Estimating the age of poorly dated fossil specimens and deposits using a total-evidence approach and the fossilized birth-death process

Supplementary materials

1 Alternative simulation conditions

In addition to the simulation datasets presented in the main text, we also studied the impact of several alternative simulation conditions. In the first case, all fossilization rates were divided by 2, which resulted in simulated trees containing between 20 and 30 fossils. In the second case, molecular sequences were simulated using a relaxed exponential clock with a mean rate of 0.005, identical to the rate used for the strict clock in the main datasets. Finally, the third simulation condition replicated one common feature of insect morphological matrices, where some precise-date fossils are very incomplete and thus contain little data. In this condition a random sample of 5% of the precise-date fossils were assigned to be incomplete. A randomly chosen 10% of the characters in the matrix had data for these incomplete fossils, all other characters were assigned the unknown character state "?". Alternative simulation conditions were simulated with different proportions of imprecise-date fossils (0.1, 0.3 or 0.5) and using an age range of 0.2 times the true age for precise-date fossils.

Results are shown in Figures 1 and 2. Overall, the alternative simulation conditions did not affect the results much, with two notable exceptions. The coverage of the fossil age estimates was improved with lower numbers of fossils compared to the standard dataset, for both precise-date and imprecise-date fossils, although the relative error was similar to the standard. Applying a relaxed clock to the molecular alignment also appeared to decrease the accuracy of the topological placement of fossil samples compared to the standard dataset, although it should be noted that a high proportion of analyses using the relaxed clock and the 0.5 proportion of imprecise-date fossils failed to converge (22 out of 100 replicates).
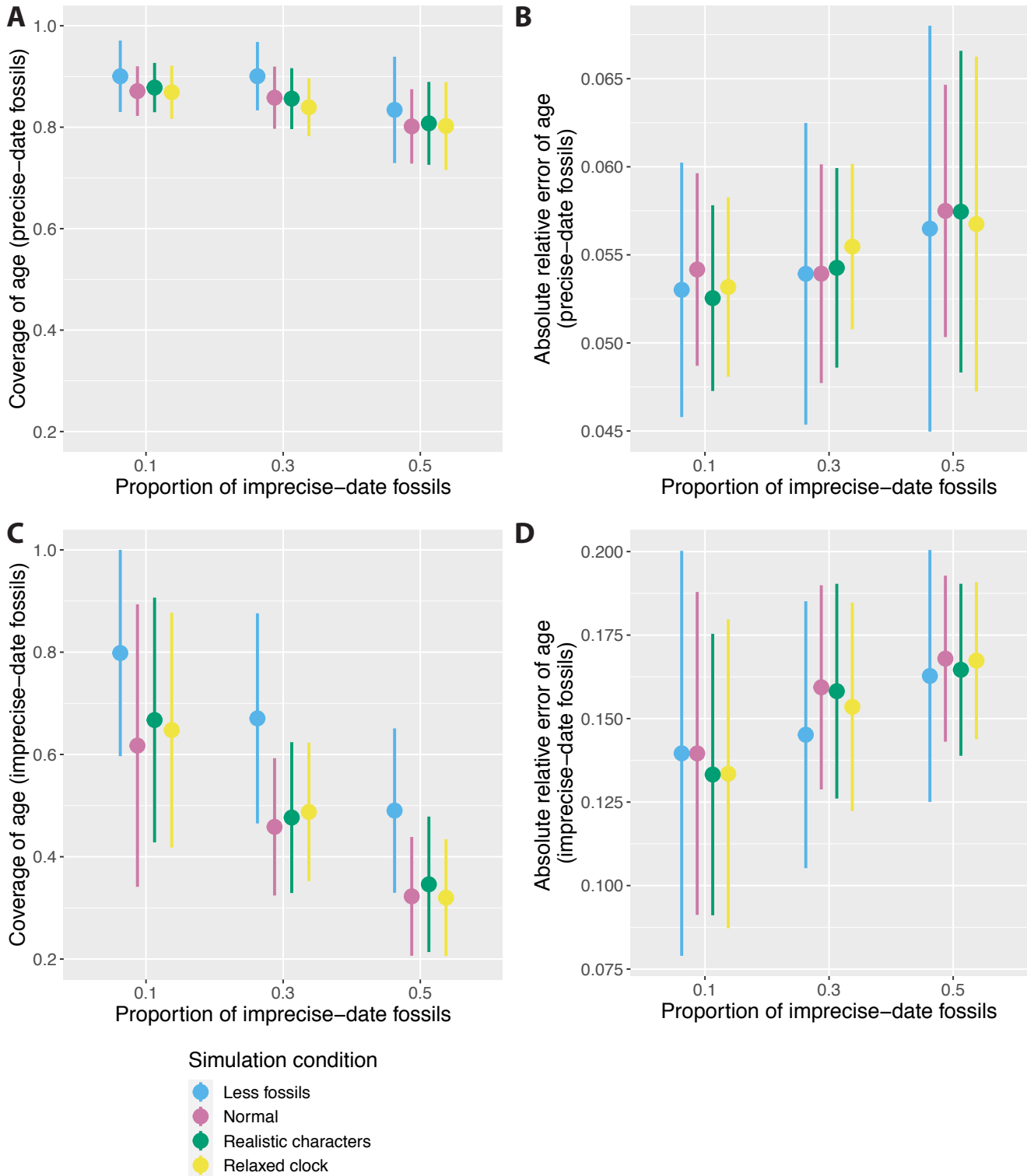
Figure 1: Absolute relative error of the median age estimate (B,D) and 95% HPD coverage (A,C) of precise-date fossils (A,B) and imprecise-date fossils (C,D) for different proportions of imprecise-date fossils, and different simulation conditions. Measures are averaged over all fossils for each replicate. The average and standard deviation across all replicates is shown.
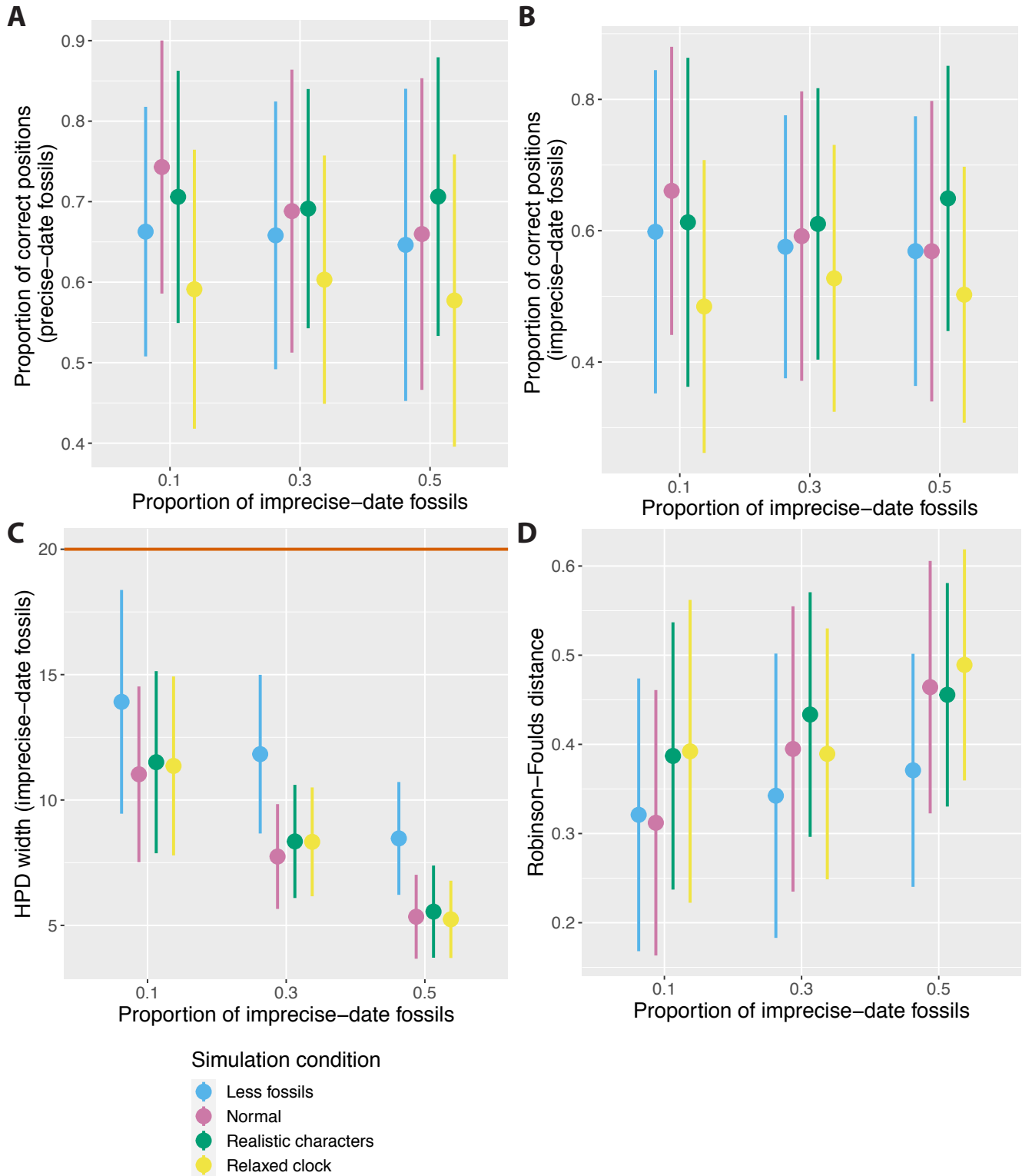
Figure 2: Proportion of posterior samples with correctly placed fossils, averaged across all precise-date fossils (A) or all imprecise-date fossils (B), width of the 95% HPD interval averaged across all imprecise-date fossils (C) and mean normalized RF distance between estimated trees and simulated tree (D), for different proportions of imprecise-date fossils and different simulation conditions. The average and standard deviation across all replicates is shown. The brown line in C shows the size of the age range set as prior for all imprecise-date fossils (*i.e.*, 20My).
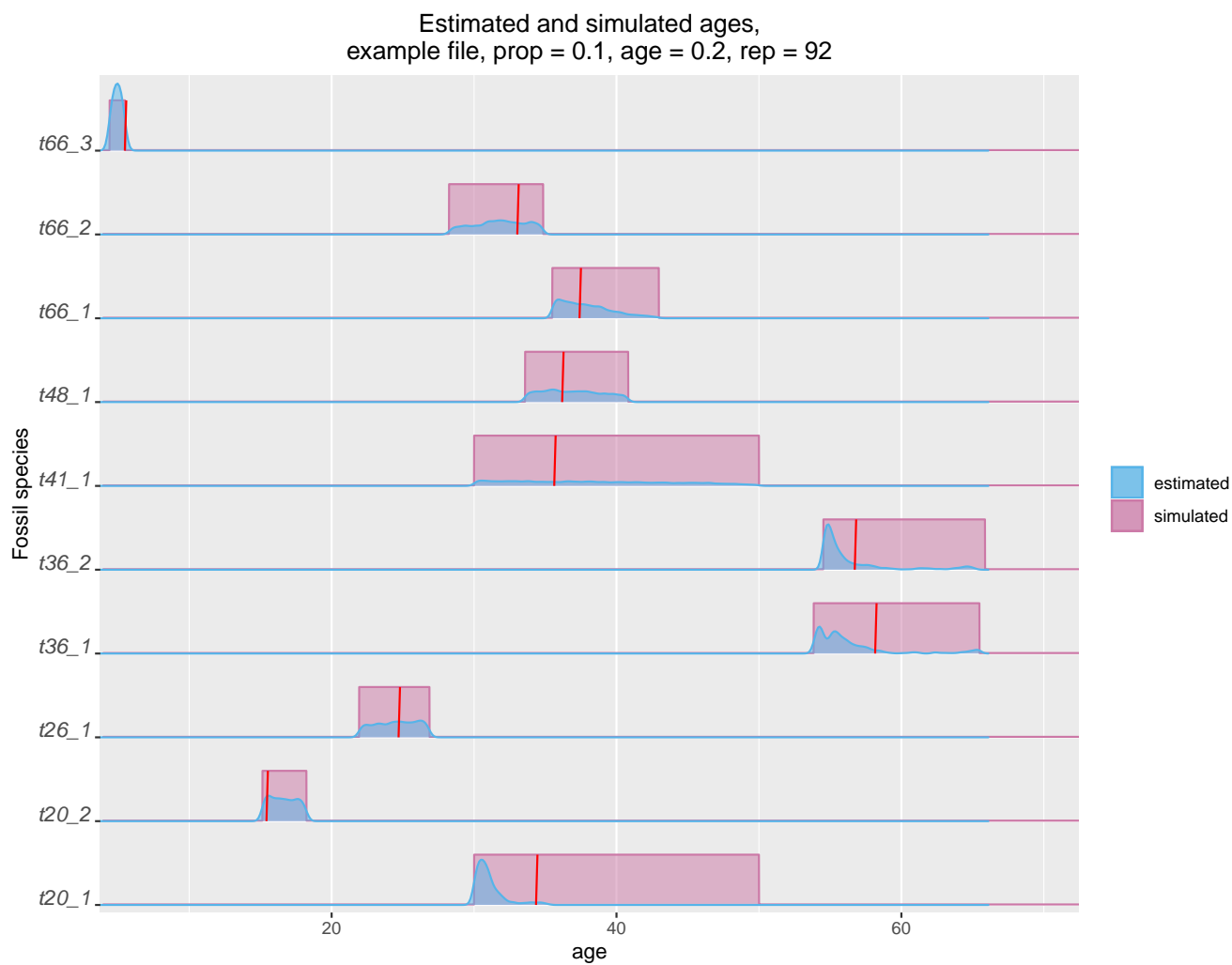
2 Supplementary figures

Figure 3: Comparison of simulated (pink) and estimated (blue) ages for one simulation replicate, with a proportion of 0.1 of imprecise-date fossils and a relative age range of 0.2 for precise-date fossils. The simulated age range is shown as a uniform distribution, while the estimated age is the inferred posterior distribution. The true age of each fossil is marked in red.