Figure S1. Clonal genealogy simulated under the coalescent and used for Figure 3.
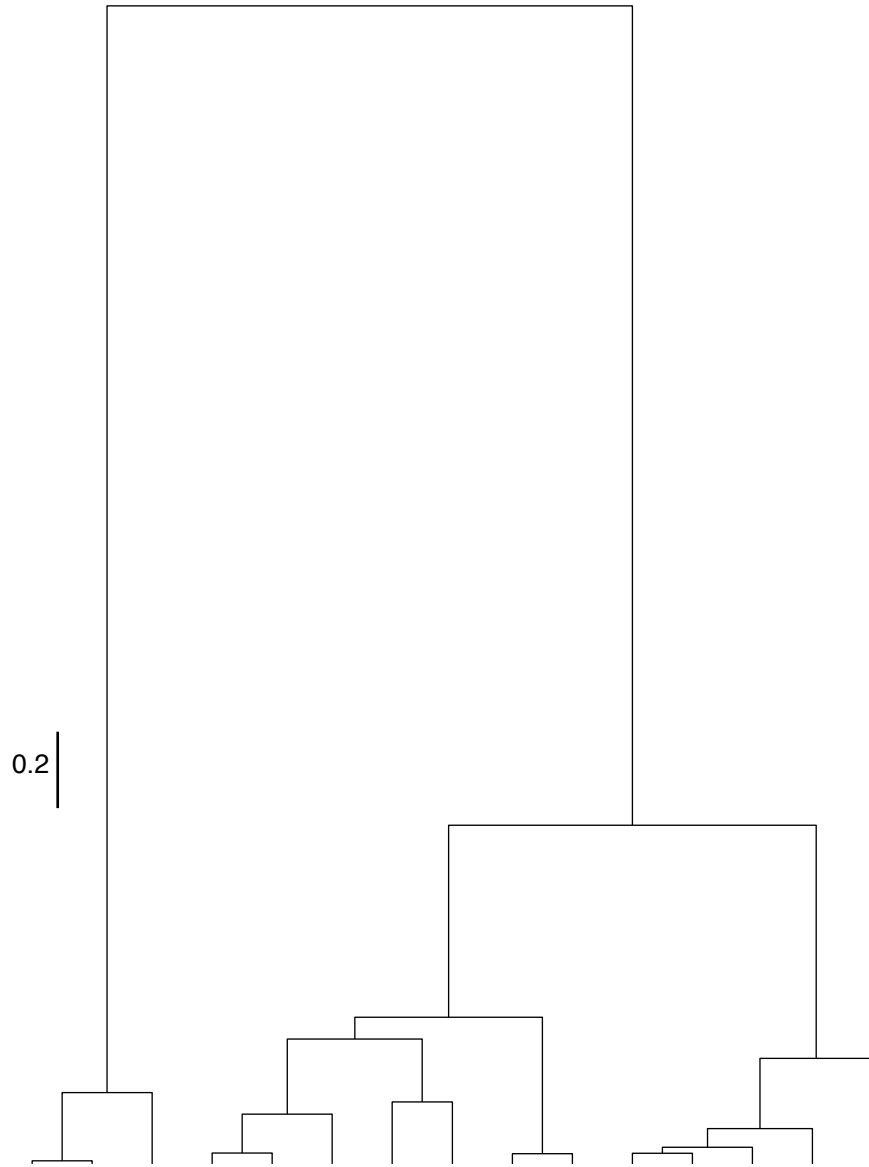
Figure S2. Clonal genealogy simulated under the coalescent and used for Figure S3.
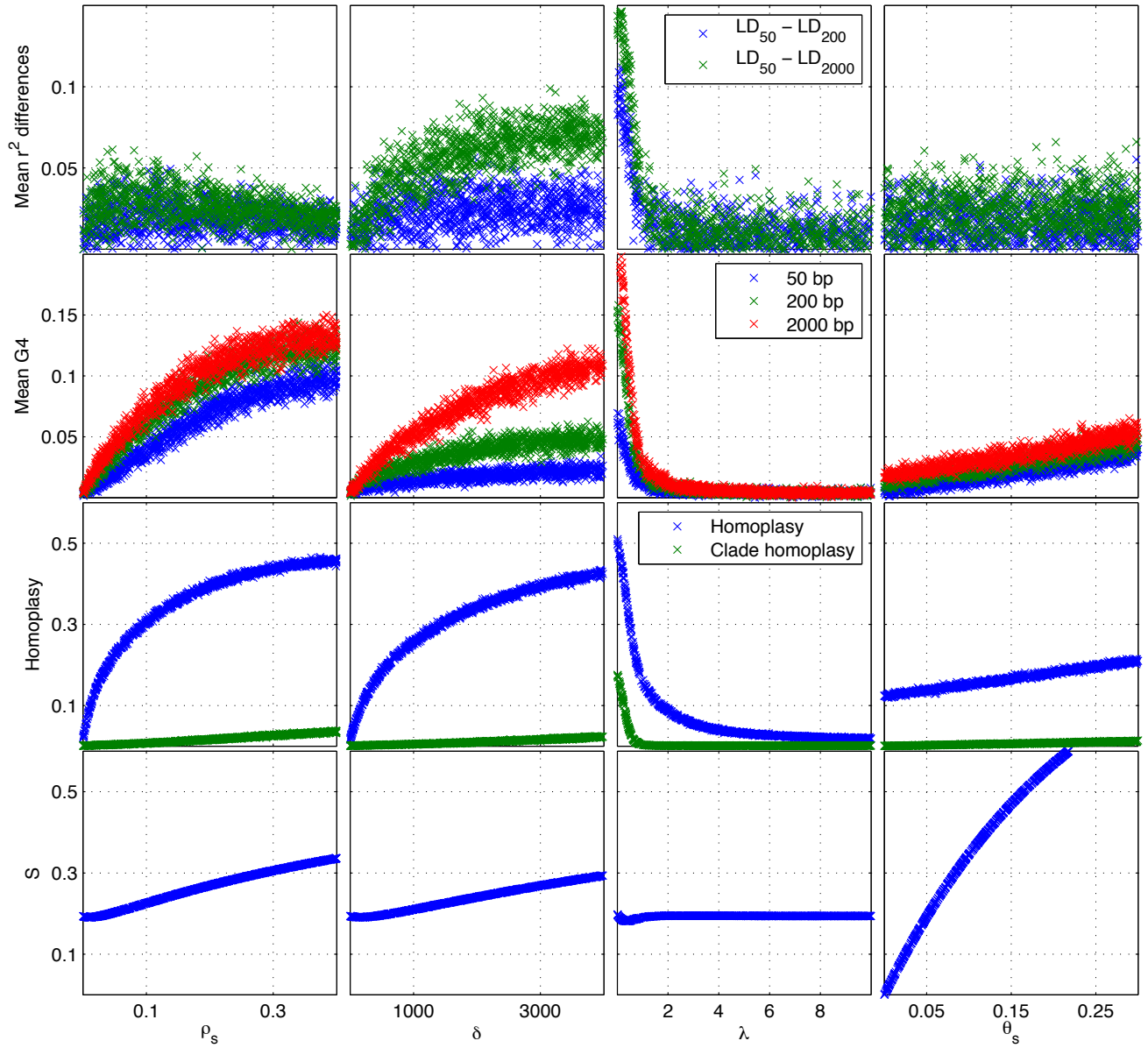
M. A. Ansari and X. Didelot

Figure S3. Relationship between the summary statistics and model parameters when simulating using the clonal genealogy in Figure S2.
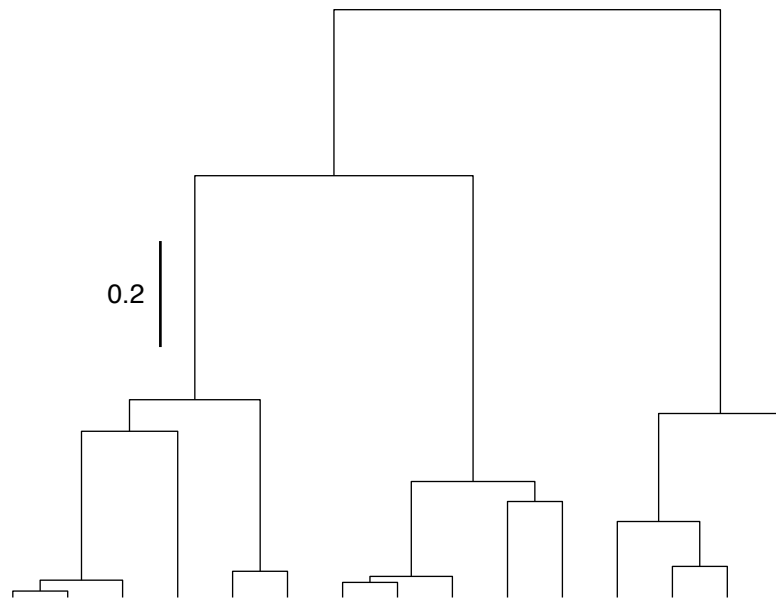
M. A. Ansari and X. Didelot

Figure S4. Clonal genealogy simulated under the coalescent and used for the application on simulated data.
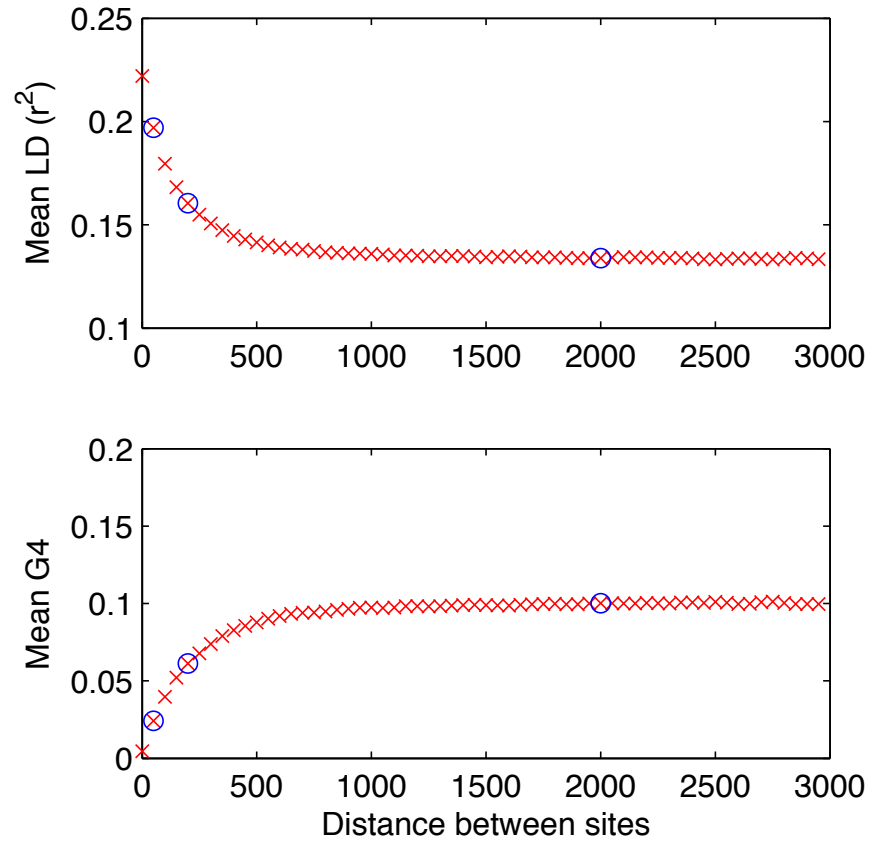
M. A. Ansari and X. Didelot

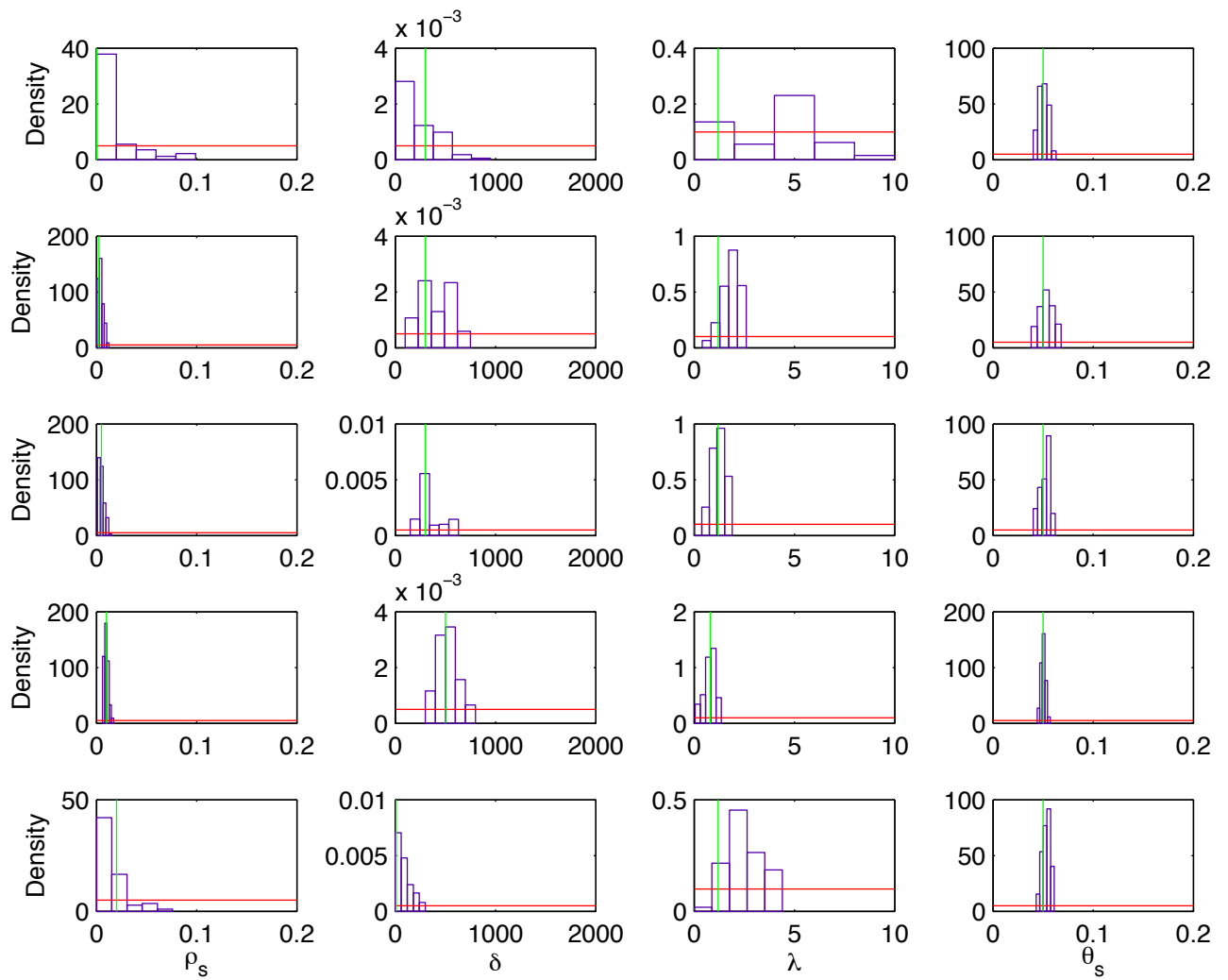**Figure S5.** LD and G4 plots for the simulated dataset.

Figure S6. Posterior marginal densities for five simulated data sets on a range of parameters. The correct value is in green, the prior in red and the posterior in blue.
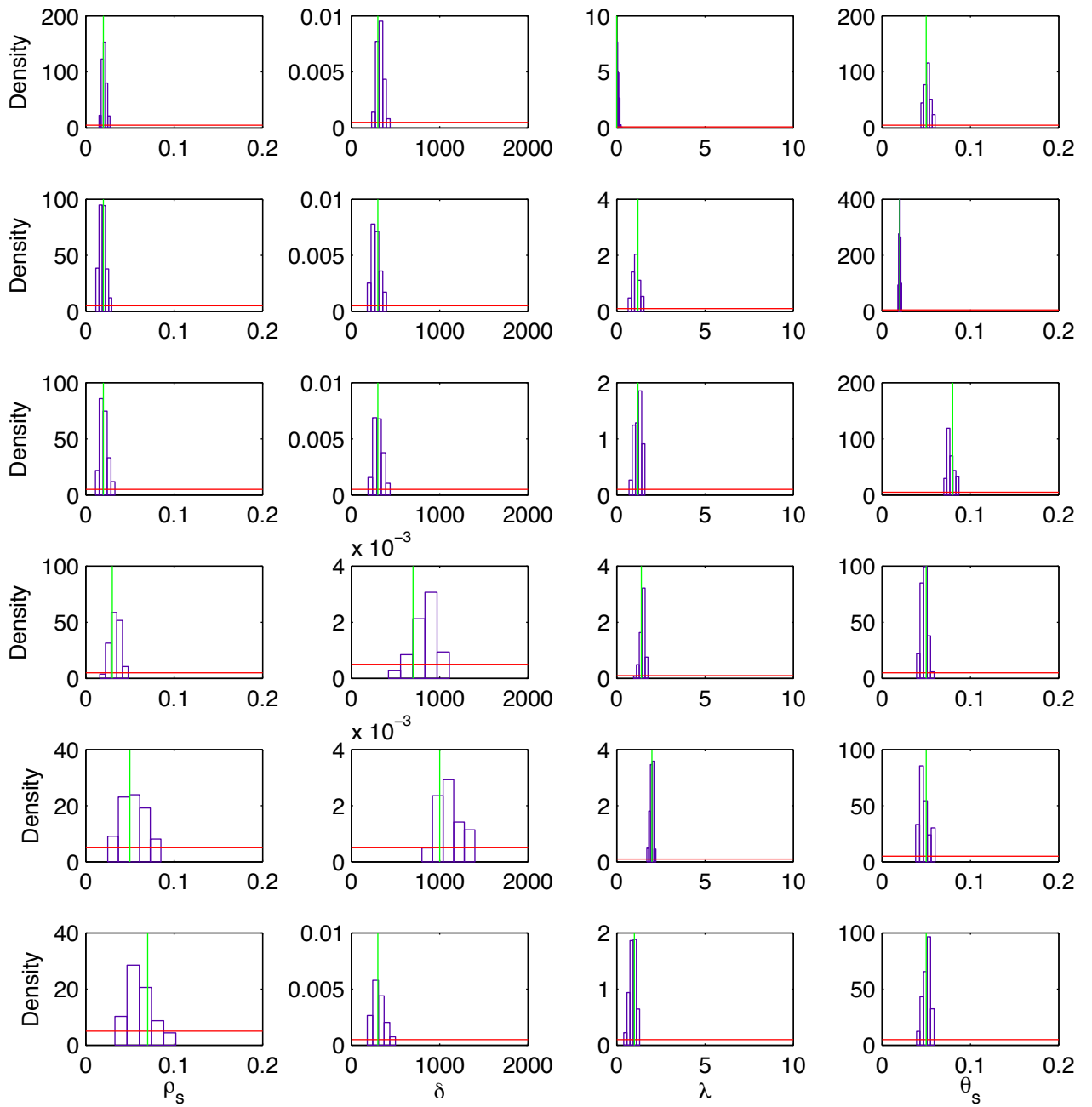
Figure S7. Posterior marginal densities for six simulated data sets on a range of parameters. The correct value is in green, the prior in red and the posterior in blue.
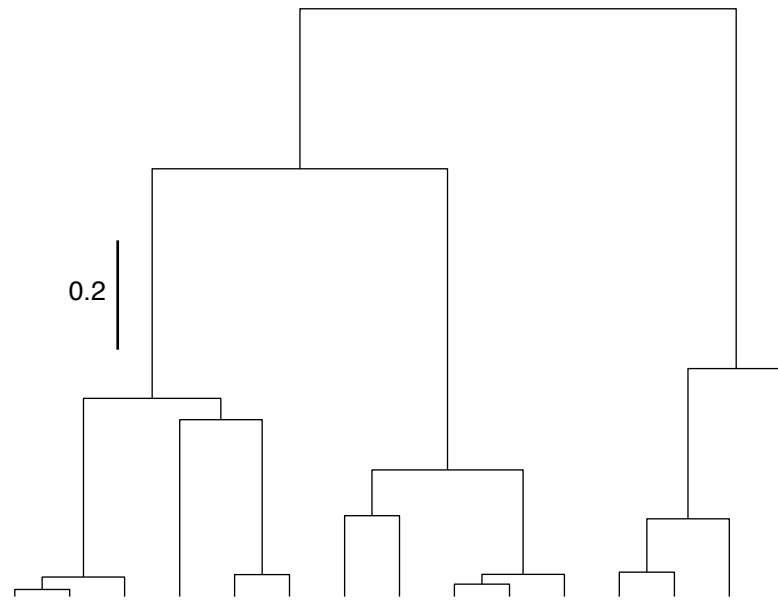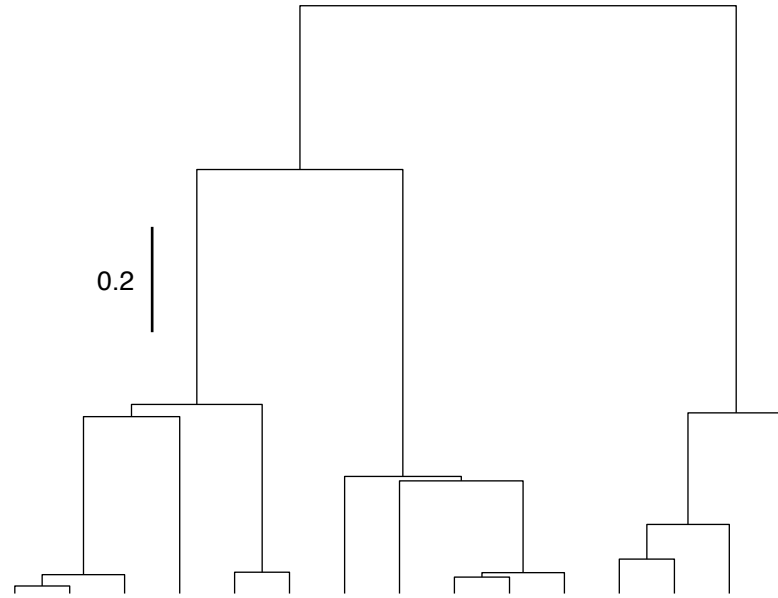
M. A. Ansari and X. Didelot

Figure S8. The two incorrect clonal genealogies used for testing robustness.
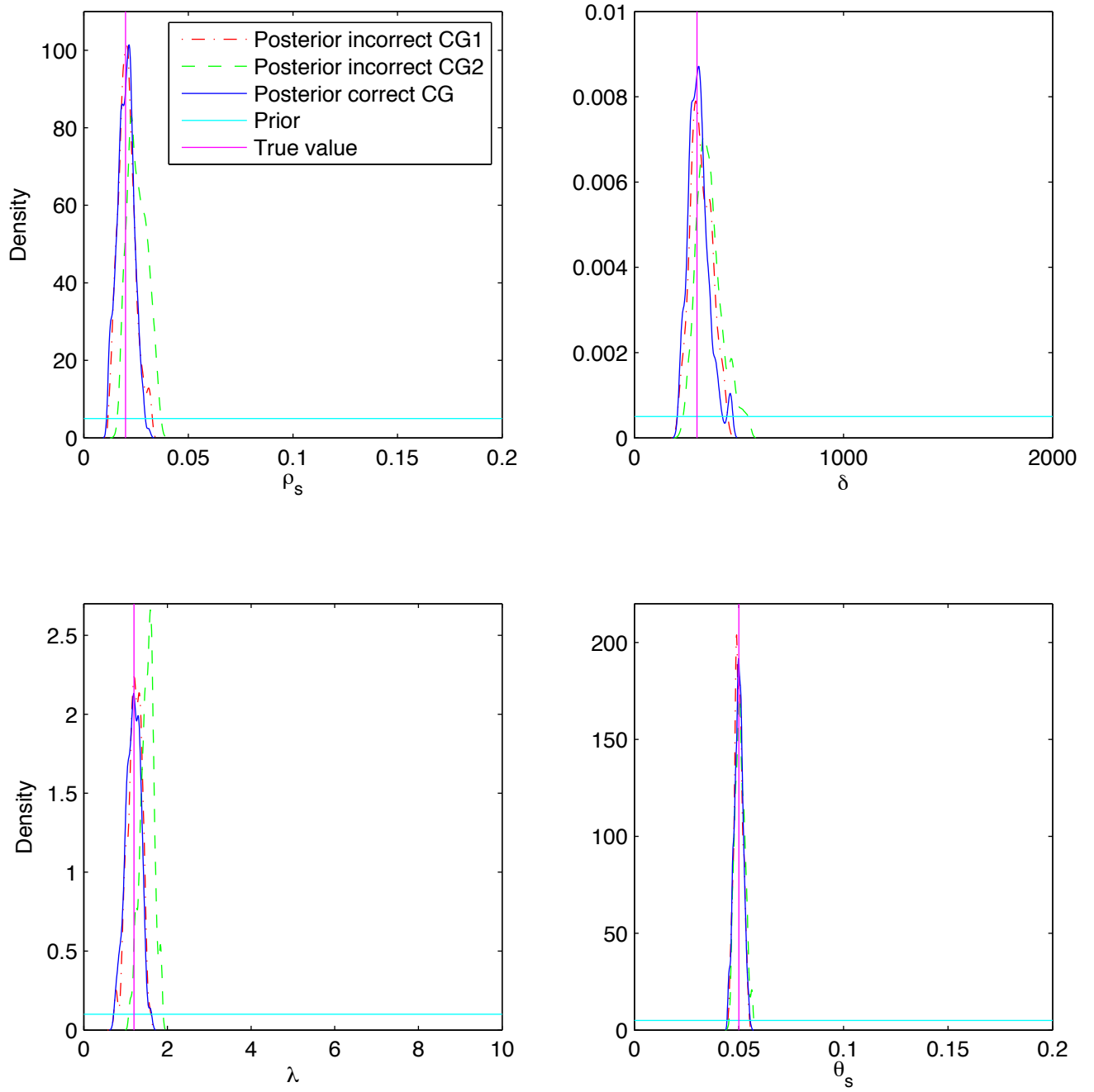
M. A. Ansari and X. Didelot

Figure S9. Comparison of posterior marginal density obtained using the correct, and the two incorrect clonal genealogies.
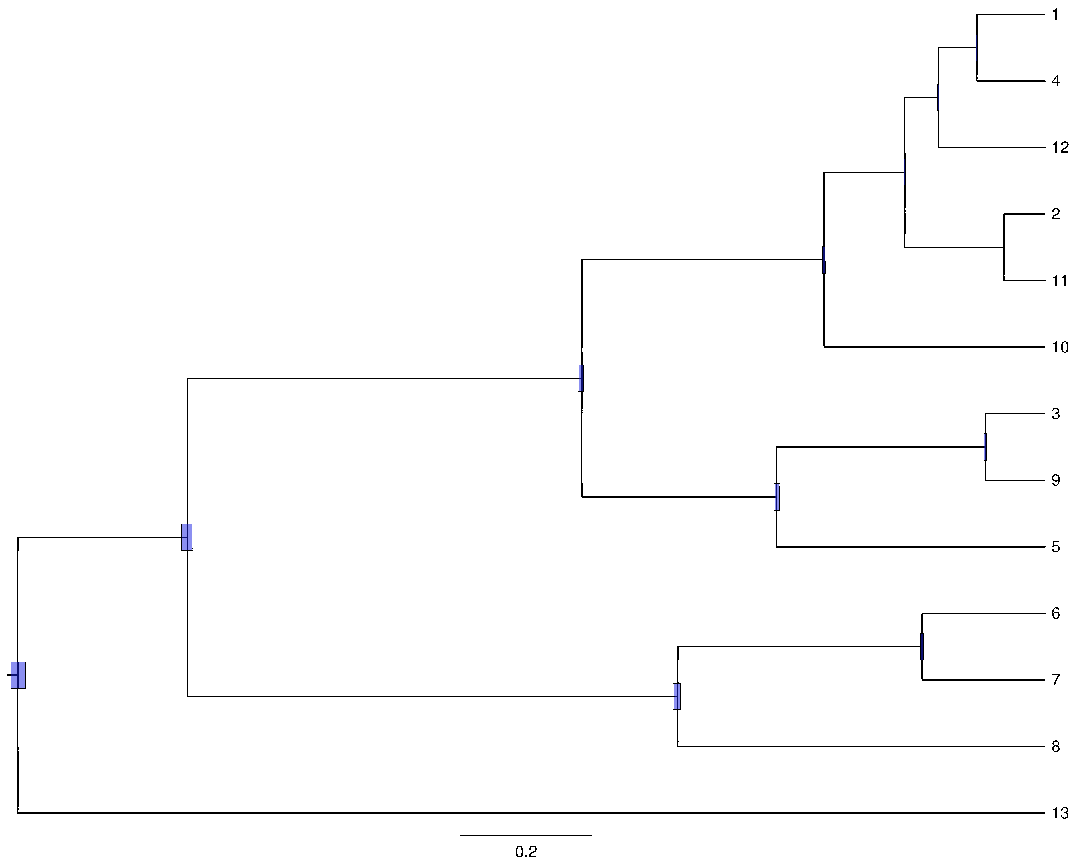
Figure S10. Clonal genealogy of the *Bacillus cereus* dataset inferred by ClonalFrame. The blue bars represent the uncertainty on the age of the internal nodes.
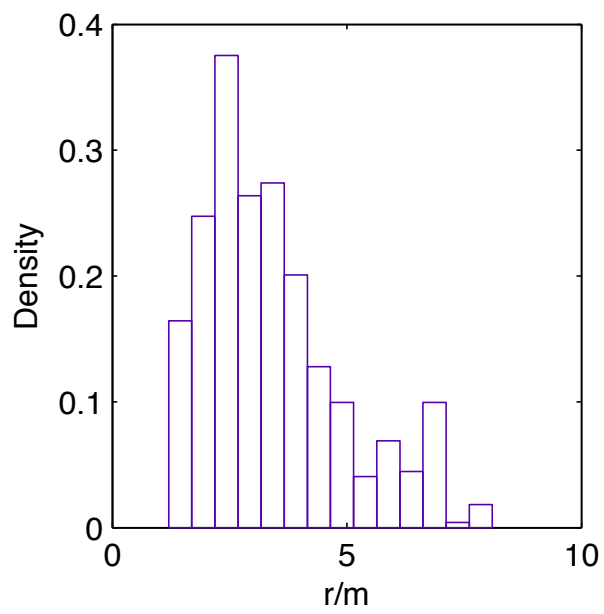
Figure S11. Posterior density of r/m for the *Bacillus cereus* dataset.
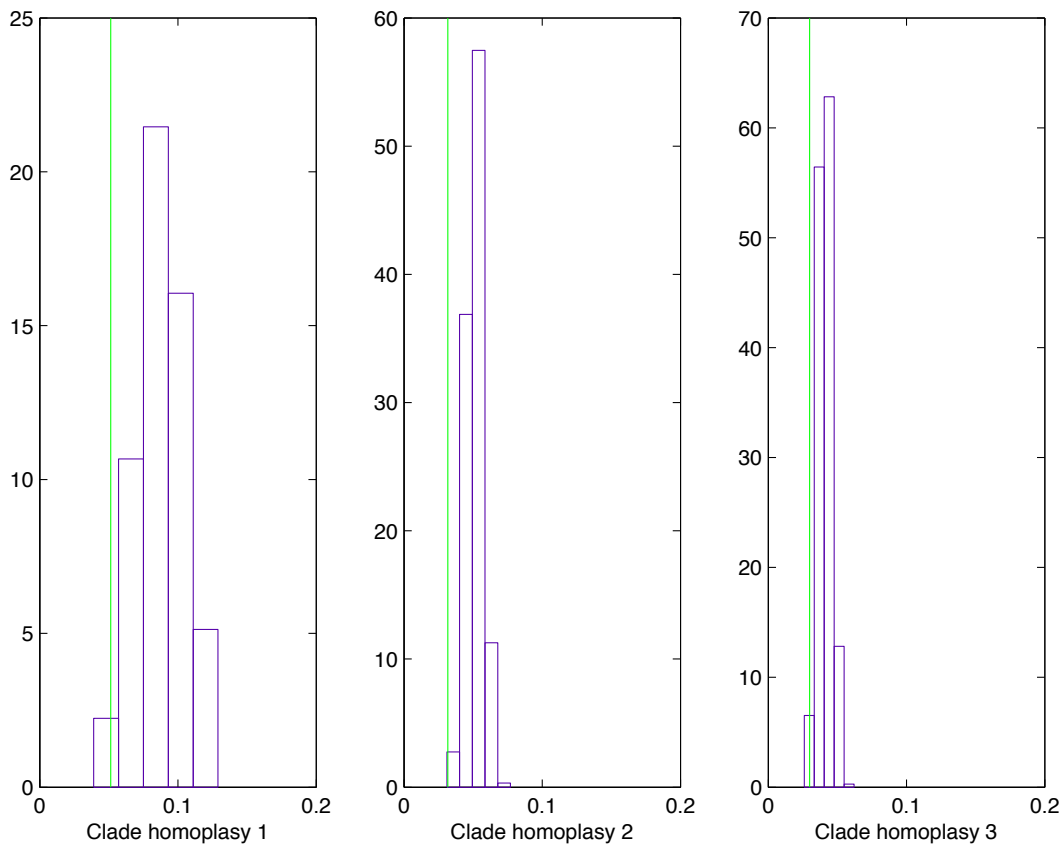
M. A. Ansari and X. Didelot

Figure S12. Posterior predictive distributions of the three additional summary statistics for the application to the *Bacillus cereus* dataset. The green lines represent the observed values.

M. A. Ansari and X. Didelot