Untangling temporal signal for calibrating the molecular clock of microbes

John H Tay[1], Arthur Kocher[2,3], Sebastian Duchene[1,2,*],

1 Peter Doherty Institute for Infection and Immunity, Department of Microbiology and Immunology, University of Melbourne, Melbourne, Australia
2 Pending pending.
3 Department of Computational Biology, Institut Pasteur, Paris, France

* sduchene@pasteur.fr

Abstract

Our understanding of the evolution of many microbes has been revolutionised by the molecular clock, a statistical tool to infer evolutionary rates and timescales. The fundamental assumption of molecular clock models is that the rate at which substitutions accumulate can be described by a statistical process. In all molecular clock models evolutionary rates and times are unidentifiable, and therefore 'calibration' information is essential to obtain estimates in calendar time.

For many microbes, the sequence sampling times themselves can be often used for calibration. Phylogenetic tests of temporal signal are often used to decide whether such calibrations are reliable. Critically, in addition to the calibration information, the full Bayesian phylogenetic model also includes the molecular clock model and a branching process (tree prior). As a result, there are multiple sources of information that are difficult to untangle.

We assessed temporal signal in three microbial data sets of human and animal diseases with a range of evolutionary characteristics and with ancient DNA sequences; *Powassan virus* (POWV), the cholera bacetrium (*Vibrio cholerae*), and the syphilis bacterium (*Treponema palladium.*). We found that the tree prior can have a substantial impact on whether temporal signal is detected. To investigate this problem we conducted extensive simulations and calculated the sensitivity and specificity of these tests under several tree priors. Highly informative sequence data sets are generally robust to the tree prior. However, in data sets with low information content, choosing a prior that is highly informative and inconsistent with the data can result in the false rejection of temporal signal.

We demonstrate that prior sensitivity analyses and prior predictive simulations are an effective means to determine whether the prior is reasonable and to improve the detection of temporal signal and leverage the information that can be drawn from molecular sequence data sets.

Author summary

Pending. [1]

Introduction

Molecular sequence data have been essential to unravel the evolutionary history of many organisms. The molecular clock is a statistical tool that posits that molecular evolution, in the form of substitutions, follows a statistical process. For example, under the simplest molecular clock model, known as the strict clock, substitutions accumulate linearly over time along a lineage, such that the evolutionary rate is constant over time [1]. At the other end of the spectrum, relaxed molecular clocks allow every lineage in a phylogenetic tree to display a different evolutionary rate ([2] and reviewed in [3]).

All molecular clock models have a fundamental limitation, where evolutionary rates and times are unidentifiable. That is, there exist an infinite number of combinations of evolutionary rates and times that are compatible with an amount of evolutionary divergence [4,5]. For this reason, external information, known as a molecular clock calibration is necessary to produce estimates in calendar units. As a case in point, consider two sequences whose genetic divergence from their most recent common ancestor is 10 subs/site. In the absence of calibrating information it is impossible to know *how rapidly* they evolve and *when* they diverged. The calibration can be a known evolutionary rate, such as 1 subs/site/year, or a divergence date, such as 1 year before present. The genetic distance can be divided by the evolutionary rate to infer the divergence time to infer a time to the most recent common ancestor of 10 years, or the genetic distance can be divided by the divergence date to infer the evolutionary rate, 10/subs/site/year in this case.

The finding that some organisms accumulate substitutions in a measurable timescale prompted the use of sequence sampling times for calibration [6,7]. The motivation behind this practice is that sequence data collected at different points in time would have accumulated a corresponding number of substitutions. With the example above, a sequence collected six months of the common ancestor would have accumulated 5 subs/site (10 subs/site/year × 0.5 years = 5 subs/site), whereas one collected after 1 year would have accrued 10 subs/site (10 subs/site/year× 1 year = 10 subs/site). As a result, sequence sampling times act as a time-calibration that is intuitively informative about the evolutionary rate.

A fundamental question about using sampling times for molecular clock calibration is the extent to which the data can be assumed to have been sampled from a measurably evolving population. A measurably evolving population is defined as to the situation where the interval of time over which the samples were taken captures an appreciable amount of evolutionary change [8]. For rapidly evolving pathogens, such as RNA viruses, a measurably evolving population may be obtained by drawing samples over weeks or months. More slowly evolving microbes, such as the tuberculosis bacterium (*Mycobacterium tuberculosis*) may require sampling over many years.

There exist several statistical tests to determine whether a population has measurably evolving behaviour, also known as tests of temporal signal. The root-to-tip regression takes a phylogenetic tree for which the branch lengths measure evolutionary distance (i.e. a phylogram) and fits a linear regression of the distance from the root to the tips as a function of their sampling time [7]. The regression slope is a crude estimate of the evolutionary rate, the x-intercept is the time to the most recent common ancestor, and the R^2 is a measure of clocklike evolution. In general, the root-to-tip regression is a powerful tool for visual inspection of the data, for example to detect outliers or identify lineages with particularly low or high evolutionary rates [9–11]. However, it the data points are not statistically independent, such that it is not a formal statistical test of temporal signal and statistics, such as p-values are invalid [12].

Date randomisation tests consist of fitting a molecular clock to the data after permuting the sampling times, a procedure that is repeated multiple times to obtain a 'null' distribution of the evolutionary rate [13]. The data are considered to have

temporal signal if the evolutionary rate estimated with the correct sampling times falls outside such 'null' distribution [13, 14].

BETS (Bayesian Evaluation of Temporal Signal) is a fully Bayesian approach for assessing temporal signal [15], by treating it as a model selection problem. The premise of the test is that the data in question should have higher statistical fit when the sampling times are included than when they are not. In practice, the data are analysed with their correct sampling times and with all samples assigned the same date (typically at present), while keeping the rest of the phylogenetic model the same, including the molecular clock, tree prior and substitution model. The log marginal likelihood is calculated in each case to compute log Bayes factors, which quantify the amount of evidence for one model over another, here that with sampling times vs that without. A major advantage of BETS is that can consider the full model and it naturally accommodates important sources of uncertainty, including that due to radio carbon dating of ancient DNA studies [16].

Most parameters of the phylogenetic model have individual prior probability distributions that can be chosen by the user, for example, the evolutionary rate, or the transition-to-transversion ratio of the HKY substitution model. The phylogenetic tree topology and branch lengths are usually assigned a branching process, such as a coalescent or birth-death process. These tree priors implicitly impose a prior probability distribution on the ages of nodes and the tree length (sum of time of all branches), and therefore may inadvertently impose highly informative calibration priors. Moreover, model selection, as is the case with BETS, can be sensitive to the choice of prior, even if the posterior is not [17, 18]. Here we investigate whether such prior information can overwhelm the signal from the data and potentially mislead Bayesian tests of temporal signal.

Results

Empirical data analyses

We collected genome sequences from three microbial data sets that have been shown to have temporal signal using Bayesian phylogenetic methods. The data sets were: *Vibrio cholerae* [19], the bacterium responsible for cholera; *Powassan virus (POWV)* [20], a tick-borne virus; and *Treponema pallidum* [21], the bacterium that causes syphilis. The *V. cholerae* and *T. pallidum* data sets involve ancient samples. We analysed the data sets using BETS under a coalescent tree prior with constant population size and two possible clock models; a strict and an uncorrelated relaxed clock with an underlying log-normal distribution. Our choice of the constant coalescent tree prior is based on statistical convenience, as it is fully parametric, and its simplicity, rather than describing a biological process. We set up our analyses in BEAST1.10 [22] and calculated log marginal likelihoods with and without sampling times for each combination of molecular clock model and tree prior.

To assess the impact of the tree prior we considered different (hyper) prior distributions for the effective population size, θ, the only parameter in the constant-size coalescent. This parameter is referred to as a *scale parameter* for time because large values imply more dispersion (the molecular clock rate is also a scale parameter), and it is typically assigned a $1/x$ prior distribution, which is the Jeffrey's prior that is uninformative and invariant to reparameterisation [23]. This prior has attractive attributes because it maximises the signal from the data, but it is an improper distribution (it does not integrate to one over its domain, because $\int_0^\infty \frac{1}{x}$ is undefined), a problem for model comparison using Bayes factors, because marginal likelihood calculations require that all priors be proper distributions [24, 25]. Instead, we selected

three prior distributions, an exponential, Γ, and log-normal, that have been used in recent literature as shown in Table 7.

Our rationale for using different prior distributions on θ is its impact on parameters that pertain to the molecular clock. Under the constant coalescent process, θ is the single parameter in most implementations. Because it is associated with the genetic diversity and population size, it is expected that large values of θ will result in a time to the most recent common ancestor that is older than if θ is small. Thus a constant coalescent tree with a prior on θ that favours large values will also favour greater node heights (older nodes). The prior on θ will also have an impact on the evolutionary rate for two key reasons. First, by impacting the overall age of the tree, it will also impact the length of time over which the sequence data evolved. Second, the default prior for the evolutionary rate in BEAST1.10 is a Gamma (Γ) distribution with shape (α) of 0.5 and beta (β, also known as the 'rate') equal to the tree length (sum of all branch lengths) [26, 27]. In this software, this prior is known as the CTMC-rate reference prior and its mean value is $0.5/T_l$, where T_l=tree length [27], meaning that it is indirectly impacted by θ.

Table 1. Prior distributions for the effective population size of the constant-size coalescent (known as θ and different to the scale parameter of the Γ distribution).

Probability distribution function	Parameters
Exponential	mean, $\mu = 1.0$
Log-normal	mean, $\mu = 1.0$; standard deviation $\sigma = 5.0$
Γ (Gamma)	shape, $\kappa = 0.001$; scale, $\theta = 1000$

The *V. cholerae* data set displayed overwhelming support for temporal signal (Table 2 and Fig 1), regardless of the molecular clock model and prior on θ, with log Bayes factors of over 200. Note that a log Bayes factor of 3.2 corresponds to a model posterior probability ≈ 0.95 [28], and is considered as 'very strong support', following Kass and Raftery [29]. Although in this data set the prior on θ did not impact model selection for detecting temporal signal, it did impact the magnitude of the Bayes factors.

For our other two data sets the impact of the prior on model selection was evident. For *Poawassan virus* the Γ and log-normal priors on θ suggested strong temporal signal, whereas the exponential prior strongly favoured the exclusion of sampling times, according to the strict and relaxed molecular clock models. In our analyses of the *T. pallidum* data set we found support for temporal signal under the strict molecular clock, according to all priors on θ, although with very strong evidence only for the exponential prior and 'positive evidence' for the Γ and log-normal priors. Under the relaxed molecular clock model all priors had very strong support against temporal signal.

Table 2. Log Bayes factors between isochronous and heterochronous models for each dataset, separated by prior on effective population size, θ

Species; Clock Model	Exponential	Gamma	Log-normal
Vibrio cholerae; *Strict Clock*	355.18	379.63	382.10
Vibrio cholerae; *Relaxed Clock*	208.97	439.63	219.60
Powassan virus (POWV); *Strict Clock*	-80.63	32.67	50.29
Powassan virus (POWV); *Relaxed Clock*	-221.94	18.79	27.23
Treponema pallidum; *Strict Clock*	105.80	2.17	1.85
Treponema pallidum; *Relaxed Clock*	-34.37	-1474.14	-34.04
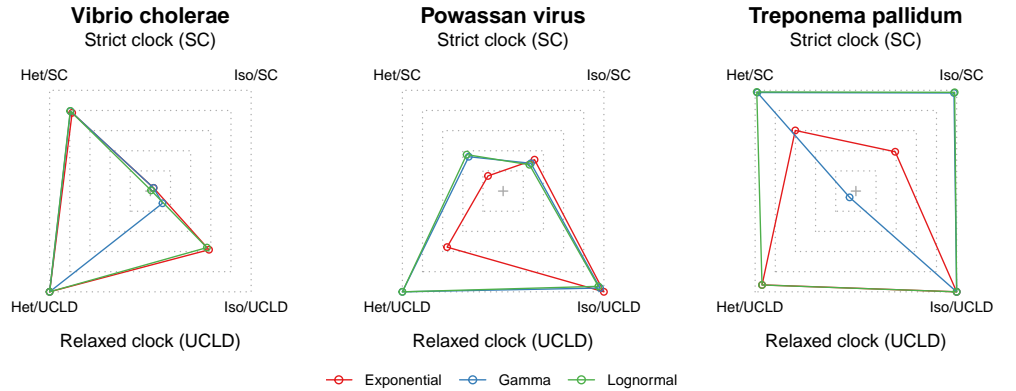
Fig 1. Relative log marginal likelihoods of empirical data sets. The polygons represent the relative log marginal likelihoods of each microbe dataset under a different effective population size (θ) prior, analysed with four different configurations. Het (heterochronous) includes sampling, while iso (isochronous) does not include any sampling times. SC is strict clock and UCLD is the uncorrelated lognormal relaxed clock. Red represents an exponential hyperprior on the effective population size, blue is a Γ hyperprior, and green is a log-normal hyperprior.

Our empirical data analyses overall demonstrate that the choice of prior has an impact on Bayesian model selection and support. The *V. cholerae* data set had a posterior that was less sensitive to the prior for key parameters, such as the evolutionary rate (Supplementary material), and in this data set the detection of temporal signal was also more robust to the prior. Moreover, this data set has been shown to have clear clocklike behaviour in root-to-tip regressions and date randomisation tests [30]. In contrast, the data sets of *Powassan virus* and *T. pallidum*, where temporal signal was not supported for all model configurations, we found that the posterior is more sensitive to the choice of prior (Supplementary information).

Simulation experiments

To understand the impact of the prior on Bayesian model selection we conducted a set of simulation experiments where the data generating process is well understood. We conducted simulations under four possible conditions; a strict or relaxed molecular clock, and where the phylogenetic time-trees were heterochronous or isochronous. Data from heterochronous trees are sampled from a measurably evolving population and are expected to display temporal signal, whereas those from isochronous trees are not from a measurably evolving population (the time-trees are ultrametric) and should not display temporal signal.

For data generated under heterochronous time-trees, we found that ten out of ten simulation replicates were correctly classified as having temporal signal, using a log Bayes factor of at least 3.2 (Table 3 and Fig 2). This perfect classification was supported regardless of the prior on θ and the molecular clock model.

For our data generated under isochronous trees, which by definition have no temporal signal, we found perfect classification under the exponential prior on θ under both clock models (Table 4 and Fig 3). For most analyses under the Γ and log-normal priors on θ we found that BETS incorrectly supported the presence of temporal signal. The exceptions were for analyses of data analysed under a relaxed clock, whether they

Table 3. Correctly classified simulation replicates under heterochronous trees. A total of ten simulations were generated in each case, under heterochronous trees, such that they are expected to display temporal signal. A number of ten represents perfect classification according to the Bayesian evaluation of temporal signal, BETS and a log Bayes factor of at least 3.2 (strong evidence for temporal signal). The rows correspond to three possible priors on the effective population size of the constant-size coalescent, θ. The 'Best clock model' is a situation where we consider the best heterochronous and isochronous model, take their log Bayes factor, and determine temporal signal if it is at least 3.2.

True clock model; clock model in analysis	Exponential	Γ	Log-normal
Strict clock; Strict clock	10	10	10
Strict clock; Relaxed clock	10	10	10
Strict clock; Best clock model	10	10	10
Relaxed clock; Strict clock	10	10	10
Relaxed clock; Relaxed clock	10	10	10
Relaxed clock; Best clock model	10	10	10
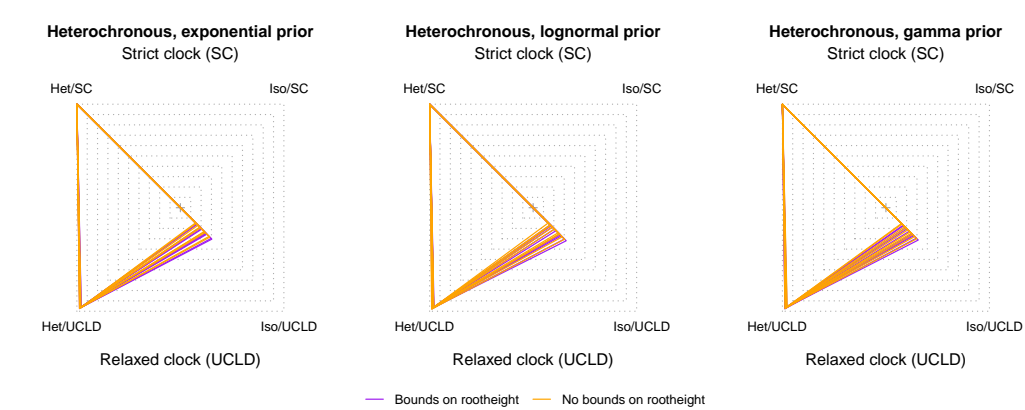
Fig 2. Relative log marginal likelihoods of simulations with temporal signal. The polygons represent the relative log marginal likelihood under three possible priors on the effective population size (θ) parameter of the constant-size coalescent tree prior. Each corner corresponds to a combination of model and sampling times, either a strict (SC) or relaxed molecular clock with an underlying log-normal distribution (UCLD), and with (heterochronous) or without (isochronous) sampling times. The correct model used to generate the data is the SC heterochronous. Each polygon is for one simulation replicate (a total of ten) and the colours denote whether we employed a hard bound on the root height of the form Uniform(0.0, 5.0) or not.

were simulated under a strict or relaxed clock. 160

A perplexing result occurs under the best molecular clock model for the Γ and 161
log-normal priors on θ. Here we take the log Bayes factor of the best heterochronous 162
model vs the best isochronous model, which produced an increase in the classification 163
error, relative to using the relaxed clock only. This phenomenon occurs because the 164

incorrect inclusion of sampling times can mislead molecular clock model selection. As a case in point one of the simulation replicates under a relaxed molecular clock and an isochronous tree (with no temporal signal) had the following log marginal likelihoods; -4109.87 for the heterochronous analyses with a strict clock, -4117.06 for the isochronous analyses with a strict clock, -4124.35 for the heterochronous analyses with a relaxed clock, and -4118.29 for the isochronous analyses with a relaxed clock. The log Bayes factors under the relaxed clock have strong evidence against temporal signal (log Bayes factor=-6.06 for heterochronous vs isochronous), whereas the opposite is true for the strict clock (log Bayes factor=7.19). However, the best heterochronous model has substantially stronger support than the best isochronous model (here the strict or relaxed molecular clock, whose log marginal likelihoods differ by only 1.3 units). It is also worthwhile to note that in general, analyses under the relaxed clock tended to have fewer classification errors (for data sets with no temporal signal) than the strict clock, regardless of the true molecular clock model used to generate the data.

Table 4. Correctly classified simulation replicates under isochronous trees. A total of ten simulations were generated in each case, under isochronous trees, such that they are not expected to support temporal signal. A number of ten represents perfect classification according to the Bayesian evaluation of temporal signal, BETS and a log Bayes factor of at most -3.2 (strong evidence against temporal signal). The rows correspond to three possible priors on the effective population size of the constant-size coalescent, θ. The 'Best clock model' is a situation where we consider the best heterochronous and isochronous model, take their log Bayes factor, and determine a lack of temporal signal if it is at most -3.2.

True clock model; clock model in analysis	Exponential	Γ	Log-normal
Strict clock; Strict clock	10	0	0
Strict clock; Relaxed clock	10	10	10
Strict clock; Best clock model	10	0	0
Relaxed clock; Strict clock	10	0	0
Relaxed clock; Relaxed clock	10	9	9
Relaxed clock; Best clock model	10	0	1

Our simulation results demonstrate that detecting temporal signal when it is not present, which is akin to a type I error, is more common under some prior configurations (here the Γ and log-normal priors on θ) than the opposite (failing to detect temporal signal when it is present). Upon inspecting the resulting phylogenetic trees and the posterior of key parameters we found a probable cause. The incorrect inclusion of sampling times produces a dramatic overestimation of the height of the tree, especially under the strict molecular clock model, which we refer to as 'tree extension' (Fig 4). Under this situation, the sampling times represent such a small proportion of the root height (the time from the most recently sampled node and the root-node), that the heterochronous tree is indistinguishable from one that is ultrametric and thus their log marginal likelihoods can be similar. In fact, the sampling times here spanned at most 0.5 units of time, such that they represent only 0.05% of the total height of a tree height of 1,000, which is not unusual when there is tree extension. Moreover, the inclusion of sampling times may improve the fit of the model by accommodating natural stochasticity in the root-to-tip distances. The phenomenon of tree extension also occurs under the relaxed molecular clock, but to a much lesser extent, and thus under this model it is easier to correctly classify isochronous data sets. The exponential tree prior favour shallower trees and this is probably the reason for why results in more accurate classification.
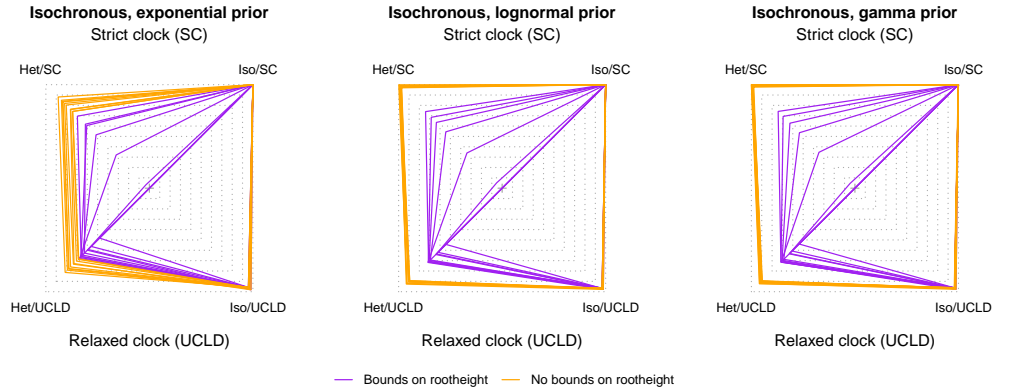
Fig 3. Relative log marginal likelihoods of simulations with no temporal signal. The polygons here represent the same information as in Fig 2. The correct model used to generate the data here is the SC isochronous (ultrametric).

In Fig 5 we show an example of phylogenetic trees from a simulation replicate with temporal signal (heterochronous). Failing to include sampling times, as in isochronous analyses, results in trees that are of similar height, so the confounding effect that we observe in isochronous data sets above does not occur. However, the time-scale of the isochronous trees is determined by the value used to fix the molecular clock rate. The typical approach is to fix this parameter to 1.0, which means that genetic distance and time are the same. An important aspect of this method is that the θ parameter of the constant-size coalescent is proportional to units of time [23, 31] (as are most other parameters of the tree prior, for example the growth rate of the exponential-growth coalescent). When time-calibrations are used θ is the population size multiplied by the generation time ($N_e \times$ generation time; note that in [23] the notation θ is used). Thus, the prior on θ for the heterochronous and isochronous analyses has different meanings, unless this parameter is scaled to match the units of the branch lengths, or the molecular clock rate in the isochronous analyses is fixed to a plausible value. In our simulations we fixed the molecular clock rate to its true value, but we also found that using a number within the expected order of magnitude of the organism in question is sufficient (e.g. 10^{-4} to 10^{-3} subs/site/year for a ssRNA virus).

Defining a suitable prior for the parameters of the tree prior requires careful attention. The prior on θ (or other parameters) should not favour implausible node heights. However, understanding the interplay between the parameters of the tree prior and the resulting tree topologies and node heights is not necessarily trivial, particularly when the tree prior involves multiple parameters (e.g. the Bayesian skyride). A pragmatic solution is to include additional prior information in the form of hard bounds on the root-height or the molecular clock rate. To this end, we investigated the effect of including a uniform prior between 0.0 and 5.0 units of time for the root height, meaning that trees that are older than 5.0 units have a prior probability of 0.0. Importantly, the trees under which we generated our simulations had root heights of around 2.0 units of time, and thus the hard bound of 5.0 allows for trees that are over twice as old as the truth.

Setting hard bounds on the root height resulted in perfect classification accuracy for
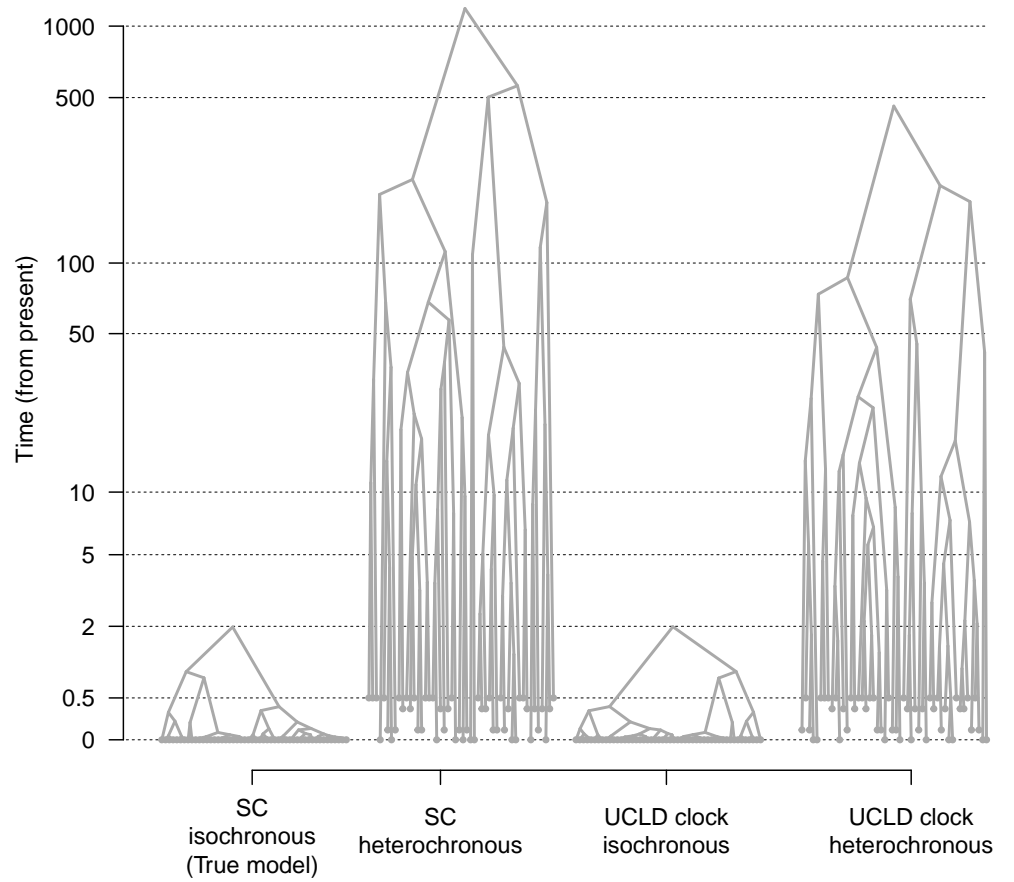
Fig 4. Phylogenetic tree extension for a simulation replicate with no temporal signal. Highest clade credibility trees from a data set simulated with no sampling times (isochronous) and under a strict molecular clock model (SC). The prior on θ is a $\Gamma(\kappa = 0.001, \theta = 1000)$, which resulted in high classification errors using BETS. The y-axis is the time from the present. Tip nodes have solid grey circles. Including sampling times that span 0.5 units of time and about 1/4 of the true root height induces dramatic overestimation of the root height, compared to the true model (SC, isochronous). This effect occurs under both molecular clock models, the SC and the relaxed molecular clock with an underlying log-normal distribution (UCLD), but it is markedly less pronounced in the UCLD. Note that the y-axis is in logarithmic scale ($\log_{10}$).

both, the heterochronous and isochronous simulation (Table 5 and Fig 2, and Table 6 and Fig 3, respectively). The improvement in classification for data sets with no temporal signal is likely because the hard bounds limit the tree extension phenomenon and thereby including sampling times imposes a penalty on the log marginal likelihood (see polygons in Fig 2).
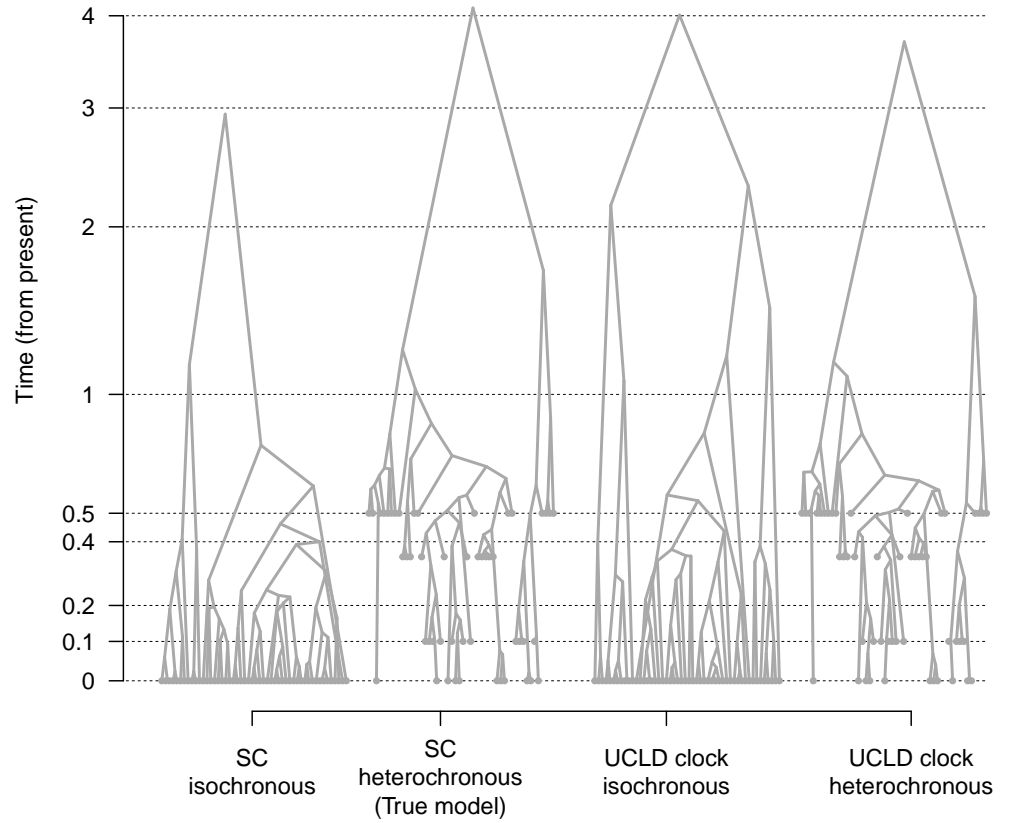
Fig 5. Phylogenetic trees from a simulation replicate with temporal signal.
Highest clade credibility trees from a data set simulated with sampling times
(heterochronous) and under a strict molecular clock model (SC). The isochronous trees
are inferred by fixing the molecular clock rate to the true value, such that the time-scale
is in the comparable units to the heterochronous analyses. Unlike the estimates for
isochronous trees (e.g. Fig 4), the height of the trees under all scenarios are comparable.
Axes and labels are the same as those of Fig 4

Prior predictive simulations and parameter correlations   233

We simulated phylogenetic trees from a prior distribution to inspect the correlation of   234
parameters and the marginal prior for those that have obvious associations (e.g. the   235
evolutionary rate has a CTMC-rate reference prior that depends on tree length). These   236
simulations are commonly known as prior predictive simulations (e.g. [32]), and referred   237
to in the phylogenetic literature as 'sampling from the prior' [33]. Initially, we set a   238
uniform prior on θ from 0 to 10^3 and recorded the evolutionary rate, the tree length and 239
root height. This prior is not generally recommended [34], and we present it here to   240
illustrate correlations between parameters and the marginal prior, instead of using it for 241
analysing our data.   242
   An obvious finding is a natural positive correlation between tree length, root height, 243

Table 5. Correctly classified simulation replicates under heterochronous trees using hard bounds on the root height. Rows and columns are identical to those of Table 3, but here the heterochronous analyses include an explicit prior on the root height, via a uniform distribution between 0 and 5.0.

| True clock model; clock model in analysis	Exponential	Γ	Log-normal
Strict clock; Strict clock	10	10	10
Strict clock; Relaxed clock	10	10	10
Strict clock; Best clock model	10	10	10
Relaxed clock; Strict clock	10	10	10
Relaxed clock; Relaxed clock	10	10	10
Relaxed clock; Best clock model	10	10	10

Table 6. Correctly classified simulation replicates under isochronous trees using hard bounds on the root height. Rows and columns are identical to those of Table 4, but here the heterochronous analyses include an explicit prior on the root height, via a uniform distribution between 0 and 5.0.

True clock model; clock model in analysis	Exponential	Γ	Log-normal
Strict clock; Strict clock	10	10	10
Strict clock; Relaxed clock	10	10	10
Strict clock; Best clock model	10	10	10
Relaxed clock; Strict clock	10	10	10
Relaxed clock; Relaxed clock	10	10	10
Relaxed clock; Best clock model	10	10	10

although the trends is not strictly linear (Fig **??**). We also observed a positive correlation between θ, tree length and tree height, which is expected because large population sizes impose long coalescent times [35]. The nature of this correlation is heteroskedastic, with the variance in these tree statistics increasing with θ. Our simulations demonstrate an inverse relationship between the evolutionary rate, θ and the tree statistics, but with a range that can span several orders of magnitude and with values ranging from 10^{-9} to 10^{-4} subs/site/time, meaning that the CTMC-rate reference prior tends to be diffuse. It is also noteworthy that the uniform prior on θ does not result in uniform marginal prior distributions for any of the parameters investigated here.

We conducted prior predictive simulations for the six prior configurations for θ under a heterochronous data set. In Fig 7 we show the resulting distribution on the root height. The exponential distribution ($\theta \sim Exponential(\mu = 1.0)$ produced trees with root heights with a mean of 2.90 units of time (95% quantile range, qr: 1.62 to 14.34), with corresponding values of 772.11 (95% qr: 1.66 to 5.06×10^5) for the log-normal (log-normal($\mu = 1.0, \sigma = 5.0$)) and 1.61 (95% qr: 1.50 to 4.87) for the Γ ($\Gamma(\kappa = 0.001, \theta = 1000)$). Clearly, the log-normal is the most vague prior here, but it produces implausible values of several orders of magnitude higher than our expectation of tree heights of around one and ten units of time (as simulated using $\theta = 1.0$).

For comparison, we also simulated trees under the same priors on θ and with hard bounds on the root height with a uniform distribution with minimum and maximum values of 0.0 and 5.0, as described above. In this case, the exponential prior on θ yielded trees with root heights of mean 2.47 (95% qr: 1.62 to 5.16), whereas those for the
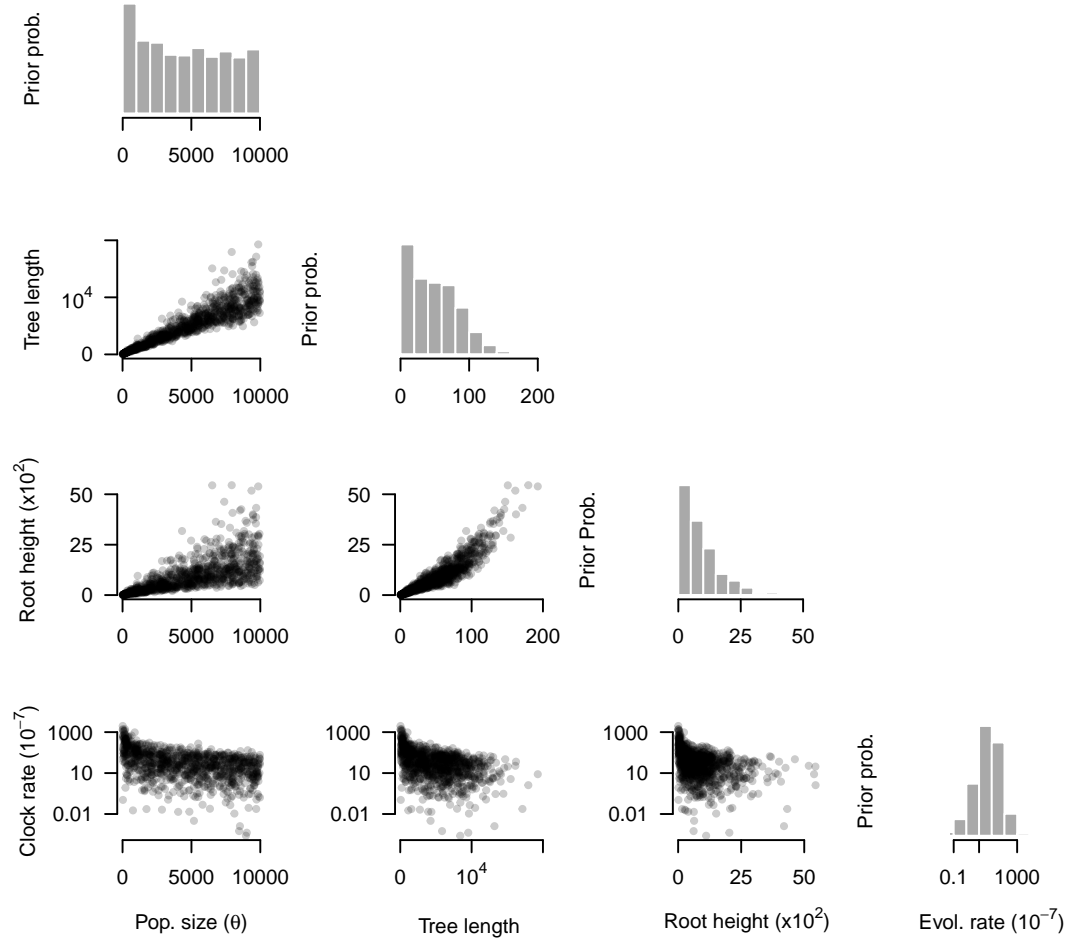
Fig 6. Marginal prior distributions and pairs plots. The grey histograms for
correspond to the parameter labelled at the bottom of each column; effective population
size (pop. size, θ), tree length, root height, and the evolutionary rate (evol. rate). The
prior for θ here is a Uniform distribution between 0 and 10^3, while that for the
evolutionary rate is a CTMC-rate reference prior. Note that the tree length and root
height have units of time, the evolutionary rate is in subs/site/time, and θ is
proportional to units of time.

log-normal and Γ were 1.61 (95% qr: 1.50 to 5.11), and 1.96 (95% qr: 1.54 to 4.42), ₂₆₇
respectively. As a result, in the log-normal prior on θ the use of hard bounds on the ₂₆₈
root height resulted in much smaller root heights, with a smaller impact on the ₂₆₉
exponential and Γ priors. ₂₇₀

Discussion ₂₇₁

Our study demonstrates that the choice of prior distribution has an impact in Bayesian ₂₇₂
model selection and that it can mislead tests of temporal signal. In general, data sets ₂₇₃
with no temporal signal are easily misclassified when the prior favours implausibly old ₂₇₄
root heights and low evolutionary rates. This is an important problem because ₂₇₅
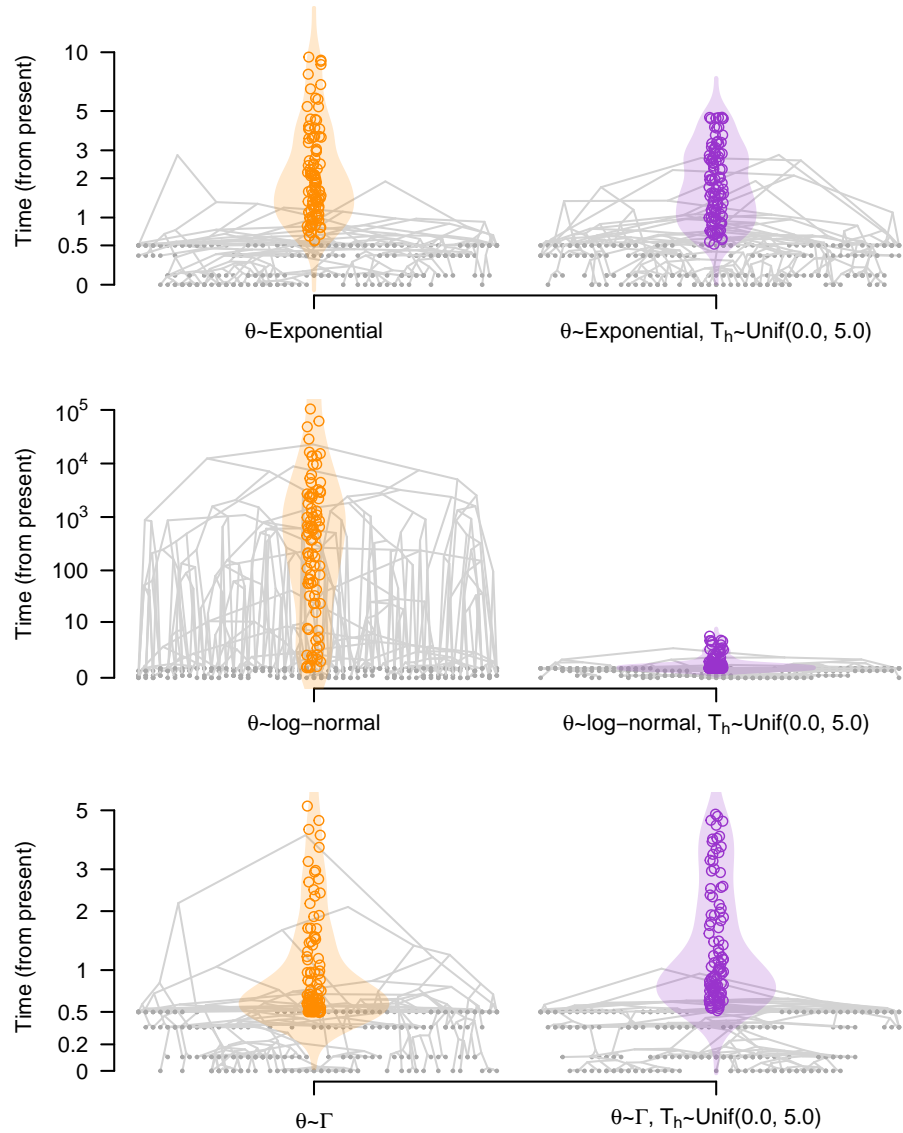
Fig 7. Prior predictive simulations and marginal priors of root heights, given the prior on the effective population size, θ. Each panel corresponds to a different prior on θ, as described on Table 7 (Exponential($\mu = 1.0$), Log-normal($\mu =1.0$, $\sigma =5.0$), and $\Gamma(\kappa = 0.001, \theta = 1000)$). We show five simulated trees from our analysis using sampling times (heterochronous analyses) and overlaid them (similar to densitree plots [36]). The violin plots show the prior densities of the root height and the hollow circles denote 100 randomly drawn samples from the prior. The y-axis is the time from the present, but note that the scales are different. Tip nodes are shown with solid grey circles. The densities and trees on the left, in orange, do not include an explicit prior on the root height (T_h), while those to the right, in purple, have a hard bound on tree height in the form of a uniform prior between 0.0 and 5.0 units of time.

incorrectly detecting temporal signal appears to result in a systematic overestimation of the evolutionary timescale and an underestimation of the molecular clock rates.

We find that tree extension is the most probable reason for the incorrect detection of temporal signal in BETS, because it reduces the sampling window relative to the root height, such that the trees with sampling times are very similar to ultrametrics trees. The phylogenetic likelihood of such overly old trees can be consistent with that from much shallower trees with no sampling times if their evolutionary rate is very low, such that their genetic distance along their branches is comparable.

the clock model. Ideally, the relaxed clock seems to be more robust because it absorbs sampling times as rate variation (we saw higher coefficients of rate variation), instead of making trees really old, but this is tricky because it might not be the best overall model for the data and we want tests of temporal signal to be consistent for among clock models.

In contrast, data sets with temporal signal are easier to classify because tree extension is not a problem. To minimise incorrect detection of temporal signal we recommend several things. Eliciting priors carefully and in the correct units for ultra and iso, so that these models only differ in the sampling times, second is doing prior predictive simulations and to check the tree height and clock rate, third is doing prior sensitivity. We find that data sets that have strong temporal signal tend to be robust to different priors. This does not need to be exhaustive, but something like what the powwassan virus paper did is good.

there are other tests of temporal signal, but they are also difficult, and mention the phylodynamic threhsold paper and the regressions. Moreover, Bayesian tests can allow us to incorporate important info, like a date uncertainty, or other stuff.

what about when there is no temporal signal

For hte ending say that This has important implications for what we classify as a measurably evolving population, for the phylodynamic threshold, and ultimately for justifying estimates of evolutionary rates and times based on tip-calibrations only.

1 ~~Our study demonstrates that the choice of prior has an impact in model selection, and can result in incorrectly selecting models for the purpose of detecting temporal signal and choosing molecular clocks.~~

2 The tree height and length are particularly important.

This has important implications for what we classify as a measurably evolving population, for the phylodynamic threshold, and ultimately for justifying estimates of evolutionary rates and times based on tip-calibrations only.

It appears that data sets that have very clear temporal signal, with clocklike behaviour, and informative sequence data, the prior is less influential. ~~However, BETS tends to incorrectly classify data sets with no temporal signal. This occurs due to tree extension and is most prevalent when we consider a strict clock. Tree extension simply occurs to compensate for the impact in the likelihood due to the inclusion of sampling times in truly ultrametric trees, such that the sampling times span a very small portion of the tree height and the tree is effectively ultrametric.~~

In the absence of internal node calibrations the tree height and tree length are largely determined by the tree prior and the hyperprior on the parameters that govern it. For this reason conducting prior predictive simulations seems essential to determine whether the resulting trees have reasonable root heights, or tend to favour very old trees. A simple way to mitigate this is to include hard bounds, but in our simulations we found that an exponential prior can be sufficient. However, some data sets include very divergent sequences and we may actually do want somewhat old trees, as appears to be the case with the powassan

virus data set. In its original analyses several lines of evidence support temporal signal, and here we find that the exponential tree prior did not favour it.

We also find that the relaxed clock can mitigate this problem of incorrect classification, but in empirical analyses it is difficult to make the decision as to whether there is temporal signal when the clock models disagree, so we recommend prior predictive simulations.

3. What about simpler tests of temporal signal, like the root-to-tip regression, or date randomisation tests?

Well... the root to tip regression is sometimes unclear (cite the BETS paper where we look at inconclusive regressions) and in the date randomisation test what we are testing is whether a particular set of sampling times produce a different estimate than others, so it can be sensitive to the number of dated samples, and may not be very useful for aDNA, where only a small number of samples are ancient.

BETS is still a solid alternative because some data sets require additional information, such as uncertainty in sampling times, monophyletic constraints and root positions. For example, the treponema data analyses did this and obtained very clear temporal signal. Our results for this data set are less conclusive and may indicate that such information is very important.

4. Future practice should involve some sort of prior predictive simulations to elicit a prior that does not favour unreasonable node heights and even rates. Importantly, the isochronous analyses should fix the clock rate to a value within the expected order of magnitude, so that the tree prior is comparable between the isochronous and heterochronous models. Choosing a reasonable prior can be difficult for pathogens that are poorly understood. However, our simulations illustrate an extreme case, where our log-normal prior favoured trees that were many orders of magnitude older than our expectation for the simulations. Similarly, prior sensitivity analyses can also be useful, where the robustness of the posterior can be compared for a subset of priors. Our empirical data demonstrate that the most informative data set, *V. cholerae* was generally robust. Similar approaches of comparing prior and posterior root heights have been proposed to assess temporal signal (cite book chapter).

5. Closing remarks. Temporal signal is important for recently emerging microbes, data sets of ancient DNA, or when the evolutionary history of a pathogen is unknown. Our study provides guidelines for determining this and for improving the reliability of estimates of evolutionary rates and times.

Whole genome sequencing and ancient DNA are two technologies that have dramatically expanded the range of organisms from which we can sample measurably evolving populations. By sequencing complete genomes, instead of individual genes, we can capture more genetic variation overall. In a microbe with an evolutionary rate of about 1×10^{-7} subs/site/year (e.g. *M. tuberculosis* [37]) sampled over a decade would yield on average one substitution for every 1M bases sequenced, with shorter stretches of sequences being mostly uninformative. As a result, many microbes whose evolution was considered too slow to be treated a measurably evolving populations, including many bacteria, are now commonly analysed using frameworks that were thought to be only applicable to rapidly evolving pathogens, such as RNA viruses [38].

Remember to mention the book chapter where we also look at how the prior on root height is affected by the presence or absence of temporal signal. Data sets with temporal signal should see more distance between the prior and the posterior.

Tests of temporal signal are contingent on priors producing reasonable estimates of key parameters, especially root height. It is difficult to predict the impact of the tree prior and its hyperpriors on the root height, which can lead to improper model selection. Thus, predictive prior simulations are essential when evaluating prior sensitivity. We tested three hyperpriors on θ to examine the relationship between parameters and the marginal priors. By placing bounds on the root height, we restrict the effect of tree extension. This allows sampling times, which are indistinguishable from noise when the trees are excessively long, to correctly calibrate true heterochronous trees.

Materials and methods

Empirical data

We selected three different datasets to evaluate temporal signal, *V. cholerae* [19], *Powassan virus* [21], and *T. pallidum* [20]. These varied in their timespans, number of informative sites, and number of sequences. The most sparse dataset was *T. pallidum*, where sampling dates spanned 481 years (1534 to 2016), with 28 sequences each containing 1,500 SNP sites. Among these samples, there were only two in the 1500's and two in the 1700's, with the bulk of the samples with the rest of the samples concentrated in the 1900's to 2000's. There were 319 sequences with 11,193 sites (the complete genome) for *Powassan virus*, spanning 24 years of sampling times (1995 to 2019). Most of the samples were collected after 2010, with only three isolated in 1995. Finally, we had 122 *V. cholera* sequences with 1,57 SNP sites across 73 years of sampling (1937 to 2010). There are 1,392 unique site patterns in *V. cholera*, 3,457 in *Powassan virus*, and 840 in *T. pallidum*.

Table 7. Unique Site Patterns for each dataset and clock model

Dataset	Empirical	Simulation (SC)	Simulations (UCLD)
Cholera	1392	580-710	640-710
Powassan Virus	3457	~4800 (iso)	~14650 (het)
Treponema	840	110-150	110-150

For the *V. cholerae* and *T. pallidum* our data consisted of SNP sites. To account for such ascertainment bias, we specified the number of constant sites of each nucleotide in our subsequent analyses. We sampled the posterior distribution using Markov chain Monte Carlo as implemented in BEAST1.10. The chain length was 10^8 steps, sampling every 10^3 to draw a total of 10^4 samples. The prior for the evolutionary rate of the strict clock was a CTMC-rate reference prior (i.e. a $\Gamma(\alpha = 0.5, \beta = T_l$, with $T_l =$ tree length, and mean$=\alpha/\beta$) [27]. For the log-normal distribution of the relaxed clock we also used the CTMC-rate reference prior for the mean rate (known as the ucld.mean in the program) and an exponential prior with mean 0.33 for the standard deviation (the ucld.stdev in the program). For the tree prior we used a constant-size coalescent with population size, θ with three possible priors, as described in Table 7. In all cases we used the HKY$+\Gamma_4$ substitution model, with the default priors in BEAST1.10.

To calculate log marginal likelihoods we used generalised stepping-stone [39, 40]. This method requires a working distribution for all parameters, including the tree prior, for which we used the matching coalescent model. We set 100 path steps between the unnormalised posterior and the working distribution, following equally spaced intervals from a $\beta(0.3, 1.0)$ distribution. For each step we ran a chain length of 2×10^6 steps.

Simulation experiments

For our simulations we considered well-understood conditions to isolate the impact of the tree prior on tests of temporal signal. We simulated phylogenetic trees under a constant-size coalescent model with a fixed population size of 1.0. The trees could be isochronous (i.e. ultrametric) or heterochronous. In the latter case we randomly assigned tip heights of of 0.0, 0.10, 0.35, or 0.50, with equal probability. This situation implies four discrete sampling periods, which resembles sequencing blitzes [41], or archaeological sampling of strata [42]. Because the coalescent model used to analyse the data conditions on sampling times we expect it to be robust to such sampling bias [43–45]. We used a JC substitution model [46] to generate sequence alignments of 1,000 nucleotides and an evolutionary rate of 0.05 subs/site/time, which resulted in around 250 unique site patterns. For our simulations under a relaxed molecular clock we sampled branch rates from a log-normal distribution with mean 0.05 and standard deviation of 0.25. The procedure for obtaining the simulated alignments consisted of specifying the model above in BEAST1.10 with an empty sequence alignment to sample phylogenetic trees. We then used NELSI [47] and Phangorn [48] to simulate evolutionary rates and sequence alignments, respectively.

We analysed the simulated data under heterochronous and isochronous models. For the heterochronous analysis we used the correct sampling times (where the correct model was the heterochronous) or assigned the tip heights above randomly (0.0, 0.10, 0.35, or 0.50) where the isocrhonous was the true model. For the evolutionary rate (clock.rate the strict and ucld.mean for the relaxed molecular clock model) we set the default CTMC-rate reference prior and the six possible configurations of the prior on θ (Table 7, plus those with hard bounds on the root height). For the isochronous analyses we did not specify sampling times and we fixed evolutionary rate to its true value of 0.05 to ensure that the branch lengths, θ and other parameters are in the correct units. We calculated log marginal likelihoods with the same procedure as for the empirical data.
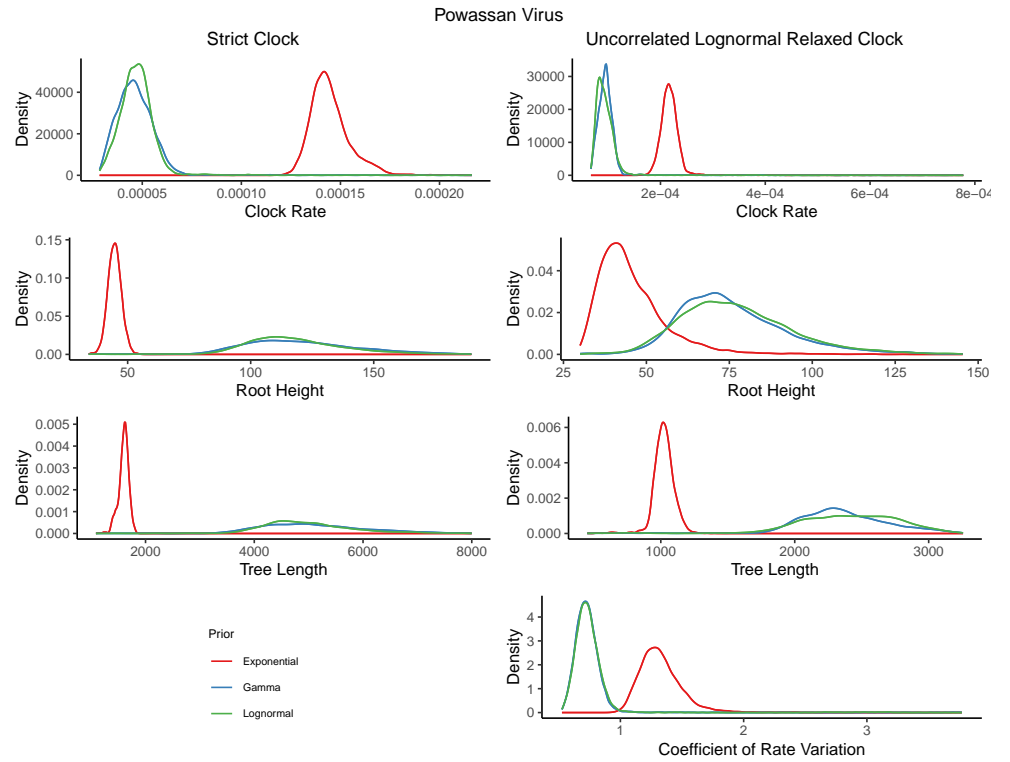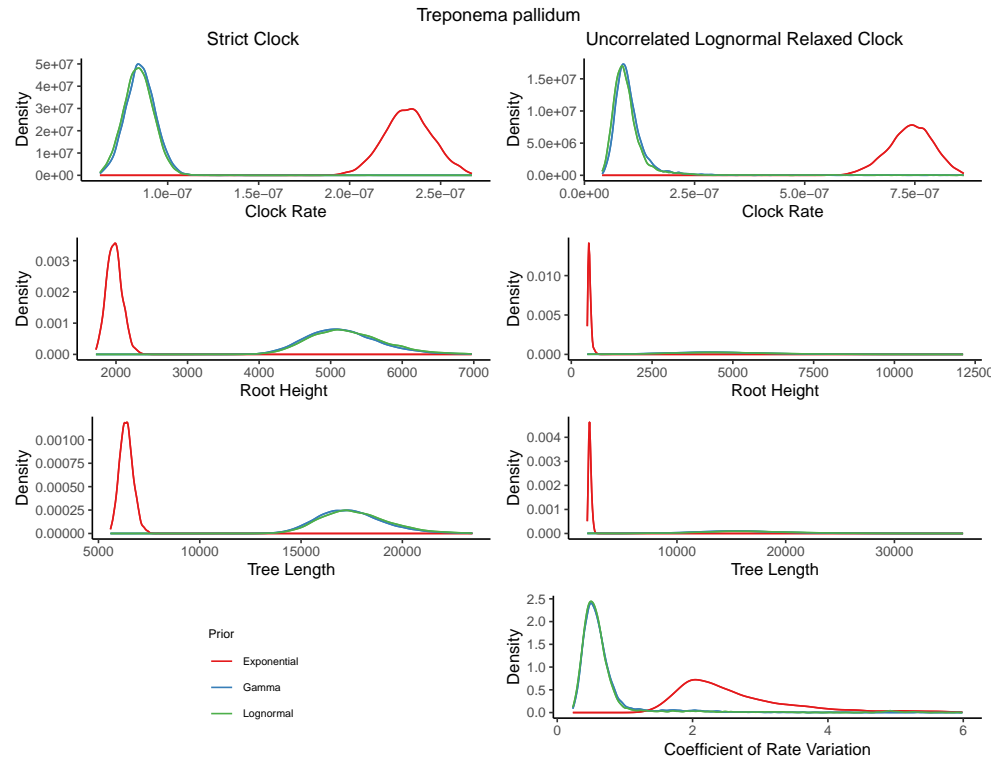
Supporting information

S1 Fig.
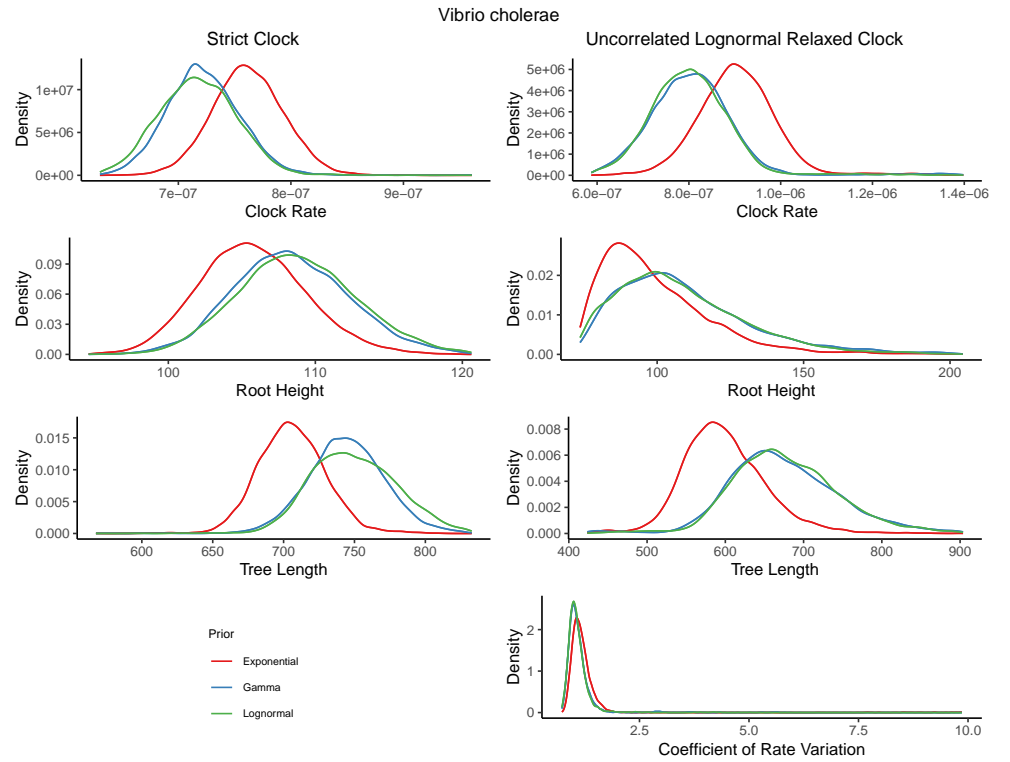
Densities of key statistics of cholera empirical data. The clock rate, root height, tree length, and coefficient of rate variation are shown under three priors on θ, exponential (red), gamma (blue), and lognormal (green).

S2 Fig.

Densities of key statistics of *Powassan virus* empirical data under. The clock rate, root height, tree length, and coefficient of rate variation are shown under three priors on θ, exponential (red), gamma (blue), and lognormal (green).

S3 Fig.

Densities of key statistics of *Treponema palladium* empirical data under. The clock rate, root height, tree length, and coefficient of rate variation are shown under three priors on θ, exponential (red), gamma (blue), and lognormal (green).

Table 8. Proportion of simulations with temporal signal under heterochronous simulated data

Species; Clock Model	Exponential	Gamma	Lognormal
Vibrio cholerae; *Strict Clock*	10	10	10
Vibrio cholerae; *Relaxed Clock*	10	10	10
Vibrio cholerae; *Best clock model*	10	10	10
Treponema pallidum; *Strict Clock*	0	0	0
Treponema pallidum; *Relaxed Clock*	0	10	8
Treponema pallidum; *Best clock model*			

Table 9. Proportion of simulations without temporal signal under isochronous simulated data

Species; Clock Model	Exponential	Gamma	Lognormal
Vibrio cholerae; *Strict Clock*	10	0	0
Vibrio cholerae; *Relaxed Clock*	0	0	0
Vibrio cholerae; *Best clock model*	0	0	0
Treponema pallidum; *Strict Clock*	10	10	10
Treponema pallidum; *Relaxed Clock*	10	0	0
Treponema pallidum; *Best clock model*			

Acknowledgments

References

1. Zuckerkandl E, Pauling L. Evolutionary divergence and convergence in proteins. In: Evolving genes and proteins. Elsevier; 1965. p. 97–166.

2. Drummond AJ, Ho SYW, Phillips MJ, Rambaut A. Relaxed phylogenetics and dating with confidence. PLoS Biology. 2006;4(5):e88.

3. Ho SY, Duchêne S. Molecular-clock methods for estimating evolutionary rates and timescales. Molecular Ecology. 2014;23(24):5947–5965.

4. Yang Z, Rannala B. Bayesian estimation of species divergence times under a molecular clock using multiple fossil calibrations with soft bounds. Molecular biology and evolution. 2006;23(1):212–226.

5. Dos Reis M, Yang Z. The unbearable uncertainty of Bayesian divergence time estimation. Journal of Systematics and Evolution. 2013;51(1):30–43.

6. Rodrigo AG, Felsenstein J. Coalescent approaches to HIV population genetics. The evolution of HIV. 1999; p. 233–272.

7. Korber B, Muldoon M, Theiler J, Gao F, Gupta R, Lapedes A, et al. Timing the ancestor of the HIV-1 pandemic strains. science. 2000;288(5472):1789–1796.

8. Drummond AJ, Pybus OG, Rambaut A, Forsberg R, Rodrigo AG. Measurably evolving populations. Trends in ecology & evolution. 2003;18(9):481–488.

9. Rambaut A, Lam TT, Max Carvalho L, Pybus OG. Exploring the temporal structure of heterochronous sequences using TempEst (formerly Path-O-Gen). Virus evolution. 2016;2(1):vew007.

10. Featherstone LA, Rambaut A, Duchene S, Wirth W. Clockor2: Inferring global and local strict molecular clocks using root-to-tip regression. BioRxiv. 2023; p. 2023–07.

11. Volz E, Frost SD. Scalable relaxed clock phylogenetic dating. Virus evolution. 2017;3(2):vex025.

12. Rieux A, Balloux F. Inferences from tip-calibrated phylogenies: a review and a practical guide. Molecular ecology. 2016;25(9):1911–1924.

13. Ramsden C, Holmes EC, Charleston MA. Hantavirus evolution in relation to its rodent and insectivore hosts: no evidence for codivergence. Molecular biology and evolution. 2009;26(1):143–153.

14. Duchêne S, Duchêne D, Holmes EC, Ho SY. The performance of the date-randomization test in phylogenetic analyses of time-structured virus data. Molecular Biology and Evolution. 2015;32(7):1895–1906.

15. Duchene S, Lemey P, Stadler T, Ho SY, Duchene DA, Dhanasekaran V, et al. Bayesian evaluation of temporal signal in measurably evolving populations. Molecular Biology and Evolution. 2020;37(11):3363–3379.

16. Molak M, Suchard MA, Ho SY, Beilman DW, Shapiro B. Empirical calibrated radiocarbon sampler: a tool for incorporating radiocarbon-date and calibration error into B ayesian phylogenetic analyses of ancient DNA. Molecular ecology resources. 2015;15(1):81–86.

17. Gelman A, Rubin DB. Avoiding model selection in Bayesian social research. Sociological methodology. 1995;25:165–173.

18. Gelman A, Carlin JB, Stern HS, Rubin DB. Bayesian data analysis; 2014.

19. Devault AM, Golding GB, Waglechner N, Enk JM, Kuch M, Tien JH, et al. Second-pandemic strain of Vibrio cholerae from the Philadelphia cholera outbreak of 1849. New England Journal of Medicine. 2014;370(4):334–340.

20. Majander K, Pfrengle S, Kocher A, Neukamm J, Du Plessis L, Pla-Díaz M, et al. Ancient bacterial genomes reveal a high diversity of Treponema pallidum strains in early modern Europe. Current Biology. 2020;30(19):3788–3803.

21. Vogels CB, Brackney DE, Dupuis AP, Robich RM, Fauver JR, Brito AF, et al. Phylogeographic reconstruction of the emergence and spread of Powassan virus in the northeastern United States. Proceedings of the National Academy of Sciences. 2023;120(16):e2218012120.

22. Suchard MA, Lemey P, Baele G, Ayres DL, Drummond AJ, Rambaut A. Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10. Virus Evolution. 2018;4(1):vey016.

23. Drummond AJ, Nicholls GK, Rodrigo AG, Solomon W. Estimating mutation parameters, population history and genealogy simultaneously from temporally spaced sequence data. Genetics. 2002;161(3):1307–1320.

24. R Oaks J, A Cobb K, N Minin V, D Leaché A. Marginal likelihoods in phylogenetics: a review of methods and applications. Systematic Biology. 2019;68(5):681–697.

25. Baele G, Li WLS, Drummond AJ, Suchard MA, Lemey P. Accurate model selection of relaxed molecular clocks in Bayesian phylogenetics. Molecular Biology and Evolution. 2013;30(2):239–243. doi:10.1093/molbev/mss243.

26. Wang Y, Yang Z. Priors in Bayesian phylogenetics. Bayesian phylogenetics: methods, algorithms, and applications. 2014; p. 5–24.

27. Ferreira MA, Suchard MA. Bayesian analysis of elapsed times in continuous-time Markov chains. Canadian Journal of Statistics. 2008;36(3):355–368.

28. Tay JH, Baele G, Duchene S. Detecting episodic evolution through Bayesian inference of molecular clock models. Molecular Biology and Evolution. 2023;40(10):msad212.

29. Kass RE, Raftery AE. Bayes factors. Journal of the American Statistical Association. 1995;90(430):773–795.

30. Duchêne S, Holt KE, Weill FX, Le Hello S, Hawkey J, Edwards DJ, et al. Genome-scale rates of evolutionary change in bacteria. Microbial genomics. 2016;2(11):e000094.

31. Ho SY, Shapiro B. Skyline-plot methods for estimating demographic history from nucleotide sequences. Molecular ecology resources. 2011;11(3):423–434.

32. Wesner JS, Pomeranz JP. Choosing priors in Bayesian ecological models by simulating from the prior predictive distribution. Ecosphere. 2021;12(9):e03739.

33. Nascimento FF, Reis Md, Yang Z. A biologist's guide to Bayesian phylogenetic analysis. Nature ecology & evolution. 2017;1(10):1446–1454.

34. Bouckaert RR. Tree priors and dating; 2021. Available from: `beast2.blogs.auckland.ac.nz/tree-priors-and-dating/`.

35. Rosenberg NA, Nordborg M. Genealogical trees, coalescent theory and the analysis of genetic polymorphisms. Nature Reviews Genetics. 2002;3(5):380–390.

36. Bouckaert RR. DensiTree: making sense of sets of phylogenetic trees. Bioinformatics. 2010;26(10):1372–1373.

37. Menardo F, Duchêne S, Brites D, Gagneux S. The molecular clock of Mycobacterium tuberculosis. PLoS pathogens. 2019;15(9):e1008067.

38. Biek R, Pybus OG, Lloyd-Smith JO, Didelot X. Measurably evolving pathogens in the genomic era. Trends in ecology & evolution. 2015;30(6):306–313.

39. Baele G, Lemey P, Suchard MA. Genealogical working distributions for Bayesian model testing with phylogenetic uncertainty. Systematic Biology. 2016;65(2):250–264.

40. Fan Y, Wu R, Chen MH, Kuo L, Lewis PO. Choosing among partition models in Bayesian phylogenetics. Molecular Biology and Evolution. 2011;28(1):523–532.

41. Porter AF, Sherry N, Andersson P, Johnson SA, Duchene S, Howden BP. New rules for genomics-informed COVID-19 responses–lessons learned from the first waves of the omicron variant in Australia. PLoS Genetics. 2022;18(10):e1010415.

42. Zhang C, Stadler T, Klopfstein S, Heath TA, Ronquist F. Total-evidence dating under the fossilized birth–death process. Systematic biology. 2016;65(2):228–249.

43. Stadler T, Vaughan TG, Gavryushkin A, Guindon S, Kühnert D, Leventhal GE, et al. How well can the exponential-growth coalescent approximate constant-rate birth–death population dynamics? Proceedings of the Royal Society B: Biological Sciences. 2015;282(1806):20150420.

44. Volz EM, Frost SD. Sampling through time and phylodynamic inference with coalescent and birth–death models. Journal of The Royal Society Interface. 2014;11(101):20140945.

45. Featherstone LA, Di Giallonardo F, Holmes EC, Vaughan TG, Duchêne S. Infectious disease phylodynamics with occurrence data. Methods in Ecology and Evolution. 2021;12(8):1498–1507.

46. Jukes TH, Cantor CR, et al. Evolution of protein molecules. Mammalian protein metabolism. 1969;3(24):21–132.

47. Ho SY, Duchêne S, Duchêne D. Simulating and detecting autocorrelation of molecular evolutionary rates among lineages. Molecular ecology resources. 2015;15(4):688–696.

48. Schliep KP. phangorn: phylogenetic analysis in R. Bioinformatics. 2011;27(4):592–593.