

1

2 Supplementary results for: The phylodynamic threshold of 3 measurably evolving populations

4 Ariane Weber^{1,*}, Julia Kende^{2,3}, Camila Duitama González^{3,4}, Sanni Översti^{1,‡} and Sebastian
5 Duchene^{3,4,5,‡,*}.

6 ¹ Max Planck Institute of Geoanthropology, Jena, Germany.

7 ² Bioinformatics and Biostatistics Hub, Institut Pasteur, Paris, France.

8 ³ Université Paris Cité, Paris, France.

9 ⁴ ED-ID unit, Dept of Computational Biology, Institut Pasteur, Paris, France.

10 ⁵ Peter Doherty Institute for Infection and Immunity, Dept of Microbiology and Immunology, University
11 of Melbourne, Melbourne, Australia.

12 *email: weber@gea.mpg.de, sduchene@pasteur.fr

13 ‡ Equal contribution to the supervision of this work.

14 In fig. S1 we show the posterior estimates of the root height for our simulations of varying sampling window
15 widths.

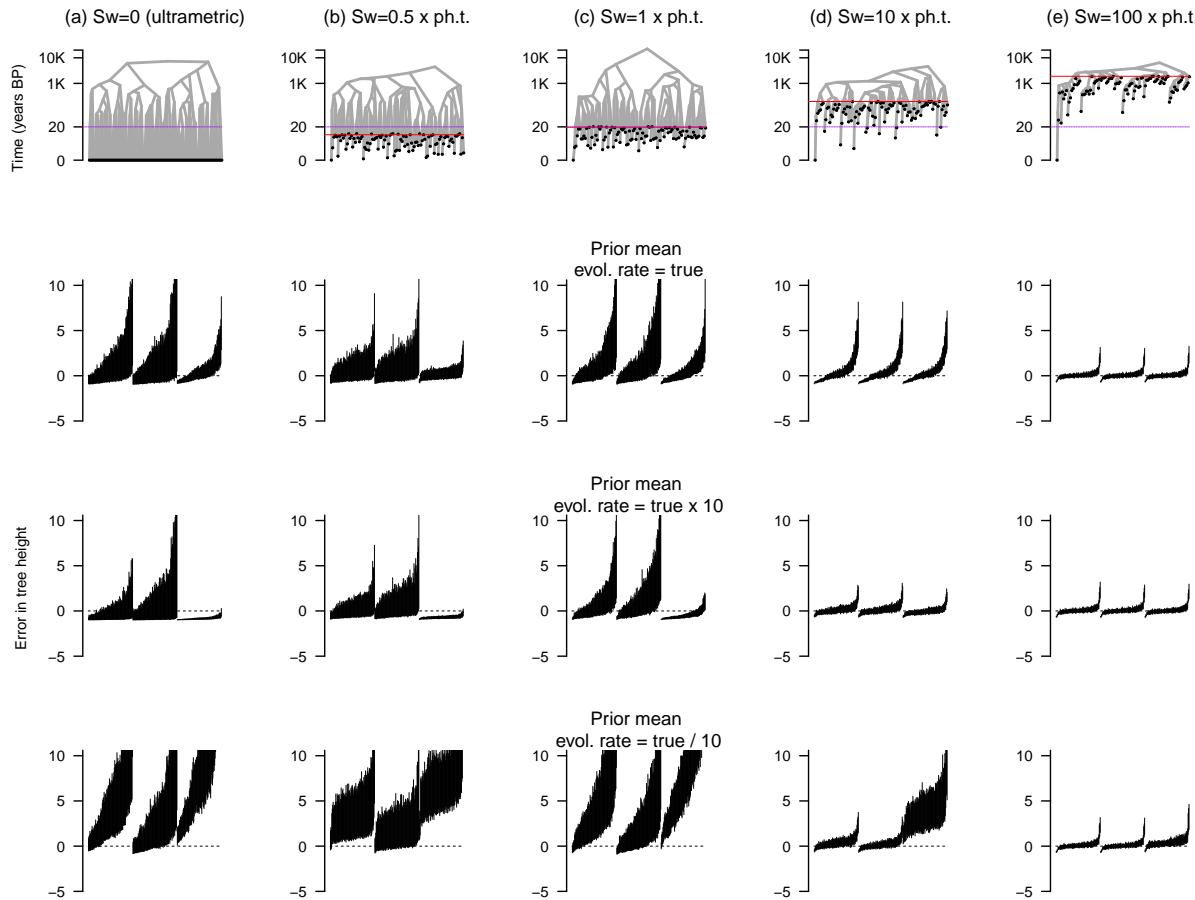


Figure 1: Estimates of the age of the root-node for simulations of varying sampling window widths. Each column corresponds to a simulation setting: (a) is for ultrametric trees where all samples are collected at the same point in time (sampling window, $sw=0$), (b) is for the situation where the sampling window is 10 years (half the expected phylodynamic threshold; sampling window, $sw=0.5 \times ph.t.$), (c) is where the sampling window is exactly the expected phylodynamic threshold of 20 years ($sw=1 \times ph.t.$). Scenarios (d) and (e) denote sampling windows that are 10 and 100 \times the expected phylodynamic threshold ($sw=10 \times ph.t$ and $sw=100 \times ph.t.$, respectively). The rows denote example phylogenetic trees and prior configurations where the mean evolutionary rate prior, M , is set to the correct value (first row), an order of magnitude higher (second row), or an order of magnitude lower, last row). Because each simulated tree has a different root height, we subtract the posterior by the truth and divide by the truth. Thus, a value of 0 is the correct value, where as one of 1 means a one-fold overestimation. The dashed line is for values of 0. Each vertical lines are the 95% posterior credible interval, with 100 replicates for each of the priors on M in table 1

Table 1: Prior configuration for the mean evolutionary rate, M of the lognormal distribution of branch rates. Note that the mean of the *Gamma* distribution here is *shape/rate* and that the true value used to generate the data is 1.5×10^{-5} subs/site/year.

Mean value of M	Prior configuration	95% quantile width / mean M
1.5×10^{-5}	$\text{Gamma}(\text{shape} = 1.5, \text{rate} = 1 \times 10^5)$	3.04
1.5×10^{-5}	$\text{Gamma}(\text{shape} = 0.3, \text{rate} = 2 \times 10^4)$	6.33
1.5×10^{-5}	$\text{Gamma}(\text{shape} = 15, \text{rate} = 1 \times 10^6)$	1.00
1.5×10^{-4}	$\text{Gamma}(\text{shape} = 1.5, \text{rate} = 1 \times 10^4)$	3.04
1.5×10^{-4}	$\text{Gamma}(\text{shape} = 0.3, \text{rate} = 2 \times 10^3)$	6.33
1.5×10^{-4}	$\text{Gamma}(\text{shape} = 15, \text{rate} = 1 \times 10^5)$	1.00
1.5×10^{-6}	$\text{Gamma}(\text{shape} = 1.5, \text{rate} = 1 \times 10^6)$	3.04
1.5×10^{-6}	$\text{Gamma}(\text{shape} = 0.3, \text{rate} = 2 \times 10^5)$	6.33
1.5×10^{-6}	$\text{Gamma}(\text{shape} = 15, \text{rate} = 1 \times 10^7)$	1.00

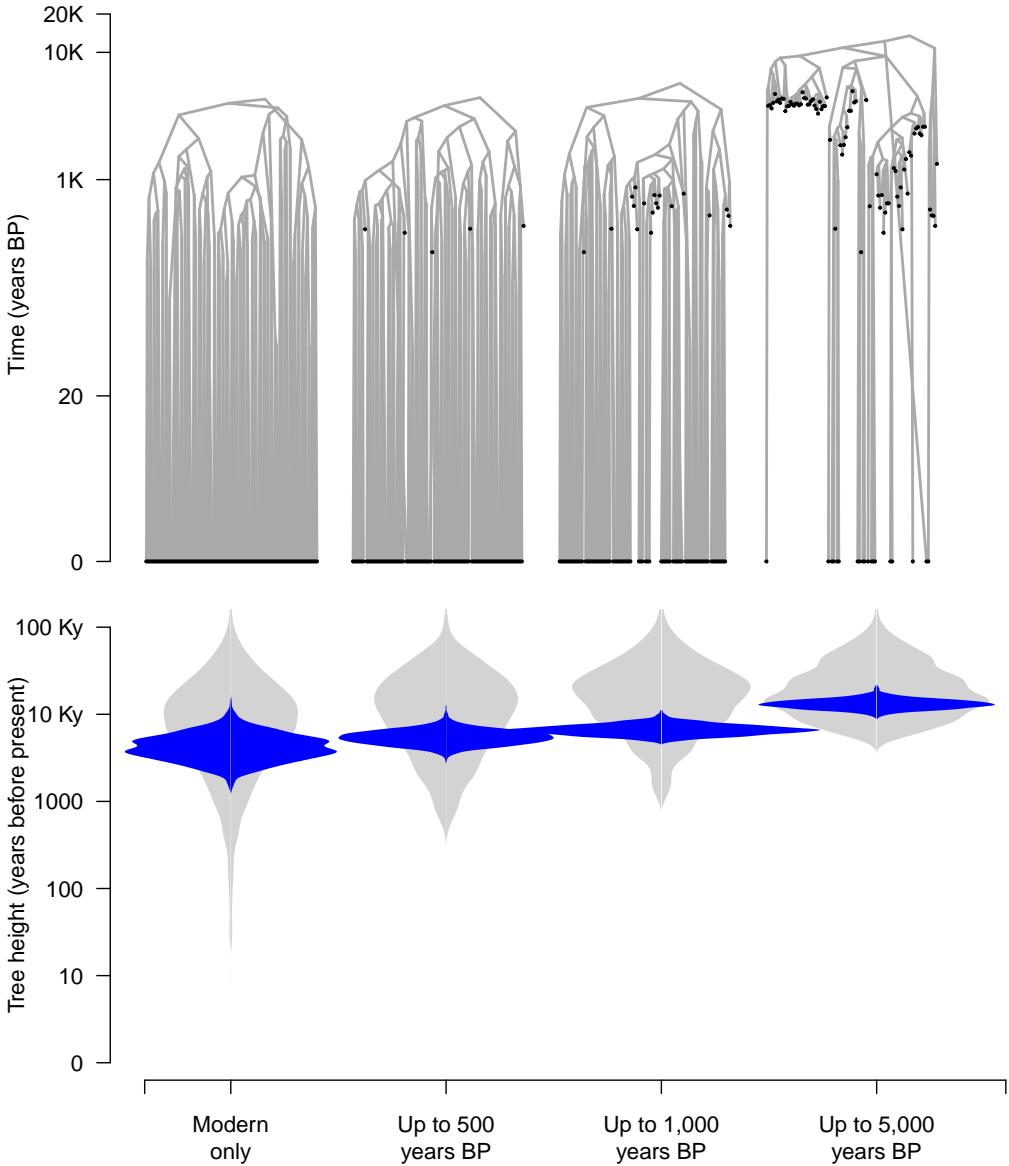


Figure 2: Results from empirical analyses of Hepatitis B virus (HBV) ancient DNA data. The phylogenetic trees correspond to highest clade credibility trees from three analyses where the data were subsampled to include an increasing number of ancient samples. First, we consider a data set for which the samples are 95% modern and the remaining 5% being the most ancient. Then, we reduce the number of modern samples to 50%, 25% and 10%, and the rest being ancient. Note that the sampling window is constant because we always retain the most ancient samples. In all cases the data sets consist of 100 genome samples. The violin plots show the posterior distribution of the tree height in blue and its corresponding prior in orange.