S2 Appendix: Well-calibrated simulation studies

Every method presented and tested in this paper was validated with a well-calibrated simulation study using the same model and operator set that which was benchmarked. This is achieved using 100 simulated datasets (each with $N = 100$ taxa and a $L = 5000$ nt alignment). The 95 % highest posterior density (HPD) interval is calculated, and the 'true' parameter values are compared with their values estimated from the simulated data.

The validation results are presented as plots. The number in the bottom right corner of each plot is the coverage; that is the percentage of simulations where the 'true' parameter value is within the 95 % HPD interval. If all parameters have coverage close to 95 % coverage, then this suggests correctness of implementation.

The parameters presented in each plot are: the Yule model birth rate λ, rate heterogeneity gamma shape Γ, HKY substitution model parameter κ, log-normal clock standard deviation σ, the frequency of the A nucleotide f_A, and the rate of the first taxon r_1. The remaining 3 frequencies and 99 taxa rates are omitted from the figures.
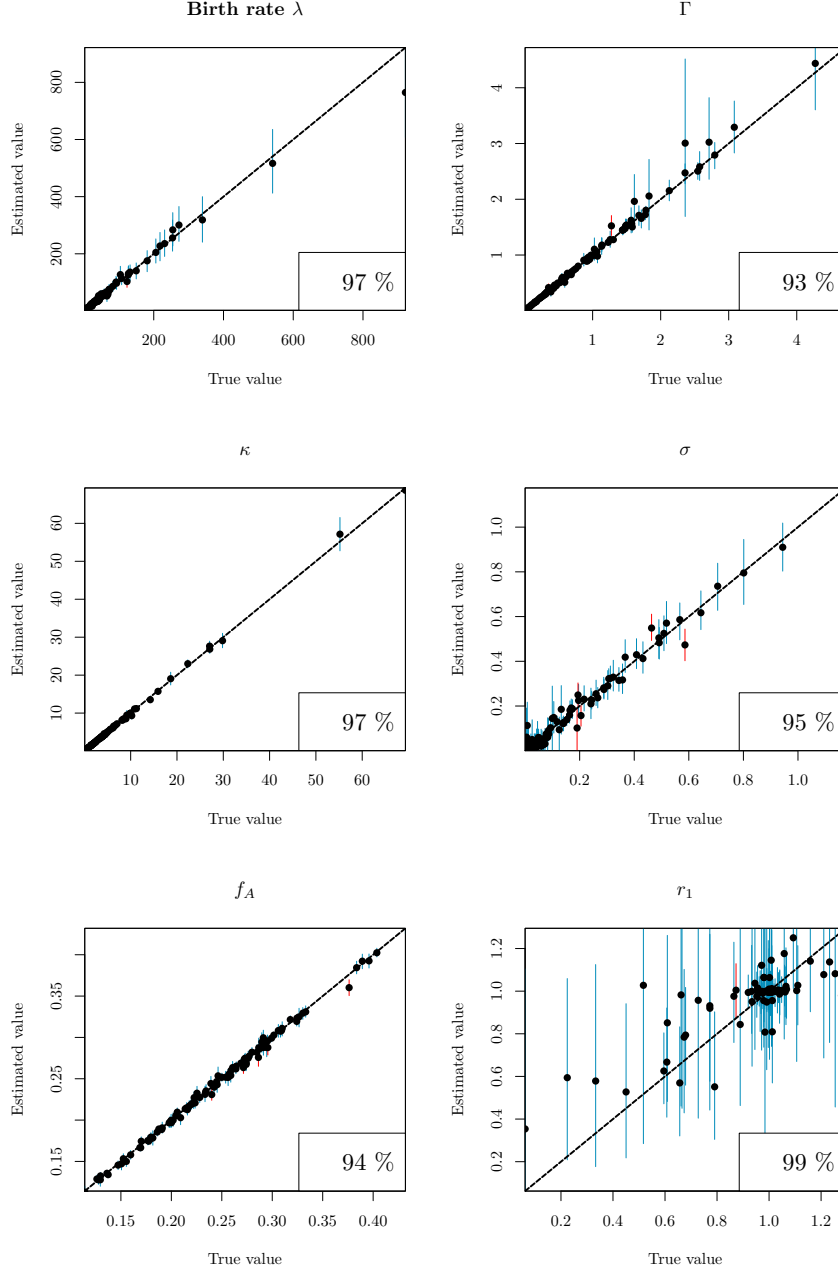
References

Figure 1: **The *cat* parameterisation.** Blue and red vertical lines are 95 % HPD intervals, where blue intervals contain the true value and red intervals do not.
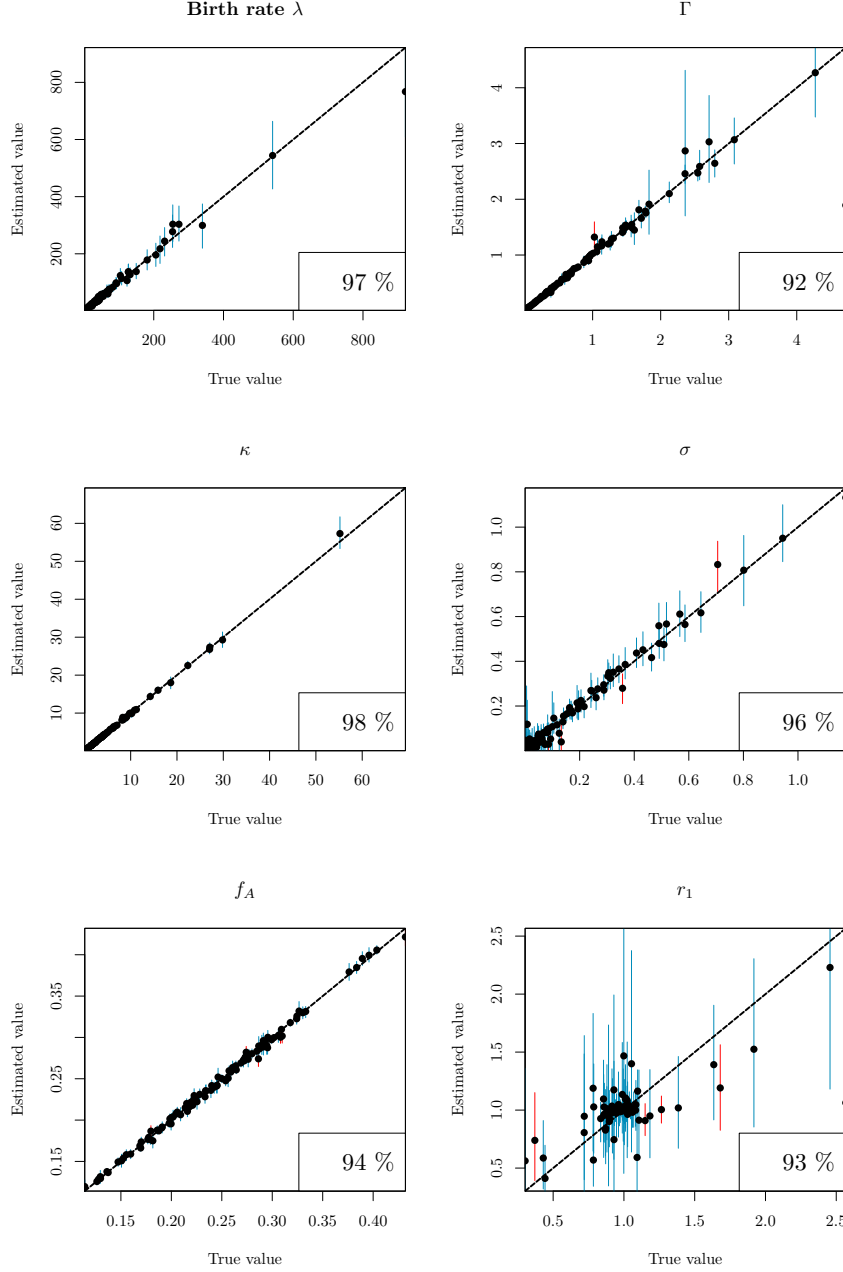
Figure 2: **The *real* parameterisation,** with the operators decribed by Zhang and Drummond 2020. The *real 2006* setting uses a subset of the operators tested in this validation.

3

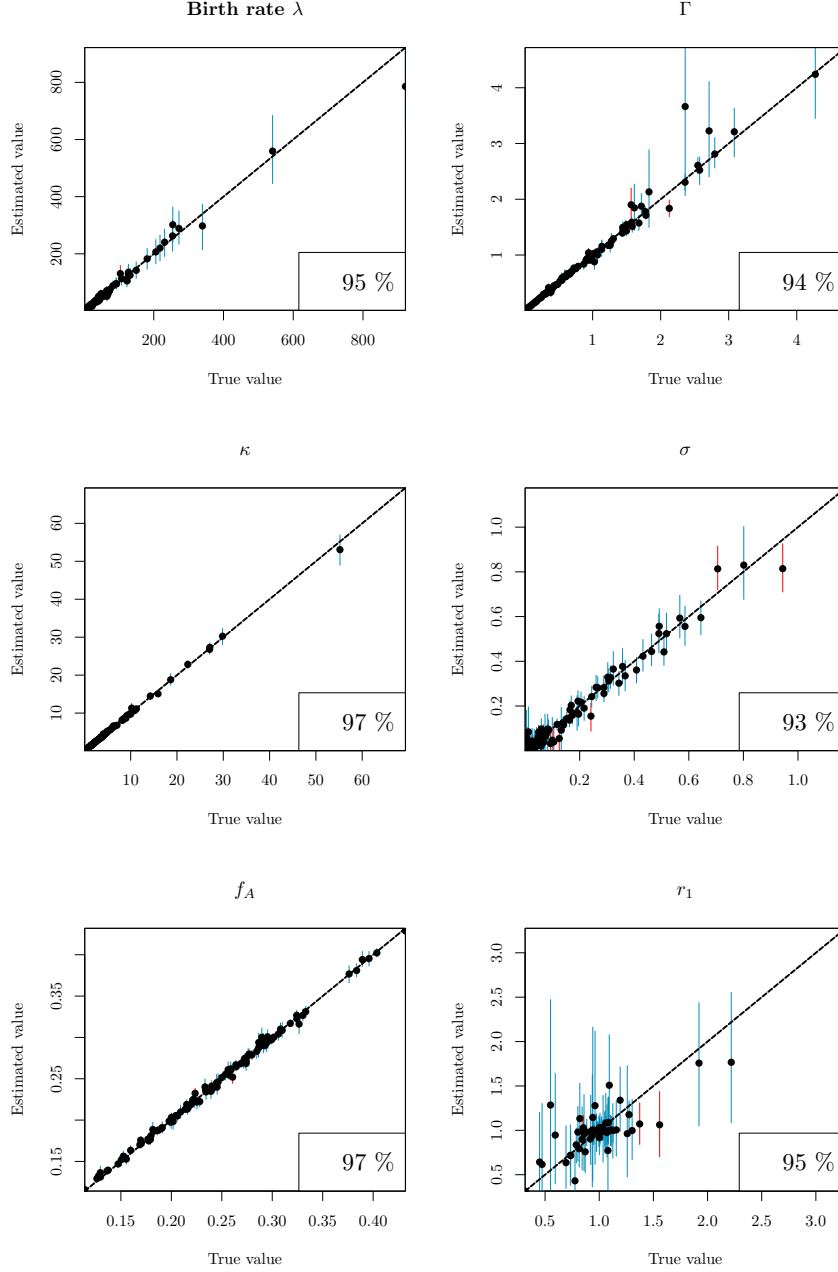Figure 3: **The *quant* parameterisation,** with the operators decribed by Zhang and Drummond 2020, and adapted for *quant* as described in **S1 Appendix**. The *quant 2006* setting uses a subset of the operators tested in this validation.
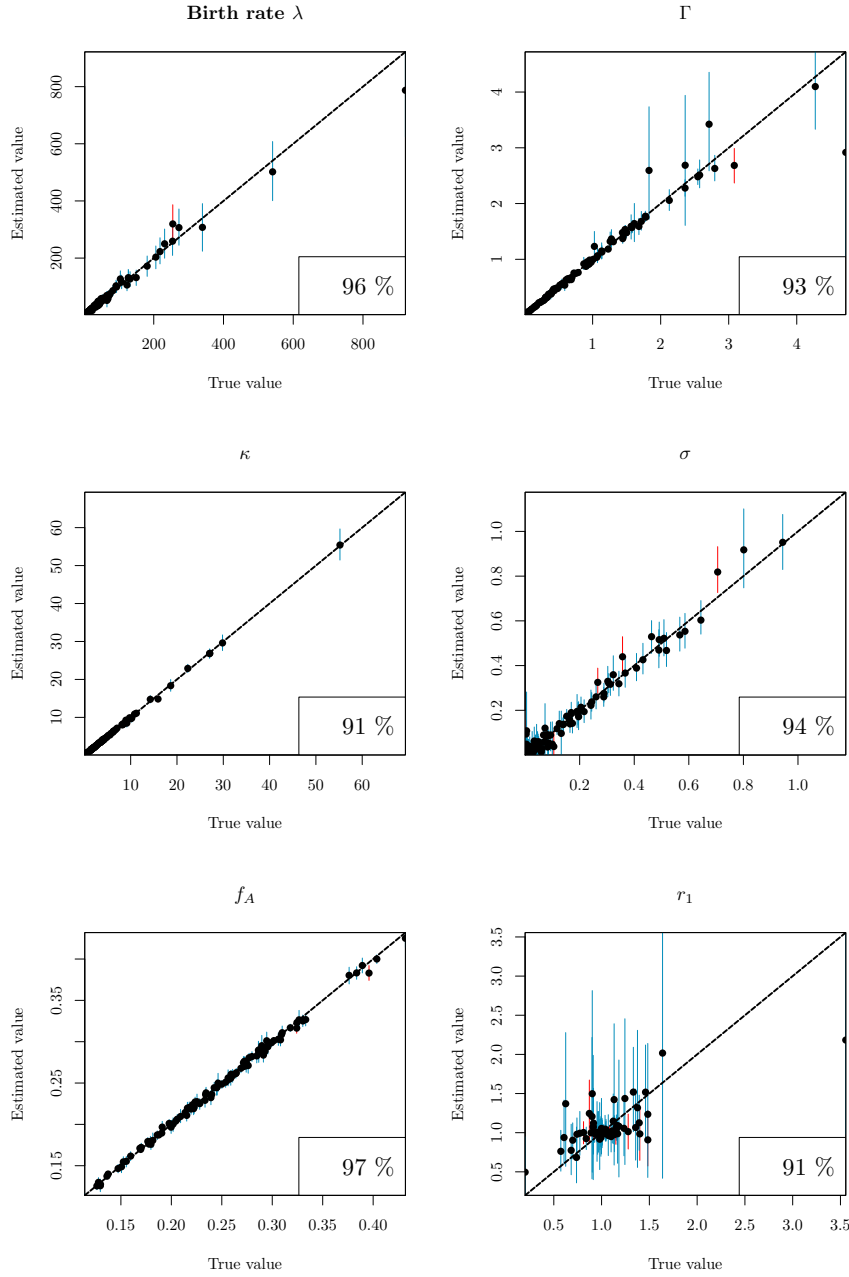
4

Figure 4: **Bactrian proposal kernel.** A Bactrian(0.95) proposal kernel is used for all clock model operators (under *quant*), as described in main article.
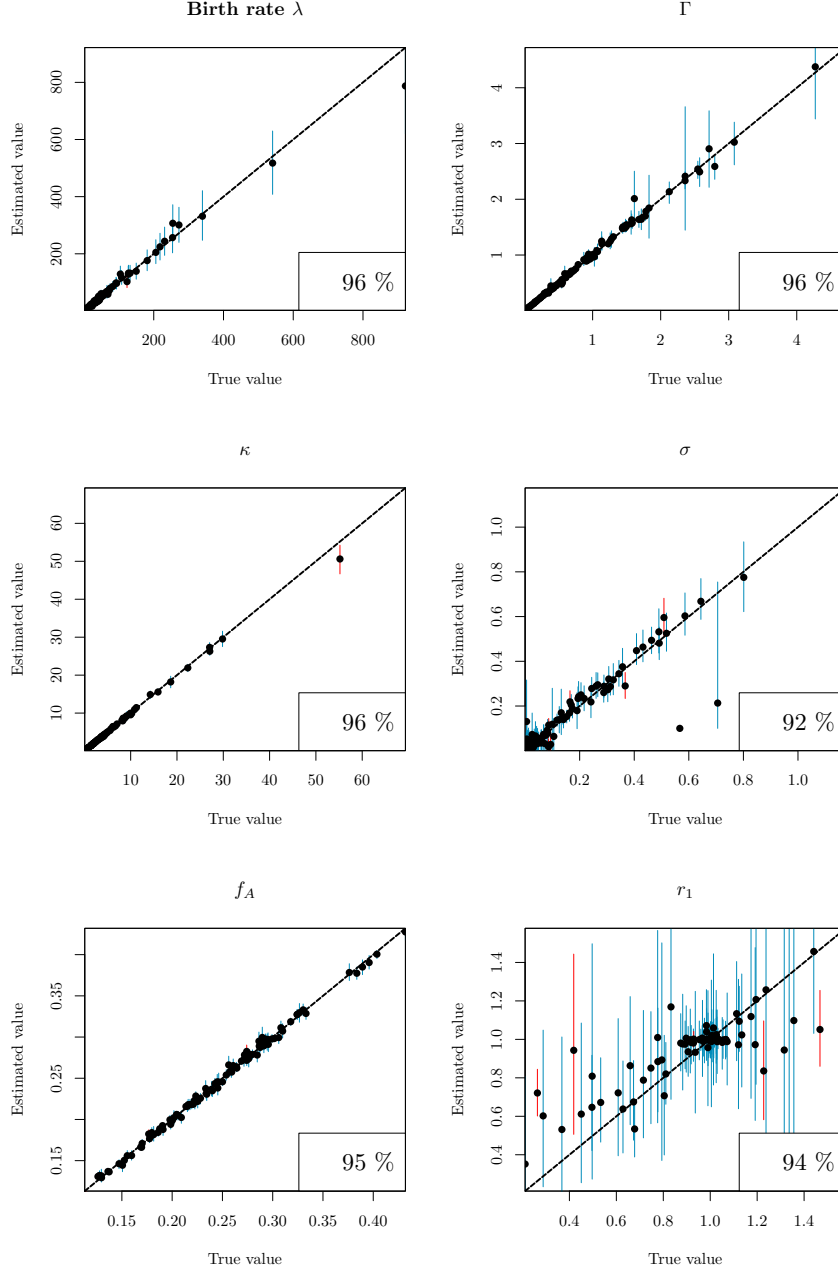
5

Figure 5: **The Narrow Exchange Rate operators under the _real_ parameterisation.** The 48 NER and 48 NERw operators, including the null Narrow Exchange operator, are tested in the same MCMC session.
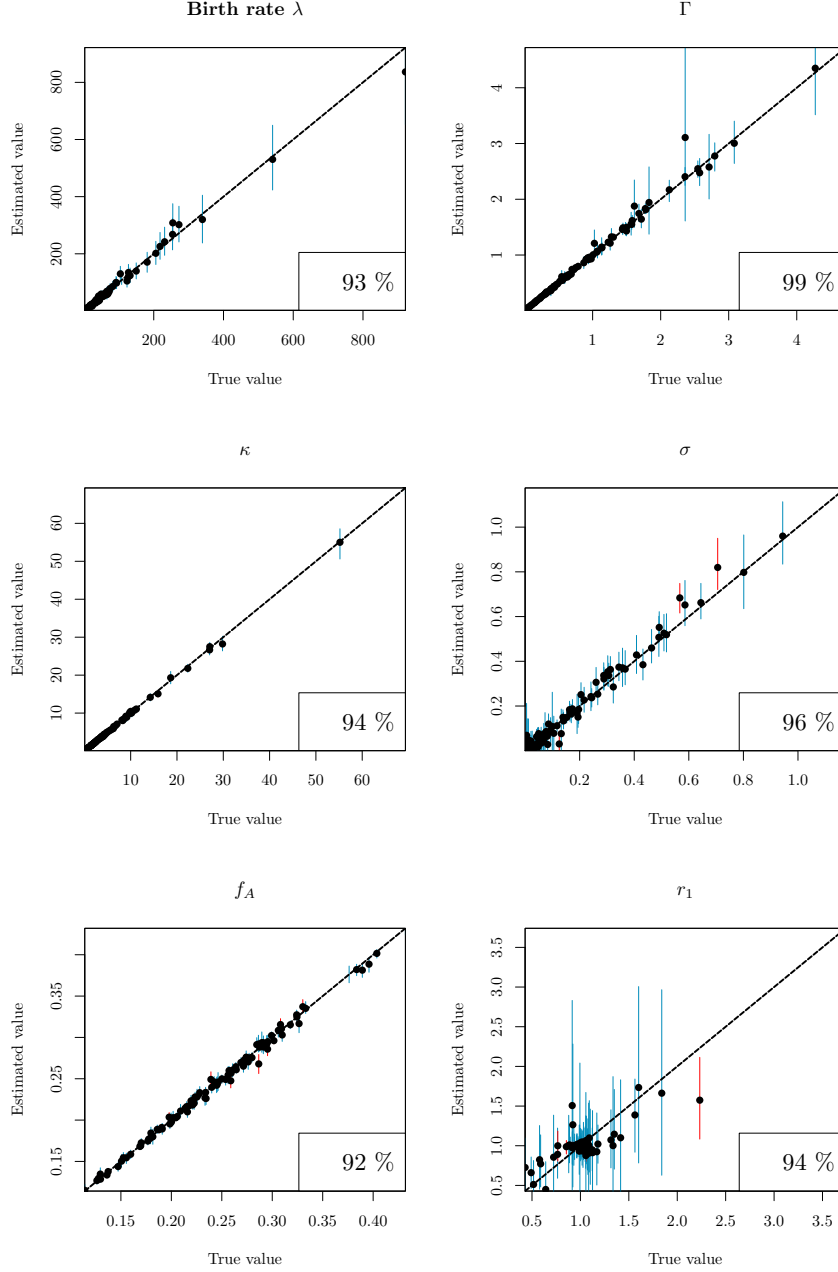
Figure 6: **The Narrow Exchange Rate operators under the *quant* parameterisation.** The 48 NER and 48 NERw operators are tested simultaneously.