pending

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.

Action: Ok, delete.

Response: Pending

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.

Action: Ok, change colours and remove points?

Response: Pending

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?

Action: Because the distribution of sampling times is not identical. This is the prior marginalised on over the sampling times.

Response: Pending

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.

Action: Ok, can do.

Response: Pending

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.

Action: They are. The tree is in \log_{10} scale, as stated in the legend.

Response: Pending

I think that the paper would be much improved by including more empirical examples to complement the simulations and the one HBV example.

Action: OK, include some flu.

Response: Pending

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.

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

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

Page 8, Line 218: change "conducted and" to "conducted an"

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

Page 11, Line 227: change "accuracy on" to "accuracy of"