

Molecular clock dating figures and tables (Dataset D)

Treponema pallidum in early modern Europe

Louis du Plessis

Last modified: 01 Jul 2020

Contents

1	Dataset	1
2	Root-to-tip regression	7
3	BEAST2 molecular clock dating	7
3.1	Clade TMRCA and sampling date estimates	7
3.2	Posterior distributions compared between models	9
3.3	MCC trees	10
4	BEAST2 Bayesian DRT	12
5	Session info	12

1 Dataset

Dataset D. That is, the complete dataset with Kampen, Nichols, NIC-2, 94A and 94B removed.

Table 1: Radiocarbon date ranges used.

Name	Sampling location	Strain	Radiocarbon date range	Mean sampling date
133	Mexico	TPE	1600–1861	1730.5
PD28	Finland	TPA	1666–1789	1727.5
SJ219	Estonia	TPA	1429–1635	1534.5
CHS119	Finland	TPE	1450–1630	1540.0

A

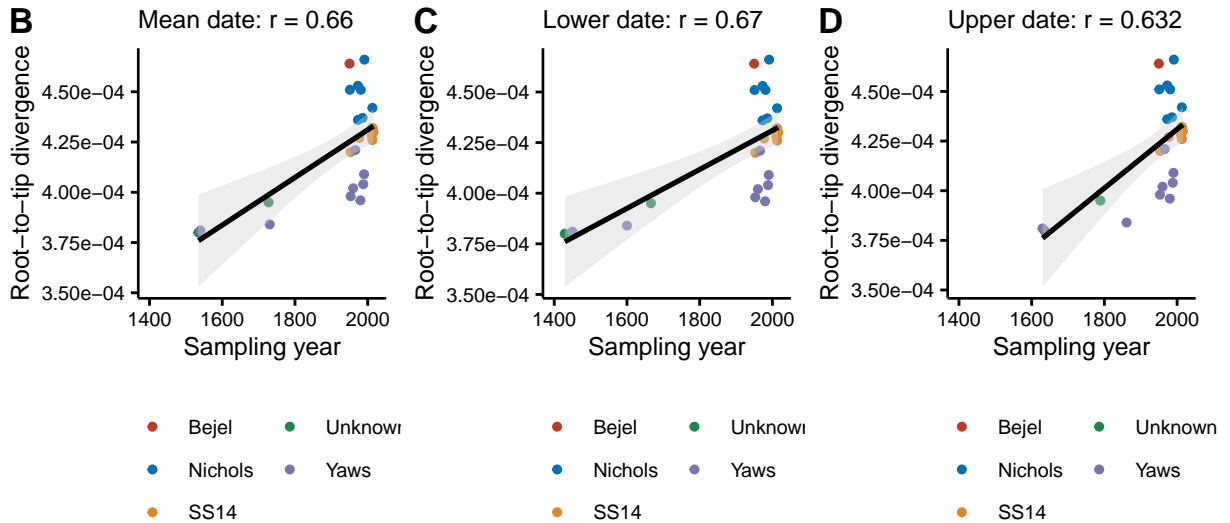
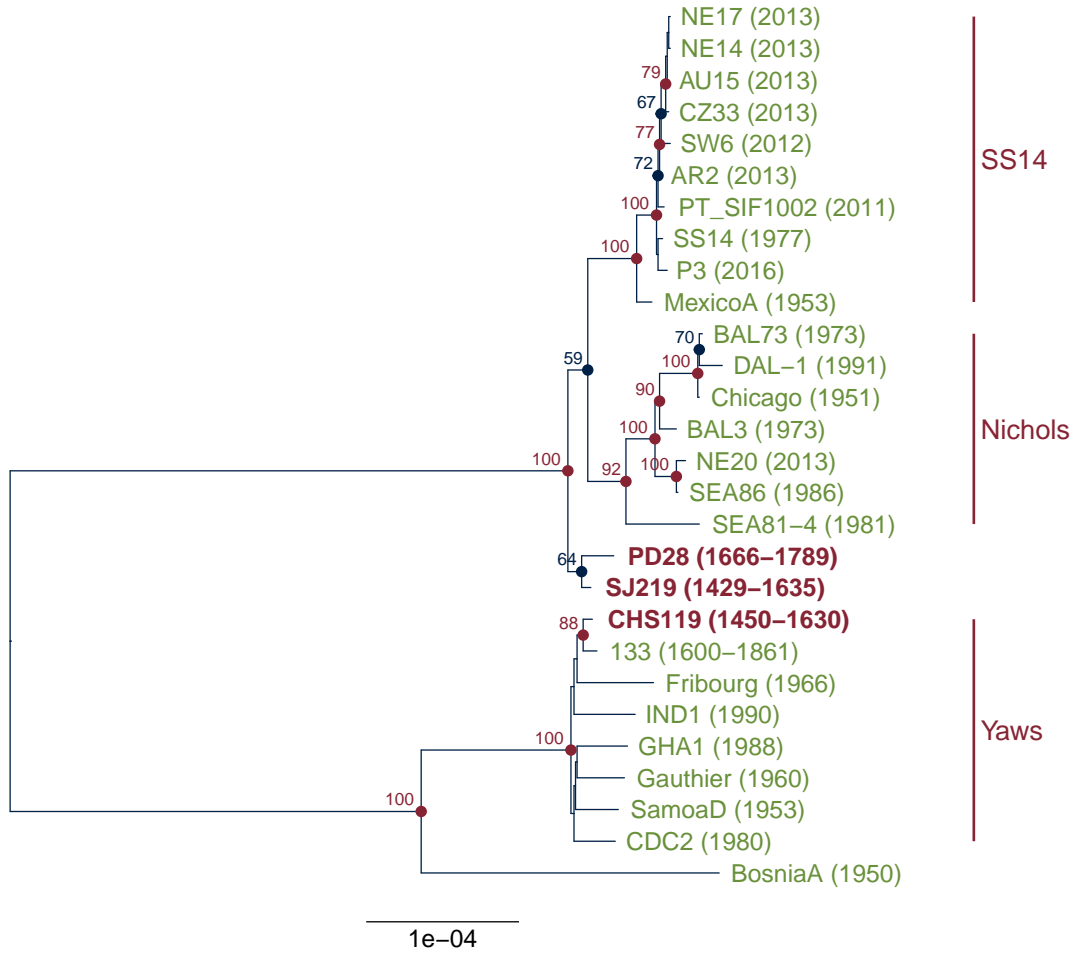


Figure 1: (A) Midpoint-rooted ML tree of **dataset D**. The sampling date or date range is given in parentheses after the sample name. Newly sequenced samples are highlighted. (B-D) Root-to-tip distance in the ML tree plotted against sampling date, respectively using the mean sampling date, lower limit of the date range and upper limit of the date range.

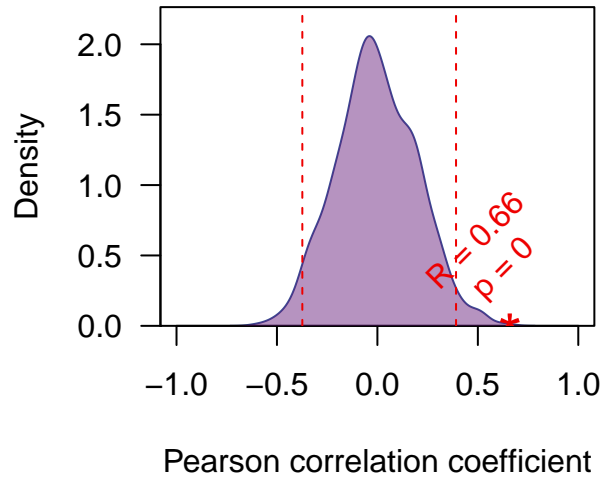


Figure 2: Null distribution of the Pearson correlation coefficient between root-to-tip distance and sampling date (using mean sampling dates) for dataset D when shuffling dates across the whole phylogeny. The dashed lines indicate 2.5 and 97.5 quantiles of the distribution and the star indicates the correlation coefficient using the true (unshuffled) sampling dates. The p-value is the proportion of replicates with the test statistic (Pearson correlation coefficient) greater than or equal to the true value.

A

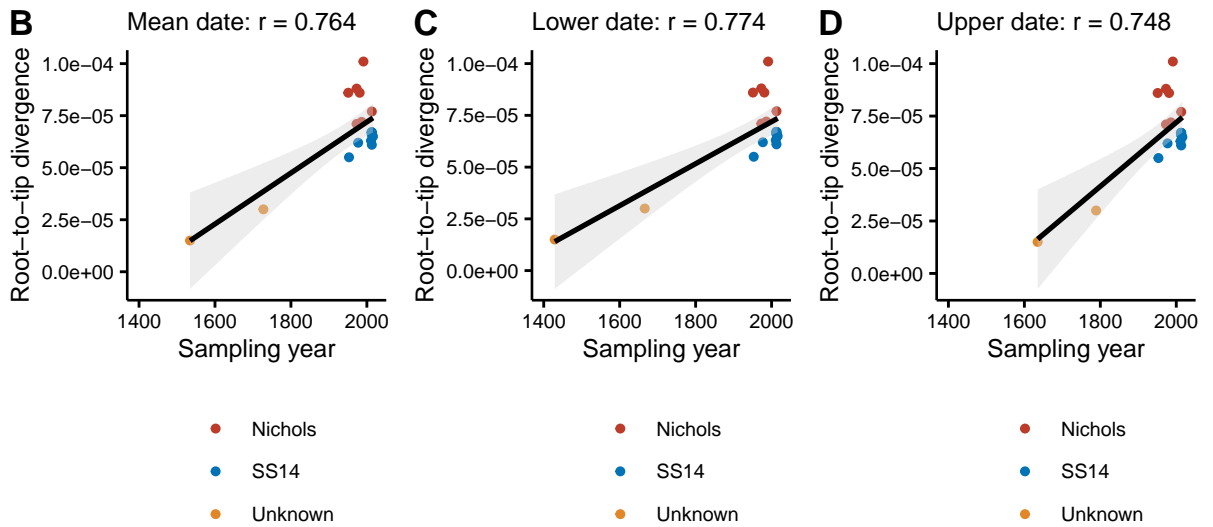
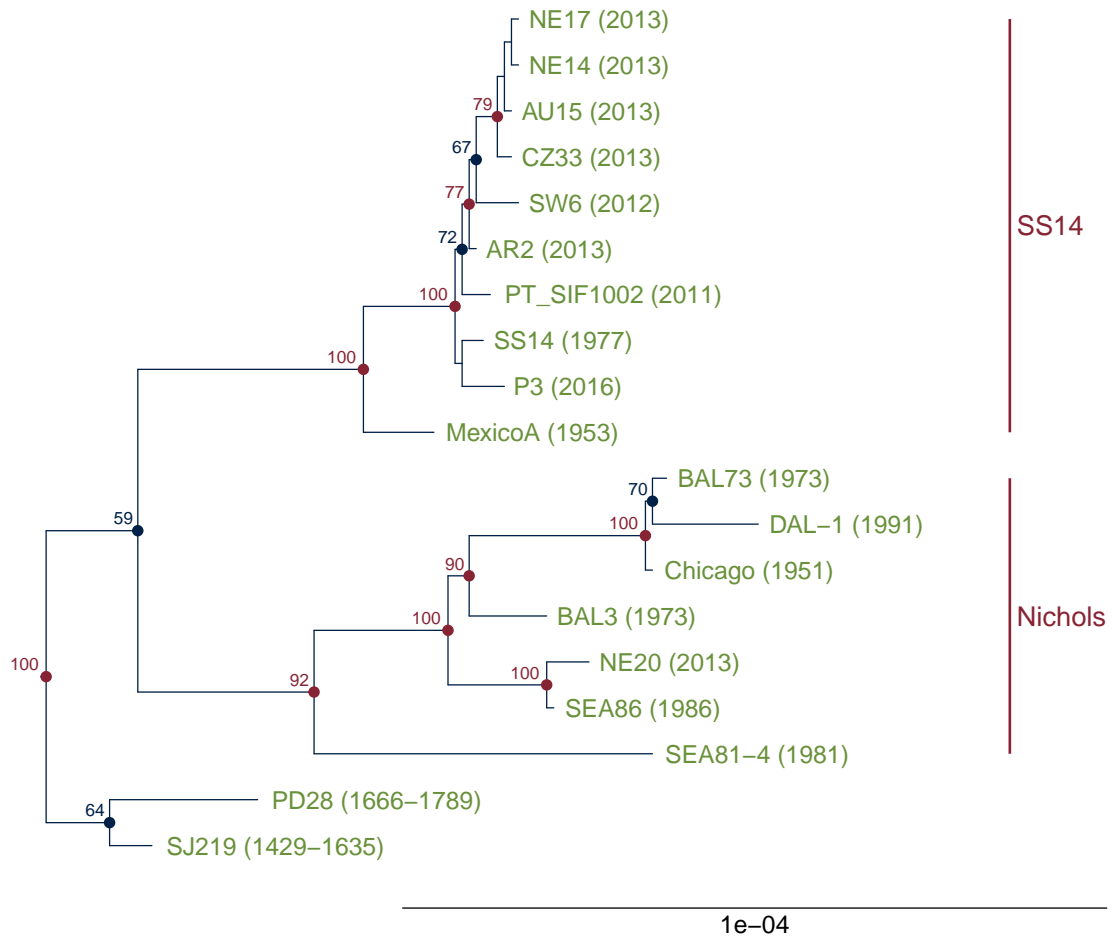
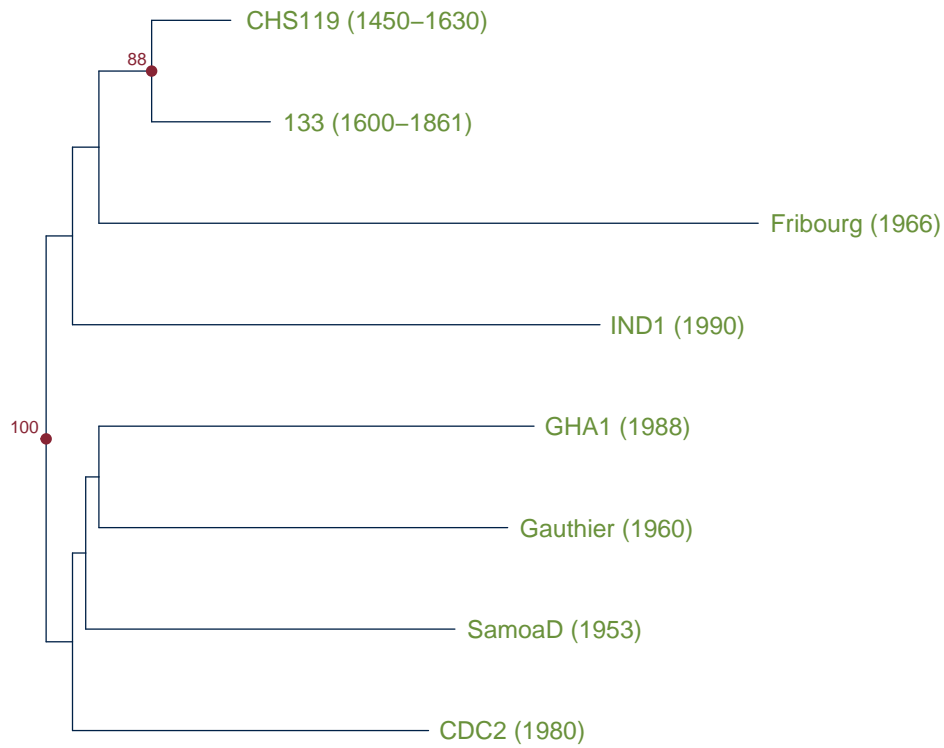


Figure 3: As the previous figure, but for the syphilis clade only.

A



1e-04

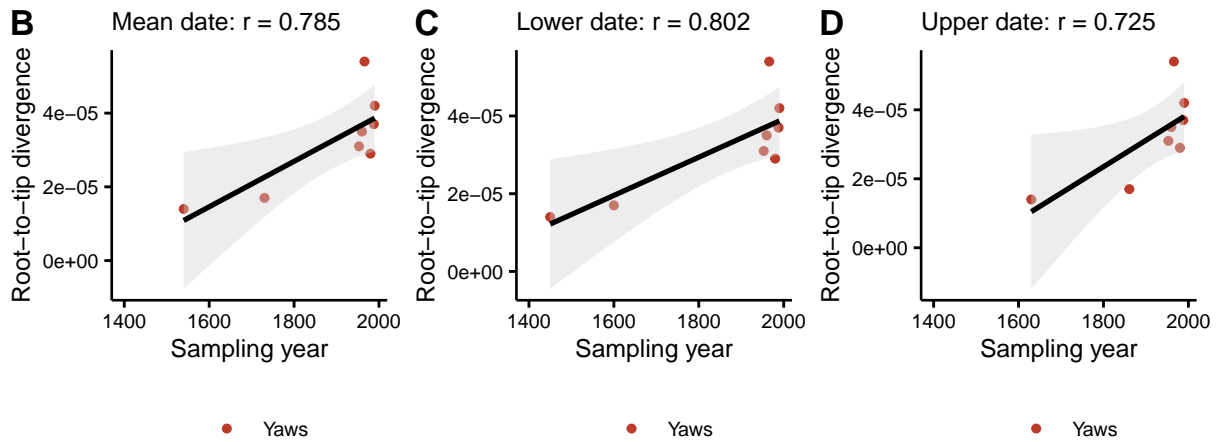


Figure 4: As the previous figure, but for the yaws clade only.

A

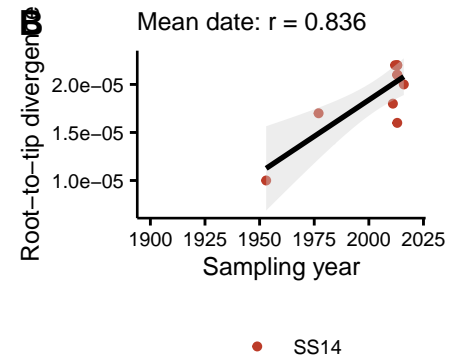
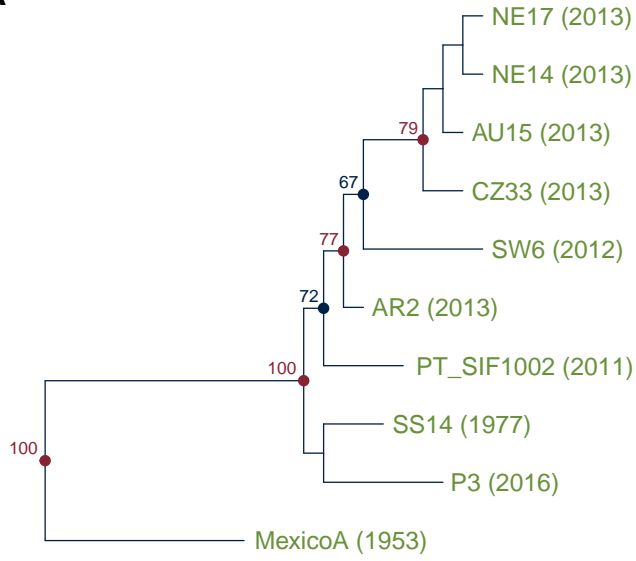
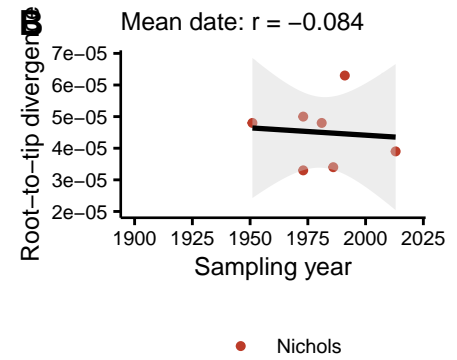
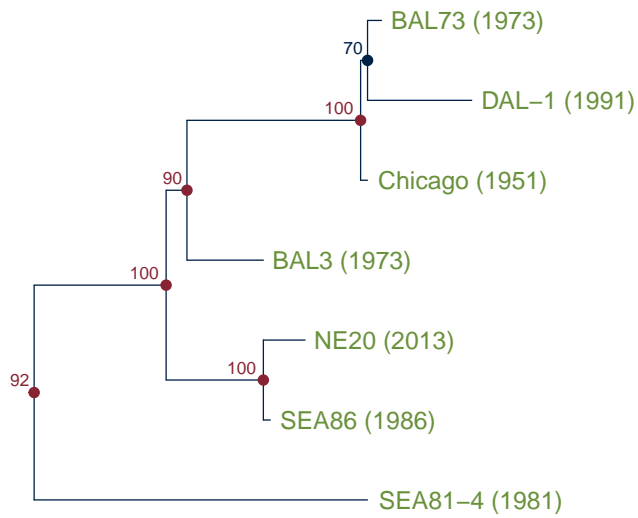


Figure 5: As the previous figure, but for the SS14 clade only.

A



1e-04

Figure 6: As the previous figure, but for the Nichols clade only.

2 Root-to-tip regression

3 BEAST2 molecular clock dating

3.1 Clade TMRCA and sampling date estimates

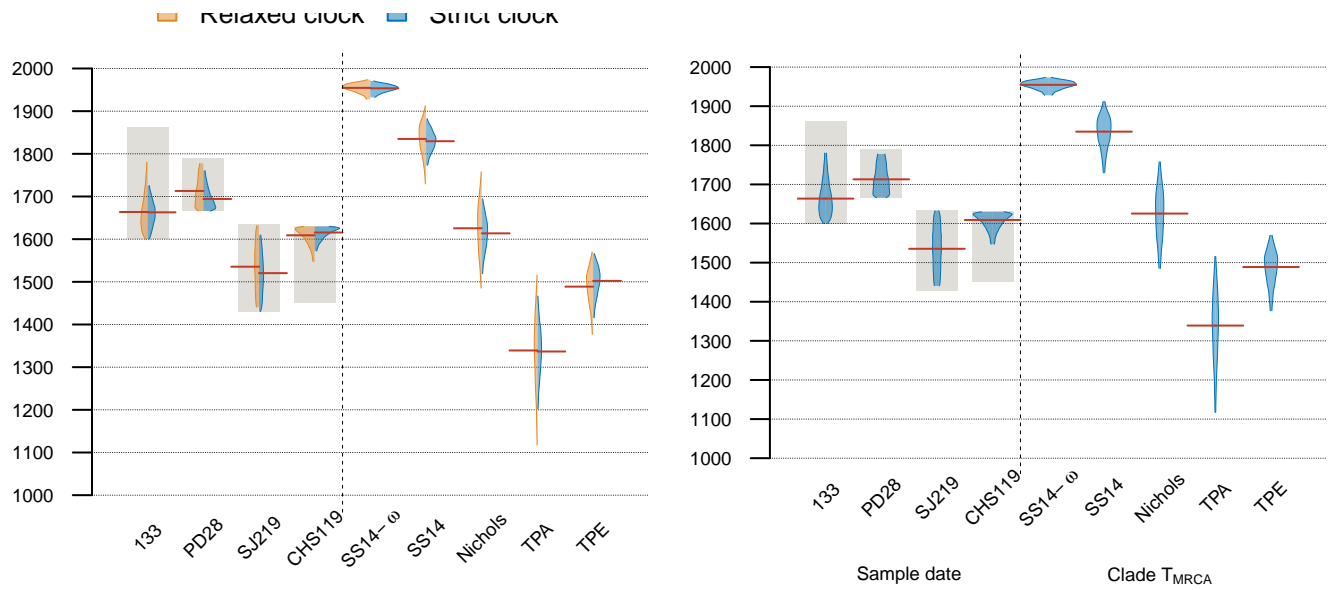


Table 2: Posterior sampling date estimates for ancient sequences. The posterior probability that a sample is pre-Columbian is calculated as the proportion of posterior samples with a date < 1493 .

Sample	Median	HPD lower	HPD upper	Pre-Columbian
133	1663.72	1600.00	1780.22	0.00
PD28	1713.01	1666.07	1777.66	0.00
SJ219	1535.41	1440.72	1632.46	0.28
CHS119	1609.09	1547.07	1629.99	0.01

Table 3: Posterior TMRCA estimates for clades. The posterior probability that a clade is monophyletic is calculated as the proportion of posterior trees where the clade is monophyletic.

Clade	Median	HPD lower	HPD upper	Monophyletic
SS14-w	1954.58	1927.80	1973.75	0.98
SS14	1834.92	1729.45	1912.34	0.98
Nichols	1625.48	1485.25	1757.74	0.97
TPA	1339.12	1117.16	1515.99	0.98
TPE	1488.78	1376.73	1570.28	0.97
TPE/TEN	472.19	-371.68	1110.10	0.96

3.2 Posterior distributions compared between models

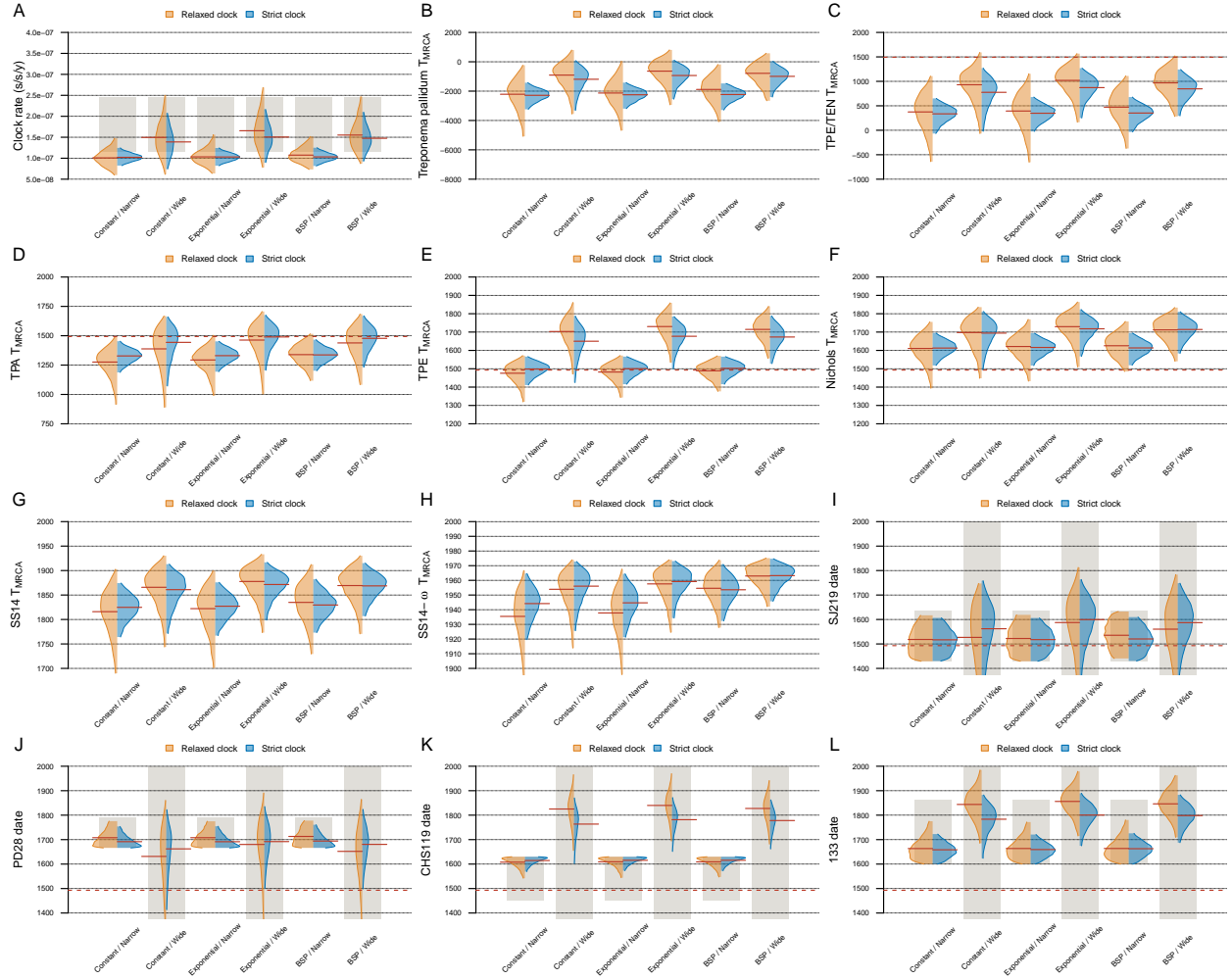


Figure 7: Posterior distributions for the clock rate, divergence dates and sampling dates under a relaxed (orange) and strict (blue) clock model with various tree priors and priors on the sampling dates of ancient sequences. The distributions are truncated at the upper and lower limits of the 95% HPD interval and the red lines indicate the median estimates.

3.3 MCC trees

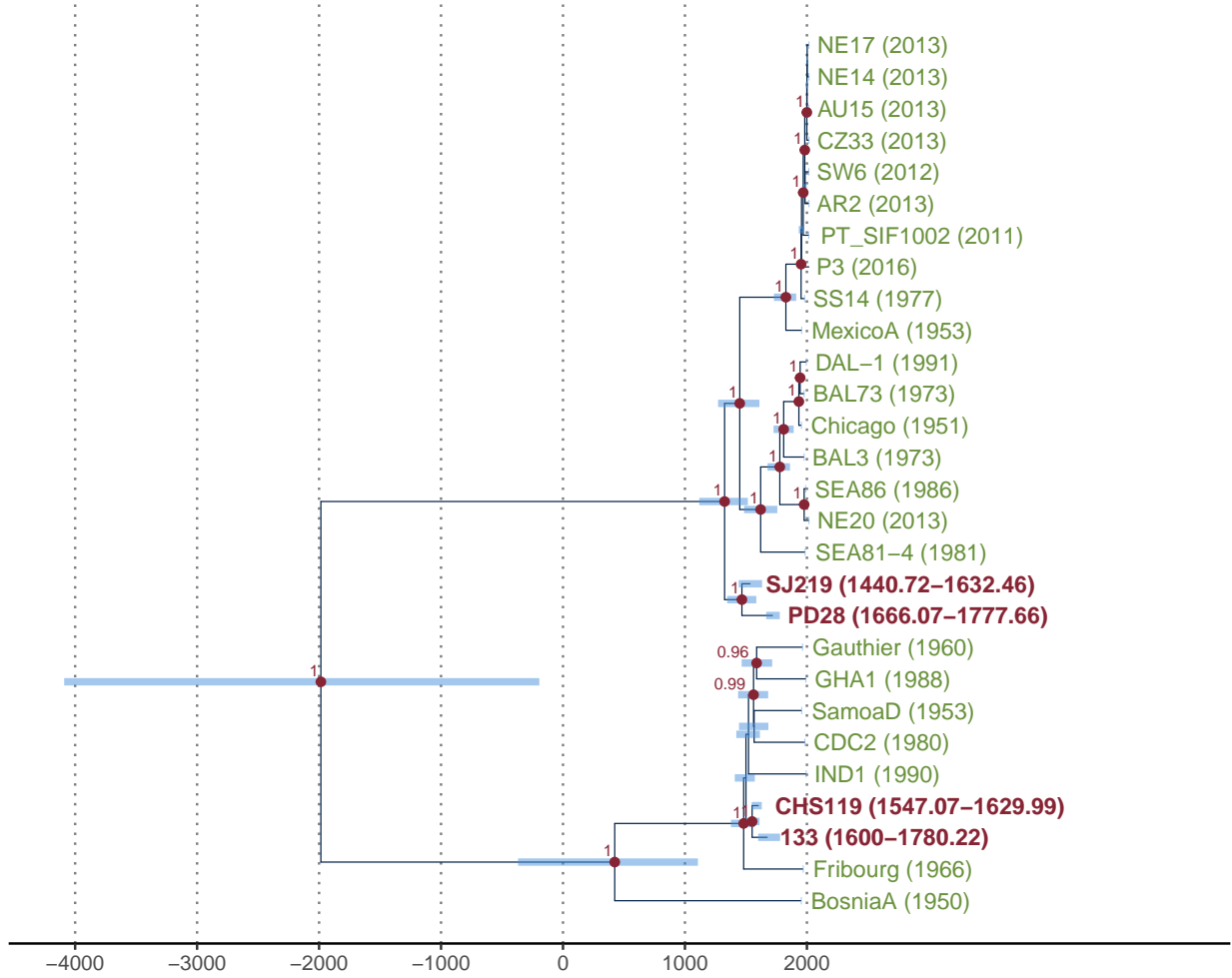


Figure 8: Maximum clade credibility tree inferred under a **relaxed clock** model, **Bayesian skyline plot** coalescent tree prior and **narrow uniform** priors on sampling dates. Sequences with estimated sampling dates are highlighted and the HPD interval of the sampling date is shown in parentheses. Nodes with posterior support of at least 90% are highlighted.

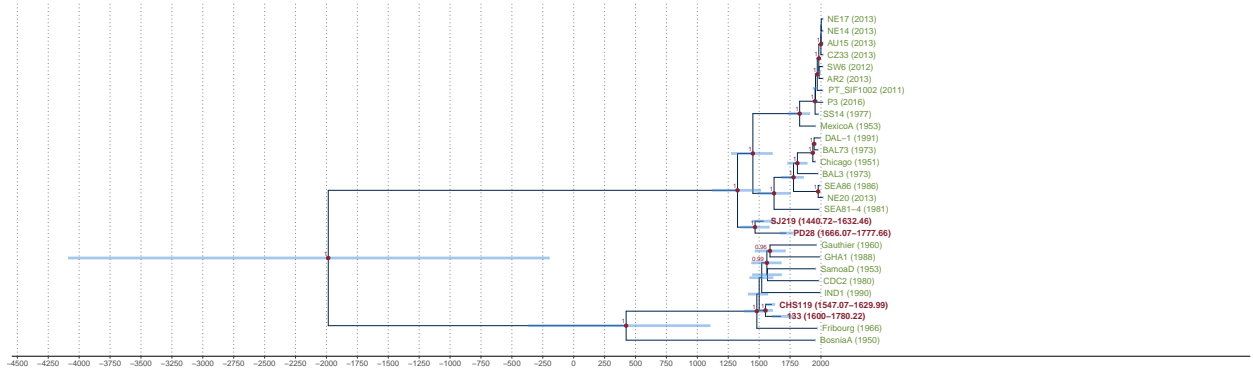


Figure 9: Maximum clade credibility tree inferred under a **relaxed clock** model, **Bayesian skyline plot** coalescent tree prior and **narrow uniform** priors on sampling dates. Sequences with estimated sampling dates are highlighted and the HPD interval of the sampling date is shown in parentheses. Nodes with posterior support of at least 90% are highlighted.

4 BEAST2 Bayesian DRT

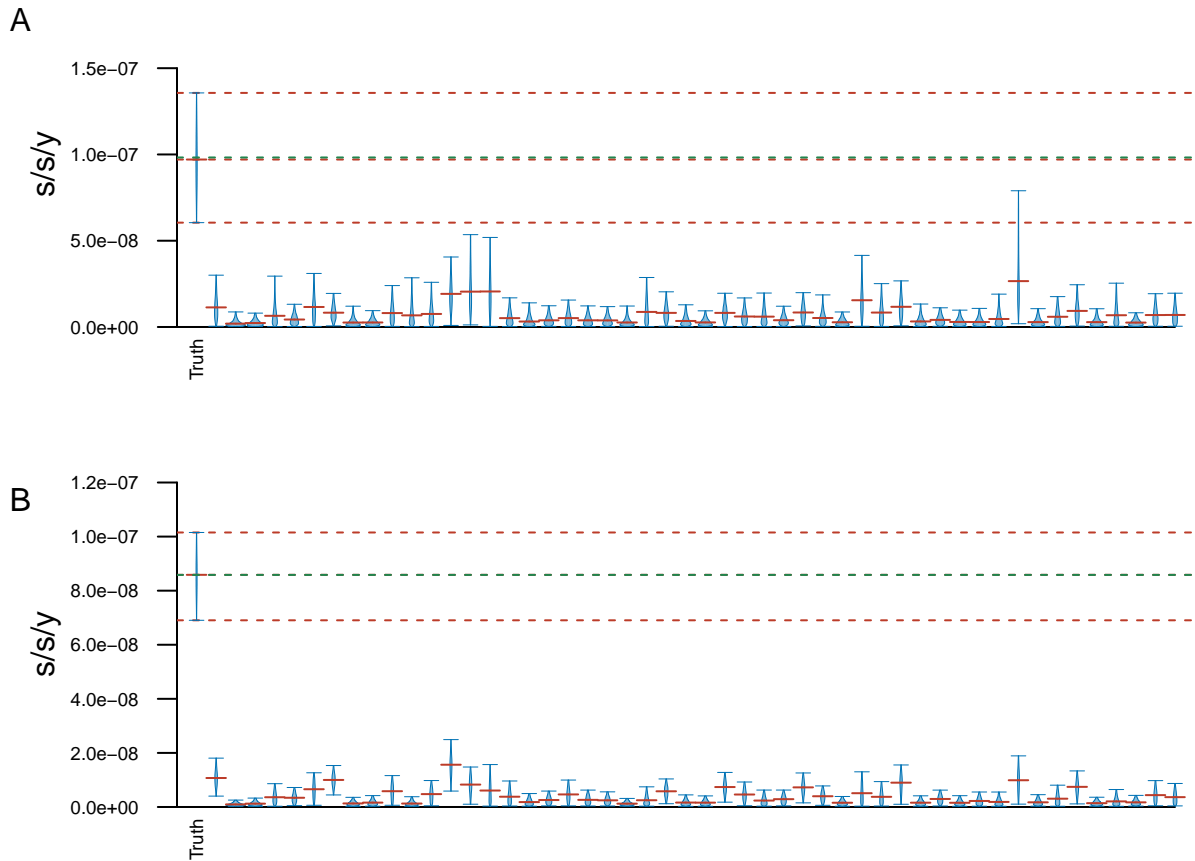


Figure 10: Date shuffling analyses performed under a **Bayesian skyline plot** coalescent tree prior and **relaxed clock** (A) and **strict clock** (B) models. The plot shows the posterior distributions for the (mean) clock rate, truncated at the upper and lower limits of the 95% HPD interval. Horizontal red lines indicate the medians of the posterior distributions. The red dashed lines indicate the median and upper and lower limits of the 95% HPD interval of the clock rate inferred under the true sampling dates.

5 Session info

```
## R version 3.5.1 (2018-07-02)
## Platform: x86_64-apple-darwin15.6.0 (64-bit)
## Running under: macOS Sierra 10.12.6
##
## Matrix products: default
## BLAS: /Library/Frameworks/R.framework/Versions/3.5/Resources/lib/libRblas.0.dylib
## LAPACK: /Library/Frameworks/R.framework/Versions/3.5/Resources/lib/libRlapack.dylib
##
## locale:
## [1] en_GB.UTF-8/en_GB.UTF-8/en_GB.UTF-8/C/en_GB.UTF-8/en_GB.UTF-8
```

```

##
## attached base packages:
## [1] stats      graphics  grDevices  utils      datasets  methods   base
##
## other attached packages:
## [1] treedater_0.5.0 limSolve_1.5.6 phangorn_2.5.5 phytools_0.6-99
## [5] maps_3.3.0      ape_5.3          ggpubr_0.2.4   magrittr_1.5
## [9] ggsci_2.9       ggplot2_3.2.1    ggtree_2.1.1   treeio_1.11.2
## [13] beastio_0.2.5   coda_0.19-3      cowplot_1.0.0
##
## loaded via a namespace (and not attached):
## [1] Rcpp_1.0.3          lattice_0.20-38     tidyr_1.0.0
## [4] gtools_3.8.1        assertthat_0.2.1    zeallot_0.1.0
## [7] digest_0.6.23       R6_2.4.1            backports_1.1.5
## [10] evaluate_0.14        highr_0.8           pillar_1.4.2
## [13] rlang_0.4.2          lazyeval_0.2.2      Matrix_1.2-18
## [16] combinat_0.0-8       rmarkdown_2.3        labeling_0.3
## [19] stringr_1.4.0        igraph_1.2.4.2       munsell_0.5.0
## [22] compiler_3.5.1       numDeriv_2016.8-1.1 xfun_0.15
## [25] pkgconfig_2.0.3      mnormt_1.5-5         htmltools_0.4.0
## [28] tidyselect_0.2.5     lpSolve_5.6.13.3     tibble_2.1.3
## [31] expm_0.999-4         codetools_0.2-16     quadprog_1.5-8
## [34] crayon_1.3.4         dplyr_0.8.3          withr_2.1.2
## [37] MASS_7.3-51.4        grid_3.5.1           nlme_3.1-143
## [40] jsonlite_1.6         gtable_0.3.0         lifecycle_0.1.0
## [43] scales_1.1.0         tidytree_0.3.1       stringi_1.4.3
## [46] farver_2.0.1         ggsignif_0.6.0       scatterplot3d_0.3-41
## [49] rvcheck_0.1.7        vctrs_0.2.1          fastmatch_1.1-0
## [52] RColorBrewer_1.1-2   tools_3.5.1          glue_1.3.1
## [55] purrr_0.3.3          plotrix_3.7-7         parallel_3.5.1
## [58] yaml_2.2.0           colorspace_1.4-1     BiocManager_1.30.10
## [61] animation_2.6        clusterGeneration_1.3.4 knitr_1.29

```