Supplementary Information

Phylodynamic signatures in the emergence of community-associated MRSA

Eike Steinig, Izzard Aglua, Sebastian Duchene, Michael T. Meehan, Mition Yoannes, Cadhla Firth, Jan Jaworski, Jimmy Drekore, Bohu Urakoko, Harry Poka, Clive Wurr, Eri Ebos, David Nangen, Elke Müller, Peter Mulvey, Charlene Jackson, Anita Blomfeldt, Hege Vangstein Aamot, Moses Laman, Laurens Manning, Megan Earls, David C. Coleman, Andrew Greenhill, Rebecca Ford, Marc Stegger, Muhammad Ali Syed, Bushra Jamil, Stefan Monecke, Ralf Ehricht, Simon Smith, William Pomat, Paul Horwood, Steven Y.C. Tong, and Emma McBryde

Eike Steinig
Email: eike.steinig@unimelb.edu.au

This PDF file includes:

Other supporting materials for this manuscript include the following:

Datasets S1 to S5

Supporting Information Text

Distribution of ST93 in Papua New Guinea. We used whole-genome sequencing data to reconstruct the evolutionary history of the first *S. aureus* genomes from Papua New Guinea, and discovered that a pediatric osteomyelitis outbreak in the remote Highlands Provinces (1) was caused by the Australian lineage ST93-MRSA-IV. Multiple lines of evidence support a wider distribution of ST93 in PNG, including two discernible introductions, sustained and long-term transmission of the outbreak since the early 2000s, occurrence of two MLST allele variants, and a heterogeneous pattern of dissemination in the remote highland towns (**Figs. 1, 2**). It remains unclear to what extent ST93-MRSA-IV has

disseminated in PNG. Our data further show that ST93-MRSA-IV is widespread in Far North Queensland and is likely the cause for the increasing rates of MRSA observed in FNQ communities over the last decade (2, 3). It is unclear what is driving the local persistence and evolution of the ST93-MRSA-IV genotype in PNG, and reservoirs in the community remain to be investigated. Data on antibiotic consumption in Simbu Province or Eastern Highland Province was not available. While antibiotic stewardship may play a role in the dissemination of ST93 in FNQ (2), and ST772 in the Islamabad-Rawalpindi metropolitan area (4), sustained circulation of virulent and transmissible clones in remote settings like PNG may also have been a result of historical transmission opportunities from the Australian East Coast after the emergence of ST93-MRSA-IV, as well as existing strain diversity and competitive interactions in the highlands, even in the absence of widespread antimicrobial consumption.

Non-synonymous mutations at the divergence of drug-resistant clades. Non-synonymous mutations in factors associated with immune response and skin colonization at the divergence of epidemic MRSA and MSSA have been detected previously in ST772 (5), and ACME and COMER elements in the USA300 clades are implicated in transmission and persistence phenotypes (6–8), but it is unclear to what degree these changes have contributed to the emergence, transmission, persistence, or fitness of resistant strains in the presence of other strains or genotypes. Given that acquisition of antibiotic resistance determinants and recruitment into population centers coincides with rapid increases in R_e across all lineages examined in this study, these factors are not likely to explain the rapid change in transmission dynamics we estimated at the divergence of resistant clades (**Fig. S8**). However, mutations may contribute to ongoing persistence in host populations before and after emergence, or constitute pre-adaptations that support successful transmission in new host populations. Canonical mutations have previously been detected at the divergence of ST772-A and have been associated with colonization ability, which may play a role in compensating for the fitness cost induced by resistance acquisition. This is supported by

preliminary phenotypic data from this lineage which suggests that there was no significant difference in biofilm formation or growth rate between MSSA and multidrug-resistant MRSA strains (5). Further rigorous experiments will need to be conducted to better understand the significance of these mutations in strains preceding the emergence of resistant clades and their interaction with resistance phenotypes.

Additional model limitations and runtime improvements. Birth-death skyline models assume that populations are well-mixed, but clear population structure is evident between the MSSA and MRSA strains as well as in other monophyletic clades, such as the introductions of ST93 into PNG, FNQ and NZ. While we attempted to employ the more parameter-rich multitype birth-death models that are inherently capable of accounting for structured populations (9), the MCMC chains ultimately did not converge. This was likely due to a combination of large bacterial genome data sets and the parameter-rich model. We therefore employed the birth-death skyline model on monophyletic subsets of the lineage-wide variant alignment, provided sufficient isolates were available (**Fig. 2B, Fig. S2**), thus reducing the potential impact on lineage-wide estimates arising from excessive population structure in the tree. We further explored a range of realistic configurations on the becoming uninfectious rate prior (1 - 10 years infectious period), for which few data are available from long-term surveillance studies, as well as weighting the reproduction number prior distribution at different levels of transmission (Methods). Prior sensitivity analysis for each lineage and clade confirms that estimates were largely driven by the available data, rather than by the prior configurations (**Figs. S7, S8**). Further improvements to enable MCMC convergence for phylodynamic models used in large bacterial populations, including multitype birth-death models supporting larger sample sizes with numeric stability and Metropolis-coupled MCMC chains (10) will be required for further investigations into the origins of community-associated *S. aureus*.

Supplementary figures

Figure S1.1 - S1.7: Lineage specific maximum-likelihood phylodynamics using Treetime: root-to-tip regression on sampling dates showing that all lineages could be considered 'measurably evolving'; in the middle the results of a date-randomisation test with 100 replicates indicating that all estimated clock rates (red line) are distinct from a distribution of rates estimated after randomizing dates across the phylogenetic tree (gray distribution); on the right is the coalescent skyline estimate of changes in the effective population size (N_e) over time.

Figure S2.1 - S2.7: Coalescent Bayesian skyline plots of changes in the effective population size (N_e) of *Staphylococcus aureus* lineages at different dimensional configurations (d = 4 and d = 8, skyline plots) with posterior distributions of dimensional configurations $d \in \{2, 4, 8, 16\}$ of MRCA (purple) and clock rate (blue) above for reference

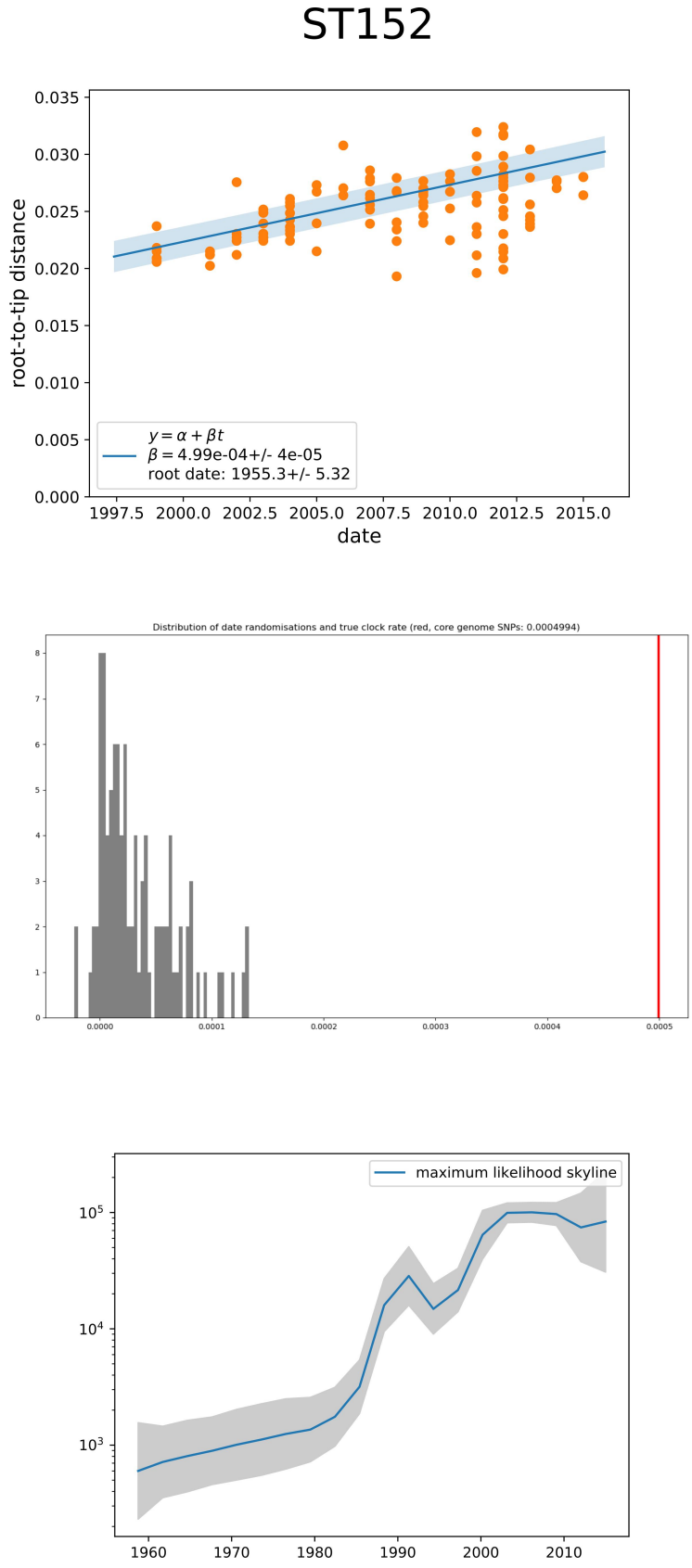

Figure S1.1: Root-to-tip regression, date randomisation and coalescent skyline estimate of the effective population size for ST152.

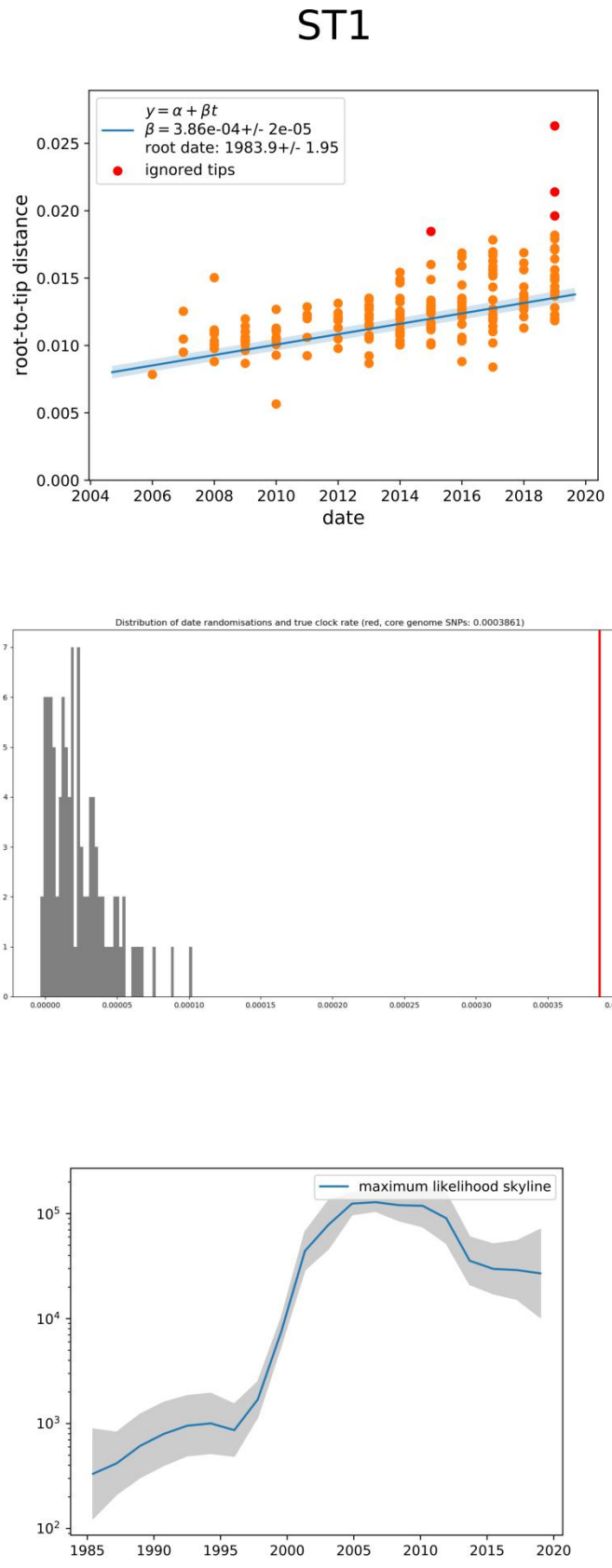

Figure S1.2: Root-to-tip regression, date randomisation and coalescent skyline estimate of the effective population size for ST1.

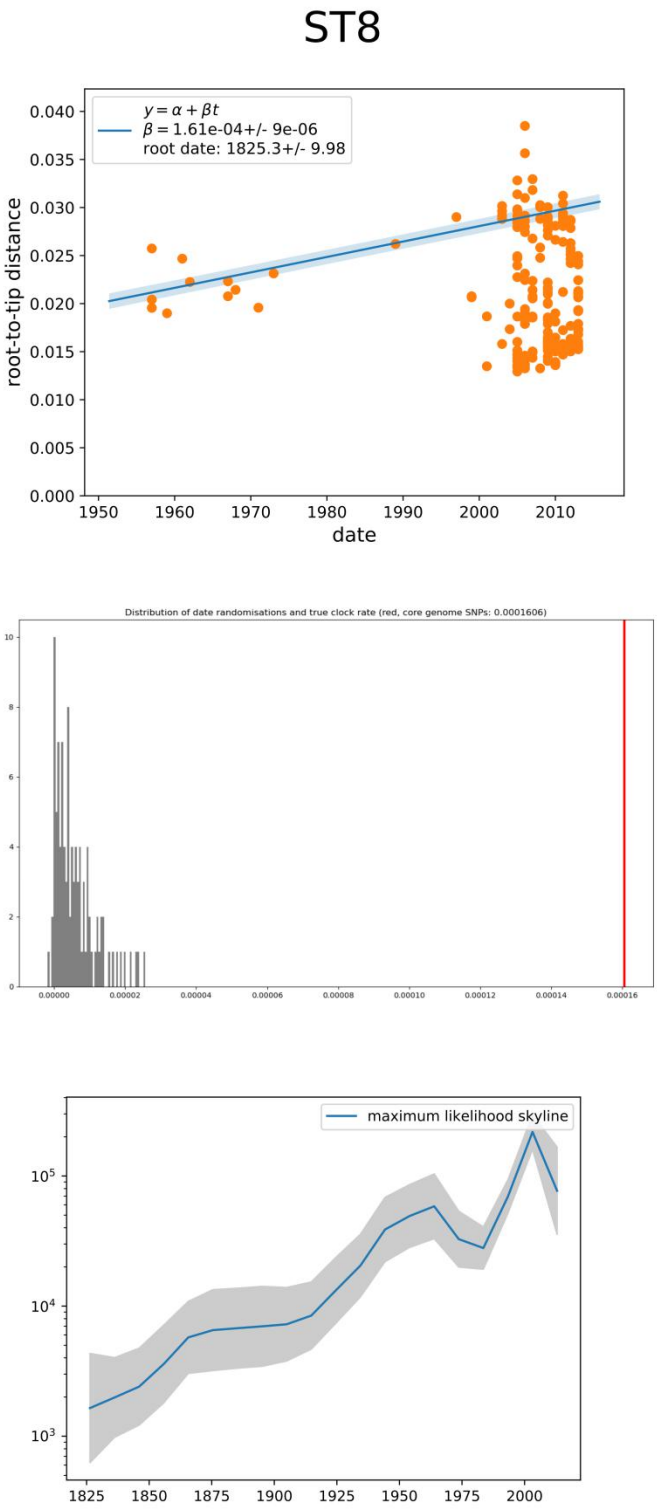

Figure S1.3: Root-to-tip regression, date randomisation and coalescent skyline estimate of the effective population size for ST8.

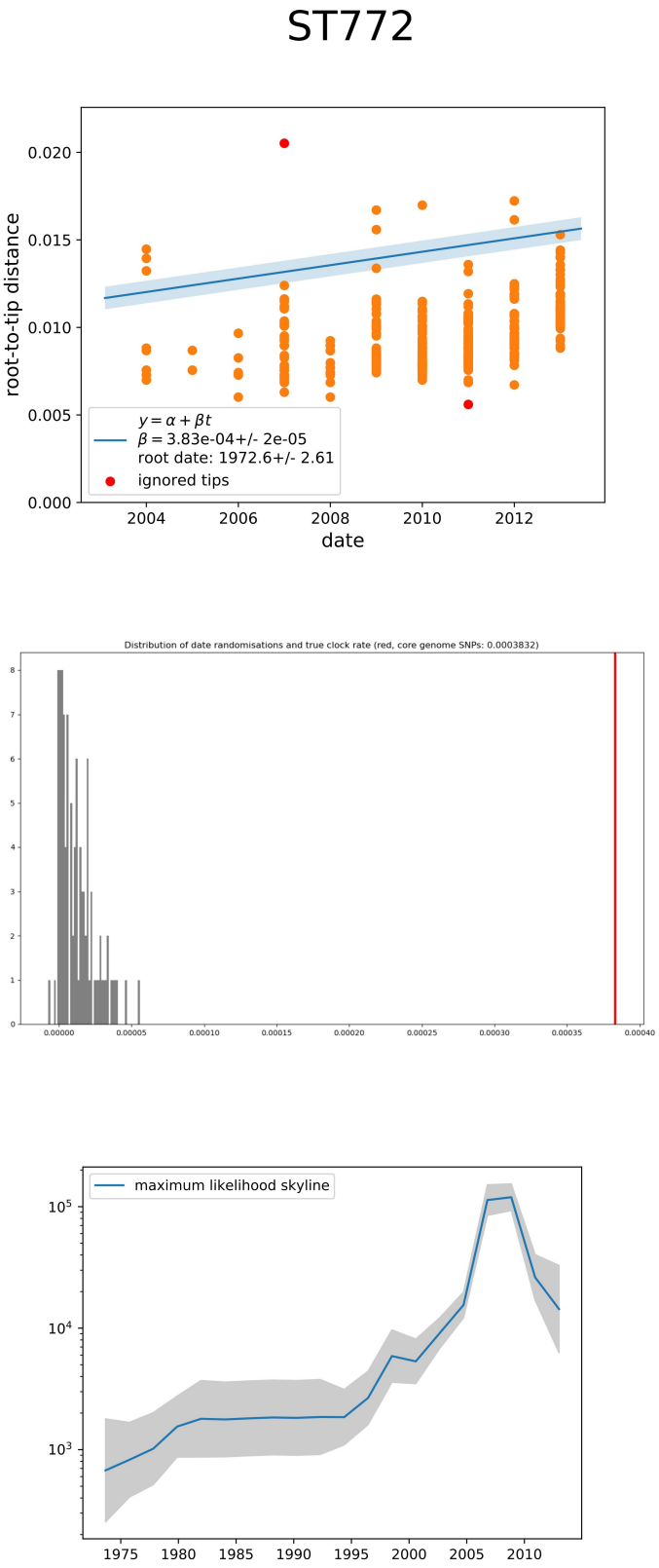

Figure S1.4: Root-to-tip regression, date randomisation and coalescent skyline estimate of the effective population size for ST772.

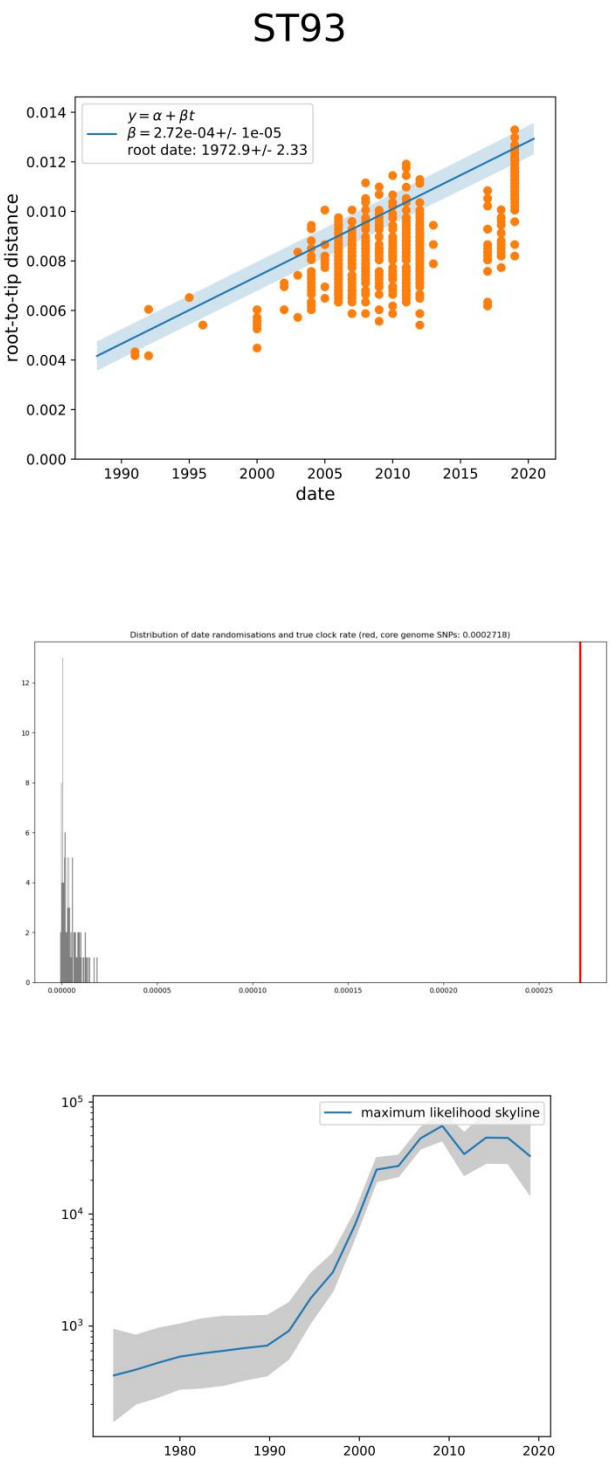

Figure S1.5: Root-to-tip regression, date randomisation and coalescent skyline estimate of the effective population size for ST93.

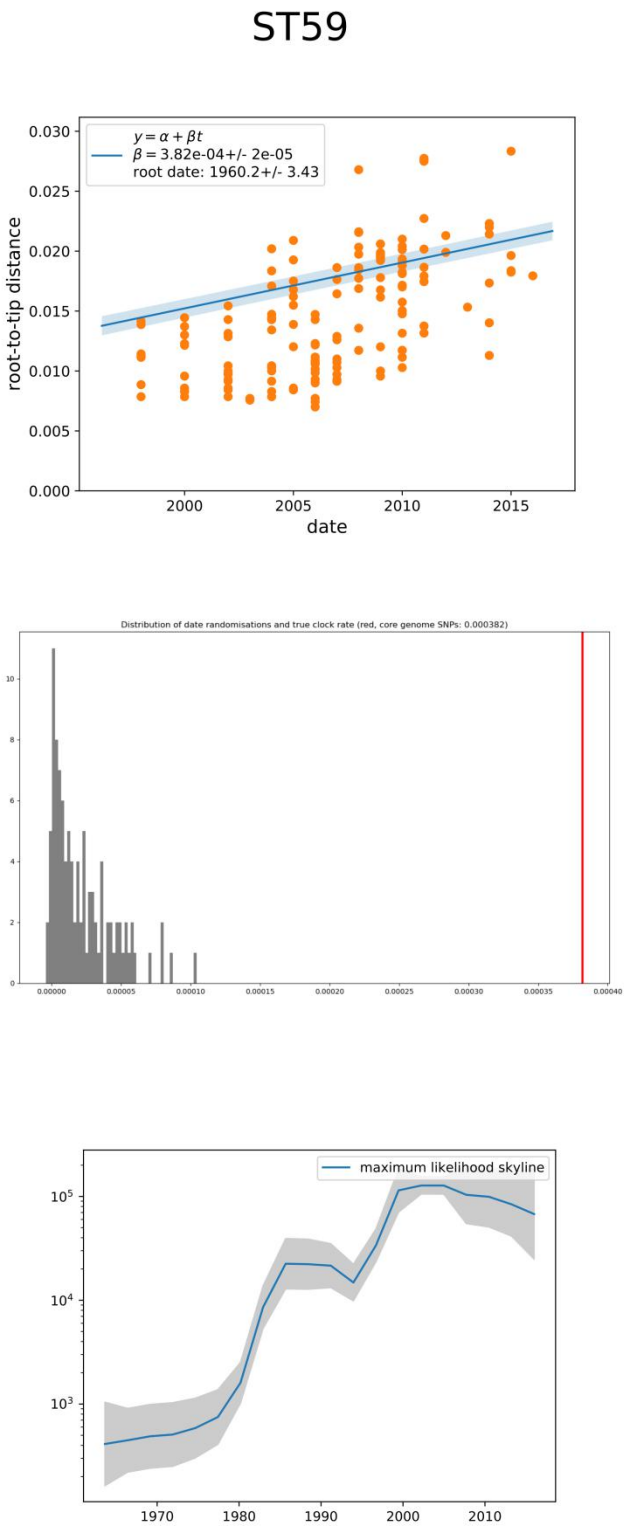

Figure S1.6: Root-to-tip regression, date randomisation and coalescent skyline estimate of the effective population size for ST59.

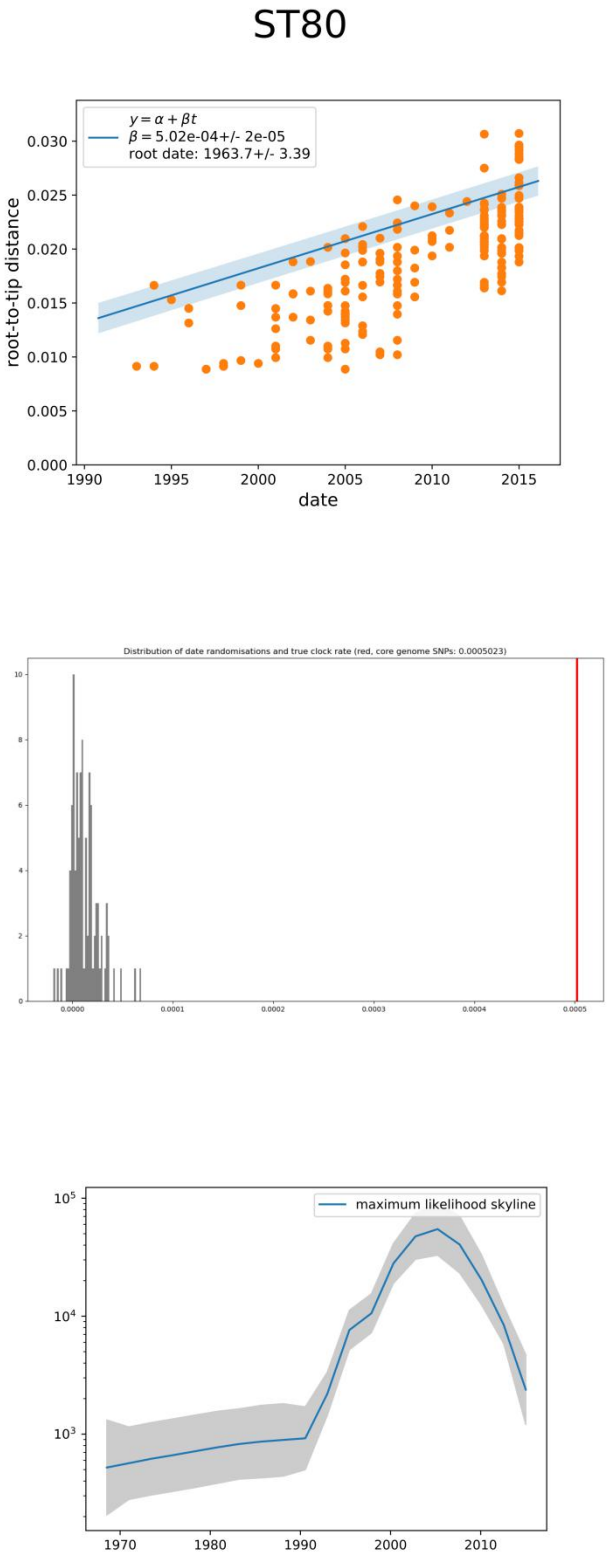

Figure S1.7: Root-to-tip regression, date randomisation and coalescent skyline estimate of the effective population size for ST80.

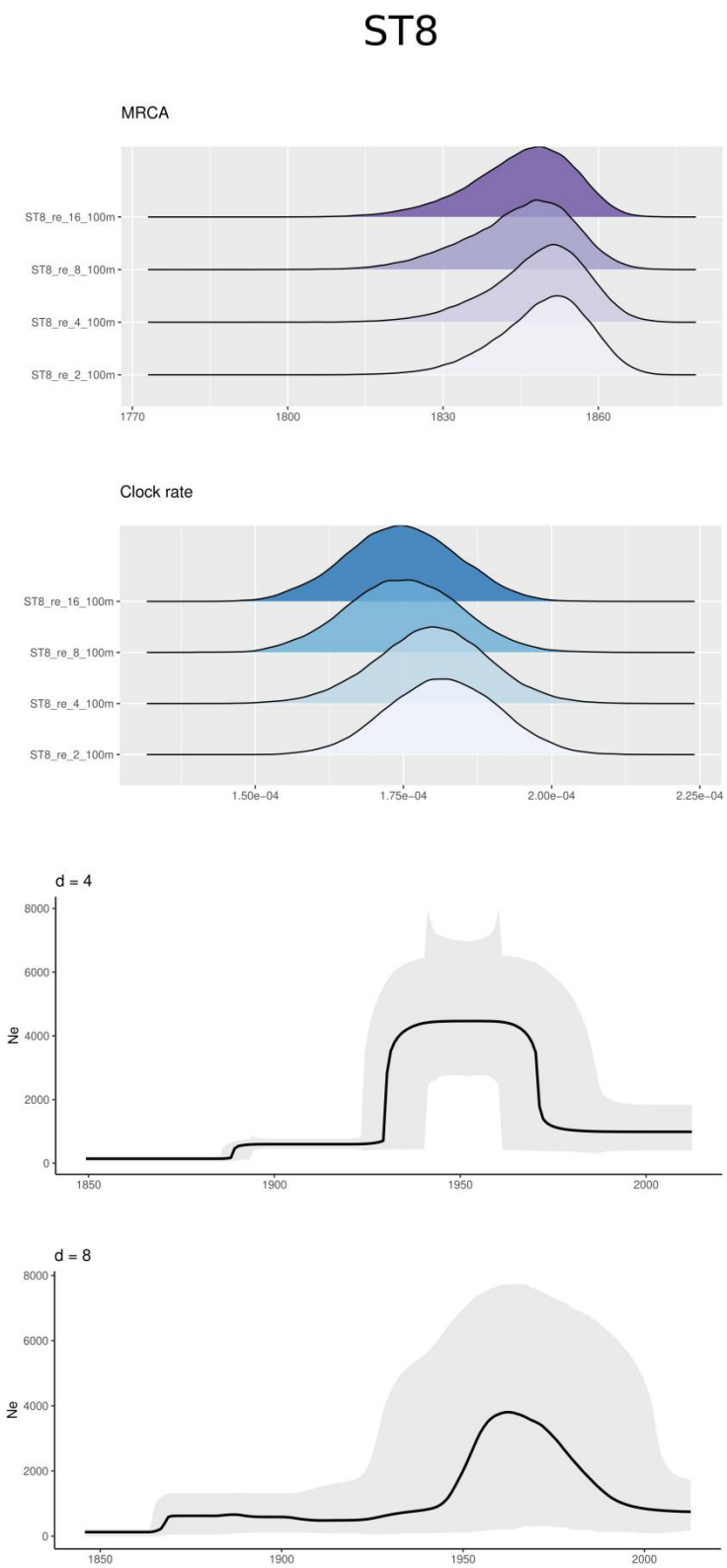

Figure S2.1: Posterior estimates and coalescent Bayesian skyline plots of changes in the effective population size at different dimensional configurations of R_e for ST8.

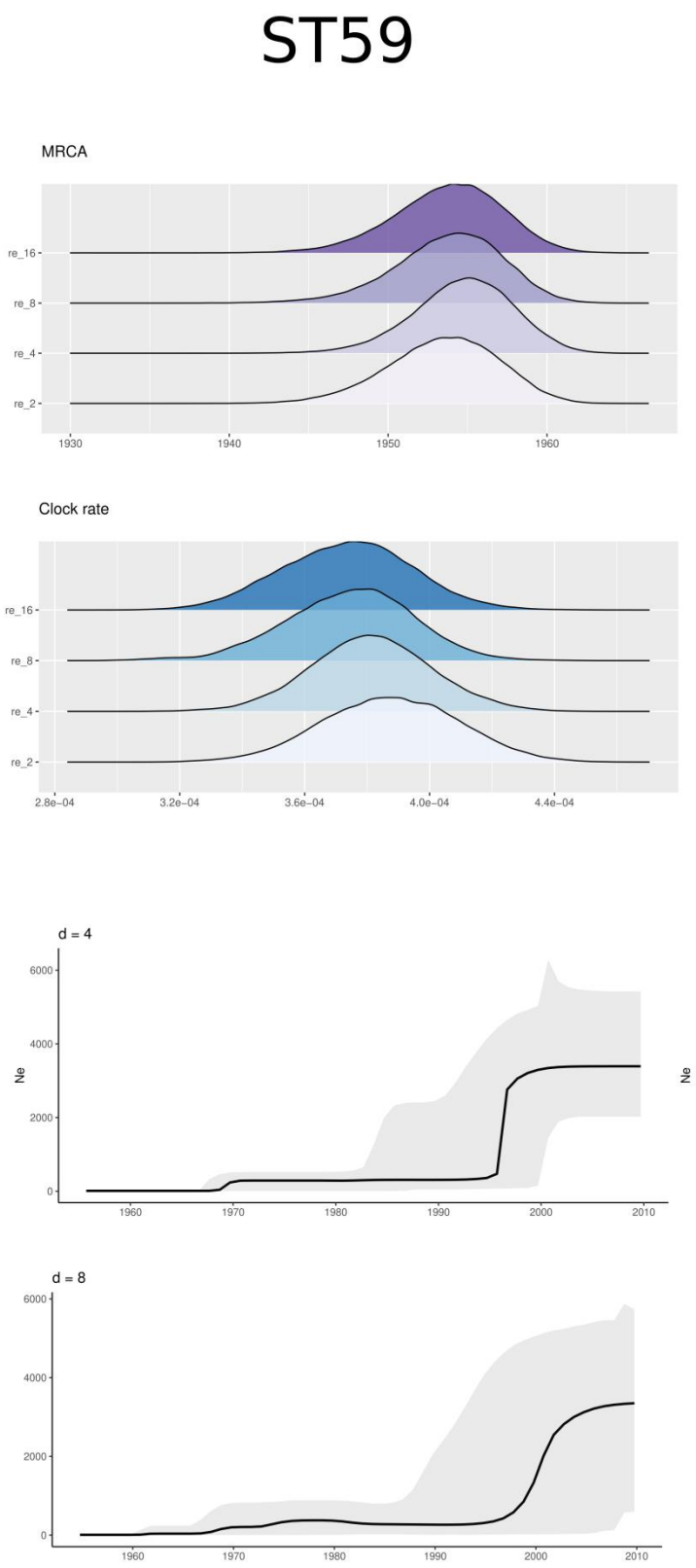

Figure S2.2: Posterior estimates and coalescent Bayesian skyline plots of changes in the effective population size at different dimensional configurations of R_e for ST59.

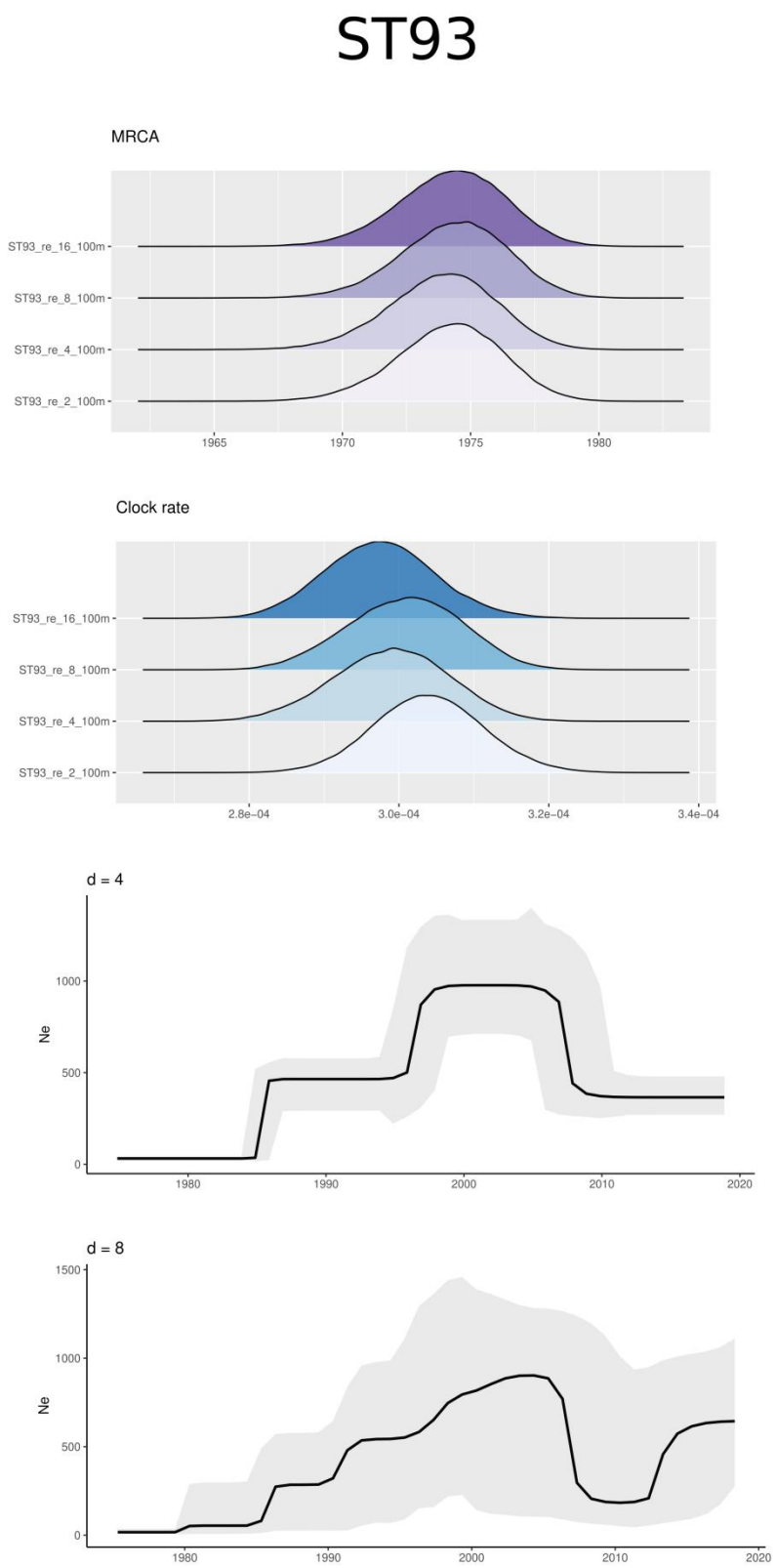

Figure S2.3: Posterior estimates and coalescent Bayesian skyline plots of changes in the effective population size at different dimensional configurations of R_e for ST93.

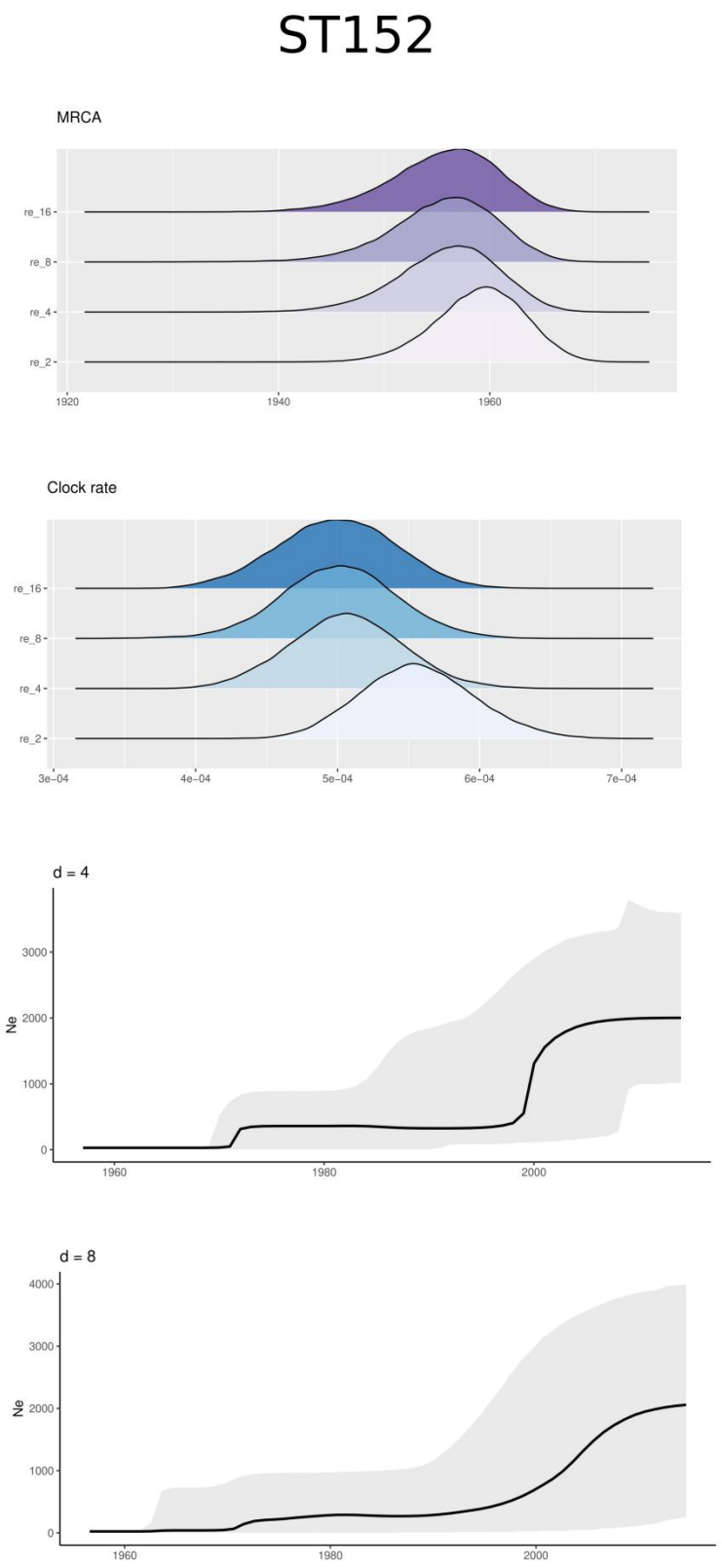

Figure S2.4: Posterior estimates and coalescent Bayesian skyline plots of changes in the effective population size at different dimensional configurations of R_e for ST152.

ST1

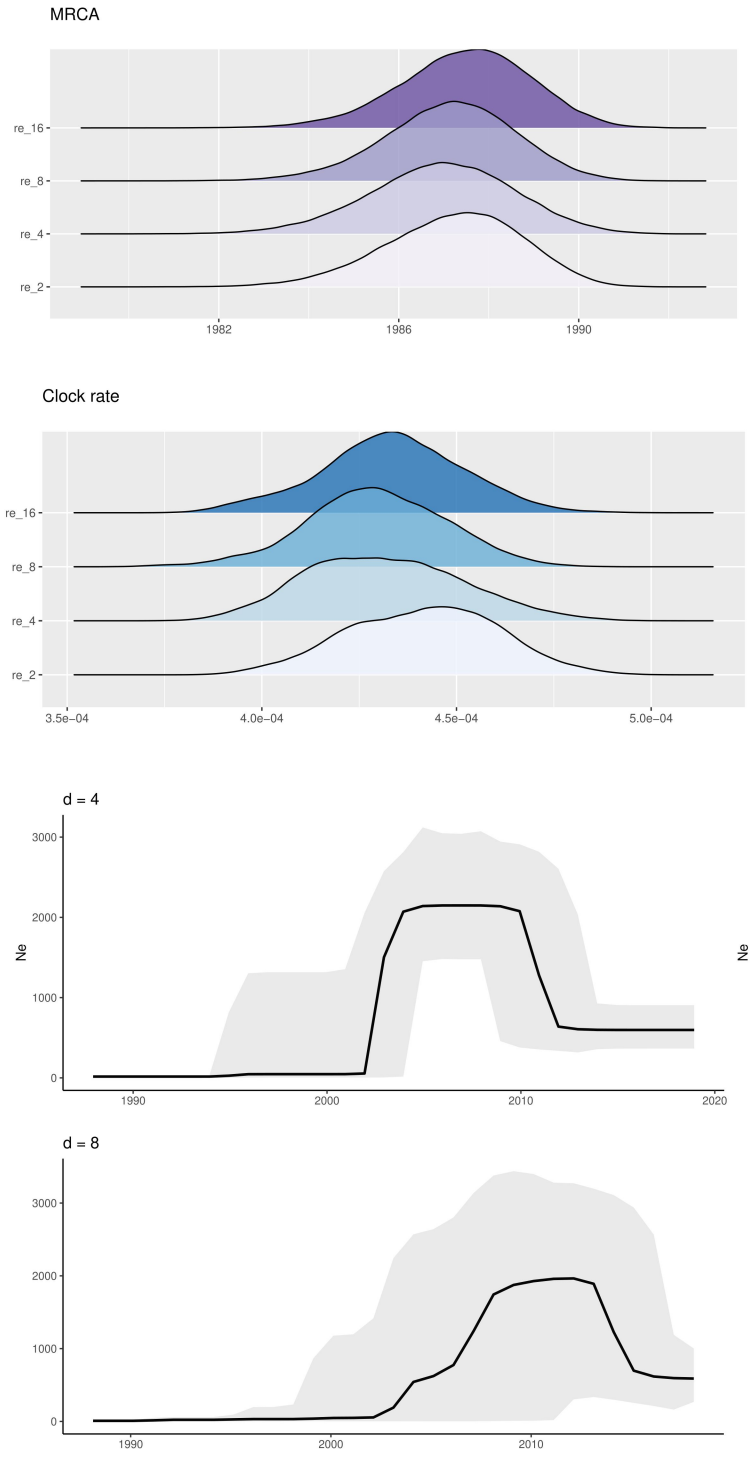

Figure S2.5: Posterior estimates and coalescent Bayesian skyline plots of changes in the effective population size at different dimensional configurations of R_e for ST1.

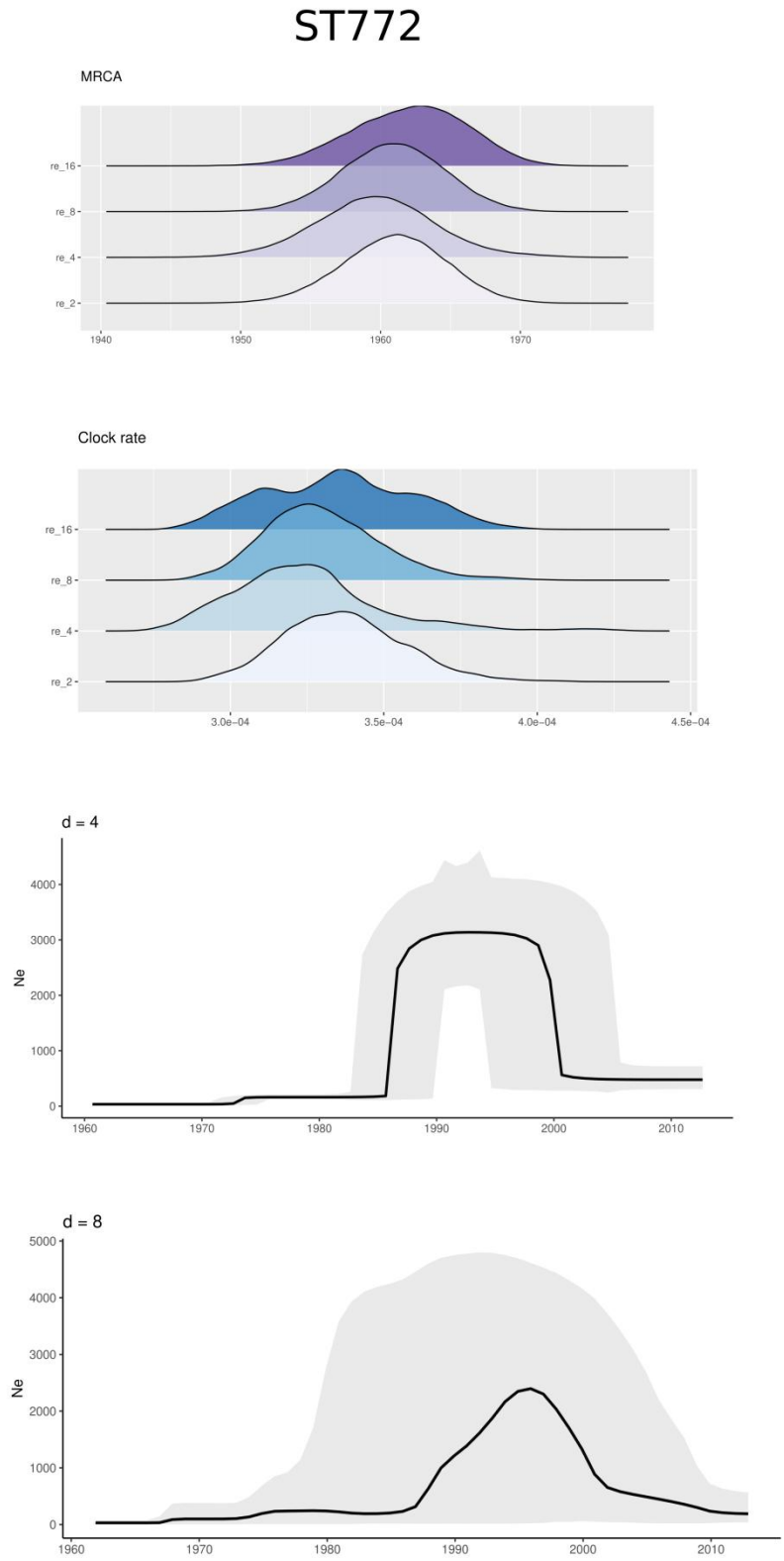

Figure S2.6: Posterior estimates and coalescent Bayesian skyline plots of changes in the effective population size at different dimensional configurations of R_e for ST772.

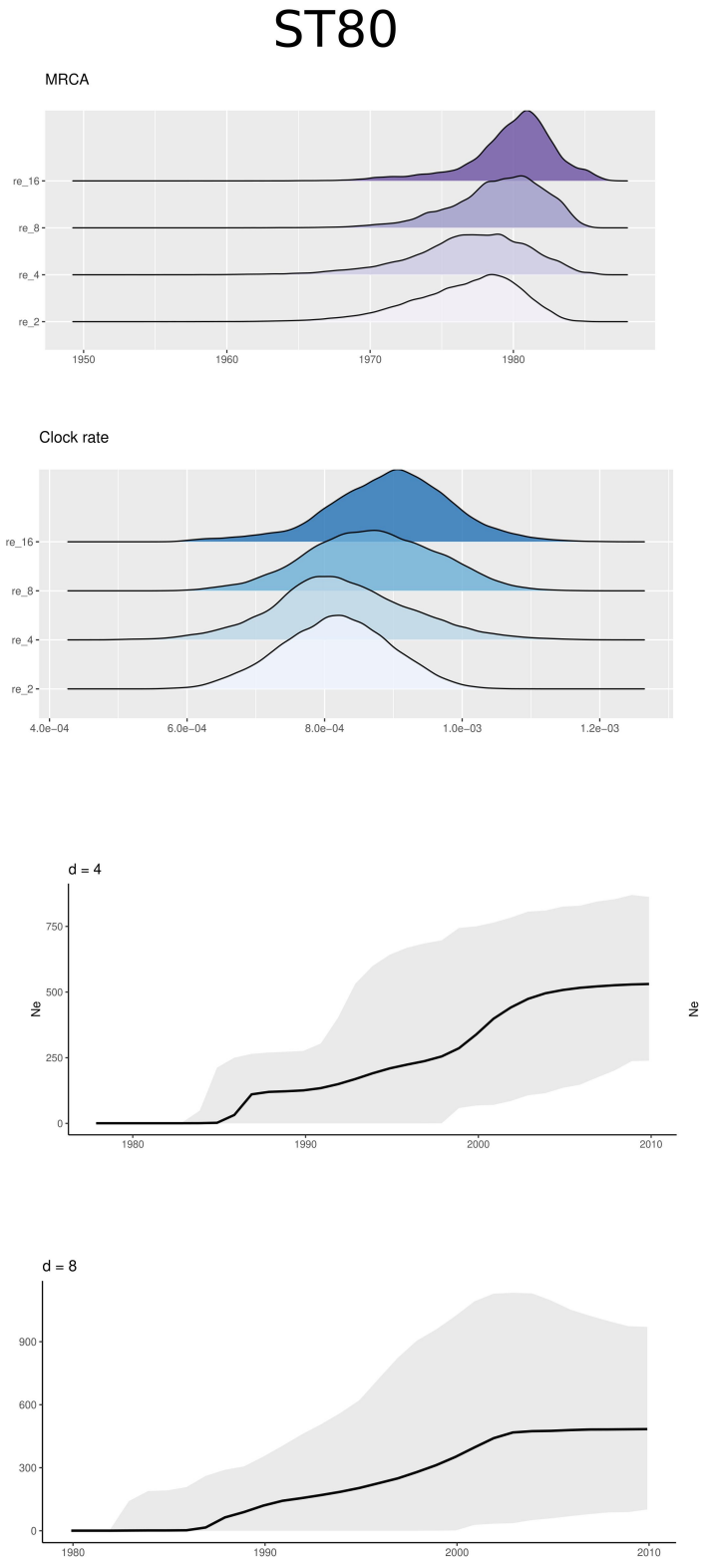

Figure S2.7: Posterior estimates and coalescent Bayesian skyline plots of changes in the effective population size at different dimensional configurations of R_e for ST80.

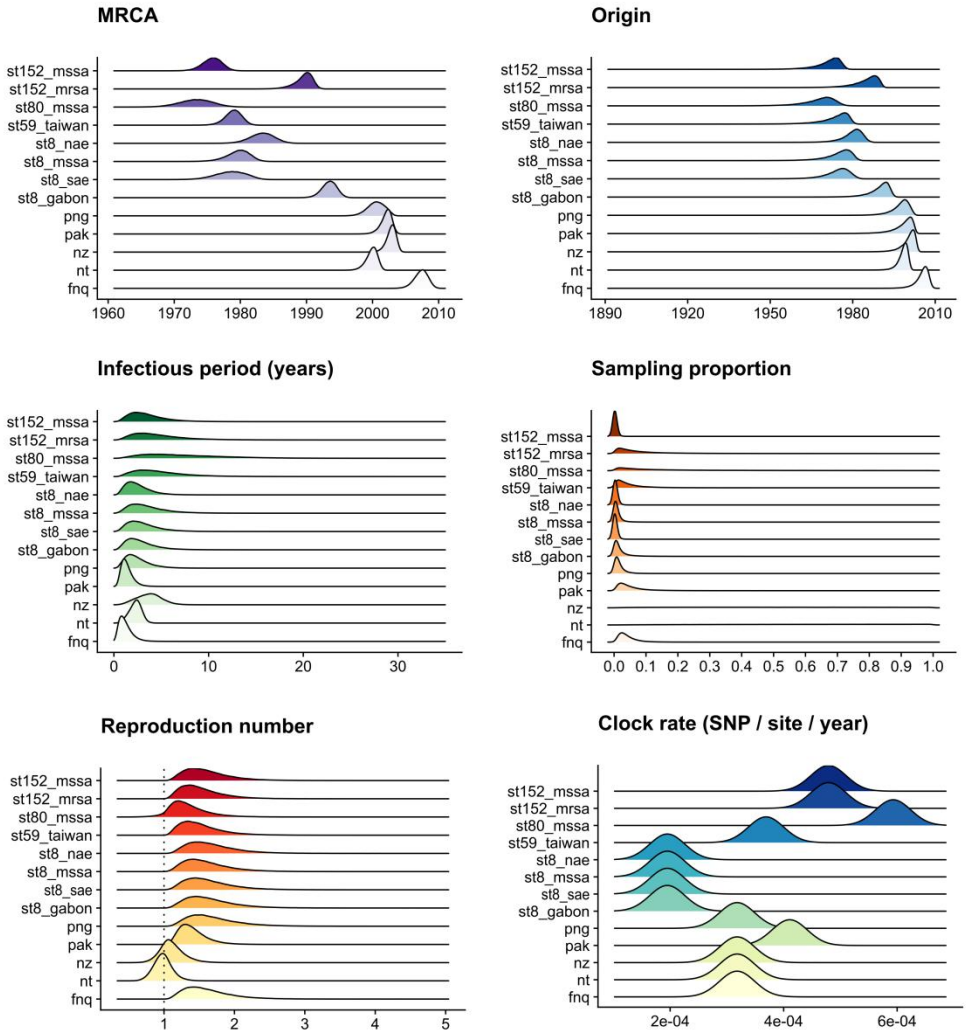

Figure S3: Birth-death skyline posteriors of outbreaks and sublineages in community-associated *Staphylococcus aureus*. Clock rates are fixed to the lineage-wide median higher posterior density interval (Table 1).

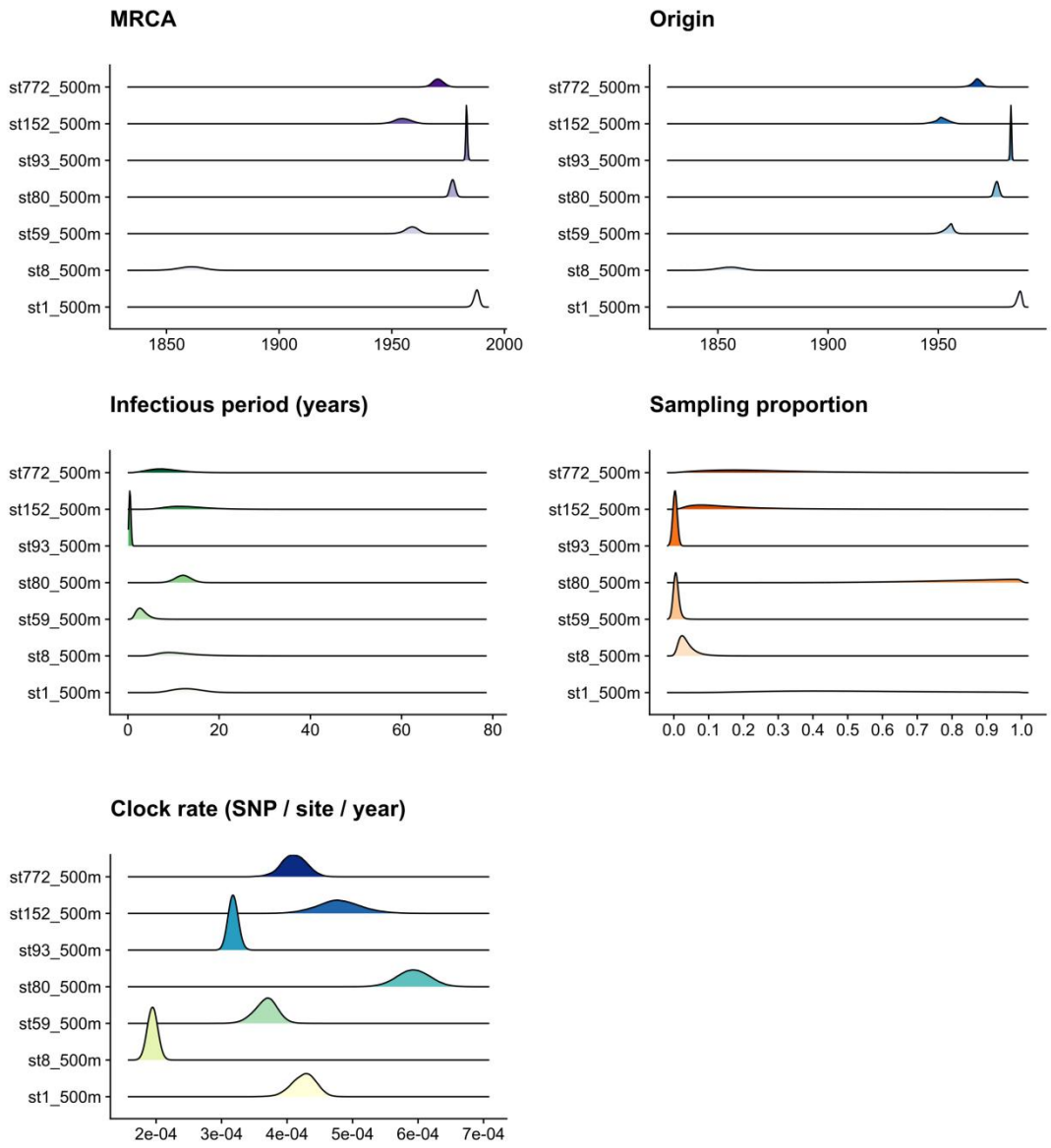

Figure S4: Birth-death skyline posteriors of community-associated *Staphylococcus aureus* lineages; clock rates are estimated, the sampling proportion (δ) prior is sliced into pre-sampling and sampling intervals, and the reproduction number (R_e) prior varies across equally distant slices over the phylogeny (Fig. 4, Table 1, models were run for 500 million iterations of the MCMC)

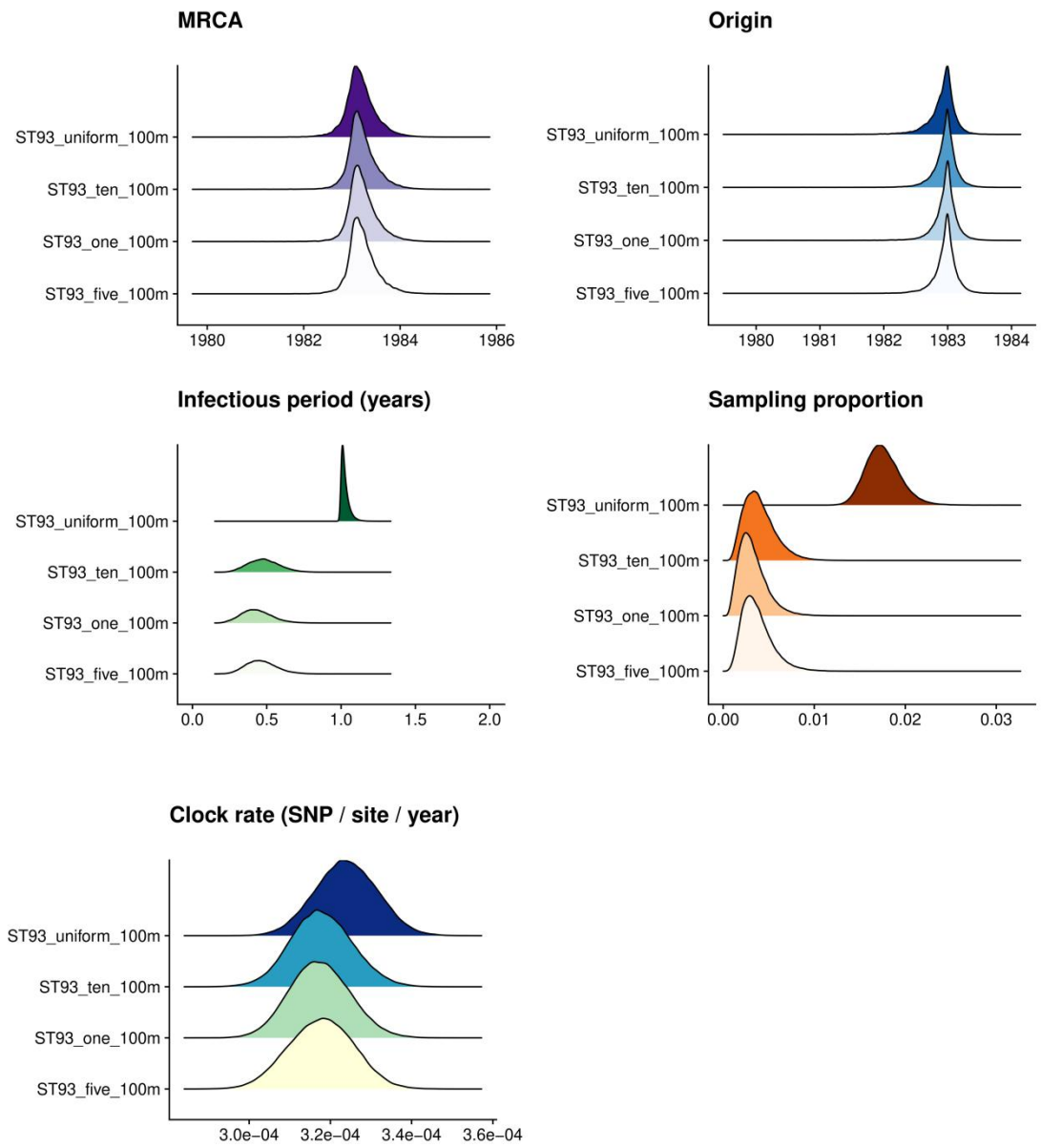

Figure S5: Posterior estimates of prior exploration in ST93 showing on the ridges the become uninfectious rate posteriors, where the prior was configured as a flat Beta (B) distribution, or with Lognormal(μ, 1.0) where μ = 0.1 (10 years infectious period), μ = 0.2 (5 years) and μ = 1.0 (1 year).

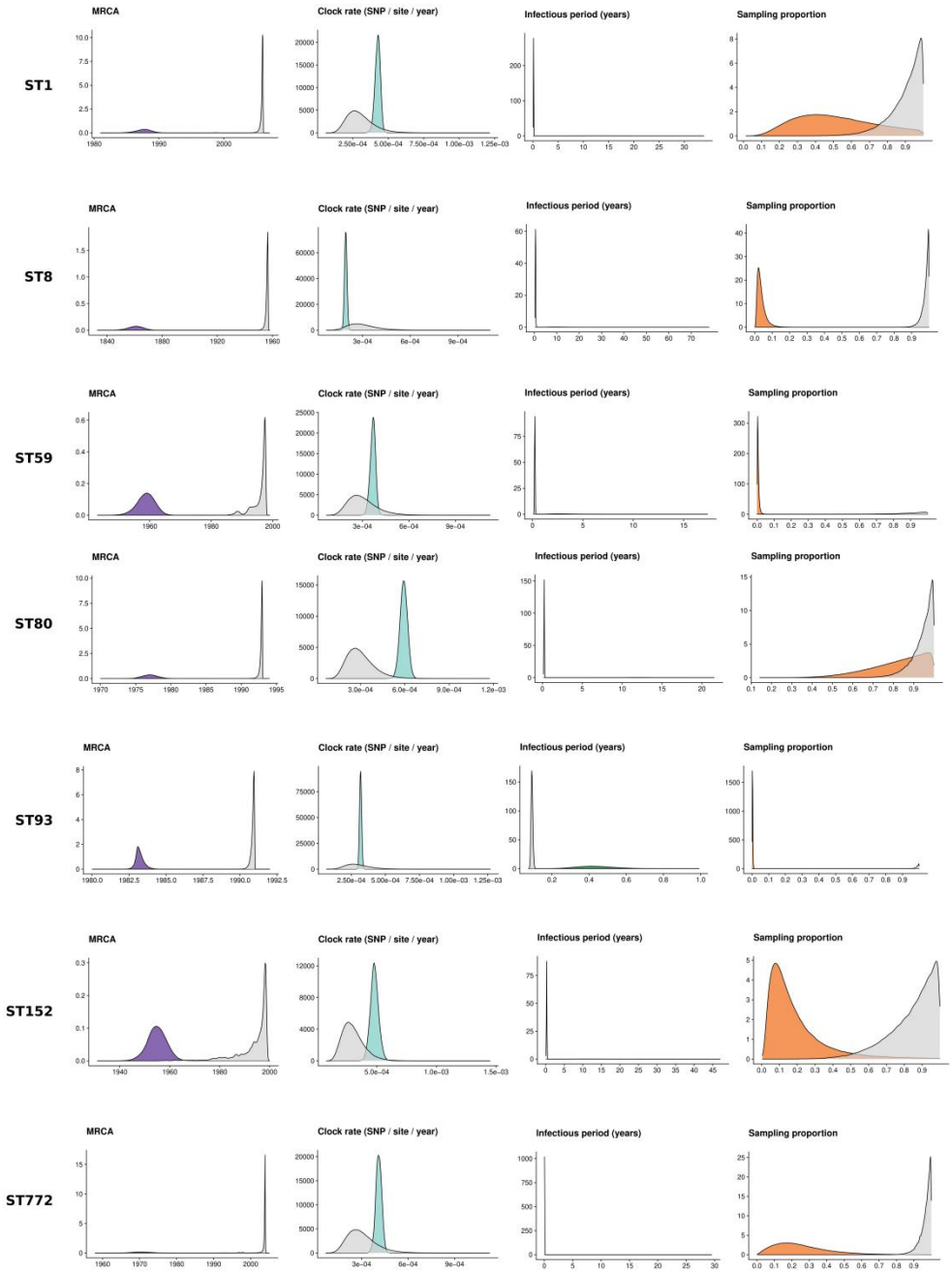

Figure S6: Prior sensitivity analysis of the birth-death skyline model for main sequence types of *Staphylococcus aureus* (Table 1), showing the posterior distributions of the MRCA, clock rate, infectious period ($\frac{1}{\delta}$) and sampling proportion (ρ) (colors), and the posterior distributions 'under-the-prior' (gray, without including the sequence alignment; necessarily including dates that can still inform the posterior estimates).

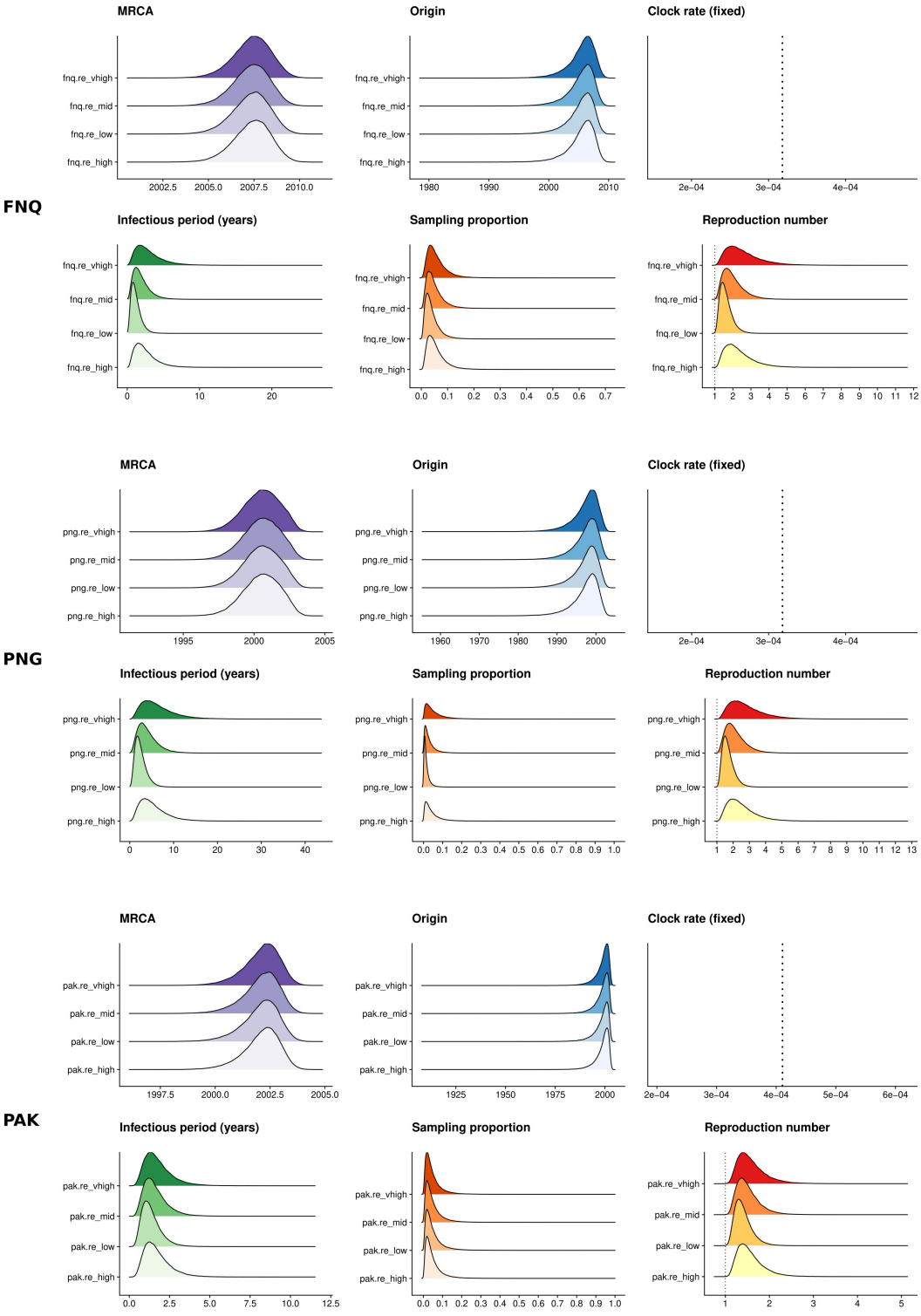

Figure S7: Birth-death skyline posteriors of three outbreaks sequenced in this study (ST93-MRSA-FNQ, ST93-MRSA-PNG, ST772-MRSA-PAK) across four configurations of the R_e prior with Gamma(2.0, θ) where labels in the plots correspond to θ = 0.5 (re_low), θ = 1.0 (re_mid), θ = 1.5 (re_high), θ = 2.0 (re_vhigh).

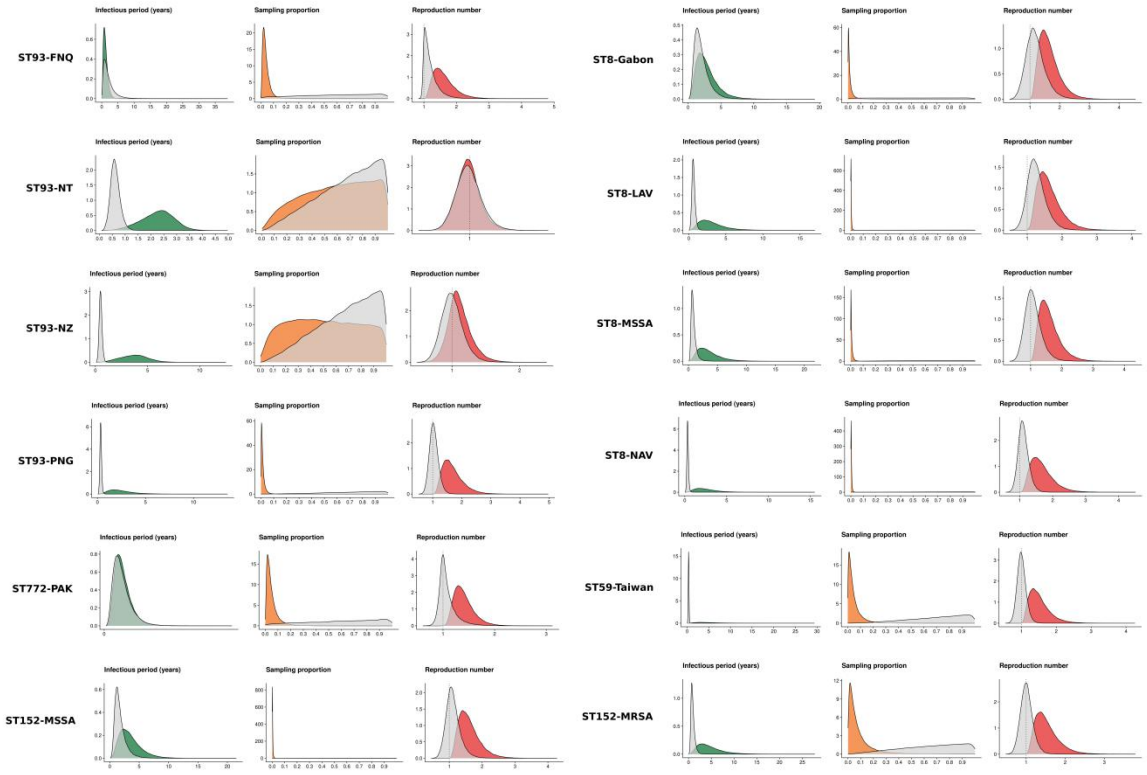

Figure S8: Prior sensitivity analysis of the birth-death skyline model for sublineages of *Staphylococcus aureus* (Table 1), showing the posterior distributions of the infectious period ($\frac{1}{\delta}$), sampling proportion (ρ) and reproduction number (R_e) (colors) and the posterior distributions 'under-the-prior' (gray, without including the sequence alignment; necessarily including dates that can still inform the posterior estimates).

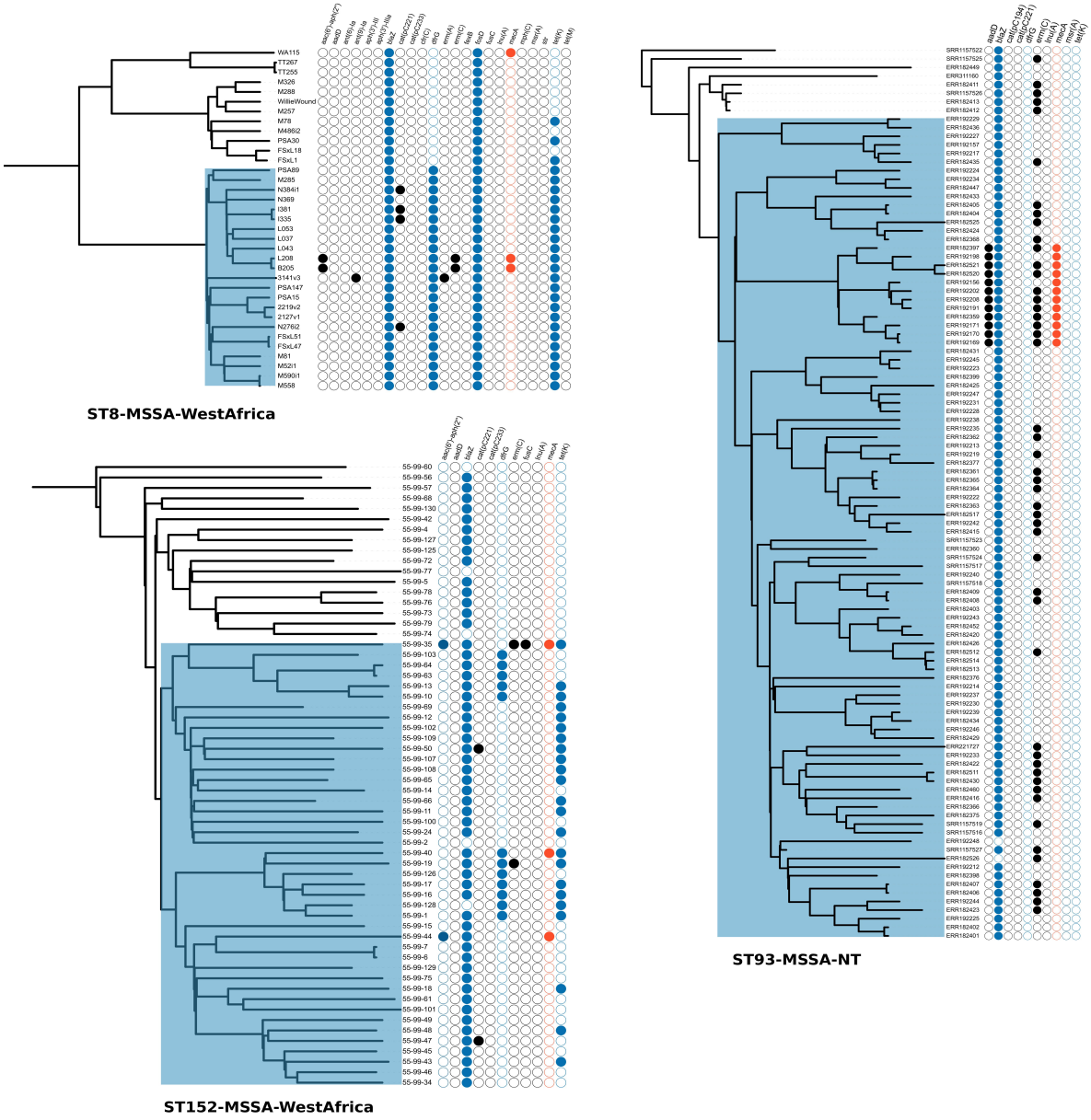

Figure S9: Acquired antimicrobial resistance determinants in community-associated methicillin-sensitive *Staphylococcus aureus* (MSSA) clades (highlighted in blue). Colors of circles indicate common (blue), sporadic (black) and methicillin-resistance (red) determinants.

Table S1: Birth-death skyline median posterior estimates for global community-associated *Staphylococcus aureus* lineages and sublineages (continued from Table 1)

Sequence types	***T***	**95% CI**	$\frac{1}{\delta}$	**95% CI**	**ρ**	**95% CI**
ST93	1982	1982 - 1983	0.47	0.24 - 0.63	0.003	0.0008 - 0.006
ST772	1967	1962 - 1972	7.81	1.89 - 15.02	0.22	0.013 - 0.56
ST80	1976	1974 - 1978	12.07	8.57 - 15.51	0.849	0.56 - 0.99
ST8	1855	1844 - 1865	10.61	4.05 - 21.33	0.308	0.004 - 0.86
ST1	1986	1984 - 1988	12.93	7.21 - 19.14	0.49	0.16 - 0.94
ST59	1955	1949 - 1958	2.93	0.95 - 5.72	0.005	0.0002 - 0.018
ST152	1951	1944 - 1957	12.69	5.37 - 22.71	0.14	0.01 - 0.52
MRSA sublineages						
ST93-MRSA-EastCoast	1991	1981 - 1995	2.72	0.504 - 6.27	0.11	0.002 - 0.407
ST93-MRSA-FNQ	2005	2000 - 2008	1.06	0.19 - 2.58	0.03	0.004 - 0.096
ST93-MRSA-PNG	1998	1990 - 2002	2.21	0.49 - 5.08	0.01	0.0001 - 0.045
ST93-MRSA-NT	1998	1994 - 2001	2.29	0.97 - 3.40	0.61	0.15 - 0.99
ST93-MRSA-NZ	2000	1994 - 2000	3.72	1.13 - 6.11	0.51	0.09 - 0.99
ST772-MRSA-Pakistan	1999	1989 - 2003	1.25	0.41 - 2.58	0.03	0.002 - 0.15
ST152-MRSA-Europe	1985	1974 - 1991	4.07	0.65 - 10.34	0.05	0.0003 - 0.296
ST8-USA300-NAE	1980	1972 - 1986	2.26	0.48 - 5.11	0.003	0.000002 - 0.03
ST8-USA300-SAE	1975	1964 - 1982	2.75	0.49 - 6.75	0.001	0.000001 - 0.01
ST8-USA300-Gabon	1990	1981 - 1995	2.52	0.41 - 6.16	0.009	0.00001 - 0.07
ST59-MRSA-Taiwan	1974	1961 - 1980	4.25	0.64 - 10.45	0.03	0.0004 - 0.16
MSSA sublineages						
ST93-MSSA-NT	1986	1974 - 1991	3.01	0.57 - 7.17	0.09	0.0009 - 0.385
ST8-MSSA-WestAfrica	1975	1964 - 1982	3.09	0.52 - 7.49	0.005	0.000002 - 0.03
ST152-MSSA-WestAfrica	1972	1960 - 1977	3.12	0.53 - 7.22	0.001	0.000001 - 0.009

Symbols and abbreviations: CI - confidence interval (higher posterior density), $\frac{1}{\delta}$ - infectious period, ρ - sampling proportion

Supplementary datasets

Dataset S1: Sample identifiers and assembly genotypes for all newly sequenced isolates from Australia, Papua New Guinea and Pakistan.

Dataset S2: Sample identifiers, dates and lineage designations for all genomes used in the lineage phylodynamic models.

Dataset S3: Sample identifiers and sublineage designations for all genomes used in the sublineage phylodynamic models.

Dataset S4: Sample identifiers and assembly genotypes for all genomes used in the lineage phylodynamic models.

Dastaset S5: Mykrobe antimicrobial resistance phenotype predictions for all genomes used in the lineage phylodynamic models.

SI References

1. I. Aglua, *et al.*, Methicillin-resistant *Staphylococcus aureus* in Melanesian children with haematogenous osteomyelitis from the Central Highlands of Papua New Guinea. *Int. J. Pediatr.* **6**, 8361–8370 (2018).
2. I. Guthridge, S. Smith, P. Horne, J. Hanson, Increasing prevalence of methicillin-resistant *Staphylococcus aureus* in remote Australian communities: implications for patients and clinicians. *Pathogen* **51**, 428–431 (2019).
3. T. M. Wozniak, *et al.*, Geospatial epidemiology of *Staphylococcus aureus* in a tropical setting: an enabling digital surveillance platform. *Sci. Rep.* **10**, 13169 (2020).
4. M. A. Syed, *et al.*, Detection and Molecular Characterization of Methicillin-Resistant *Staphylococcus aureus* from Table Eggs in Haripur, Pakistan. *Foodborne Pathog. Dis.* **15**, 86–93 (2018).
5. E. J. Steinig, *et al.*, Evolution and Global Transmission of a Multidrug-Resistant, Community-Associated Methicillin-Resistant *Staphylococcus aureus* Lineage from the Indian Subcontinent. *MBio* **10** (2019).

6. B. A. Diep, *et al.*, The Arginine Catabolic Mobile Element and Staphylococcal Chromosomal Cassette *mec* Linkage: Convergence of Virulence and Resistance in the USA300 Clone of Methicillin-Resistant *Staphylococcus aureus*. *J. Infect. Dis.* **197**, 1523–1530 (2008).

7. P. J. Planet, *et al.*, Emergence of the epidemic methicillin-resistant *Staphylococcus aureus* strain USA300 coincides with horizontal transfer of the arginine catabolic mobile element and *speG*-mediated adaptations for survival on skin. *MBio* **4**, e00889–13 (2013).

8. P. J. Planet, *et al.*, Parallel Epidemics of Community-Associated Methicillin-Resistant *Staphylococcus aureus* USA300 Infection in North and South America. *J. Infect. Dis.* **212**, 1874–1882 (2015).

9. D. Kühnert, T. Stadler, T. G. Vaughan, A. J. Drummond, Phylodynamics with Migration: A Computational Framework to Quantify Population Structure from Genomic Data. *Mol. Biol. Evol.* **33**, 2102–2116 (2016).

10. N. F. Müller, R. R. Bouckaert, Adaptive Metropolis-coupled MCMC for BEAST 2. *PeerJ* **8**, e9473 (2020).