

Fast and Slow Implementations of Relaxed-Clock Methods Show Similar Patterns of Accuracy in Estimating Divergence Times

Fabio U. Battistuzzi,¹ Paul Billing-Ross,¹ Aditya Paliwal,^{1,2} and Sudhir Kumar*,^{1,2}

¹Center for Evolutionary Medicine and Informatics, The Biodesign Institute, Arizona State University

²School of Life Sciences, Arizona State University

*Corresponding author: E-mail: s.kumar@asu.edu.

Associate editor: Koichiro Tamura

Abstract

Phylogenetic analyses are using increasingly larger data sets for estimating divergence times. With this increase in data sizes, the computation time required is becoming a bottleneck in evolutionary investigations. Our recent study of two relaxed-clock programs (BEAST and MultiDivTime [MDT]) showed their usefulness in time estimation; however, they place a significant computational time burden on biologists even for moderately small data sets. Here, we report speed and accuracy of another relaxed-clock program (MCMCTree, MC2T). We find it to be much faster than both MDT and BEAST while producing comparable time estimates. These results will encourage the analysis of larger data sets as well as the evaluation of the robustness of estimated times to changes in the model of evolutionary rates and clock calibrations.

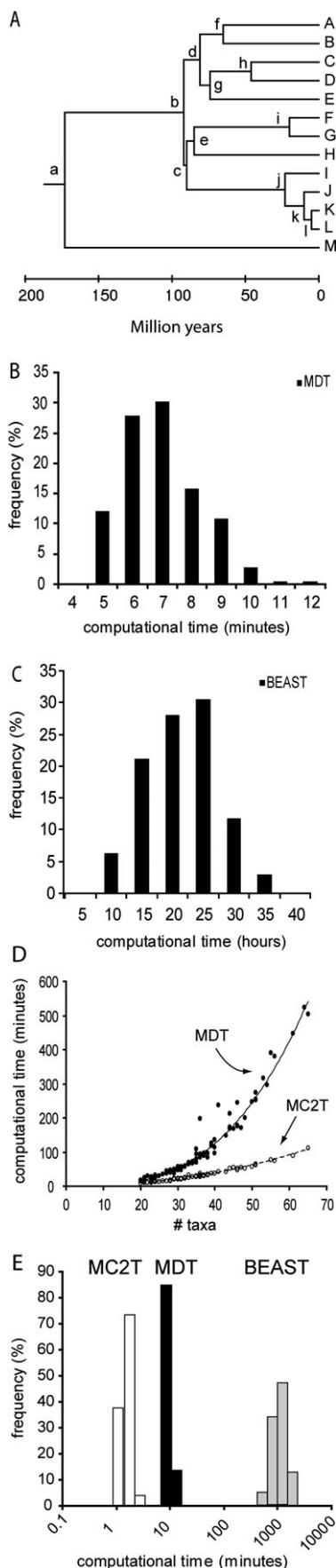
Introduction

In Battistuzzi et al. (2010), we reported that the MultiDivTime (MDT) and BEAST software packages are accurate in estimating divergence times when the model describing the change in evolutionary rate among lineages matched the underlying assumptions (autocorrelated [AR] or random rates [RR]) (Thorne and Kishino 2002; Drummond and Rambaut 2007). However, the computational time needed to complete calculations was rather large for even a small phylogeny (e.g., in BEAST) and exorbitant for large phylogenies (fig. 1). Such requirements not only slow down the pace of discovery but ultimately discourage efforts to assess the robustness of inferred times to assumptions made regarding evolutionary rate models and calibrations used. Consequently, the quality of scientific discoveries also suffers. Therefore, a tool that performs as well as MDT and BEAST but takes much less time is desired.

Here, we explored the use of the MCMCTree program, MC2T, which is based on a Bayesian framework for time estimation in the PAML package (Yang 2007). MC2T takes as input a tree topology, a sequence alignment, and bounds on clock calibrations. It first estimates branch lengths, which is followed by the inference of divergence times for each node in the tree based on the selection of prior distributions on clock calibrations (e.g., Gamma and skewed Normal) and the model of evolutionary rate change in the tree (AR or RR) (Battistuzzi et al. 2010). Using simulated sequences and perfect calibration points (± 1 My around the true time), we directly compared the computational speed and time accuracy of MC2T with those of MDT and BEAST. All three tools use independent algorithms and programming implementations of relaxed-clock methods, and their direct comparison is interesting because of their applicability to estimating species divergence times.

We first used alignments resulting from the concatenation of ten individually simulated genes, which were the primary focus of our previous study (alignment length = $\sim 8,000$ – $23,000$ basepairs) (Battistuzzi et al. 2010). In both the best- and worst-case scenarios (model match and model violation), MC2T completes analyses at a fraction of the time taken by MDT or BEAST (fig. 1E). At an average, MC2T took less than 2 minutes to generate divergence times with default parameters, but MDT consumed at least 3-fold the time. BEAST required at least two orders of magnitude larger computational time. Furthermore, the rate of increase in computational time requirements increases much more slowly for MC2T than for MDT (fig. 1D); BEAST computations did not complete for large data sets (supplementary fig. S1, and Supplementary Material online).

Does MC2T speed come at a price of decreased accuracy of inferred time estimates? To answer this question, we evaluated its accuracy in estimating divergence times under the cases where the rate change models assumed by the program match or violate the assumptions for generating sequences. In the model-match analyses, the normalized difference from true time of time estimates reported by MC2T shows a central tendency with an absolute mean that differs from the true time by 7.5% (fig. 2A and B). MC2T shows greater difficulty in correctly estimating time when the lineage rates are autocorrelated (AR simulated data sets) as compared with RR simulated data sets (fig. 2A and B histograms). This difficulty is also reflected in the success rate of the credibility intervals (CrIs) in including the true time: the success rate for AR data sets is only 86% as compared with 98% for RR data sets. In both cases, the width of the 95% CrIs was approximately one-third of the true time.



As expected, divergence time estimates are less accurate when an incorrect model is chosen to describe the among lineage heterogeneity (model violation). The normalized distribution of estimated times is still symmetric but has an absolute mean 20% higher than the model-match case (fig. 2C and D). Again, MC2T had greater difficulty with accuracy for AR cases (fig. 2C and D histograms) with Crls success rates at 82% and 94% for AR and RR, respectively. The width of the 95% Crl was similar to the model match case.

Overall, the above results are encouraging, as the estimated times were within 10% of the true time at an average, and the Crls were not overly liberal. However, time estimates based on multiple concatenated gene sets are likely to be good because different sets of lineages evolve slower or faster than average in different genes, which will cancel (or at least reduce) rate differences across lineages (Battistuzzi et al. 2010). But, some empirical data sets will have genome-wide biases, which will not benefit from homogenization of the rate variations across lineages. Therefore, we evaluated the accuracy of MC2T for such a scenario and found divergence times to be, on average, 15% different from the true time (model match scenarios). Success rates were very high (96%), but the average width of Crls were much larger (73% of the estimated time) because statistical methods are not expected to overcome uncertainty in the time estimate introduced by large degrees of rate variation present across lineages.

Next, we compared MC2T results with those produced by MDT and BEAST. The implementation of AR and RR models in MC2T allows for a direct comparison of this fast method with these two slower methods. In the model match cases, accuracies of estimated times for both the fast and slow methods show similar trends, with the slow method being 2.5% more accurate than the fast method (7.5% and 5.1% for fast and slow methods, respectively; these values are averages of AR and RR cases, see fig. 2A and B). This is primarily because divergence times for AR cases are estimated less accurately in the fast method (fig. 2A and B, compare line and bar graphs). Fast and slow methods show overall similar success rates (92% and 95%, respectively), but Crls are ~20% larger in the fast method (26.5% and 34.5% for slow and fast methods, respectively).

In contrast to the model match case, success rates for the slow methods (MDT and BEAST) are worse than those for the fast method (81% vs. 88%) when the lineage rate model is violated. This is most likely due to larger Crl widths for the fast method (35% relative to the true time) as compared with the slow methods (27%). Instead, the average differences of time estimates from true times for fast and

FIG. 1. Computational times of MDT, BEAST, and MC2T for a 14 sequence phylogeny (model-match case). (A) Phylogeny used to simulate sequences; calibration node pairs are *h* and *k* (46 and 10 Ma), *h* and *d* (46 and 81 Ma), and *j* and *k* (23 and 10 Ma). Distribution of computational times for (B) MDT and (C) BEAST. (D) Relationship between the number of sequences and the computational time for MDT and MC2T. (E) Overall, comparison of computational times on a log scale for MC2T, MDT, and BEAST.

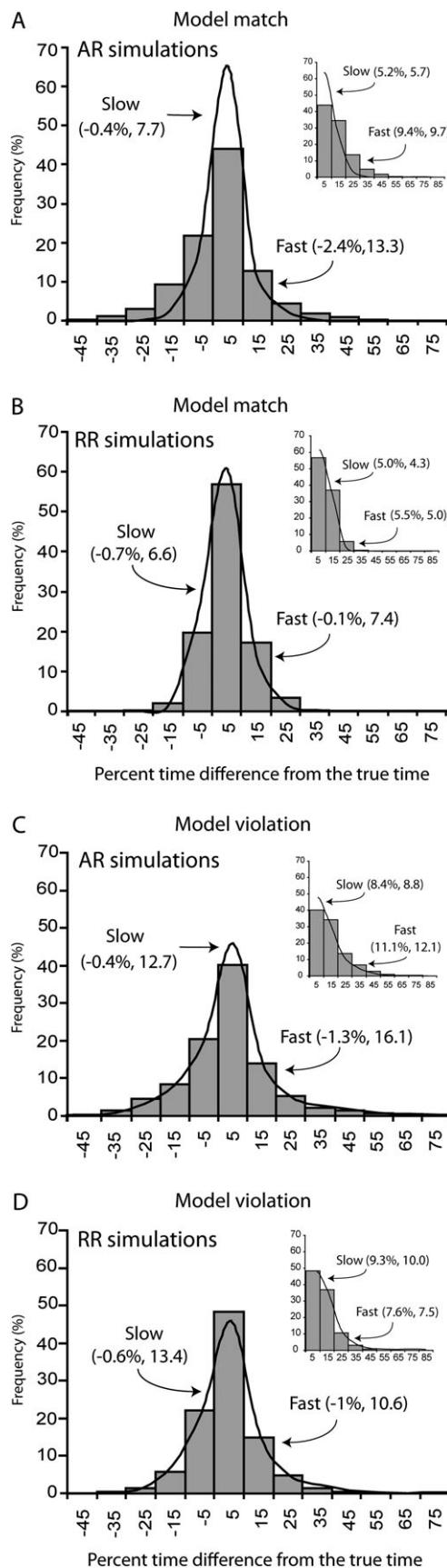


FIG. 2. Distributions of the normalized difference (%) of estimated times from true times for slow (MDT and BEAST; lines) and fast methods (MC2T; histograms) when (A, B) the correct lineage rate

slow methods are similar (9.4% vs. 8.9%; fig. 2C and D). Thus, one could conclude that slow and fast methods have their individual strengths, but they are overall similar in accuracies despite differences in calculation time needed.

Finally, we examined the performance of composite credibility intervals (cCrIs) yielded by fast and slow relaxed-clock methods because the actual model of rate variation among lineages is never known in real data analysis. For calculating cCrI for a given node in the tree, we use relaxed-clock methods with AR and with RR assumption to generate two CrIs. Then, the upper and lower bound for the cCrI is equal to the minimum of the two individual CrIs and the maximum of the two CrIs, respectively (Battistuzzi et al. 2010). Because of the fast computational speed of MC2T, this approach is more practical. The true time is included in cCrIs produced using MC2T in 89–98% of the simulated data sets (average = 96%), an improvement that is accompanied by a small widening of the confidence interval (0–12%; average = 4.5%). These results are similar to those reported previously for slower methods (97% accuracy with an average 10% CrI size increase).

In summary, slow and fast relaxed-clock methods produce comparable results at least for data sets with small number of sequences. This enables researchers to evaluate the robustness of the estimated divergence times not only to the assumptions made regarding the model describing the inequality of lineage rates but also to the taxon sampling and calibration boundaries. A more thorough application of molecular clocks will ultimately increase the confidence that scientists have in the estimated divergence times and result in more accurate timelines of species divergences and gene duplications.

Methods

We used previously simulated sequences, phylogeny, and calibration points (Battistuzzi et al. 2010) for evaluating the accuracy and speed of performance of the MC2T program (Yang 2007). MC2T analyses were carried out on a series of Intel Q8400 (2.66 GHz; 6–8 GB of memory; Windows 7) using the MC2T version with soft boundaries on calibrations; very similar results were obtained for the version that implements hard boundaries. The calculation times for MC2T included the time taken in all steps from sequence and tree inputs into PAML to the final timetree output. These MC2T computational times were compared with those of MDT and BEAST (v1.4.8), which were run on comparable computers (Intel Xeons 2.66–3.20 GHz; 4–24 GB of memory; UNIX). For these methods, computational times included the time taken for estimating branch length and divergence

variation (model match) and (C, D) the incorrect model (model violation) were used. Graphs in the inset show histograms for the absolute values of the normalized differences. Numbers in parentheses are the mean and standard deviations of the distribution (in order). In the main panels, a negative value on the x axis shows the degree of underestimation, and positive values indicate overestimations.

times. For making direct comparisons, the true topology was fully constrained in BEAST, MC2T, and MDT analyses, such that only the branch length and divergence times were optimized, using an exact and an approximate likelihood estimation in BEAST and in MC2T/MDT, respectively. In these analyses, chain lengths for the three methods were kept at default values (MC2T at 50,000; MDT at 1,000,000; and BEAST at 10,000,000) because it is not possible to have prior knowledge of more appropriate values in practice. All programs reported reaching convergence with these defaults. We also estimated the accuracy of time estimates by increasing the numbers of independent samples (effective sample size values), which did not improve estimated times for the simulated data in our case (see *supplementary information, Supplementary Material* online).

Following Battistuzzi et al. (2010), we used three pairs of calibrations selected at different heights in the tree and described by perfect boundaries (uniform distribution of ± 1 My around the true time) (fig. 1A). MC2T was used with approximate likelihood, and parameters coincided with those used to simulate the sequences. In MC2T, multiple values for the birth–death priors were tested following Yang (2006), and the combination of values that produced intermediate results and simulated more closely the birth–death process in MDT was chosen ($\lambda = 2$, $\mu = 2$, $\rho = 0.1$) (Yang 2006; Yang and Rannala 2006; Rannala and Yang 2007). The fine-tuning step was repeated until values were within the range of 0.2–0.4. Other parameters (*rgene_gamma*, *sigma2_gamma*) correspond to the values used in MDT and BEAST, when applicable, and were optimized for each replicate independently.

We also conducted some initial exploration of the effects of prior distribution on the estimated times because of their importance in the Bayesian framework. We compared the prior distributions of MC2T with those of MDT for a subset of our simulations (ten concatenated gene data sets and ten rate-synchronized data sets). As reported by others, MC2T showed older priors for the deepest nodes and younger priors for the shallower nodes as compared

with MDT (Inoue et al. 2010), with 95% CrIs of the priors being narrower in MDT by an average of 20%. However, the posterior distributions were similar for these two methods.

Supplementary Material

Supplementary material and figure S1 are available at *Molecular Biology and Evolution* online (<http://www.mbe.oxfordjournals.org/>).

Acknowledgments

We thank Alan Filipski for helpful comments and discussions and Carol Williams for text editing support. This study was supported by the National Science Foundation (DBI-0850013) and National Institute of Health (HG002096-09) to S.K. and by funds from the Howard Hughes Medical Institute through the Undergraduate Science Education Program and from the Arizona State University School of Life Sciences.

References

- Battistuzzi FU, Filipski A, Hedges SB, Kumar S. 2010. Performance of relaxed-clock methods in estimating evolutionary divergence times and their credibility intervals. *Mol Biol Evol*. 27:1289–1300.
- Drummond AJ, Rambaut A. 2007. Beast: Bayesian evolutionary analysis by sampling trees. *BMC Evol Biol*. 7:214.
- Inoue J, Donoghue PCJ, Yang Z. 2010. The impact of representation of fossil calibrations on bayesian estimation of species divergence times. *Syst Biol*. 59:74–89.
- Rannala B, Yang Z. 2007. Bayesian estimation of species divergence times from multiple loci using multiple calibrations. *Syst Biol*. 56:453–466.
- Thorne JL, Kishino H. 2002. Divergence time and evolutionary rate estimation with multilocus data. *Syst Biol*. 51:689–702.
- Yang ZH. 2006. Computational molecular evolution. Oxford: Oxford University Press.
- Yang ZH. 2007. Paml 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol*. 24:1586–1591.
- Yang ZH, Rannala B. 2006. Bayesian estimation of species divergence times under a molecular clock using multiple fossil calibrations with soft bounds. *Mol Biol Evol*. 23:212–226.