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.

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, 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 investigate the impact of the tree prior we considered different (hyper) prior distributions for the effective population size, Φ, 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 $1/\Phi$ 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 9.

Our rationale for using different prior distributions on Φ is its impact on parameters that pertain to the molecular clock. Under the constant coalescent process, Φ 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 Φ will result in a time to the most recent common ancestor that is older than if Φ is small. Thus a constant coalescent tree prior with a prior for Φ that favours large values will elicit old node times. The prior on Φ 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 $0.5/T$, where T=tree length, meaning that it is indirectly impacted by Φ.

Table 1. Prior distributions for the effective population size, Φ parameter

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

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 Φ, with log Bayes factors of over 200. Note that a log Bayes factor of 3.2 corresponds to a model posterior probability $\approx$0.95 [28], and is considered as 'very strong support', following Kass and Raftery [29]. Although in this data set the prior on Φ 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 Φ 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 Φ, 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, Φ

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

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
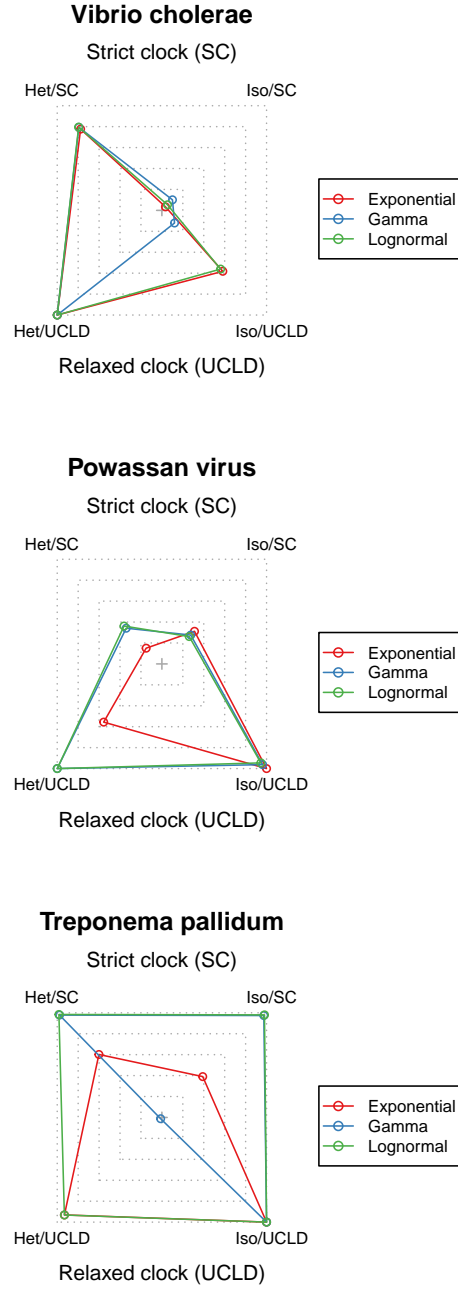
Vibrio cholerae

Strict clock (SC)

Powassan virus

Strict clock (SC)

Treponema pallidum

Strict clock (SC)

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 (Φ) 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.

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].

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 10 out of 10 simulation replicates were correctly classified as having temporal signal, using a log Bayes factor of at least 3.2 (Table 3). This perfect classification was supported regardless of the prior on Φ and the molecular clock model.

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, Φ. 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

For our data generated under isochronous trees, which by definition have no temporal signal, we found perfect classification under the exponential prior on Φ under both clock models (Table 4). For most analyses under the Γ and log-normal priors on Φ 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 were simulated under a strict or relaxed clock.

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

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 priorsomes on Φ) 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 overestimate of the height of the tree, especially under the strict molecular clock model (Fig 2). Under this situation, the sampling times represent such a small proportion of the tree height, that the heterochronous tree is indistinguishable from one that is ultrametric and thus their log marginal likelihoods can be similar. In fact, the inclusion of sampling times may improve the fit of the model by accommodating natural stochasticity in the root-to-tip distances. This phenomenon of tree extension also occurs under the relaxed molecular clock, but to a much lesser extent, such that 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.

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, Φ. 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

In Fig 3 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. A problem with this method is that the Φ 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
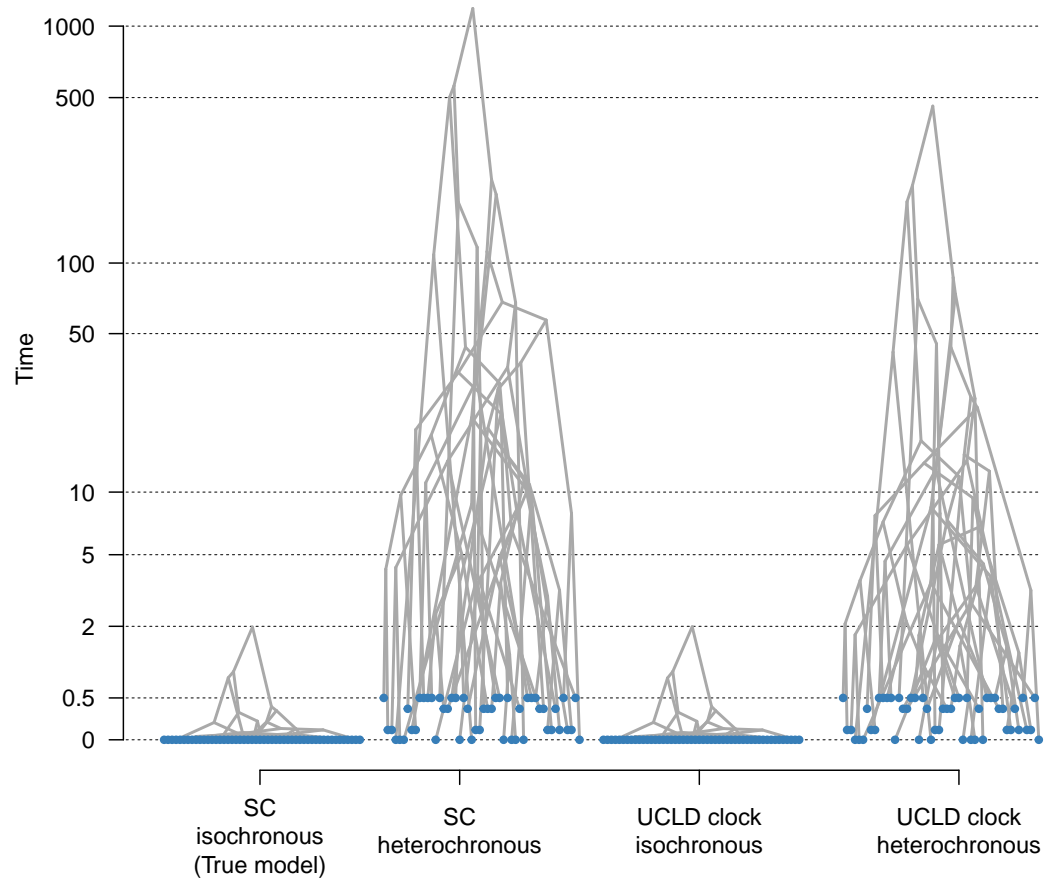
Fig 2. Distortion of phylogenetic node times from a simulation replicate with no temporal signal. Highest clade credibility trees from a data set simulated with no sampling times (isochronous) and under a strict molecular clock model (SC). The prior on Φ is a $\Gamma(\kappa = 0.001, \theta = 1000)$, which resulted in high classification errors. Tip nodes are labelled with solid blue circles. Including sampling times that span 0.5 units of time and about 1/4 of the true tree height result in a dramatic overestimation of the tree height. This effect occurs under 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 ($\log_{10}$).

coalescent). Thus, the prior on Φ 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).

Now discuss the inclusion of bounds as a potential solution and describe those results

The next section needs to be about prior sensitivity and prior predictive simulations

198
199
200
201
202
203
204
205

Fig 3. Distortion of phylogenetic node times 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 2), the root height of the trees with under all scenarios are comparable. Axes and labels are the same as those of Fig 2

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. 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.

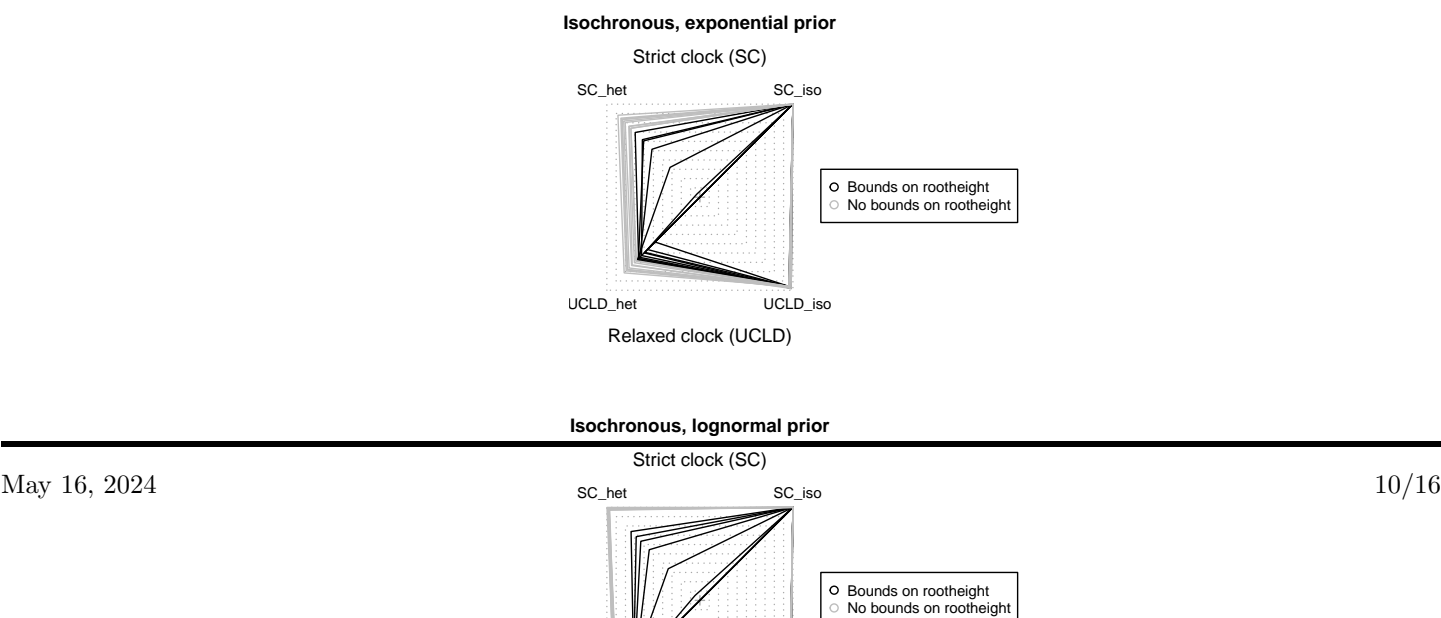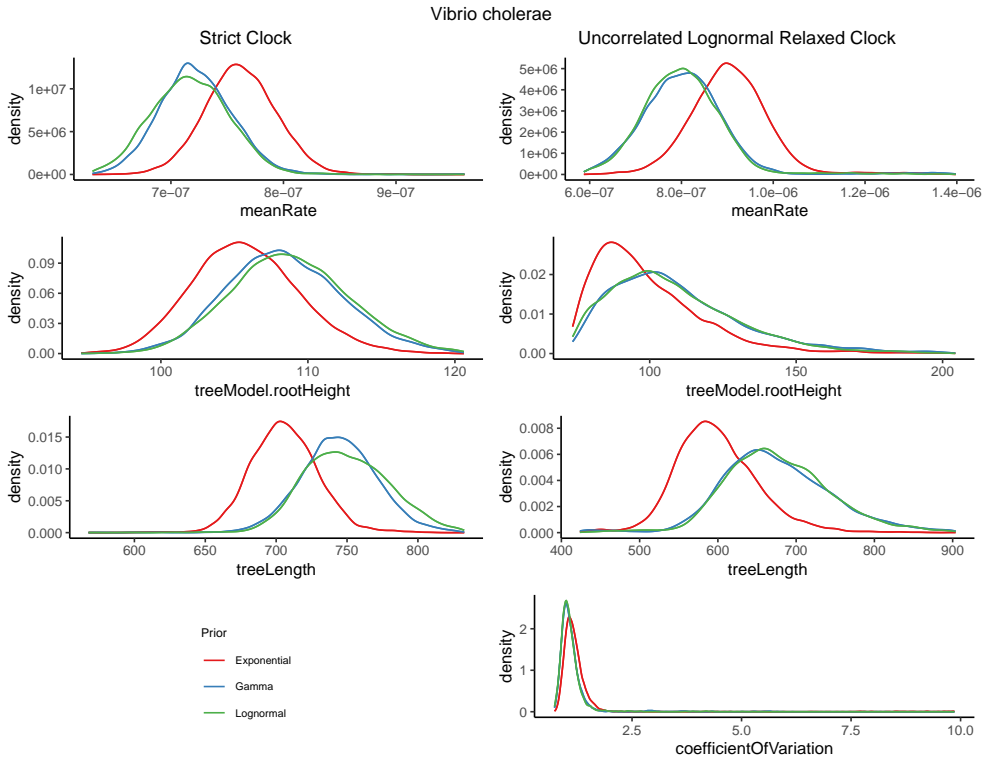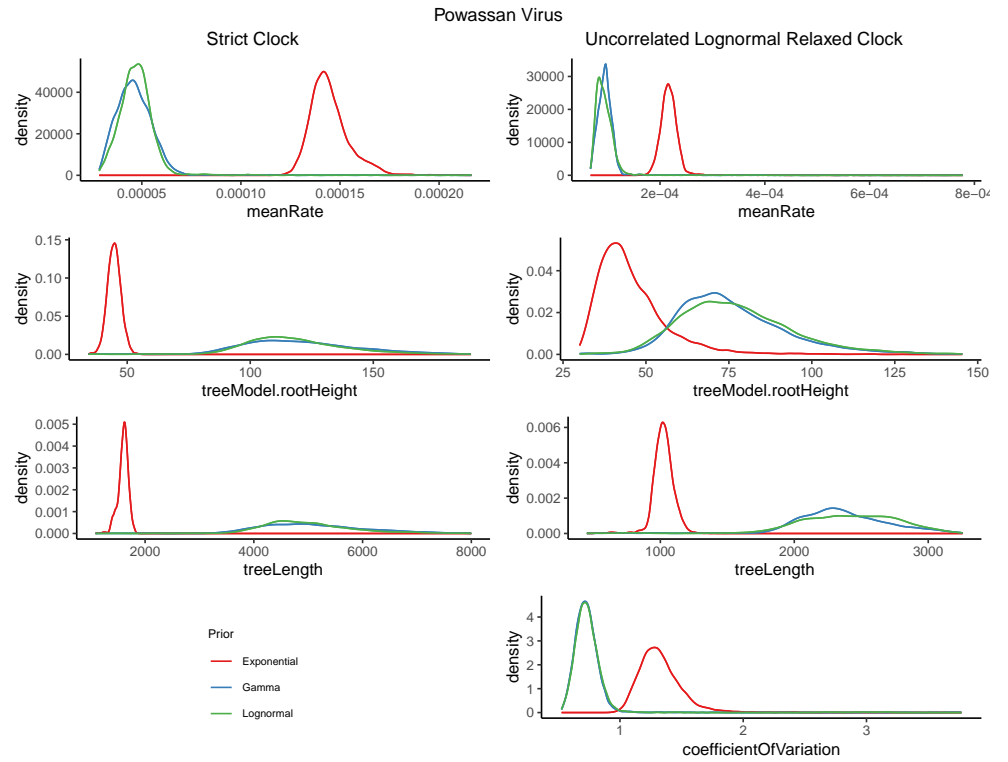| True clock model; clock model in analysis	Exponential	Γ	Log-normal
Strict clock; Strict clock	10	10	10
Strict clock; Relaxed clock	10	10	10
Strict clock; Best clock model	10	10	10
Relaxed clock; Strict clock	10	10	10
Relaxed clock; Relaxed clock	10	10	10
Relaxed clock; Best clock model	10	10	10

Table 7. 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
Orthoflavivirus powassanense; *Strict Clock*			
Orthoflavivirus powassanense; *Relaxed Clock*			
Treponema pallidum; *Strict Clock*	0	0	0
Treponema pallidum; *Relaxed Clock*	0	10	8
Treponema pallidum; *Relaxed Clock*			

Isochronous, exponential prior

Strict clock (SC)

SC_het   SC_iso

○ Bounds on rootheight
○ No bounds on rootheight

UCLD_het   UCLD_iso

Relaxed clock (UCLD)

Isochronous, lognormal prior

Strict clock (SC)

SC_het   SC_iso

○ Bounds on rootheight
○ No bounds on rootheight

Table 8. Proportion of simulations without temporal signal under isochronous simulated data

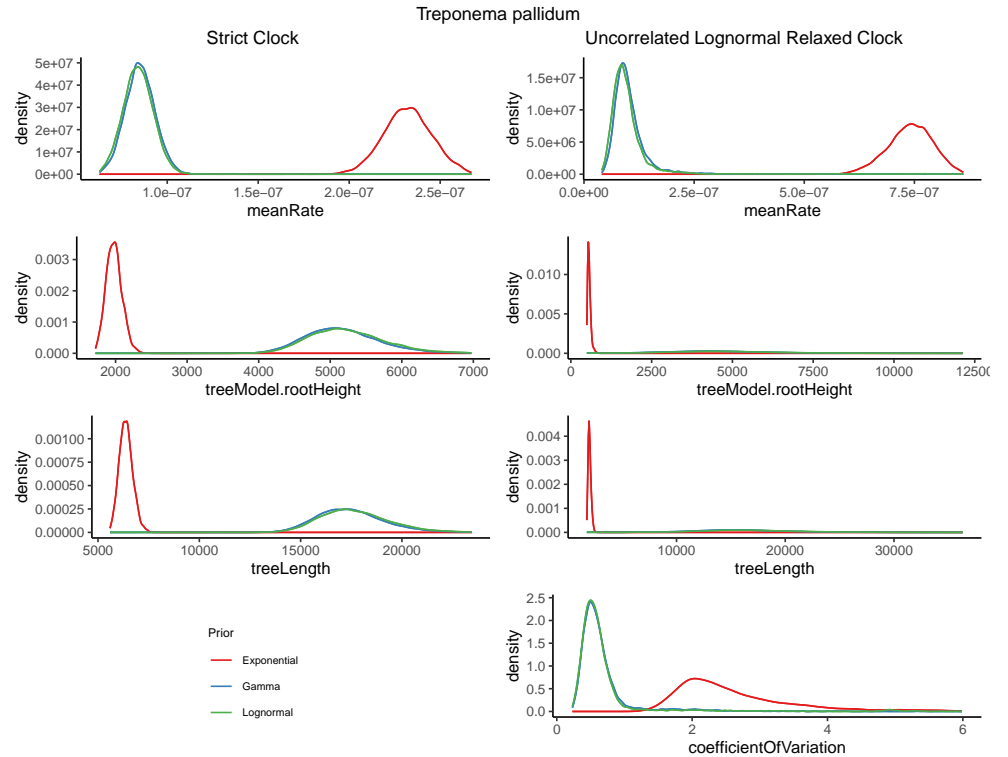| Species; Clock Model	Exponential	Gamma	Lognormal
Vibrio cholerae; *Strict Clock*	10	0	0
Vibrio cholerae; *Relaxed Clock*	0	0	0
Vibrio cholerae; *Best clock model*	0	0	0
Orthoflavivirus powassanense; *Strict Clock*			
Orthoflavivirus powassanense; *Relaxed Clock*			
Treponema pallidum; *Strict Clock*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

208

To do: Plot densities of clock rates, root height, tree length, coefficient of rate variation. One panel per data set, show density for the three priors and the same colours as the fig above.

209
210
211

212

Density plots of the clock rate, root height, tree length, and coeffcient of rate variation of cholera.

213
214

Density plots of the clock rate, root height, tree length, and coeffcient of rate variation of powassan virus.

Density plots of the clock rate, root height, tree length, and coeffcient of rate variation of treponema palladium.

Discussion

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* [32]) 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 [33].

Ancient DNA has had a profound impact in our understanding of the long-term evolution of many microbial pathogens [34, 35]. The inclusion of ancient DNA in molecular evolutionary analyses can substantially broaden the sampling window, and therefore provide invaluable data points for molecular clock calibration because the data can effectively be treated as coming from a measurably evolving population [36].

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.

Materials and methods

Empirical data

We selected three different datasets to evaluate temporal signal, cholera (), powassan virus (), and treponema (). These varied in their timespans, number of informative sites, and number of sequences. The most sparse dataset was treponema, where sampling dates spanned 481 years, with 28 sequences each containing 1500 sites. There were 319 sequences with 11193 sites for powassan virus, spanning 5 years of sampling times. Finally, we had 122 cholera sequences with 1757 sites across 73 years of sampling. For the empirical data there are 1392 unique site patterns in cholera, 3457 in powassan virus, and 840 in treponema. In the simulated data, there are 580-710 (SC) and 640-710 (UCLD) in cholera, there are 110-150 in both SC and UCLD for treponema. For the powassan virus SC, there are ∼4800 in the isochronous trees and ∼14650 in the heterchronous trees.

Table 9. 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

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 $\Gamma(\alpha = 0.5, \beta = T$, with $T =$ tree length, and mean$=\alpha/\beta$). 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, Φ with three possible priors, as described in Table 9. In all cases we used the HKY+Γ_4 substitution model, with the default priors in BEAST1.10.

To calculate log marginal likelihoods we used generalised stepping-stone [37, 38]. 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

Sebastian, describe small simulations. That is, that the population size was 1.0, and what the clock rates were. For the analyses say that in the isochronous the clock rate was fixed to the correct value to ensure that the units are correct, etc... Also describe how we designed the bounds here and what they do

Supporting information

S1 Fig. **Bold the title sentence.** Add descriptive text after the title of the item (optional).

S2 Fig. **Lorem ipsum.** Maecenas convallis mauris sit amet sem ultrices gravida. Etiam eget sapien nibh. Sed ac ipsum eget enim egestas ullamcorper nec euismod ligula. Curabitur fringilla pulvinar lectus consectetur pellentesque.

S1 File. **Lorem ipsum.** Maecenas convallis mauris sit amet sem ultrices gravida. Etiam eget sapien nibh. Sed ac ipsum eget enim egestas ullamcorper nec euismod ligula. Curabitur fringilla pulvinar lectus consectetur pellentesque.

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;370(4).

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

33. 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.

34. Duchêne S, Ho SY, Carmichael AG, Holmes EC, Poinar H. The recovery, interpretation and use of ancient pathogen genomes. Current Biology. 2020;30(19):R1215–R1231.

35. Spyrou MA, Bos KI, Herbig A, Krause J. Ancient pathogen genomics as an emerging tool for infectious disease research. Nature Reviews Genetics. 2019;20(6):323–340.

36. Ho SY, Duchêne S. Dating the emergence of human pathogens. Science. 2020;368(6497):1310–1311.

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

38. 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.