

BMC Evolutionary Biology

Improving the performance of Bayesian phylogenetic inference under relaxed clock models --Manuscript Draft--

Manuscript Number:	EVOB-D-19-00107R1	
Full Title:	Improving the performance of Bayesian phylogenetic inference under relaxed clock models	
Article Type:	Methodology article	
Section/Category:	Phylogenetics and Phylogeography	
Funding Information:	Chinese Government Scholarship (201706990021)	Mr. Rong Zhang
	Marsden Fund (16-UOA-277)	Professor Alexei Drummond
Abstract:	<p>Background: Bayesian MCMC has become a common approach for phylogenetic inference. But the growing size of molecular sequence data sets has created a pressing need to improve the computational efficiency of Bayesian phylogenetic inference algorithms.</p> <p>Results: This paper develops a new algorithm to improve the efficiency of Bayesian phylogenetic inference for models that include a per-branch rate parameter. In a Markov chain Monte Carlo algorithm, the presented proposal kernel changes evolutionary rates and divergence times at the same time, under the constraint that the implied genetic distances remain constant. Specifically, the proposal operates on the divergence time of an internal node and the three adjacent branch rates. For the root of a phylogenetic tree, there are three strategies discussed, named Simple Distance, Small Pulley and Big Pulley. Note that Big Pulley is able to change the tree topology, which enables the operator to sample all the possible rooted trees consistent with the implied unrooted tree.</p> <p>To validate its effectiveness, a series of experiments have been performed by implementing the proposed operator in the BEAST2 software.</p> <p>Conclusions: The results demonstrate that the proposed operator is able to improve the performance by giving better estimates for a given chain length and by using less running time for a given level of accuracy. Measured by effective samples per hour, use of the proposed operator results in overall mixing that is up to an order of magnitude more efficient than the current operators in BEAST2 on both real and simulated data sets.</p>	
Corresponding Author:	Rong Zhang The University of Auckland School of Computer Science NEW ZEALAND	
Corresponding Author E-Mail:	rzha419@aucklanduni.ac.nz	
Corresponding Author Secondary Information:		
Corresponding Author's Institution:	The University of Auckland School of Computer Science	
Corresponding Author's Secondary Institution:		
First Author:	Rong Zhang	
First Author Secondary Information:		
Order of Authors:	Rong Zhang	
	Alexei Drummond	
Order of Authors Secondary Information:		
Response to Reviewers:	<p>Response to the editor and reviewer</p> <p>We greatly appreciate the editor and the reviewer for the efforts and the valuable suggestions and hope that deficiencies pointed out in the original submission are overcome in the revised version. Our responses of the Referee's Report are given below.</p>	

Response to reviewer #1

General comments

1. p2, l7-8 It is not clear here if better estimates of divergence times have been obtained in the referenced papers or if this is simply an expectation

Author's Response:

Thanks for your comment.

To emphasize the proposed operator is applied to a relaxed clock model, we expect that relaxed models can help better explain genetic distances by allowing rates to vary throughout the tree, after reading the referred literatures. In the revised manuscript, we have rewritten the statements to avoid confusions.

2. p6, l44-45 On a dataset with sampling-through-time, the leaf node does not have to be the younger child. Does the operator still work in that case or is it a requirement that $tO > tY$?

Author's Response:

Thank you for pointing out this special case.

After careful considerations and tests, we confirmed that the operator should still work properly. So, we modified the requirements both in the code and revised manuscript. To be specific, in an asymmetric tree, we use the term "extinct child" to represent the node having no child nodes, which may be a sampled ancestor or a tip. It is denoted by node Y and its node time is tY . Similarly, the term "extant child" refers to the node having child nodes, which is denoted by node O and its node time is tO . Therefore, there is no deterministic relationship between tO and tY .

However, in order to make the node times valid after the operation, it is necessary to require that $tY < tO' < tX'$, because node Y will become the child node of node O in the proposed tree.

3. There is a mismatch between Figures 12 and 13 which show results for S3 and Table 5 which contains the standard deviation of the clock, which according to Figure 11 is S1. Similarly p9, l26 it is unclear which of S1 or S3 was actually measured.

Author's Response:

We are sorry for making these mistakes.

In Figure 8 (Figure 11 in original manuscript), we used s_1 to denote the standard deviation of the clock (abbreviated to $Ucldstdev$), and used s_3 to denote the standard deviation of the LogNormal prior of s_1 , so that the framework of well-calibrated simulation study can be simple enough for readers to understand. In the revised manuscript, we have added more descriptions to explain the notations below the figure.

Additionally, in order to better identify what parameters are compared in Figure 9 and Table 1 (Figure 12, Figure 13 and Table 5 in original manuscript), we used the more detailed terms such as "Ucldstdev", "Kappa" and "Birth rate".

To make it clear which parameter is actually measured when comparing the efficiency, we also used the detailed term, such as "ucld.mean" (the mean of clock) and "tree.length" (total branch lengths in substitutions), in Figure 10 in the revised manuscript.

4. p9 l33-34 I suggest adding some details on why configuration 1 was chosen as a comparison.

Author's Response:

Thank you for your professional suggestion.

Configuration 1 uses discrete rate categories to approximate the branch rates under certain distribution, which is default setting in BEAST2. But the proposed operators work on absolute continuous rates. We compared the efficiency under these two configurations to show the proposed operators is able to provide better performance than current operators in BEAST2.

In the revised manuscript, we have added more details to clarify the configurations in Section Performance comparison.

5. p10, l16 I don't understand what "the sampled trees were filtered by the shared common ancestor of each taxa" means. Later l22 it appears that the authors discarded the trees which did not match the reference topology, but it's unclear if that is the same thing as the previously mentioned filtering or an entirely different step of the process.

Author's Response:

Thank you for your comment.

In this section, we ran a MCMC chain with length 200000000 using BEAST2 and got 2001 sampled trees from the whole chain. Then, we used the maximum clade credibility tree (Figure 11 (a)) to filter the 2001 sampled trees. Namely, we obtained the trees that have the same topology and clades as the maximum clade credibility tree, which means the trees that do not match the maximum clade credibility tree are discarded. Hence, the correlation analysis was conducted by using the branch rates and node times in the trees that match the maximum clade credibility tree.

To make it clear, we have rewritten the statement of the process to avoid multiple lines describing similar content.

6. p10, l37-42 Is it really unexpected that the results obtained using the new operators would be consistent with the correlations enforced by these new operators ? Overall it's unclear what the conclusions from this section are.

Author's Response:

Thank you for your comment and question.

First of all, the correlations are not enforced by the proposed operators. Instead, the correlations are reflected by the data set, since the genetic distances are provided by the sequence alignment. And according to the molecular models, the genetic distances are product of evolutionary rates and divergence times, which gives the correlations. In this section, we analyzed posterior distributions sampled by the proposed operators. The results show that the correlations are apparent, which means the proposed operators are better at sampling the rates and times.

7. p11, l8-23 It is very unclear what was actually done in this section. For instance, were the assigned divergence times and rates fixed in the analysis? If so, how were they calculated, and if not, what are the values shown in figure 17? How were the summary trees obtained?

Author's Response:

Thank you for your question.

In this section, we aim to investigate the performance of the proposed operator by giving the genetic distances directly. What we did in this section includes: 1) getting the unrooted phylogenetic tree from sequence alignment, which provides genetic distances among taxa; 2) sampling the unrooted tree by using the proposed operators; 3) comparing the results of sampling the unrooted tree with the results of sampling the sequence alignment.

Since the operators maintain the genetic distances among taxa and the underlying unrooted tree, they should be able to provide the consistent rates and divergence times with those obtained by sampling from the sequence alignment, if the genetic distances and unrooted tree are accurate.

More specifically, we used an online programme PhyML 3.0 to estimate the maximum-likelihood tree based on the ratites data set. Figure 12(a) (Figure 17(a) in original manuscript) shows the unrooted estimated maximum-likelihood tree with genetic distances on the branches.

Then, as an initial state of the MCMC chain, we assigned the root, divergence times and rates on the estimated maximum-likelihood tree to make it a valid rooted time tree, and make sure that the genetic distances of the rooted time tree match those of the estimated maximum-likelihood tree. Hence, these initially assigned divergence times and rates are not fixed and will be sampled by the proposed operators.

Afterwards, we used the proposed operators only to sample the initialised rooted time tree, so that the rates and divergence times are sampled, but the genetic distances are always the same as the estimated maximum-likelihood tree.

Finally, after the simulation finished, we used the programme TreeTraceAnalysis and obtained the three unique tree topologies from the sampled trees, which is shown in Figure 12 (b), (c), (d). Moreover, we used the programme TreeAnnotator to summarise the sampled trees having the three unique tree topologies. The posterior mean of each branch rate and the 95% HPD of each divergence time are also labelled in Figure 12 (b), (c), (d).

In the revised manuscript, we also cited several similar works that infer unrooted trees and then use those as data, including Bayesian MCMC and non-Bayesian methods.

8. Figures 12 and 13 are identical.

Author's Response:

In the latest manuscript, the well-calibrated simulation results of 20-taxa data set

(Figure 12 in the original manuscript) has been removed and the results of 120-taxon data set has been updated (Figure 13 in the original manuscript).

9. Figure 16b I suggest adding a legend showing the range of plotted values.

Author's Response:

Thank you for your valuable advice.

In the revised manuscript, we have updated Figure 11(b) (Figure 16(b) in the original manuscript) by plotting a legend that shows the range of Pearson correlation coefficients.

10. I would put a time axis on figure 17, as the credible intervals shown are not very meaningful without it.

Author's Response:

Thank you for your comment.

In the revised manuscript, we have put a time axis below each tree in Figure 12 (Figure 17 in the original manuscript).

Minor comments

11. p1, l30 It is noticed that -> Note that

12. p2, l39 Preliminaries -> Preliminaries

13. p4, l37 max(tj, tk) -> max(tL, tR)

14. p4, l53 the total genetic distance dL and dR -> the total genetic distance dL + dR

15. p6, l54 what happens if tO' < min(tG1, tG2) ?

16. p9, l26 MCMC simulation -> MCMC inference

17. p10, l48 becomes -> become

18. Figure 1 remove the box which only contains =, Figure 1 H3,R and H4,R should be H3,L and H4,L

19. Figure 11 substitution tree -> substitution tree

20. Figure 16 compare -> comparison

Author's Response:

After carefully reviewing the original manuscript, the following mistakes that you pointed out have been corrected in the revised manuscript. (Figure 11 in the original manuscript becomes Figure 8 in the revised manuscript.) (Figure 16 in the original manuscript becomes Figure 11 in the revised manuscript.)

Response to reviewer #2

Major Issues

1. More background on the operator is required.

(1) It is necessary to explain why preserving genetic distances is the goal of the operator. That is, to point out that the transition probability matrix for a branch is $\exp(Qrt)$ so holding $d = rt$ constant does not change the likelihood along that branch, and thus requires no re-computation of any partial likelihoods, speeding up MCMC.
(2) Similarly, a brief introduction to the notion of an underlying unrooted phylogenetic tree would be useful for understanding the Pulley operators.

Author's Response:

Thanks for your comment.

In the revised manuscript, we have added the necessary backgrounds you mentioned. The details are shown as follows.

(1) The reason why the proposed operator maintains genetic distances is explained in Section Tree proposals in the revised manuscript.

(2) To explain the underlying unrooted tree, we added a new subfigure Figure 3(a) in the revised manuscript.

2. Small Pulley and Big Pulley can only be used on reversible CTMC models where unrooted trees can be used in inference. This is not a huge limitation in practice, but it should be mentioned.

Author's Response:

Thanks for pointing out this detail.

In the revised manuscript, we claimed this limitation in Discuss section.

3. The description of the asymmetric case in Big Pulley appears to assume that the younger child is a tip, but this is only a given if the tree has no heterochronous samples (which are increasingly common in real datasets). At a quick glance, it appears that the

move could still work in this case but would require $t_Y < t_O' < t_X'$ and not just $t_O' < t_X'$.

Author's Response:

Thanks for your feedback.

After careful thoughts, we confirm that it is not necessary to assume the younger child node to be a tip. To make it clear, we made a statement that node O refers to the node having child nodes and node Y refers to the node having no child nodes, in the revised manuscript. And we use the term "extant" and "extinct" to describe node O and Y respectively, instead of "older" and "younger".

Besides, we have also modified the requirement of the proposed node times so that $t_Y < t_O' < t_X'$ should be satisfied.

Finally, the plots of asymmetric tree shapes in Figure 5 and Figure 7 have been revised so that node Y does not look like a tip.

4. Additional information is required about the simulation study.

(1) What priors were used for inference? Especially important is the prior on the root age.

(2) What other operators were used on the tree and the branch rates? These are the only operators that can change the underlying unrooted phylogeny, which makes them crucial to performance.

Author's Response:

Thanks for your professional questions.

(1) The priors used in the well-calibrated simulation study are basically presented in the framework in Figure 8. To be more specific, the priors include: (a) a Yule model tree prior where the birth rate has a $\text{LogNormal}(M2=10, s2=0.3)$ distribution as prior, (b) base frequency having a Dirichlet ($\text{alpha}=10$) distribution as prior, (c) kappa having a $\text{LogNormal}(m4=1.0, s4=0.2)$ distribution as prior, (d) branch rates having a $\text{LogNormal}(M1=1, s1)$ distribution as prior, and (e) standard deviation of rates prior having a $\text{LogNormal}(m3=-1.5, s3=0.35)$ as hyper prior.

To answer your question, the root age (t_{root}) is the sum of interval lengths between two speciation events (tao_i). Under a Yule model, tao_i follows an Exponential distribution with rate $i * \lambda$, where the birth rate λ has a $\text{LogNormal}(M2=10, s2=0.3)$ prior distribution. Therefore, the prior distribution of the root age can be represented by

$t_{\text{root}} = \sum(tao_i)$, where $i=1,2,\dots,N$ and $tao_i \sim \text{Exponential}(i * \lambda)$.

(2) There are two other operators used to sample the branch rates, i.e. a random walk operator and a swap operator.

The underlying unrooted phylogeny is changed by the following operators: a SubtreeSlide operator, a WideExchange operator, a NarrowExchange operator, a WilsonBalding operator.

(3) For more details of the well-calibrated simulation study, readers can visit our GitHub repository and find the corresponding .xml file from the link below.

(https://github.com/Rong419/OperatorPaper/validation/calibrated/cal_val_120_template.xml)

5. More information is needed when discussing the performance of the new operator.

(1) What were p and q (from Figure 1), the proportion of root operations for Simple Distance and Small Pulley?

(2) Without discussing operator weights, it is difficult to interpret the change in run time cost due to the Constant Distance operator. Discussing time required per operator may be clearer still, allowing comparison directly between node age proposals.

Author's Response:

Thank you for your valuable advice.

(1) In the original manuscript, p and q were used to denote the proportion of weights of Simple Distance and Small Pulley. To avoid confusions, we have removed p and q from Figure 1 in the revised manuscript. But we also claimed that we assigned equal weights on operations to all internal nodes (including the root) in the Discussion section.

(2) To give details about operator weights, we have added a new Table 7 in Appendix section in the revised manuscript to show weights on operators in the simulations of analysed data sets.

6. Figures 12 and 13 appear to be completely identical, it would appear that the 20-taxon figure was duplicated.

Author's Response:
We have deleted the redundant figure of the 20-taxa results, and the revised manuscript only shows the results of 120-taxa data set.

Minor Issues

7. In the preliminaries, there are some issues with switching between parameterizations in terms of node times, t, and in terms of the tree, g.
(1) The change from $\text{Pr}(g)$ in equation 1 to $\text{Pr}(t | \Phi)$ in equation 2 is a bit jarring and equation 2 is less general. $\text{Pr}(t | \Phi)$ assumes independence between tree topology and divergence times, which is not always the case (for example the model of Barido-Sottani et al. (2018)).
(2) Page 3, lines 52-54 refer to proposing a tree g' , whereas page 3 line 38 states the operator works on times.
(3) Readers will have an easier time if one parameterization is used consistently. I personally see no strong argument in favor of $\text{Pr}(t | \Phi)$, $\text{Pr}(g | \Phi)$ still allows the use of the vector of node times, t.

Author's Response:

Thank you for helping us find the issues.

In the latest manuscript, we have carefully dealt with these issues. The details are listed below.

- (1) We have modified Equation 2 and make it more appropriate.
Firstly, we introduce the notations of probability density in Equation 1. Then, Equation 2 is written by using the forms of conditional probability.
(2) To make it clear, we claim that g represents the phylogenetic tree in current state, and g' represents the phylogenetic tree in proposal state. And the phylogenetic tree is an edge graph E and a set of node times t , i.e. $g = \{E, t\}$. After the proposed operations, g' could be different because the times are different or the edge graph, or both. In the revised manuscript, we have rewritten the statement that the operations on internal node propose one node time in the set of times and three branch rates in the set of rates. And the tree topology remains the same (E does not change). Moreover, we have also clearly stated what is exactly proposed by the operator in the rest of manuscript.
(3) To make our manuscript more readable, we have used parameters r, t, g to represent the branch rates, divergence times and tree topology in the original state respectively. And parameters r', t', g' are used to represent the branch rates, divergence times and tree topology in proposed state respectively. In the revised manuscript, we have also used the bold style to represent the vector of parameters, for instance, the vector of all divergence times is denoted by t .

8. In Small Pulley there are some issues with clarity.

- (1) The statement "Small Pulley proposes a new genetic distance of a branch on one side of the root" is somewhat misleading, as it in fact proposes new distances on both sides of the root (by proposing a single number and using it to change both).
(2) It would help to introduce $D = d_L + d_R$ around page 4 line 53 and then state that d_R will be adjusted simultaneously so as to preserve D .

Author's Response:

Thank you for helping us make our manuscript clearer and straightforward.

It is true that Small Pulley proposes one genetic distance (d_L) and changes distance of the other branch (d_R), so as to maintain the sum of the two distances ($d_L + d_R$). In the revised manuscript, we have modified the statement and introduced $D = d_L + d_R$.

9. In Big Pulley there are some issues with clarity.

- (1) Explaining Exchange() before the moves is important, but the sentence "Firstly, a method called Exchange is designed to propose a new tree topology" is confusing when in fact calling Exchange() is step 3.
(2) The description of symmetric tree step 3 (page 6 lines 6-7) is confusing, as 50% of the time we will apply the method to L and either child of R.
(3) In equation 8, presumably d_1 is d_H1 , but this is not stated. Equation 10 uses d_G1 instead of d_1 , which seems more clear.

Author's Response:

Thank you for your comment and suggestions.

- (1) In the revised manuscript, we have modified the descriptions when introducing the Exchange method.
(2) We have eliminated the confusing description and made it clear that the Exchange

method will be applied to the selected node and one of its sibling's child nodes.
(3) In the revised manuscript, we have replaced the unclear notations of distances "d1, d2" with "d_H1, d_H2", so that it is explicit to understand the notations in Equation 8.

10. In the section, "Correlation analysis of rates and node times," there are some issues.

(1) A statement of motivation for this section is needed: what purpose does this experiment serve?

(2) The current comparison scheme is difficult to interpret. The rate-to-rate and age-to-age correlations do not seem to be important, but take up more of the figure than the important comparisons. It would be simpler to directly compare branch lengths to the rates of those branches, perhaps by taking the Pearson correlation coefficient of length and rate across the posterior. Branches could be matched across trees much as they currently are. The results could be presented as a histogram or a heatmap as is currently done.

(3) The statement, "With full length genomes now available, this limiting case might be approached in some data sets," ignores the complexities involved in inferring trees from genomes and requires assuming both a single topology across all loci in a genome (ignoring, for example, incomplete lineage sorting) and shared branch lengths at different loci (which need not be the case partitioning the dataset for analysis, see for example Lanfear et al. (2012)).

Author's Response:

Thank you for your professional comment.

(1) In the manuscript, we have claimed the motivation of the conducted correlation analysis in the beginning of subsection Correlation analysis of rates and branch lengths. And we have explained our motivation when discussing the results.

(2) In the revised manuscript, we have updated correlation analysis by plotting the coefficient between branch length and rates.

(3) We have corrected our original statement. In the revised manuscript, we have cited the referred work to show that there exist details in inferring trees from genomes.

Nevertheless, this paper uses this approach as a simple test to demonstrate that the operators are useful in sampling rates and divergence times in relaxed clock models.

11. In the appendix there are some issues with clarity.

(1) The relationship between son/dau and L/R is unclear. This makes understanding Algorithm 1 difficult.

(2) The section on sampling from the prior needs an overview to explain, briefly, the motivation, design, and goals of the experiments.

Author's Response:

Thank you for your comment.

(1) In the manuscript, we removed notations "dau/son" and use "L/R" to denote the two child nodes of the root in Big Pulley, so that the notations are consistent throughout the whole manuscript and easier for readers to understand.

(2) We have added a paragraph to briefly explain the motivation, design, and goals of the experiments in Section Sampling from the prior in the revised manuscript.

12. The numbering on the figures and tables is perplexing. A number of tables and figures are only referenced from the appendix but have lower numbers than main-text figures and tables. This makes it seem as if one has accidentally skipped portions of the manuscript when reading through it.

Author's Response:

We apologize for the disordered figures and tables in the original manuscript.

In the revised manuscript, we have made the numbering on the figures and tables consistent with the referred order in the main text and appendix. Namely, Figure 1 – Figure 12 and Table 1 – Table 2 belong the main text, Figure 13 – Figure 24 and Table 3 – Table 7 belong to Appendix.

13. The proposal to infer unrooted trees and then use those as data is interesting. Some discussion of related approaches (see below) is in order.

(1) Thorne and Kishino (1998), Guindon (2010), and dos Reis and Yang (2011) perform a pre-MCMC step to approximate the likelihood surface of the underlying unrooted phylogeny, bypassing the need for the pruning algorithm but allowing for changes to the genetic distances.

(2) Non-Bayesian methods such as TreeTime (Sagulenko et al. 2018), r8s (Sanderson

2003), and LSD (To et al. 2015) use an unrooted phylogeny as data to estimate the time tree.

Author's Response:

Thank you for providing us these important literatures.

We have added a new paragraph in "Section Sampling a fixed unrooted tree" about these referred works in the revised manuscript.

Typos and Other Minor Comments

14. (1) While the operators as discussed in this paper are, to my knowledge, novel, others have used operators similar to the proposal on internal node heights (e.g. <https://github.com/revbayes/revbayes/blob/master/src/core/moves/compound/RateAgeBetaShift.cpp>)

(2) I wonder if there may be efficiency gains by employing proposals other than a uniform, such as a bactrian proposal (Yang and Rodriguez 2013)

Author's Response:

Thank you for providing us a new idea.

(1) After reviewing the code in the referred link and the descriptions in the referred paper, we found that the authors define a certain distribution for the internal node height, such as a Beta distribution and a Bactrian distribution, which indicates the probability of the new node height.

(2) However, the operators introduced in our manuscript use a Uniform distribution to move the node height uniformly on the branch. To make comparisons, we ran the simulations by using the three different proposals. The results are shown in the Appendix3.4 in the revised manuscript.

15. The proposed operator is discussed in the context of uncorrelated clock models, but it should also be applicable to autocorrelated models like that of Thorne and Kishino (1998).

Author's Response:

Thank you for your comment.

We agree that the proposed operator is also able to work in auto-correlated models. In the revised manuscript, we claimed that the proposed operator can be applied to any relaxed clock models.

16. The choice of kappa in the simulation study is somewhat strange, as usually the transition-transversion rate-ratio is expected to be above 1.

Author's Response:

Thank you for your suggestion.

In the revised manuscript, we have updated the results of calibrated-simulation study after rerunning the simulations by choosing a proper prior of kappa, i.e. LogNormal($m4=1.0$, $s4=0.2$), the mean of which is around 2.77.

17. It is somewhat perplexing that fewer of the 120-taxon simulations had the mean rate in the 95% CI.

Author's Response:

Thank you for comment.

In the revised manuscript, the well-calibrated simulation for 120 taxa was performed by using the latest code. The result shows that the mean rate has 100 percent coverage.

18. (1) Page 2 lines 7-8: The sentence "By allowing rates" is somewhat unclear as currently phrased.

(2) Page 2 line 24, the statement "since each step in the chain requires a likelihood calculation" is somewhat misleading, with cached partial likelihoods many moves only require parts of the likelihood to be re-evaluated.

Author's Response:

Thanks for your suggestion.

We have deleted the unclear and misleading statement in the revised manuscript. The details are as follows.

(1) "By allowing rates" is deleted. We stated that relaxed clock models where rates vary in the phylogenetic tree have been investigated.

(2) "since each step in the chain requires a likelihood calculation" is deleted. We explained the constant likelihood by fixing the genetic distance, which avoids recalculating any partial likelihood.

19. (1) In "Simple Distance" (page 4 line 38), t_i and t_j should be t_R and t_L.

(2) Page 5 line 20 should "rooted" be "unrooted"?

(3) Page 10 line 18, taxa should be taxon

(4) The axis label "number of runs" for Figures 12 and 13 might be more clear as something like "replicate" or "simulation number."

Author's Response:

Thanks for your correction.

In the revised manuscript, we have corrected the mistakes that have been found.

20. Page 10 line 15 states "After analyzing the ratite dataset," but this dataset has not been previously mentioned.

Author's Response:

Thanks for your comment.

In the revised manuscript, we have added a brief introduction of the ratite data set before describing the analysing process.

21. In figures 14 and 15, the same color scheme is used but the meanings of the colors are different. It would be easier to follow if different colors were used in these figures.

Author's Response:

Thank you for your suggestion.

We have paid attention to modifying the colour schemes in the revised manuscript.

(1) Figure 14 and Figure 15 in the original manuscript that shows the efficiency have become Figure 10 in the revised manuscript. Moreover, we compared the efficiency using 6 data sets and all the sampled parameters. Therefore, 6 different colours are used to represent different data sets.

(2) In Figure 17 – Figure 21, we used another colour scheme to represent the two configurations.

(3) In Figure 22 – Figure 24, we used the third colour scheme to represent the three different proposal methods (See Response 14(2)).

We greatly appreciate the reviewer for the valuable suggestions. We try our best to overcome the deficiencies pointed out in the original submission. If there are any problems in the revised version, please do not hesitate to point out. We will revise the submission according to reviewer's suggestions.

Additional Information:

Question

Response

Has this manuscript been submitted before to this journal or another journal in the BMC series</ a>?

No

Click here to view linked References
Zhang and Drummond

1
2
3
4

METHODOLOGY ARTICLE

Improving the performance of Bayesian phylogenetic inference under relaxed clock models

Rong Zhang^{1*} and Alexei Drummond^{1,2}

*Correspondence:
rzha419@aucklanduni.ac.nz
¹School of Computer Science,
University of Auckland, Auckland,
New Zealand
Full list of author information is
available at the end of the article

Abstract

Background: Bayesian MCMC has become a common approach for phylogenetic inference. But the growing size of molecular sequence data sets has created a pressing need to improve the computational efficiency of Bayesian phylogenetic inference algorithms.

Results: This paper develops a new algorithm to improve the efficiency of Bayesian phylogenetic inference for models that include a per-branch rate parameter. In a Markov chain Monte Carlo algorithm, the presented proposal kernel changes evolutionary rates and divergence times at the same time, under the constraint that the implied genetic distances remain constant. Specifically, the proposal operates on the divergence time of an internal node and the three adjacent branch rates. For the root of a phylogenetic tree, there are three strategies discussed, named Simple Distance, Small Pulley and Big Pulley. Note that Big Pulley is able to change the tree topology, which enables the operator to sample all the possible rooted trees consistent with the implied unrooted tree. To validate its effectiveness, a series of experiments have been performed by implementing the proposed operator in the BEAST2 software.

Conclusions: The results demonstrate that the proposed operator is able to improve the performance by giving better estimates for a given chain length and by using less running time for a given level of accuracy. Measured by effective samples per hour, use of the proposed operator results in overall mixing that is up to an order of magnitude more efficient than the current operators in BEAST2 on both real and simulated data sets.

Keywords: Bayesian MCMC; Bayesian phylogenetics; Proposal kernel; Genetic distances; Divergence times; Evolutionary rates

Background

Bayesian phylogenetics puts an emphasis on estimating a probability distribution over parameters of interest, including the phylogenetic tree topology and divergence times, given the data. The Metropolis-Hastings Markov chain Monte Carlo (MCMC) [1, 2] algorithm has been the primary computational tool used in Bayesian phylogenetics for sampling from the posterior distribution. This paper is aimed at improving the performance of the relaxed clock model in Bayesian phylogenetic analysis.

Early implementations of Bayesian phylogenetic inference assumed a strict molecular clock where the evolutionary rates are the same at every branch [3]. This was the preferred method for estimating divergence times [4, 5]. The introduction of relaxed molecular clocks allowed for the estimation of divergence times [6] and phylogeny [7] in the presence of rate heterogeneity among branches. Since then, the

63
64
65

6 relaxed clock model has been widely applied, such as the study of Nothofagus [8]
7 and flowering plants [9]. Many aspects of the performance and accuracy of relaxed
8 clock models have subsequently been investigated (e.g. [10], [11]).

9 Bayesian phylogenetic inference via MCMC is computationally intensive for large
10 data sets. Two approaches to improve efficiency are (i) by making faster likelihood
11 calculations, and (ii) by incorporating more effective proposal kernels. Calculating
12 the phylogenetic likelihood is computationally expensive. Hence, researchers have
13 tried many ways to tackle the computation burden in the likelihood calculations,
14 such as detection of repeating sites [12], approximate methods (e.g. [13, 14]) and
15 the use of parallelisation strategies (e.g. BEAGLE [15]).

16 However the overall efficiency of the sampling process also depends strongly on
17 the construction of the proposal mechanism. An effective proposal mechanism is
18 proficient at exploring the posterior distribution, and can do so with fewer steps
19 in the MCMC chain. Therefore fewer likelihood calculations are required, since
20 each step in the chain that changes the tree or substitution parameters requires a
21 likelihood calculation.

22 A major limitation in Bayesian MCMC analysis of phylogeny lies in the efficiency
23 with which operators sample the tree space [16, 17]. Fast and reliable estimation
24 is dependent on a good mixture of operators, since the posterior distribution often
25 exhibits correlations between the tree and other random variables.

26 In this paper, we present a novel operator that works alongside standard operators
27 by proposing moves within a subspace of constant genetic distances. Namely, the
28 proposed operator changes both divergence times of nodes and neighbouring branch
29 rates so that the implied genetic distances are not changed. For time-reversible
30 substitution models the phylogenetic likelihood will also be unchanged under this
31 operation. The proposed operator has been implemented and tested in BEAST2
32 [18].

41 Preliminaries

42 Bayesian MCMC

43 Let \mathbf{D} , g and Φ denote the data, phylogenetic time-tree and a set of evolutionary
44 parameters respectively. The time-tree $g = \{E, \mathbf{t}\}$ consists of a directed edge graph,
45 E , defining a rooted tree topology on a set of labelled taxa and a set of associated
46 divergence times \mathbf{t} (for details see e.g. [19]). The posterior probability density can
47 be calculated using equation 1. It consists of prior distributions for the tree and the
48 parameters, a phylogenetic likelihood that conveys information from data, and the
49 posterior distribution to be inferred. These are denoted in the form of probability
50 densities by $p(g)$, $p(\Phi)$, $p(\mathbf{D}|g, \Phi)$, $p(g, \Phi|\mathbf{D})$ respectively. From a Bayesian per-
51 spective, the phylogenetic trees and the parameters are random variables described
52 by a posterior probability distribution given the observed data \mathbf{D} .

53

$$54 p(g, \Phi|\mathbf{D}) = \frac{p(\mathbf{D}|g, \Phi)p(g)p(\Phi)}{p(\mathbf{D})} \quad (1)$$

55

56 However, due to the state space being high dimensional and the marginal like-
57 lihood being infeasible to calculate, MCMC is adopted to sample the posterior

distribution. Specifically, MCMC algorithms construct a Markov chain whose stationary distribution is the posterior distribution $p(g, \Phi | \mathbf{D})$, in such a way that the computation of the marginal likelihood $p(\mathbf{D})$ is avoided.

Tree proposals

We use the term “operator” to describe an algorithm that can be used to draw a new state θ' given an existing state $\theta = \{g, \Phi\}$ from a specific proposal kernel $q(\theta'|\theta)$ and also return the Hastings-Green ratio for the proposed state transition [2, 20].

Standard naïve operators such as the random walk operator propose the new state θ' by adding a random variate to a component of the current state θ [21]. Similarly, scale operators multiply a subset of the current state by a random scale factor [22]. They are suitable for working on a single random variable, or a single component of the model, for example the population size parameter of the coalescent tree prior. Standard operators for the tree topology and divergence times include the subtree slide operator, Wilson-balding and narrow exchange operators [19, 23].

In this paper, the novelty of the proposed operators lies in maintaining the genetic distance d while changing the rate r and divergence time t . The reason is that the likelihood along one branch is constant if its distance is fixed, i.e. $d = r \times t$, noting that the likelihood is calculated based on transition probability matrix for each branch of $e^{\mathbf{Q}d_i}$, where d_i is the branch length in units of substitutions per site for branch i . In this way, the joint distribution on rates and divergence times can be explored without proposing states that would adversely affect the phylogenetic likelihood.

Uncorrelated relaxed clock model

Molecular clocks model how molecular sequences evolve along branches in the phylogenetic tree, so that a time tree can be reconciled with the genetic distances between sequences. In this paper, uncorrelated relaxed clock models are adopted, where the rates are drawn independently and identically from a given prior distribution, such as the log-normal distribution [7]. As a result, the rates can vary markedly between parent and child branches.

Referring to the Bayesian framework in equation (1), the joint inference of evolutionary rates \mathbf{r} and the time tree g can be obtained by the conditional distribution in equation 2:

$$p(g, \mathbf{r}, \Phi | \mathbf{D}) = \frac{p(\mathbf{D}|g, \mathbf{r}, \Phi)p(\mathbf{r})p(g)p(\Phi)}{p(\mathbf{D})}, \quad (2)$$

where $p(\mathbf{r})$ is the prior for rates specified in uncorrelated relaxed clock model. In the constructed Markov chain, the operator proposes a new state $\theta' = (\mathbf{r}', g', \Phi')$, from the original state $\theta = \{\mathbf{r}, g, \Phi\}$.

While the proposed operator is introduced based on uncorrelated clock models, it could equally be applied to any other relaxed clock that applies a rate parameter to each branch, such as autocorrelated clock models [6].

6

Methods

7
8
9
10
11
12
13
In this section, we define the Constant Distance Operator. Figure 1 illustrates the
flow chart of the proposed operators. In a phylogenetic tree, the node to operate on
is denoted by \mathbf{X} and the Constant Distance Operator works differently on internal
nodes and the root node. The details of the operations are introduced step by step
in the following subsections.

14

Operations on internal nodes

15
16
17
18
Figure 2 represents the tree (or subtree) with the node \mathbf{X} that is randomly selected
from among the internal nodes. Let g be the tree in the current state. The following
steps propose a new divergence time in g' and three rates in \mathbf{r}' .

19
20
Step 1 Identify the parent node and two child nodes of \mathbf{X} , denoted by \mathbf{P} , \mathbf{L} and
 \mathbf{R} respectively.

22
23
Step 2 Denote the nodes times of \mathbf{X} , \mathbf{P} , \mathbf{L} and \mathbf{R} by t_X , t_P , t_L , t_R respectively.
Denote the rates on the branches above the nodes by r_X , r_L and r_R respectively.

24
25
26
27
Step 3 Propose a new node time for \mathbf{X} by $t_{X'} \leftarrow t_X + a$, where a follows a Uniform
distribution with a symmetric window size w , i.e. $a \sim \text{Uniform}[-w, +w]$, for some
window size w . Make sure that the proposed time is valid, i.e. $\max\{t_L, t_R\} < t_{X'} <$
 t_P holds. Otherwise, we reject the proposal.

29
30
Step 4 Propose new rates by using equation 3.

31
32
$$r_{X'} = \frac{r_X \times (t_P - t_X)}{t_P - t_{X'}} \quad r_{L'} = \frac{r_L \times (t_X - t_L)}{t_{X'} - t_L} \quad r_{R'} = \frac{r_R \times (t_X - t_R)}{t_{X'} - t_R} \quad (3)$$

34
35
36
Step 5 Return the Green ratio α_{IN} (Refer to *Calculating the Green Ratio* in the
following subsection).

38

Operations on the root

39
40
41
42
43
44
45
46
47
48
49
50
We present three strategies for proposing the new rates and a new divergence time
for the special case when \mathbf{X} is the root node. i) The Simple Distance operator only
proposes a new root time. ii) Small Pulley adjusts the distances of branches on
both sides of the root. iii) Big Pulley proposes a new tree topology by rearranging
the root, without perturbing the unrooted tree. As is illustrated in Figure 3(a), all
the operations on the root, including Big Pulley that changes the tree topology, do
not change the underlying unrooted tree. For instance, no matter where the root X
is (either on branch EF or AE), the operators maintain the distances (d_{AB} , d_{AC} ,
 d_{AD} , d_{BC} , d_{BD} , d_{CD}) and preserve the unrooted tree at the same time.

51

Simple Distance

52
53
54
55
56
57
58
59
Figure 3 (b), (c) and (d) show the trees that are rooted at the node \mathbf{X} . The original
tree g in the current state is shown in Figure 3(b). Similar to the operations on
internal nodes, we will use the following steps to propose a new root time in g'
and two rates in \mathbf{r}' , as is illustrated in Figure 3(c). At the same time, the genetic
distances of two branches linked to the root, i.e. d_L and d_R , are kept constant
constant.

60
61
62
Step 1 Identify the child nodes of the root \mathbf{X} , denoted by \mathbf{L} and \mathbf{R} . Their corre-
sponding node times and branch rates are t_X , t_L , t_R and r_L , r_R .

63
64
65

6 Step 2 Propose a new node time for the root \mathbf{X} by $t_{\mathbf{X}'} \leftarrow t_{\mathbf{X}} + a$, where $a \sim$
7 Uniform[-w, +w]. Make sure that $t_{\mathbf{X}'} > \max\{t_L, t_R\}$ holds. Otherwise, we reject
8 the proposal.

9 Step 3 Propose new rates for branches on both sides of the root by using equation
10 4.
11

$$12 \quad r_L' = \frac{r_L \times (t_X - t_L)}{t_{\mathbf{X}'} - t_L} \quad r_R' = \frac{r_R \times (t_X - t_R)}{t_{\mathbf{X}'} - t_R} \quad (4)$$

13 Step 4 Return the Green ratio α_{SD} .

14 Small Pulley

15 In contrast to Simple Distance, Small Pulley changes genetic distances of branches
16 on both sides of the root. As is illustrated in Figure 3(d), two new rates in \mathbf{r}' are
17 proposed based on those in the original tree g . In order to maintain the total genetic
18 distance $d_L + d_R$ of the two branches linked to the root, after d_L' is proposed, d_R
19 will be adjusted simultaneously. In other words, Small Pulley keeps $D = d_L + d_R$
20 constant. The detailed process includes the following 4 steps.

21 Step 1 Identify the child nodes of the root \mathbf{X} , denoted by \mathbf{L} and \mathbf{R} . Their cor-
22 responding node times and branch rates are t_X , t_L , t_R and r_L , r_R . The implied
23 genetic distances of the two branches linked to the root can be calculated by:
24

$$25 \quad d_L = r_L \times (t_X - t_L) \quad d_R = r_R \times (t_X - t_R) \quad (5)$$

26 Step 2 Propose a new genetic distance for d_L by adding a random number that
27 follows a Uniform distribution, i.e. $d_L' \leftarrow d_L + b$, where $b \sim \text{Uniform}[-v, +v]$, for
28 some window size v . Make sure that $0 < d_L' < D$ holds. Otherwise, we reject the
29 proposal.

30 Step 3 Propose new rates for branches on each side of the root:

$$31 \quad r_L' = \frac{d_L'}{t_X - t_L} \quad r_R' = \frac{D - d_L'}{t_X - t_R} \quad (6)$$

32 Step 4 Return the Green ratio α_{SP} .

33 Big Pulley

34 Big Pulley resamples the rates and times while maintaining the implied unrooted
35 tree in units of genetic distance. So the genetic distances between the taxa are held
36 constant, but the location of the root in the time tree is readjusted.

37 Before describing the detailed steps, we introduce a method *Exchange* that pro-
38 poses a new root position. In Figure 4, let (i) \mathbf{X} denote the root of tree g , (ii) \mathbf{C}
39 and \mathbf{N} denote the two child nodes of \mathbf{X} , (iii) \mathbf{S} and \mathbf{M} denote the two child nodes
40 of \mathbf{C} . The *Exchange*(\mathbf{M} , \mathbf{N}) method involves the following steps:

- 41 • Swap the two nodes by pruning and regrafting, i.e. cutting \mathbf{M} (\mathbf{N}) at its
42 original position and attaching it to the original position of \mathbf{N} (\mathbf{M}).
- 43 • Propose $d_C' \leftarrow d_C + b$, where $b \sim \text{Uniform}[-v, +v]$. Make sure that $0 < d_C' <$
44 D holds, where $D = d_C + d_N$. Otherwise, we reject the proposal.

- 1
2
3
4
5 • The distances on the other three branches, i.e. d_S , d_M and d_N , will be ad-
6 justed:
7
8
9

10 $d_S' = d_S \quad d_M' = d_M - d_C' \quad d_N' = d_N + d_C \quad (7)$
11
12

13 As can be seen from the above descriptions, the method *Exchange*(\mathbf{M} , \mathbf{N}) swaps
14 two nodes and adjusts distances (d_S , d_M , d_N and d_C) on the four branches so as
15 to maintain the implied genetic distances among three taxa **S**, **M** and **N**.
16
17

18 Additionally, operations in Big Pulley vary depending on the shape of phylogenetic
19 tree. In Figure 5, a symmetric tree is shown on the left, in which both the child
20 nodes of the root have child nodes, i.e. **L** having children **H1**, **H2** and **R** having
21 children **H3**, **H4**. But in the asymmetric tree on the right, only one of the child
22 nodes of the root has child nodes below it, i.e. **O** having children **G1**, **G2**. But the
23 other child node **Y** doesn't have any offsprings, which may be a tip or a sampled
24 ancestor. The corresponding operations are detailed in the following two parts.
25
26
27
28

29 *Symmetric tree* For the symmetric tree in Figure 5, the operations are illustrated
30 in Figure 6, after which one of the four possible trees (① ② ③ ④) will be proposed.
31 The detailed process is described in Algorithm 1.
32
33

34 **Algorithm 1** Proposal for symmetric trees in Big pulley

35 {Step 1: Identify the child nodes of the root **X**, denoted by **L** and **R**. Correspondingly, the node times
36 are denoted by t_X , t_L , t_R . And the child nodes below them are denoted by **H1**, **H2**, **H3** and **H4**.}
37 Let **X** be the root of the tree.
38 Let **L** and **R** be the left child and right child of **X**, respectively.
39 {Step 2: Propose a new node time for the root **X**.}
40 $a \sim \text{Uniform}[-w, +w]$
41 $t_X' \leftarrow t_X + a$
42 {Step 3: Propose a new node time either for **L** or **R**, and adjust adjacent rates.}
43 if $\sigma_1 \sim \text{Uniform}(0, 1) < 0.5$ then
44 Pick **L** and propose a new node time by $t_L' \leftarrow t_L + a_1$, where $a_1 \sim \text{Uniform}[-w, +w]$.
45 if $t_R < t_L' < t_X'$ then
46 if $\sigma_2 \sim \text{Uniform}(0, 1) < 0.5$ then
47 Apply *Exchange* (**H1**, **R**) and propose tree ①.
48 else
49 Apply *Exchange* (**H2**, **R**) and propose tree ②.
50 end if
51 else
52 Reject the proposal.
53 end if
54 else
55 Pick **R** and propose a new node time by $t_R' \leftarrow t_R + a_2$, where $a_2 \sim \text{Uniform}[-w, +w]$.
56 if $t_L < t_R' < t_X'$ then
57 if $\sigma_3 \sim \text{Uniform}(0, 1) < 0.5$ then
58 Apply *Exchange* (**H3**, **L**) and propose tree ③.
59 else
60 Apply *Exchange* (**H4**, **L**) and propose tree ④.
61 end if
62 else
63 Reject the proposal.
64 end if
65 end if
66 {Step 4: Update the rates on the corresponding branches.}
67 {Step 5: Return the Green ratio α_{BP} .}

6 For example, suppose we are going to propose tree ①. After the new node times
 7 for the root **X** and **L** are proposed, we apply the method by *Exchange* (**H1**, **R**), so
 8 that four distances are adjusted, as follows:
 9

$$10 \quad 11 \quad 12 \quad d_{H1}' = d_{H1} - d_L' \quad d_{H2}' = d_{H2} \quad d_L' = d_L + b \quad d_R' = d_L + d_R \quad (8)$$

13
14
15 Finally, in this example the new rates would be updated by:
16
17
18

$$19 \quad 20 \quad r_{H1}' = \frac{d_{H1}'}{t_{X'} - t_{H1}} \quad r_{H2}' = \frac{d_{H2}'}{t_{L'} - t_{H2}} \quad r_L' = \frac{d_L'}{t_{X'} - t_{L'}} \quad r_R' = \frac{d_R'}{t_{L'} - t_R} \quad (9)$$

21
22
23 Asymmetric tree For an asymmetric tree such as in Figure 5 we would operate
 24 as illustrated in Figure 7, in which there are three possible trees (⑤ ⑥ ⑦). The
 25 operations are detailed in Algorithm 2.
 26
27

Algorithm 2 Proposal for asymmetric trees in Big pulley

29 {Step 1: Identify the extant child node of the root **X**, which has two child nodes below and is denoted
 30 by **O**. The extinct child node of the root, which does not have any child nodes, is denoted by **Y**.
 31 The node times of the root **X**, **Y**, **O** and its child nodes are denoted by t_X , t_Y , t_O , t_{G1} and t_{G2}
 32 respectively.}
 Let **X** be the root of the tree.
 Let **O** be the child of **X** that has children, and let **Y** be the child of **X** that does not have children.
 {Step 2: Propose a new node time for the root **X**.}
 $a \sim \text{Uniform}[-w, +w]$
 $t_{X'} \leftarrow t_X + a$
 {Step 3: Propose a new node time for the node **O**.}
 $a_3 \sim \text{Uniform}[-w, +w]$
 $t_{O'} \leftarrow t_O + a_3$
 if $t_{O'} < t_Y$ or $t_{O'} > t_{X'}$ then
 Reject the proposal.
 end if
 {Step 4: Adjust the distances according to the tree corresponding topologies.}
 if $t_{O'} > \max\{t_{G1}, t_{G2}\}$ or $t_{G1} = t_{G2}$ then
 if $\sigma_4 \sim \text{Uniform}(0, 1) < 0.5$ then
 Apply *Exchange* (**G1**, **Y**) and propose tree ⑤.
 else
 Apply *Exchange* (**G2**, **Y**) and propose tree ⑥.
 end if
 else if $\min\{t_{G1}, t_{G2}\} < t_{O'} < \max\{t_{G1}, t_{G2}\}$ then
 Exchange the older child of **O** and **Y**. (For the asymmetric tree in Figure 5, we apply *Exchange*
 (**G1**, **Y**) and propose tree ⑦).
 else if $t_{O'} < \min\{t_{G1}, t_{G2}\}$ then
 Reject the proposal.
 end if
 {Step 4: Update the rates on the corresponding branches.}
 {Step 5: Return the Green ratio α_{BP} .}

53
54 To give an example, assume we are going to propose tree ⑤. Firstly, $t_{X'}$ and $t_{O'}$
 55 are proposed in Step 3 and Step 4. Then, in Step 4, the method *Exchange* (**G1**, **Y**)
 56 is applied, after which the four distances are adjusted as follows:
 57
58

$$61 \quad 62 \quad d_{G1}' = d_{G1} - d_{O'} \quad d_{G2}' = d_{G2} \quad d_O' = d_O + b \quad d_Y' = d_Y + d_O \quad (10)$$

And the four rates are updated as follows:

$$r_{G1}' = \frac{d_{G1}'}{t_X' - t_{G1}} \quad r_{G2}' = \frac{d_{G2}'}{t_O' - t_{G2}} \quad r_O' = \frac{d_O'}{t_X' - t_O'} \quad r_Y' = \frac{d_Y'}{t_O' - t_Y} \quad (11)$$

Calculating the Green ratio

MCMC operators must use reversible proposal distributions to satisfy the detailed balance requirements of the MCMC algorithm (Refer to Appendix section 1 for more details). Therefore, all four of our operators involve a final step of calculating the Green ratio for the proposal.

According to the third and fourth steps in the operations for internal nodes, three rates on the branches linked to the selected internal node are proposed by one random number a that is used to change the node time. There are four parameters involved in this proposal, comprised of a 3-dimensional rate space and a 1-dimensional time space. The proposed operator utilises one random number in time space and makes changes in both time and rate space, which leads to a dimension-matching problem. To solve this dimension-matching problem, as is mentioned in Green's paper [20], it is necessary to construct the Jacobian matrix. In equation (12), \mathbf{J}_1 deals with the parametric spaces before the proposal in vector $\mathbf{IN} = [t_X, r_X, r_L, r_R]$ and after the proposal in vector $\mathbf{OUT} = [t_X', r_X', r_L', r_R']$.

$$\mathbf{J}_1 = \begin{bmatrix} \frac{\partial \mathbf{f}}{\partial t_X} & \frac{\partial \mathbf{f}}{\partial r_X} & \frac{\partial \mathbf{f}}{\partial r_L} & \frac{\partial \mathbf{f}}{\partial r_R} \end{bmatrix} = \begin{bmatrix} \frac{\partial f_1}{\partial t_X} & \frac{\partial f_1}{\partial r_X} & \frac{\partial f_1}{\partial r_L} & \frac{\partial f_1}{\partial r_R} \\ \frac{\partial f_2}{\partial t_X} & \frac{\partial f_2}{\partial r_X} & \frac{\partial f_2}{\partial r_L} & \frac{\partial f_2}{\partial r_R} \\ \frac{\partial f_3}{\partial t_X} & \frac{\partial f_3}{\partial r_X} & \frac{\partial f_3}{\partial r_L} & \frac{\partial f_3}{\partial r_R} \\ \frac{\partial f_4}{\partial t_X} & \frac{\partial f_4}{\partial r_X} & \frac{\partial f_4}{\partial r_L} & \frac{\partial f_4}{\partial r_R} \end{bmatrix}, \quad (12)$$

where the functions f_1, f_2, f_3 and f_4 represent how the operator makes a proposal. After substituting equation (3) in equation (12), the Green ratio for the internal nodes can be derived:

$$\alpha_{IN} = \frac{p(-a)}{p(a)} |\mathbf{J}_1| = \frac{t_P - t_X}{t_P - t_X'} \times \frac{t_X - t_L}{t_X' - t_L} \times \frac{t_X - t_R}{t_X' - t_R}, \quad (13)$$

where the proposal density $p(-a)$ is equal to $p(a)$ since the random number a is drawn from Uniform distribution.

Likewise, the Green ratio for Simple Distance, Small Pulley and Big Pulley can be obtained:

$$\alpha_{SD} = \frac{t_X - t_L}{t_X' - t_L} \times \frac{t_X - t_R}{t_X' - t_R}, \quad (14)$$

$$\alpha_{SP} = 1, \quad (15)$$

$$\alpha_{BP} = \mu \times \frac{t_X' - t_C}{t_X' - t_C'} \times \frac{t_C - t_S}{t_C' - t_S} \times \frac{t_C - t_{N1}}{t_X' - t_{N1}} \times \frac{t_X - t_{N2}}{t_C' - t_{N2}}, \quad (16)$$

6 where $\mu = p(g', g)/p(g, g')$ is defined as the proposal ratio of topology change and
7 is obtained by Algorithm 3. More details of how to calculate the determinant of the
8 Jacobian matrix are explained in Appendix section 1.

10 Results

11 To validate the correctness and determine the efficiency, we conducted a series of
12 experiments by implementing the Constant Distance operator in BEAST2 [18].

13 First, we perform a well-calibrated simulation study, which tests our operator
14 alongside existing operators. Correctness was further confirmed by sampling trees
15 from the prior distribution i.e. without data (Refer to Appendix section 2 for more
16 details). By comparing effective sample sizes (ESS) [24] and running times, it is
17 demonstrated that the performance is improved when including our proposed op-
18 erator. Finally, the posterior correlation of rates and node times are discussed.

22 Well-calibrated simulation study

23 A well-calibrated simulation study is a powerful tool for evaluating and validating
24 the implementation of a Bayesian model [25].

25 Figure 8 shows the Bayesian model used in this study, which includes the evolution-
26 ary model and the prior distributions of parameters. As is shown in the figure,
27 the sequence alignment is simulated by a phylogenetic continuous-time Markov
28 chain in BEAST2. It contains a substitution rate matrix given by the HKY85 [26]
29 model and a substitution tree determined by an uncorrelated relaxed clock model
30 and Yule model. More specifically, base frequencies π follow a Dirichlet distribution
31 and the transition-transversion ratio κ follows a log-normal prior distribution. The
32 distribution of node times is described in a Yule tree ψ with hyperparameter birth
33 rate λ following a log-normal distribution. The rates r_i follow a log-normal distribu-
34 tion with mean of 1 and standard deviation s_1 following a hyperprior distribution.

35 First, we sampled parameters and trees from the full model 100 times. The ran-
36 dom parameters included: standard deviation of rates across branches s_1 , birth rate
37 λ , base frequencies π and transition-transversion bias κ . Second, we simulated nu-
38 cleotide alignments using the simulated parameters. In total, 100 data sets were
39 simulated, each with 120 taxa. Third, we used BEAST2 with the Constant Dis-
40 tance operator to infer the tree and parameters from each of the 100 simulated data
41 sets in turn. Finally, the posterior estimates of the parameters were compared with
42 the real values that were used to simulate the corresponding sequence alignment.
43 The comparisons are shown in Figures 9.

44 These results show that the true values of the parameters are within the 95%
45 highest posterior density (HPD) interval approximately 95% of the time (Table 1).
46 This well-calibrated simulation study formed part of the validation of our imple-
47 mentation of the Constant Distance operator.

56 Performance comparison

57 To evaluate the performance of Constant Distance operator in a Bayesian phylo-
58 genetic analysis, we explored the time required to adequately sample the posterior
59 distribution. This was achieved by examining (i) the total time taken by BEAST2
60 to complete the MCMC inference (running time), and (ii) the effective sample size
61

62
63
64
65

(ESS) of the sampled parameters. The effective sample size of a parameter is the number of effectively independent samples from the posterior distribution. Larger ESS indicates a better approximation of the marginal posterior distribution of the parameter. We used Tracer [24] to compute ESS.

For each dataset, we compared two operator configurations. 1) Using the current operators in BEAST2 to sample discrete rate categories (Category). 2) Using the Constant Distance operator to sample continuous rates specified by an uncorrelated related clock model (Cons). The Category configuration is the default setting in BEAST 2.5. We aim to compare the performance of the Constant Distance operator to that of the existing operator schedule. In each configuration, the data set was analyzed 20 times with the prior distributions and all other model specifications held constant. The details of operator weights used are given in Appendix 3.1.

We performed the analysis on two sets of simulated sequence alignment (See Appendix 3.2 for more details). The simulated data sets both have 20 taxa but different sequence lengths, i.e. one data set containing 500 sites, the other containing 1,000 sites. Moreover, we used four real data sets to further evaluate the performance of Constant Distance operator, including a primate data set [27] and three other data sets (Anolis [28], RSV2 [29, 30] and HIV-1 [31]) in BEAST2 [32].

The ESS and running time are summarised in Figure 10 and Table 2. To be more specific, we measure the efficiency by ESS per hour, which is calculated by the ESS of parameters in one simulation divided by the running time in hours. Then we compare the efficiency of two configurations by calculating the ratios of ESS per hour for simulations in the two configurations. If the ratio is larger than 1, then ESS per hour of Cons configuration is larger than that of Category configuration. As is shown in Figure 10, most ratios of the parameters are above the red line (larger than 1), which indicates that Cons configuration provides larger ESS per hour for most parameters. Although there are several parameters sampled by Cons configuration having smaller ESS per hour in some data sets, it should be noticed that the ratio is calculated by choosing random simulations in the two configurations (See Appendix 3.3 for more details). Additionally, it is worth noting that the efficiency is improved more obviously in simulated data set having 1000 sites, compared with the data sets having 500 sites. This means the proposed operators sample rates and node times more efficiently if the genetic distances are more accurate.

Table 2 lists the average running time of the data sets. It can be seen that Cons configuration finished simulations with less time in most cases. Moreover, Table 2 also shows the parameter that has the smallest ESS in Category configuration, and is compared with the corresponding ESS in Cons configuration. After calculating the ESS per hour, we conclude that Cons configuration improved the efficiency of the worst estimated parameter in Category configuration by a factor of 2.3 to 15.8.

Correlation analysis of rates and branch lengths

In this section, we conduct a pairwise comparison between rates and branch lengths in units of time. We used a data set of ratite mitochondrial genomes [33]. This data set includes 7 species of ratites and an alignment of 10767 sites. After analysing the ratites data set in BEAST2 using the Constant Distance operator, we calculated the Pearson coefficient between the rates and the times across branches to investigate the posterior correlation of these parameters.

6 The results are summarised in Figure 11. Figure 11(a) presents the topology of the
7 maximum clade credibility tree. We utilised the programme TreeStat2 [34] to obtain
8 the filtered trees that have the same topology as the maximum clade credibility tree
9 from the sampled trees in MCMC chain. This means the trees that have different
10 shared common ancestors of each taxon from the reference tree are filtered out.
11

12 Afterwards, Figure 11(b) shows the pairwise comparison of the 12 branch rates
13 and 12 branch lengths (in time) on these filtered trees. As can be seen from the
14 diagonal, the rate on one branch is negatively correlated with the length of that
15 branch, which indicates that an older divergence time will lead to a smaller rate.
16 This is because the primary signal in the data is genetic distance, so that there
17 will be a range of rates and divergence times that are consistent with the genetic
18 distances, but the products of these quantities will vary less than the individual
19 parameters. The consequence is that there will tend to be a negative relationship
20 between rate r_i and branch length l_i i.e. $r_i = d_i/l_i$. At the same time, there will
21 tend to be a positive relationship between rate r_i and its parent's branch length l_{ip} ,
22 since a larger l_{ip} leads to a smaller l_i . Moreover, for cherries that share the same
23 branch length in the tree, they will tend to have the same correlation pattern. Take
24 ANDI and DIGI as an example. r_1 and l_1 are negatively correlated, but r_1 and l_8
25 are positively correlated, which is also the correlation of r_2 , l_2 and l_8 .
26
27

28 It is precisely this form of correlation structure in the posterior that our operator
29 anticipates, and these correlations are the reason that our operator performs better
30 than naive alternatives.

31 Sampling a fixed unrooted tree

32 A limiting case for the relaxed molecular clock model (and one exploited in some
33 of our validation tests) occurs for long sequences, when the branch lengths of the
34 unrooted tree, in units of expected substitutions per site, become known without
35 error. With full length genomes now available, although inferring trees from genomes
36 involves complexities and assumptions such as a good partition scheme [35], this
37 limiting case might be approached in some data sets. As a simple test in this paper,
38 this gives rise to an alternative approach to analysis, where divergence times, a root
39 position and the branch rates are random variables, and the data are a set of branch
40 lengths in units of substitution on a known unrooted tree topology.

41 Previous work done by Reis and Yang [14] also tried to approximate the likelihood
42 of such an unrooted tree in Bayesian phylogenetic inference. Similar researches in
43 [6, 13] show that these methods can account for rate changes in a relaxed clock
44 model, but the genetic distances are not fixed, for example Stéphane Guindon used
45 a Gibbs sampling algorithm [13]. Outside of the Bayesian MCMC formalism, least-
46 squares criteria [36] and maximum likelihood [37, 38], can also be applied to estimate
47 substitution rates and divergence times in unrooted trees.

48 In this section, we investigated this approach on a fixed substitution tree recon-
49 structed from whole mitochondrial genomes from a set of ratite species [33]. Since
50 no uncertainty is admitted in the genetic distances and the proposed operator don't
51 change the genetic distances, the phylogenetic likelihood is no longer needed and
52 the unrooted tree becomes the data, rather than a multiple sequence alignment.
53
54

6 First of all, we used the ratites data set to construct an unrooted tree with PhyML
7 3.0 [39, 40]. Figure 12(a) shows the unrooted tree with the genetic distances on the
8 branches which are fixed in the subsequent relaxed clock analysis in BEAST2.

9 As an initial starting point, the root is assigned using the midpoint method. After
10 that, according to the genetic distances among seven taxa and the position of the
11 root, consistent divergence times are specified and assigned to each ancestral node,
12 so that a valid rooted time tree is obtained. Once divergence times are determined,
13 rates on the branches are also calculated so that the products match the unrooted
14 substitution tree.

15 Then we used Constant Distance operator to sample a Markov chain initiated
16 by this starting tree. The resulting posterior distribution is shown in Figure 12(b)-
17 (d). As can be seen, despite that there is some uncertainty in the root position,
18 the most probable tree in Figure 12(b) is consistent with previous analyses of this
19 data (see Figure 2 in Ref. [33]). For large data sets of long sequences, the proposed
20 operators may prove useful to provide faster divergence time estimates based on the
21 assumption of known unrooted topology and branch lengths in units of expected
22 substitutions per site.

23 Discussion

24 We have demonstrated that the presented operator is valid and able to improve the
25 efficiency of phylogenetic MCMC for relaxed clock models. The overall performance
26 of a Bayesian phylogenetic analysis will be affected by the proportion of MCMC
27 steps that this operator is chosen to make the proposal. In the BEAST2 software,
28 this can be changed by modifying the relative weights operators in the operator
29 schedule. The ideal proportion is non-trivial to determine for an arbitrary data
30 set. In this study, we assigned equal weights on operations to all internal nodes
31 (including the root). How to assign weights to achieve better performance is not
32 studied in this paper, and users may assign different weights in practice. Hence, an
33 optimal method of assigning weights still needs further investigation.

34 The key idea of the presented operator (to maintain the genetic distances) shows
35 a novel direction for more efficient proposals in Bayesian phylogenetic MCMC. For
36 example, the operations on the internal nodes, in the current study, involve one
37 random internal node, one node time and three branch rates. If two or more nodes
38 are selected, then more associated rates and node times can be sampled in one
39 proposal, which may achieve even better efficiency. Another possible approach is
40 to make small changes to the genetic distances as well. To minimise the number
41 of changes to genetic distances, a two-dimensional random draw will be used to
42 change four parameters (one divergence time and one rate changed directly, the
43 other two rates derived so as to minimise changes to genetic distances). What's
44 more, it should be pointed out that Small Pulley and Big Pulley can only be applied
45 to reversible continuous-time Markov chain models where unrooted trees can be
46 used in inference, because these operators require the underlying unrooted tree to
47 be unchanged. Future work could elaborate a larger class of operators along these
48 lines.

49 As data sets have increased in size the impetus to improve efficiency of Bayesian
50 phylogenetic inference algorithms has steadily increased. Besides more effective pro-
51 posal mechanisms within Metropolis-Hastings MCMC, completely novel approaches

6 to Bayesian phylogenetics have also begun to get some attention. Variational meth-
7 ods are one alternative for approximating Bayesian posterior distributions [41].
8 These approaches make inference an optimisation problem and take advantage of
9 tractable variational distributions that approximate the posterior distribution, thus
10 decreasing the computational cost by avoiding high-dimensional integrals in MCMC
11 sampling schemes. Recent work has investigated the potential for applying varia-
12 tional methods to phylogenetics [42, 43]. Our improved MCMC methods provide a
13 performance baseline for these new approaches.
14

16 **Conclusions**

17 As data sets have increased in size, the need for computational efficiency of Bayesian
18 phylogenetic analyses has also increased. In this paper, we have discussed a new
19 tree proposal that substantially increases the efficiency of Bayesian phylogenetic
20 inference under a popular class of relaxed molecular clock models.

21 We demonstrate the correctness of this algorithm with a series of tests including
22 a well-calibrated simulation study. Based on both simulated and real data sets,
23 the proposed operator is more efficient than current algorithms implemented in
24 BEAST2. Performance improvements of greater than an order of magnitude increase
25 in ESS/hour were measured on both real and simulated data. The proposed operator
26 is available for use as a package of BEAST2.

31 Abbreviations

32 MCMC	Markov chain Monte Carlo
33 ESS	effective sample sizes
34 HPD	highest posterior density
35 Cateogry	Using the current operators in BEAST2 to sample discrete rate categories
36 Cons	Using the Constant Distance operator to sample continuous rates specified by an uncorrelated related clock model

39 Declarations

40 Ethics approval and consent to participate
41 Not applicable

42 Consent for publication
43 Not applicable

44 Availability of data and material

45 The source code of the proposed operator and the data sets analysed during the current study are available in the
46 Github repository (<https://github.com/Rong419/ConstantDistanceOperator.git>).

47 Competing interests

48 The authors declare that they have no competing interests.

51 Funding

52 The author(s) would like to acknowledge support from a Royal Society of New Zealand Marsden award
53 (#UOA1611; 16-UOA-277). This work is also partially supported by scholarship from China Scholarship Council
54 (No.201706990021).

55 Authors' contributions

56 RZ developed the operator and was a major contributor in writing the manuscript. AJD supervised the
57 implementation of the operator and the writing process of the manuscript. All authors read and approved the final
58 manuscript.

Acknowledgements

The author(s) wish to acknowledge the use of New Zealand eScience Infrastructure (NeSI) high performance computing facilities, consulting support and/or training services as part of this research. New Zealand's national facilities are provided by NeSI and funded jointly by NeSI's collaborator institutions and through the Ministry of Business, Innovation & Employment's Research Infrastructure programme. URL <https://www.nesi.org.nz>.

Author details

¹School of Computer Science, University of Auckland, Auckland, New Zealand. ²School of Biological Sciences, University of Auckland, Auckland, New Zealand.

References

1. Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N., Teller, A.H., Teller, E.: Equation of state calculations by fast computing machines. *The journal of chemical physics* **21**(6), 1087–1092 (1953)
2. Hastings, W.K.: Monte carlo sampling methods using markov chains and their applications. *Biometrika* **57**(1), 97–109 (1970)
3. Zuckerkandl, E., Pauling, L.: Evolutionary divergence and convergence in proteins, 97–166 (1965)
4. Yang, Z., Rannala, B.: Bayesian phylogenetic inference using dna sequences: a markov chain monte carlo method. *Molecular biology and evolution* **14**(7), 717–724 (1997)
5. Rannala, B., Yang, Z.: Bayes estimation of species divergence times and ancestral population sizes using dna sequences from multiple loci. *Genetics* **164**(4), 1645–1656 (2003)
6. Thorne, J.L., Kishino, H., Painter, I.S.: Estimating the rate of evolution of the rate of molecular evolution. *Mol Biol Evol* **15**(12), 1647–57 (1998). doi:[10.1093/oxfordjournals.molbev.a025892](https://doi.org/10.1093/oxfordjournals.molbev.a025892)
7. Drummond, A.J., Ho, S.Y., Phillips, M.J., Rambaut, A.: Relaxed phylogenetics and dating with confidence. *PLoS biology* **4**(5), 88 (2006)
8. Knapp, M., Stöckler, K., Havell, D., Delsuc, F., Sebastiani, F., Lockhart, P.J.: Relaxed molecular clock provides evidence for long-distance dispersal of nothofagus (southern beech). *PLoS Biology* **3**(1), 14 (2005)
9. Smith, S.A., Beaulieu, J.M., Donoghue, M.J.: An uncorrelated relaxed-clock analysis suggests an earlier origin for flowering plants. *Proceedings of the National Academy of Sciences*, 201001225 (2010)
10. Ho, S.Y., Phillips, M.J., Drummond, A.J., Cooper, A.: Accuracy of rate estimation using relaxed-clock models with a critical focus on the early metazoan radiation. *Molecular Biology and Evolution* **22**(5), 1355–1363 (2005)
11. Lepage, T., Bryant, D., Philippe, H., Lartillot, N.: A general comparison of relaxed molecular clock models. *Molecular biology and evolution* **24**(12), 2669–2680 (2007). doi:[10.1016/B978-1-4832-2734-4.50017-6](https://doi.org/10.1016/B978-1-4832-2734-4.50017-6)
12. Kober, K., Stamatakis, A., Flouri, T.: Efficient detection of repeating sites to accelerate phylogenetic likelihood calculations. *Systematic biology* **66**(2), 205–217 (2017)
13. Guindon, S.: Bayesian estimation of divergence times from large sequence alignments. *Molecular Biology and Evolution* **27**(8), 1768–1781 (2010)
14. Reis, M.d., Yang, Z.: Approximate likelihood calculation on a phylogeny for bayesian estimation of divergence times. *Molecular Biology and Evolution* **28**(7), 2161–2172 (2011)
15. Ayres, D.L., Darling, A., Zwickl, D.J., Beerli, P., Holder, M.T., Lewis, P.O., Hulsenbeck, J.P., Ronquist, F., Swofford, D.L., Cummings, M.P., et al.: Beagle: an application programming interface and high-performance computing library for statistical phylogenetics. *Systematic biology* **61**(1), 170–173 (2011)
16. Lakner, C., Van Der Mark, P., Hulsenbeck, J.P., Larget, B., Ronquist, F.: Efficiency of markov chain monte carlo tree proposals in bayesian phylogenetics. *Systematic biology* **57**(1), 86–103 (2008)
17. Höhna, S., Drummond, A.J.: Guided tree topology proposals for bayesian phylogenetic inference. *Syst Biol* **61**(1), 1–11 (2012). doi:[10.1093/sysbio/syr074](https://doi.org/10.1093/sysbio/syr074)
18. Bouckaert, R., Heled, J., Kühnert, D., Vaughan, T., Wu, C.-H., Xie, D., Suchard, M.A., Rambaut, A., Drummond, A.J.: Beast 2: a software platform for bayesian evolutionary analysis. *PLoS computational biology* **10**(4), 1003537 (2014)
19. Drummond, A., Nicholls, G., Rodrigo, A., Solomon, W.: Estimating mutation parameters, population history and genealogy simultaneously from temporally spaced sequence data. *Genetics* **161**, 1307–1320 (2002)
20. Green, P.J.: Reversible jump markov chain monte carlo computation and bayesian model determination. *Biometrika* **82**(4), 711–732 (1995)
21. Suchard, M.A.: Stochastic models for horizontal gene transfer: taking a random walk through tree space. *Genetics* (2005)
22. Higuchi, T.: Monte carlo filter using the genetic algorithm operators. *Journal of Statistical Computation and Simulation* **59**(1), 1–23 (1997)
23. Höhna, S., Defoin-Platel, M., Drummond, A.J.: Clock-constrained tree proposal operators in bayesian phylogenetic inference. In: BioInformatics and BioEngineering, 2008. BIBE 2008. 8th IEEE International Conference On, pp. 1–7 (2008). IEEE
24. Rambaut, A., Suchard, M., Xie, D., Drummond, A.: Tracer v1. 6. 2014 (2015)
25. Dawid, A.P.: The well-calibrated bayesian. *Journal of the American Statistical Association* **77**(379), 605–610 (1982)
26. Hasegawa, M., Kishino, H., Yano, T.-a.: Dating of the human-ape splitting by a molecular clock of mitochondrial dna. *Journal of molecular evolution* **22**(2), 160–174 (1985)
27. Finstermeier, K., Zinner, D., Bräuer, M., Meyer, M., Kreuz, E., Hofreiter, M., Roos, C.: A mitogenomic phylogeny of living primates. *PloS one* **8**(7), 69504 (2013)
28. Jackman, T.R., Larson, A., De Queiroz, K., Losos, J.B.: Phylogenetic relationships and tempo of early diversification in anolis lizards. *Systematic Biology* **48**(2), 254–285 (1999)
29. Zlateva, K.T., Lemey, P., Vandamme, A.-M., Van Ranst, M.: Molecular evolution and circulation patterns of human respiratory syncytial virus subgroup a: positively selected sites in the attachment g glycoprotein. *Journal of virology* **78**(9), 4675–4683 (2004)

- 1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
30. Zlateva, K.T., Lemey, P., Moës, E., Vandamme, A.-M., Van Ranst, M.: Genetic variability and molecular evolution of the human respiratory syncytial virus subgroup b attachment g protein. *Journal of virology* **79**(14), 9157–9167 (2005)
 31. Shankarappa, R., Margolick, J.B., Gange, S.J., Rodrigo, A.G., Upchurch, D., Farzadegan, H., Gupta, P., Rinaldo, C.R., Learn, G.H., He, X., et al.: Consistent viral evolutionary changes associated with the progression of human immunodeficiency virus type 1 infection. *Journal of virology* **73**(12), 10489–10502 (1999)
 32. BEAST2 Data Sets. <https://github.com/CompEvol/beast2/tree/master/examples/nexus>
 33. Cooper, A., Lalueza-Fox, C., Anderson, S., Rambaut, A., Austin, J., Ward, R.: Complete mitochondrial genome sequences of two extinct moas clarify ratite evolution. *Nature* **409**(6821), 704 (2001)
 34. TreeStat2. <https://github.com/alexeid/TreeStat2>
 35. Lanfear, R., Calcott, B., Ho, S.Y., Guindon, S.: Partitionfinder: combined selection of partitioning schemes and substitution models for phylogenetic analyses. *Molecular biology and evolution* **29**(6), 1695–1701 (2012)
 36. To, T.-H., Jung, M., Lycett, S., Gascuel, O.: Fast dating using least-squares criteria and algorithms. *Systematic biology* **65**(1), 82–97 (2015)
 37. Sagulenko, P., Puller, V., Neher, R.A.: Treetime: Maximum-likelihood phylodynamic analysis. *Virus evolution* **4**(1), 042 (2018)
 38. Sanderson, M.J.: r8s: inferring absolute rates of molecular evolution and divergence times in the absence of a molecular clock. *Bioinformatics* **19**(2), 301–302 (2003)
 39. PhyML3.0: New Algorithms, Methods and Utilities. <http://www.atgc-montpellier.fr/phym/>
 40. Guindon, S., Dufayard, J.-F., Lefort, V., Anisimova, M., Hordijk, W., Gascuel, O.: New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML3.0. *Systematic biology* **59**(3), 307–321 (2010)
 41. Beal, M.J.: Variational Algorithms for Approximate Bayesian Inference, p. 281. university of London London, England (2003)
 42. Zhang, C., IV, F.A.M.: Variational bayesian phylogenetic inference. In: International Conference on Learning Representations (2019). <https://openreview.net/forum?id=SJVmjR9FX>
 43. Dang, T., Kishino, H.: Stochastic variational inference for bayesian phylogenetics: A case of cat model. *Molecular biology and evolution* **36**(4), 825–833 (2019)
 44. TreeTraceAnalysis. <https://github.com/CompEvol/beast2/blob/master/src/beast/evolution/tree/TreeTraceAnalysis.java>
 45. TreeAnnotator. <https://beast2.blogs.auckland.ac.nz/treeannotator/>
 46. Peskun, P.H.: Optimum monte-carlo sampling using markov chains. *Biometrika* **60**(3), 607–612 (1973)
 47. Pybus, O.G., Rambaut, A.: Genie: estimating demographic history from molecular phylogenies. *Bioinformatics* **18**(10), 1404–1405 (2002)
 48. Yang, Z., Rodriguez, C.E.: Searching for efficient markov chain monte carlo proposal kernels. *Proceedings of the National Academy of Sciences* **110**(48), 19307–19312 (2013). doi:[10.1073/pnas.1311790110](https://doi.org/10.1073/pnas.1311790110)
 49. RateAgeBetaShift. <https://github.com/revbayes/revbayes/blob/master/src/core/moves/compound/RateAgeBetaShift.cpp>

Figure Legends

[width=12cm]Fig01-flowchart.eps

Figure 1 The flow chart of the Constant Distance operator.

[width=12cm]Fig02-internalnode.eps

Figure 2 Illustration of the operation on an internal node. The operator proposes $t_{X'}$, $r_{X'}$, r_L' and r_R' , during which d_L , d_R , d_X are kept constant.

[width=12cm]Fig03-rootstrategy.eps

Figure 3 Illustration of operations on root. (a) An example of a 4-taxa unrooted tree and two possible rooted trees for the operator to sample, during which the unrooted tree can not be changed. Based on the original tree in (b), Simple Distance proposes a node time in g' and two rates in r' and keeps d_L , d_R constant in (c). Small Pulley proposes two rates in r' and $D = d_L + d_R$ remains constant in (d).

Figure 4 Illustration of Exchange (M,N) method. This method is applied to tree g and proposes g' by swapping \mathbf{M} and \mathbf{N} , so that the three distances are adjusted to maintain the distances among \mathbf{S} , \mathbf{M} and \mathbf{N} . That is, $d_{C'} = d_C + b$, $d_{N'} = d_C + d_N$ and $d_{M'} = d_M - d_C'$, where $b \sim U[-v, +v]$.

Figure 5 Two different tree shapes. The symmetric tree is on the left and the asymmetric tree is on the right. The dashed triangles represent the potential subtrees rooted at the nodes.

Figure 6 Illustration of operations on the symmetric tree in Figure 5. The proposed operator will propose one of the four possible trees, each with 0.25 probability.

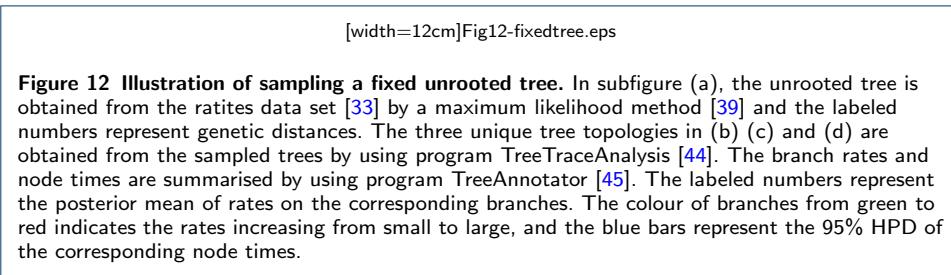
Figure 7 Illustration of operations on the asymmetric tree in Figure 5. The proposed operator will propose one of the three possible trees. If $t_o < t_{G1}$, ⑦ has 1 probability, otherwise ⑤ and ⑥ have 0.5 probability each.

Figure 8 The models and prior distributions to simulate the sequence data. The sequence alignment (SA) is simulated through a phylogenetic continuous-time Markov Chain (PhyloCTMC) that consists of a substitution model (HKY) and an uncorrelated relaxed clock model (UCRelaxedClockModel). The random variables in HKY model construct the mutation rate matrix (Q), including base frequencies ($\pi = \{\pi_A, \pi_C, \pi_G, \pi_T\}$) and kappa (κ). The time trees (ψ) and branch rates (r_i for each branch i in ψ) construct the substitution tree (ST). The branch rates have a LogNormal prior with fixed mean 1 and certain standard deviation (denoted by s_1). And the time trees have a Yule model prior with birth rate (λ) having a LogNormal prior. The other prior distributions include a Dirichlet distributions on π , a LogNormal distribution on κ , and a LogNormal distribution on s_1 . For notations in LogNormal distributions, the uppercase letters represent the parameters in real space, and the lowercase letters represent the parameters in log space. In all the simulations, the number of taxa is fixed at 120 (n = 120).

Figure 9 Well-calibrated simulation study with 120 taxa. Each point is a separate simulated dataset.

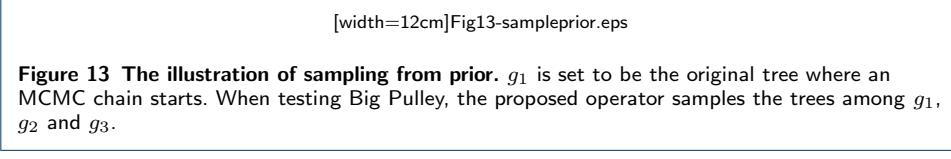
Figure 10 Comparison of ESS and running time. There are 6 data sets analysed, including 4 real data sets and 2 simulated data sets with different number of sites, as is shown in the legend. The red line represent the position where the ratio of ESS per hour is equal to 1. The horizontal axis represents the names of sampled parameters.

Figure 11 Correlation analysis in the ratites tree. l represents the length of a branch, that is the time difference between a parent node and a child node, where $l_1 = l_2 = t_1 - 0$, $l_3 = l_4 = t_2 - 0$, $l_5 = t_3 - 0$, $l_6 = t_4 - 0$, $l_7 = T - 0$, $l_8 = t_5 - t_1$, $l_9 = t_3 - t_2$, $l_{10} = t_4 - t_3$, $l_{11} = t_5 - t_4$ and $l_{12} = T - t_5$. The rates and branch lengths are converted into log space and then Pearson's coefficients are computed, which range from -1 to 1. Blue indicates positive correlations and red indicates negative correlations. The darker the colour, the stronger the correlation.



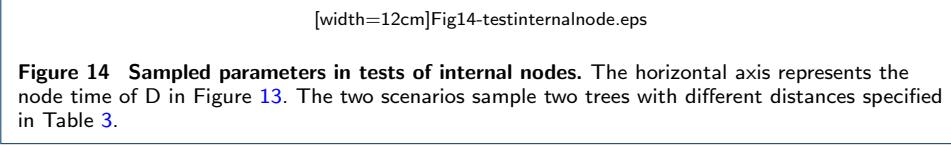
[width=12cm]Fig12-fixedtree.eps

Figure 12 Illustration of sampling a fixed unrooted tree. In subfigure (a), the unrooted tree is obtained from the ratites data set [33] by a maximum likelihood method [39] and the labeled numbers represent genetic distances. The three unique tree topologies in (b) (c) and (d) are obtained from the sampled trees by using program TreeTraceAnalysis [44]. The branch rates and node times are summarised by using program TreeAnnotator [45]. The labeled numbers represent the posterior mean of rates on the corresponding branches. The colour of branches from green to red indicates the rates increasing from small to large, and the blue bars represent the 95% HPD of the corresponding node times.



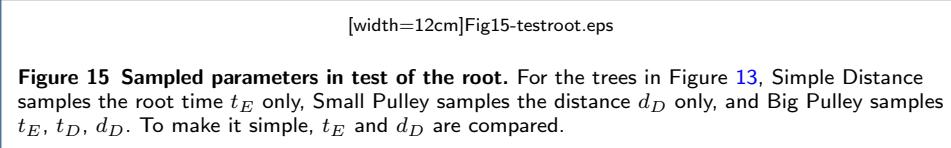
[width=12cm]Fig13-sampleprior.eps

Figure 13 The illustration of sampling from prior. g_1 is set to be the original tree where an MCMC chain starts. When testing Big Pulley, the proposed operator samples the trees among g_1 , g_2 and g_3 .



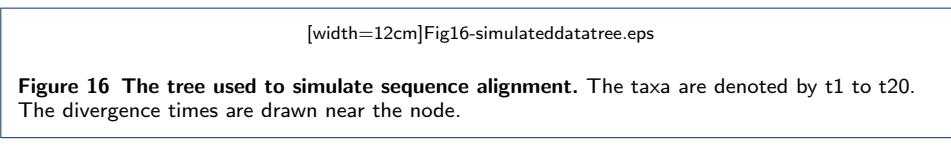
[width=12cm]Fig14-testinternalnode.eps

Figure 14 Sampled parameters in tests of internal nodes. The horizontal axis represents the node time of D in Figure 13. The two scenarios sample two trees with different distances specified in Table 3.



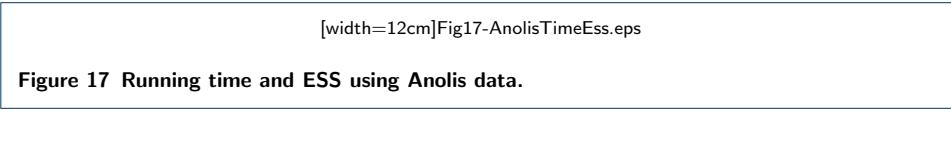
[width=12cm]Fig15-testroot.eps

Figure 15 Sampled parameters in test of the root. For the trees in Figure 13, Simple Distance samples the root time t_E only, Small Pulley samples the distance d_D only, and Big Pulley samples t_E , t_D , d_D . To make it simple, t_E and d_D are compared.



[width=12cm]Fig16-simulateddatatypee.eps

Figure 16 The tree used to simulate sequence alignment. The taxa are denoted by t1 to t20. The divergence times are drawn near the node.



[width=12cm]Fig17-AnolisTimeEss.eps

Figure 17 Running time and ESS using Anolis data.



[width=12cm]Fig18-RSV2TimeEss.eps

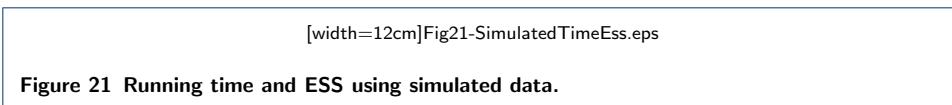
Figure 18 Running time and ESS using RSV2 data.

[width=12cm]Fig19-ShankarappaTimeEss.eps

Figure 19 Running time and ESS using HIV-1 data.

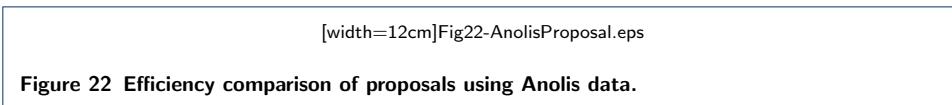
[width=12cm]Fig20-PrimatesTimeEss.eps

Figure 20 Running time and ESS using primates data.



[width=12cm]Fig21-SimulatedTimeEss.eps

Figure 21 Running time and ESS using simulated data.



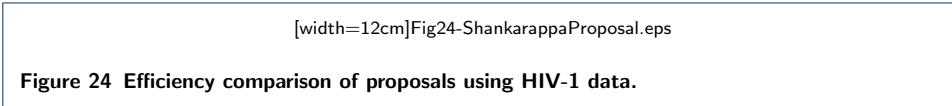
[width=12cm]Fig22-AnolisProposal.eps

Figure 22 Efficiency comparison of proposals using Anolis data.



[width=12cm]Fig23-RSV2Proposal.eps

Figure 23 Efficiency comparison of proposals using RSV2 data.



[width=12cm]Fig24-ShankarappaProposal.eps

Figure 24 Efficiency comparison of proposals using HIV-1 data.

Tables

Parameters	Coverage	Parameters	Coverage
Tree height	89	Ucldstdev	91
Tree length	91	π_A	94
Kappa	97	π_C	96
Birth rate	99	π_G	95
Rate mean	100	π_T	97

12
13
14
Table 1 Percentage of real values lying in the 95% HPD in Figure 9

Data	Configuration	Average running time	Parameter	ESS
Anolis	Category	0.9053	rate.coeff	486.32
	Cons	0.6231		2237.97
RSV2	Category	5.1009	rate.coeff	383.98
	Cons	4.4077		2379.33
HIV-1	Category	5.3436	prior	360.86
	Cons	5.2128		795.21
Primates	Category	12.5615	ucl.stdev	57.47
	Cons	12.3820		164.52
Simualted 500 sites	Category	0.4644	rate.coeff	793.27
	Cons	0.4805		3047.09
Simualted 1000 sites	Category	1.9270	rate.coeff	1599.42
	Cons	0.4805		6255.03

26
27
28
Table 2 Summary of ESS and running time

	genetic distances (fixed)				t_D initial	t_E (fixed)	initial rates			
	d_j	d_k	d_x	d_i			r_j	r_k	r_x	r_i
Scenario 1	0.1	0.2	0.4	0.27	1	10	0.1	0.2	0.04	0.03
Scenario 2	0.4	0.8	2.4	1.6	0.4	0.8	1	2	3	4

32
33
34
Table 3 Initial settings for testing operations on internal nodes

	Chain Length	Sample from MCMC			Integral curve			Plot
		Mean	Err	St.dev	Mean	Err	St.dev	
Senario 1	10000000	3.2727	8.3e-3	0.5467	3.2669	1.3e-06	0.5553	Figure 14(a) Figure 14(b)
	20000000	3.271	6.1e-3	0.5616				
Senario 2	10000000	0.4677	3.9e-04	0.0265	0.4667	3.5e-05	0.0262	Figure 14(c) Figure 14(d)
	20000000	0.4672	2.8e-04	0.0262				

41
42
43
Table 4 Results of sampling the internal node

Strategy	genetic distances				t_D	t_E	initial rates			
	d_j	d_k	d_x	d_i			r_j	r_k	r_x	r_i
Simple Distance	0.1	0.2	0.4	0.27	1	10	0.1	0.2	0.04	0.03
Small Pulley	0.1	0.2		0.67	1	10	0.1	0.2	0.04	0.03
Big Pulley	0.5	0.5		0.5	5	10	0.1	0.1	0.03	0.04

48
49
50
51
Table 5 Initial settings for operations on the root

Strategy	Variable	Sample from MCMC		Integral curve		Plot
Simple Distance	t_E	Mean	St.dev	Mean	St.dev	Figure 15(a)
Small Pulley	d_i	0.3480	0.0492	0.3476	0.0494	Figure 15(b)
Big Pulley	d_i	0.1016	0.0766	0.0960	0.0760	Figure 15(c)
	t_E	3.3017	0.6908	3.3095	0.6912	Figure 15(d)

52
53
54
55
56
57
58
59
60
61
62
63
64
65
Table 6 Results of sampling the root

Table 7 Operator weights in MCMC chains

Operator class	Name	Simulated data Category	Anolis Category	RSV2 Category	HIV-1 Category	Primates Category
		Cons	Cons	Cons	Cons	Cons
rates times	ConstantDistance Operator	0.1919	-	0.2248	-	0.2228
	Rate Normal Operators ¹	0.1919	0.2879	0.1349	0.1337	0.1850
	UcidSdev Scale Operator ²	0.0288	0.0288	0.0270	0.0267	0.2775
	UcidMean Scale Operator	-	-	-	0.0357	0.0278
	UcidMean Tree UpperDown Operator	-	-	-	0.0446	-
	InternalNodeTime Scale Operator	0.0480	0.0960	0.0270	0.0267	0.0267
	RootAge Scale Operator	0.0480	0.0480	0.0270	0.0267	0.0463
	AllNodeTimes Uniform Operator	0.0480	0.0960	0.1799	0.1337	0.0463
	SubtreeSlide Operator	0.1440	0.1440	0.0989	0.1349	0.0463
	NarrowExchange Operator	0.1440	0.1440	0.0989	0.1349	0.0463
Tree	WideExchange Operator	0.0480	0.0480	0.0270	0.0267	0.0463
	WilsonBalding Operator	0.0480	0.0480	0.0270	0.0267	0.0463
	BirthRate Scale Operator	0.0480	0.0480	0.0629	0.0270	0.0278
	DeathRate Scale Operator	-	-	0.0629	-	-
	PopulationSize Scale Operator	-	-	-	0.0802	-
	Kappa Scale Operator	0.0096	0.0096	0.0009	0.0009	0.0185
	Frequencies DeltaExchange Operator	0.0019	0.0019	0.0009	0.0009	0.0009

Note

1: Random walk operator and Swap operator in Cons configuration, Random walk operator, Scale operator and Swap operator in Category configuration.

2: The operator introduced in Appendix section 4 is used in Cons configuration, a Scale operator is used in Category configuration .

-: The parameter is not sampled and no operator is assigned.

1
2
3
4
5
Appendix

6 1. The Green ratio

7 When developing an operator for MCMC, the proposal function must be reversible. In other words, the probability
8 that the operator propose a new state from the current state is required to be equal to the probability that the
9 proposed state goes back to current state. To be specific, let $\pi(x)$ be the target probability distribution and
10 $p(x, x')$ be the transition kernel in the continuous Markov chain. The reversibility condition requires that
11 $\pi(x)p(x, x') = \pi(x')p(x', x)$. And an operator provides a proposal $q(x, x')$ with some probability $\alpha(x, x')$ that
12 the proposal is accepted. Thus, the reversibility condition is rewritten as
13 $\pi(x)q(x, x')\alpha(x, x') = \pi(x')q(x', x)\alpha(x', x)$.

14 Considering the subspace φ_1 on x and subspace φ_2 on x' , it is assumed that there is a symmetric measure on the
15 combined parametric space $\varphi = \varphi_1 \times \varphi_2$, so that $\pi(x)q(x, x')$ has a density with respect to a single measure on
16 φ . Then, Green suggested that the reversibility condition should be satisfied by detailed balance [20], as represented
17 by equation (17). And according to Peskun' proof, it is optimal to take equation (18) as the acceptance probability
18 to retain the detailed balance [46].

$$\int_A \pi(x) dx \int_B q(x, x') \alpha(x, x') dx = \int_B \pi(x') dx' \int_A q(x', x) \alpha(x', x) dx', \quad (17)$$

21 where $A \in \varphi_1$ and $B \in \varphi_2$ are two Borel sets. $q(x, x')$ denotes the probability that the operator proposes a new
22 state x' given the current state x .

$$\alpha_H(x, x') = \min \left\{ 1, \frac{\pi(x')p(x', x)}{\pi(x)p(x, x')} \right\}, \quad (18)$$

27 where $p(x', dx)/p(x, dx')$ is known as the Hastings ratio.

28 However, for operators that do not have a symmetric measure, it is necessary to include the Jacobian matrix \mathbf{J} in
29 order to deal with the dimension matching problem, as is discussed in Green's paper [20]. In this case, equation (18)
30 is extended, as is shown in equation (19).

$$\alpha_G(x, x') = \min \left\{ 1, \frac{\pi(x')p(x', x)}{\pi(x)p(x, x')} |\mathbf{J}| \right\}, \quad (19)$$

34 where $\mathbf{J} = \nabla h(x, x')$ represents a vector differential matrix of deterministic function h . $\alpha = \frac{p(x', x)}{p(x, x')} |\mathbf{J}|$ is defined
35 as the Green ratio, and \mathbf{J} ensures that the proposal have a symmetric measure on each subspace in state x and x' .

37 1.1 Calculating the Green ratio for operations on internal nodes

38 The Constant Distance Operator firstly proposes a new time for the randomly selected internal node (equation
39 (20a)), and then proposes three rates by the original distances and new node times(equation (20b)~equation
40 (20d)).

$$f_1 : t_{X'} = t_X + a \quad (20a)$$

$$f_2 : r_{X'} = \frac{r_X \times (t_P - t_X)}{t_P - t_{X'}} \quad (20b)$$

$$f_3 : r_{L'} = \frac{r_L \times (t_X - t_L)}{t_{X'} - t_L} \quad (20c)$$

$$f_4 : r_{R'} = \frac{r_R \times (t_X - t_R)}{t_{X'} - t_R} \quad (20d)$$

56 Substituting equation (20) in the Jacobian matrix \mathbf{J}_1 (equation (12)), we can get equation (21), so that the
57 determinant of \mathbf{J}_1 can be obtained by equation (22).

$$\mathbf{J}_1 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ \frac{-r_X}{t_P - t_{X'}} & \frac{t_P - t_X}{t_P - t_{X'}} & 0 & 0 \\ \frac{r_L}{t_X' - t_L} & 0 & \frac{t_X - t_L}{t_{X'} - t_L} & 0 \\ \frac{r_R}{t_{X'} - t_R} & 0 & 0 & \frac{t_X - t_R}{t_{X'} - t_R} \end{bmatrix} \quad (21)$$

$$\begin{aligned}
 |\mathbf{J}_1| &= 1 \times \begin{vmatrix} \frac{t_P - t_X}{t_P - t_{X'}} & 0 & 0 \\ 0 & \frac{t_X - t_L}{t_{X'} - t_L} & 0 \\ 0 & 0 & \frac{t_X - t_R}{t_{X'} - t_R} \end{vmatrix} \\
 &= \frac{t_P - t_X}{t_P - t_{X'}} \times \begin{vmatrix} \frac{t_X - t_L}{t_{X'} - t_L} & 0 \\ 0 & \frac{t_X - t_R}{t_{X'} - t_R} \end{vmatrix} \\
 &= \frac{t_P - t_X}{t_P - t_{X'}} \times \frac{t_X - t_L}{t_{X'} - t_L} \times \frac{t_X - t_R}{t_{X'} - t_R}
 \end{aligned} \tag{22}$$

1.2 Calculating the Green ratio for Simple Distance

Simple Distance proposes two rates by using equation (23b) and equation (23c), according the new root time in equation (23a). So the Jacobian matrix can be obtained as is shown in equation (24).

$$t_{X'} = t_X + a \tag{23a}$$

$$r_L' = \frac{r_L \times (t_X - t_L)}{t_{X'} - t_L} \tag{23b}$$

$$r_R' = \frac{r_R \times (t_X - t_R)}{t_{X'} - t_R} \tag{23c}$$

$$\mathbf{J}_2 = \begin{bmatrix} \frac{\partial t_{X'}}{\partial r_{X'}} & \frac{\partial t_{X'}}{\partial r_{X'}} & \frac{\partial t_{X'}}{\partial r_{R'}} \\ \frac{\partial r_L}{\partial r_{X'}} & \frac{\partial r_L}{\partial r_{X'}} & \frac{\partial r_L}{\partial r_{R'}} \\ \frac{\partial r_R}{\partial r_{X'}} & \frac{\partial r_R}{\partial r_{X'}} & \frac{\partial r_R}{\partial r_{R'}} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ \frac{r_L}{t_{X'} - t_L} & \frac{t_X - t_L}{t_{X'} - t_L} & 0 \\ \frac{r_R}{t_{X'} - t_R} & 0 & \frac{t_X - t_R}{t_{X'} - t_R} \end{bmatrix} \tag{24}$$

So the determinant of \mathbf{J}_2 is calculated by equation (25)

$$|\mathbf{J}_2| = \frac{t_X - t_L}{t_{X'} - t_L} \times \frac{t_X - t_R}{t_{X'} - t_R} \tag{25}$$

Calculating the Green ratio for Small Pulley

Small Pulley proposes a new genetic distance of a branch on one side of the root by adding a random number b , which is equal to adding a random number b to the original product of rate and time on that branch. As a result, a new rate is proposed by equation (26a). Similarly, a new rate on another branch is proposed by equation (26b), because the total distance of the two branches linked to the root should remain constant.

$$r_L' = \frac{r_L \times (t_X - t_L) + b}{t_X - t_L} \tag{26a}$$

$$r_R' = \frac{[r_R \times (t_X - t_R) + r_L \times (t_X - t_L)] - [r_L \times (t_X - t_L) + b]}{t_X - t_R} = \frac{r_R \times (t_X - t_R) - b}{t_X - t_R} \tag{26b}$$

Then, as is illustrated in equation (27), the Jacobian matrix \mathbf{J}_3 is simply obtained, which makes the determinant $|\mathbf{J}_3| = 1$.

$$\mathbf{J}_3 = \begin{bmatrix} \frac{\partial r_L'}{\partial r_{L'}} & \frac{\partial r_L'}{\partial r_{R'}} \\ \frac{\partial r_R}{\partial r_{L'}} & \frac{\partial r_R}{\partial r_{R'}} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \tag{27}$$

1
2
3
4
5
6 1.3 Calculating the Green ratio for Big Pulley

7 Two new node times are proposed in Big Pulley. One is the root time (equation (28a)), the other is the node time
8 of the child node of the root. It can be either children of the root, i.e. **son** and **dau**. So $t_{C'}$ is used to denote the
9 node time proposed, as is seen in equation (28b). In addition, the distances are adjusted by the method *Exchange*
10 (**M**, **N**), dependent on which nodes are chosen. As a result, the four rates are proposed, as is shown in equation
11 (28c)~equation (28f)

$$t_X' = t_X + a \quad (28a)$$

$$t_C' = t_C + a_{1,2,3} \quad (28b)$$

$$r_C' = \frac{r_C \times (t_X - t_C) + b}{t_X' - t_C'} \quad (28c)$$

$$r_S' = \frac{r_2 \times (t_C - t_S)}{t_C' - t_S} \quad (28d)$$

$$r_M' = \frac{r_M \times (t_C - t_M) - [r_C \times (t_X - t_C) + b]}{t_X' - t_M} \quad (28e)$$

$$r_N' = \frac{r_C \times (t_X - t_C) + r_N \times (t_X - t_N)}{t_C' - t_N} \quad (28f)$$

36 where $a_{1,2,3}$ is the random number to propose a new node time for the child node of the root. Depending on which
37 child node is selected, the notation is different, i.e. a_1 , a_2 , a_3 . Here, to make it a general case, a_x is used.

38 Therefore, the Jacobian matrix \mathbf{J}_4 for the six parameters in equation (28) is obtained by equation (29). And the
39 determinant of \mathbf{J}_4 is calculated shown in equation (30).

$$\mathbf{J}_4 = \begin{bmatrix} \frac{\partial t_X'}{\partial t_X} & \frac{\partial t_X'}{\partial t_C} & \frac{\partial t_X'}{\partial r_C} & \frac{\partial t_X'}{\partial r_S} & \frac{\partial t_X'}{\partial r_M} & \frac{\partial t_X'}{\partial r_N} \\ \frac{\partial t_C'}{\partial t_X} & \frac{\partial t_C'}{\partial t_C} & \frac{\partial t_C'}{\partial r_C} & \frac{\partial t_C'}{\partial r_S} & \frac{\partial t_C'}{\partial r_M} & \frac{\partial t_C'}{\partial r_N} \\ \frac{\partial r_C'}{\partial t_X} & \frac{\partial r_C'}{\partial t_C} & \frac{\partial r_C'}{\partial r_C} & \frac{\partial r_C'}{\partial r_S} & \frac{\partial r_C'}{\partial r_M} & \frac{\partial r_C'}{\partial r_N} \\ \frac{\partial r_S'}{\partial t_X} & \frac{\partial r_S'}{\partial t_C} & \frac{\partial r_S'}{\partial r_C} & \frac{\partial r_S'}{\partial r_S} & \frac{\partial r_S'}{\partial r_M} & \frac{\partial r_S'}{\partial r_N} \\ \frac{\partial r_M'}{\partial t_X} & \frac{\partial r_M'}{\partial t_C} & \frac{\partial r_M'}{\partial r_C} & \frac{\partial r_M'}{\partial r_S} & \frac{\partial r_M'}{\partial r_M} & \frac{\partial r_M'}{\partial r_N} \\ \frac{\partial r_N'}{\partial t_X} & \frac{\partial r_N'}{\partial t_C} & \frac{\partial r_N'}{\partial r_C} & \frac{\partial r_N'}{\partial r_S} & \frac{\partial r_N'}{\partial r_M} & \frac{\partial r_N'}{\partial r_N} \end{bmatrix} \quad (29)$$

$$= \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ \frac{r_C}{t_X' - t_C'} & \frac{-r_C}{t_X' - t_C'} & \frac{t_X' - t_C}{t_X' - t_C'} & 0 & 0 & 0 \\ 0 & \frac{r_S}{t_X' - t_C'} & 0 & \frac{t_C - t_S}{t_C' - t_S} & 0 & 0 \\ \frac{-r_C}{t_X' - t_M} & \frac{r_N + r_C}{t_X' - t_M} & \frac{-(t_X - t_C)}{t_X' - t_M} & 0 & \frac{t_C - t_M}{t_X' - t_M} & 0 \\ \frac{r_C + r_S}{t_C' - t_N} & \frac{-(r_C + r_S)}{t_C' - t_N} & \frac{t_X - t_C}{t_C' - t_N} & 0 & 0 & \frac{t_X - t_N}{t_C' - t_N} \end{bmatrix}$$

$$|\mathbf{J}_4| = \frac{t_X' - t_C}{t_X' - t_C'} \times \frac{t_C - t_S}{t_C' - t_S} \times \frac{t_C - t_M}{t_X' - t_M} \times \frac{t_X - t_N}{t_C' - t_N} \quad (30)$$

59 Last but not least, due to the change of tree topology in *Exchange* (**M**, **N**), the probability of the proposed tree
60 going back to the original tree $p(g|g')$, as well as the probability of making the proposal $p(g'|g)$, should be
61 considered. As the ratio of $p(g|g')/p(g'|g)$ is defined as μ , the calculation of μ is detailed in the following
62 algorithm.

Algorithm 3 Calculation of μ for Big pulley

```

1
2
3
4
5
6
7 Original tree is symmetric:
8   if the node that has been exchanged with L or R has child nodes then
9      $\alpha = \beta = 0.25$ 
10    else if  $t_R > t_L$  then
11       $\alpha = 1, \beta = 0.5$ 
12    else if  $t_R < t_L$  then
13       $\alpha = 0.5, \beta = 1$ 
14    else if  $t_R = t_L$  then
15       $\alpha = \beta = 1$ 
16    end if
17    if Proposed tree belongs to ① or ② then
18      Return  $\mu = \frac{\alpha}{0.25}$ 
19    end if
20    if Proposed tree belongs to ③ or ④ then
21      Return  $\mu = \frac{\beta}{0.25}$ 
22    end if
23
24 Original tree is asymmetric:
25   if the node that has been exchanged with O has child nodes then
26      $\gamma = 0.25$ 
27   else
28      $\gamma = 0.5$ 
29   end if
30   if Proposed tree belongs to ⑤ or ⑥ then
31     Return  $\mu = \frac{\gamma}{0.5}$ 
32   end if
33   if Proposed tree belongs to ⑦ then
34     Return  $\mu = \frac{0.25}{1}$ 
35   end if
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

```

2. Sampling from the prior

In this section, we aim to validate the correctness of the proposed operators. To be more specific, we firstly run the simulations by sampling from prior distributions in BEAST2. Since the prior distributions are deterministic, we can analytically calculate the theoretical joint-distributions of sampled parameters in MCMC chains. By comparing the sampled distributions with the analytical results, we demonstrate whether the proposed operators are able to sample parameters correctly.

In Figure 13, a tree with three taxa A, B and C (plus one internal node D, and root E) is used as a small example in the experiments in this section. In the figure, g_1 is set as the initial tree. Firstly, a LogNormal distribution is used as the rate prior in the uncorrelated relaxed clock model, given by equation (31).

$$r = \{r_A \ r_B \ r_C \ r_D\} \sim \text{LogNormal}(m = -3, s = 0.25) \quad (31)$$

In addition, a Coalescent model [47] with constant population size ($N = 0.3$) is used to describe the tree prior. Hence, for the tree in Figure 13, the probability of node times is calculated by equation (32).

$$p(t = \{t_E, t_D\}) = \left(\frac{1}{N} \times e^{-\frac{1}{N}(t_E - t_D)}\right) \times \left(\frac{1}{N} \times e^{-\frac{3}{N}t_D}\right) \quad (32)$$

After the priors are specified, the distribution to sample can be exactly known, since the samples are drawn from the prior distributions. In other words, as the rates are functions of its genetic distance and times, the joint distribution to sample can be represented by equation (33).

$$\begin{aligned} p(r, t) &= p(t_E, t_D) \times p(r_D) \times p(r_A) \times p(r_B) \times p(r_C) \\ &= p(t_E, t_D) \times p\left(\frac{d_D}{t_E - t_D}\right) \times p\left(\frac{d_A}{t_D - t_A}\right) \times p\left(\frac{d_B}{t_D - t_B}\right) \times p\left(\frac{d_C}{t_E - t_C}\right), \end{aligned} \quad (33)$$

where $p(\cdot)$ is the probability of certain rate values in the LogNormal distribution. Therefore, the whole probability can be obtained by conducting numerical integration on equation (33), which shows the probability distribution over all the possible values of parameters.

2.1 Test the operator on internal nodes

The genetic distances, node times and rates for g_1 in Figure 13 are given in Table 3. To test roundly, two scenarios are designed. In each scenario, the genetic distances are fixed, the node time t_D starts from the initial value and will be changed by the proposed operator during the sampling process. Essentially, the proposed operator makes node D

move between node A and E . Besides, to make sure that the result is robust, two different MCMC chain lengths are performed in each scenario, i.e. 10 million and 20 million.

The mean, mean error and the standard deviation of the MCMC samples are summarised in Table 4. Besides, according to equation (33), the actual joint distribution is obtained by using equation (34), and is used to evaluate the results, which is also included in Table 4. Moreover, the histograms of MCMC samples that indicate the sampled distributions, as well as the curves of the numerical integration of equation (34), are shown in Figure 14. From Table 4 and Figure 14, it can be seen that the red curves well fit the black histograms, and the mean values and standard deviations are consistent, which makes it safe to conclude that the proposed operator samples the internal node correctly.

$$p(r, t) = \int_{t_D=0}^{t_E} p(t_E, t_D) \times p\left(\frac{d_A}{t_D}\right) \times p\left(\frac{d_B}{t_D}\right) \times p\left(\frac{d_D}{t_E - t_D}\right) \times p\left(\frac{d_C}{t_E}\right) dt_D \quad (34)$$

2.2 Test the operator on root

Still starting from g_1 in Figure 13, the initial settings for testing the root are given in Table 5. And the three strategies are tested separately in the following parts.

2.2.1 Using Simple Distance The root time t_E is sampled by Simple Distance, which ranges from 1 to positive infinity theoretically. Namely, all the genetic distances and the node time t_D are fixed. Similar to equation (34), the joint distribution of t_E and rates to sample can be obtained by equation (35).

$$p(r, t) = \int_{t_E=1}^{+\infty} p(t_E, t_D) \times p\left(\frac{d_A}{t_D}\right) \times p\left(\frac{d_B}{t_D}\right) \times p\left(\frac{d_D}{t_E - t_D}\right) \times p\left(\frac{d_C}{t_E}\right) dt_E \quad (35)$$

The results are given in Table 6 and Figure 15(a). As can be seen, the mean and the standard deviation of MCMC samples and numerical integration are close to each other, which confirms that the two distribution are the same. Thus, Simple Distance is proved to be correct.

2.2.2 Using Small Pulley Although both d_x and d_i are changed during the sampling process when using Small Pulley, the sum of d_D and d_C are kept 0.67 in this test, as the initial setting shown in Table 5. To make it simple, only d_D is compared.

Then, based on equation (33), the exact distribution of d_i can be obtained by equation (36), which is compared with the sampled distribution in Table 6 and Figure 15(b). Even though there exist some errors, the sampled parameters can be considered to follow the same distribution. So the Small Pulley is also able to provide correct samples.

$$p(r, t) = \int_{d_D=1}^{0.67} p(t_E, t_D) \times p\left(\frac{d_A}{t_D}\right) \times p\left(\frac{d_B}{t_D}\right) \times p\left(\frac{d_D}{t_E - t_D}\right) \times p\left(\frac{0.67 - d_D}{t_E}\right) dd_D \quad (36)$$

2.2.3 Using Big Pulley For g_1 in Figure 13, a new tree, together with the root time t_E and node time of its older child t_D , as well as a genetic distance d_i , is proposed by Big Pulley. In this case, the initial tree g_1 will either go to g_2 or g_3 , as is shown in Figure 13. So the samples are repeatedly drawn from the 3 trees. Besides, according to the initial settings in Table 5, the genetic distances remain unchanged during the process, i.e. $d_{AB} = 1$, $d_{AC} = 1$ and $d_{BC} = 1$ hold. Hence, the distribution we are about to achieve can be calculated by equation (37).

$$\begin{aligned} p(r, t) &= \int_{t_E=0}^{+\infty} \int_{t_D=0}^{t_E} \int_{d_D=0}^{0.5} p(t_E, t_D) \times p\left(\frac{0.5}{t_D}\right) \\ &\quad \times p\left(\frac{0.5}{t_D}\right) \times p\left(\frac{d_D}{t_E - t_D}\right) \times p\left(\frac{0.5 - d_D}{t_E}\right) dd_D dt_D dt_E \end{aligned} \quad (37)$$

The statistical measurements, i.e. mean and standard deviation, are compared in Table 6. The histograms of samples and numerical curves of d_D and t_E are pictured in Figure 15(c) and Figure 15(d). It is shown that the two distributions are consistent within the acceptable error range. Therefore, Big Pulley can also give the right combinations of rates and node times, under the condition that the genetic distances among taxa are constant.

3. Performance analysis of operators

This section provides the details of the results presented in *Performance comparison* section.

3.1 Operator weights The weights on operators for the simulations when comparing efficiency are listed in Table 7. Although how to assign weights to achieve better performance is not studied in this paper, we maintain the percentage of weights on three operator class in Category and Cons configurations. But we modified some weights on the operators inside the same class, and we assigned different weights for different data sets.

3.2 Simulated data sets We simulated two sets of sequence alignment on the same tree with 20 taxa that is shown in Figure 16. We used HKY model as substitution model with $\kappa = 2.4751$, and the base frequencies are $\pi = (0.21930.22680.30070.2531)$. In the uncorrelated relaxed clock model, the standard deviation of the branch rates ($Ucldstdev$) is 0.1803. The models and prior distributions are the same as is described in Figure 8.

3.3 Efficiency measured by ESS per hour Since we compare the efficiency based on ESS per hour using two configurations, i.e. Category and Cons, the ratio of ESS per hour is calculated by a random simulation in the two configurations, as is shown in Figure 10. Then Table 2 lists the average running time and ESS of particular parameters in the simulations using different data sets. Here, we present the detailed running time and ESS of the simulations, which can be seen in Figure 17 to Figure 21. Overall, we conclude that the proposed operators are able to provide better performance, because the figures suggest that Cons configuration requires less running time and have larger ESS for most parameters in most simulations. Especially, for those poorly estimated parameters in Category configuration, the improvement is more obvious. For data sets such as primates and simulated data with 500 sites, the running time is slightly larger in Cons configuration, but the ESS are much larger, which makes it acceptable to reduce the MCMC chain length and get the same performance.

3.4 Efficiency measured by proposals The operators introduced in the paper utilise a random walk proposal for the new node time, which draws a random number from a uniform distribution and moves the node uniformly on the branch. However, others proposals, such as a Bactrian proposal [48] and a Beta proposal [49], assign a specific distribution on the new node time so that it is more probable to move to a certain height on the branch, either far away from or close to its original position. This section applied Random walk proposal (the operators in this paper), Bactrian proposal and Beta proposal to the three data sets, and the results are compared to those using Category configuration.

The comparisons are shown in Figure 22, Figure 23 and Figure 24. It is indicated that Beta proposal achieved worst performance in the three analysed data sets. The performance of the Constant Distance operator (Random walk) and Bactrian proposal varies depending on different data sets. However, these two proposal methods are both more efficient than the Category configuration. Therefore, it still needs further investigation to demonstrate the effectiveness of different proposals.

4. UcldstdevScaleOperator: a scale operator on standard deviation

It should be noted that the proposed ConstantDistance operator parameterises branch rates as continuous random variables, instead of discrete rate categories as is used in current BEAST2 settings. In uncorrelated relaxed clock model, branch rates are assumed to have a lognormal prior distribution, where the real mean is fixed to 1 and the standard deviation (denoted by Ucldstdev) is usually sampled with a hyper prior such as gamma($\alpha = 0.5396$, $\beta = 0.3819$). When a new Ucldstdev is proposed in one state during MCMC sampling by normal operators, the probability of all rates change as well under the new log normal distribution. Therefore, the authors implemented a separate operator working on Ucldstdev, which is able to solve this problem properly. The first step is to propose a new Ucldstdev by a scale operation, which multiplies current Ucldstdev by a random factor, as is shown in equation (38).

$$Ucldstdev' = Ucldstdev \times \text{scale} \quad (38)$$

where $\text{scale} = \text{Factor} + [\xi \times (\frac{1}{\text{Factor}} - \text{Factor})]$ and ξ is a random variable from a $Uniform(0, 1)$, Factor is a user-defined parameter to specify how bold the proposal is.

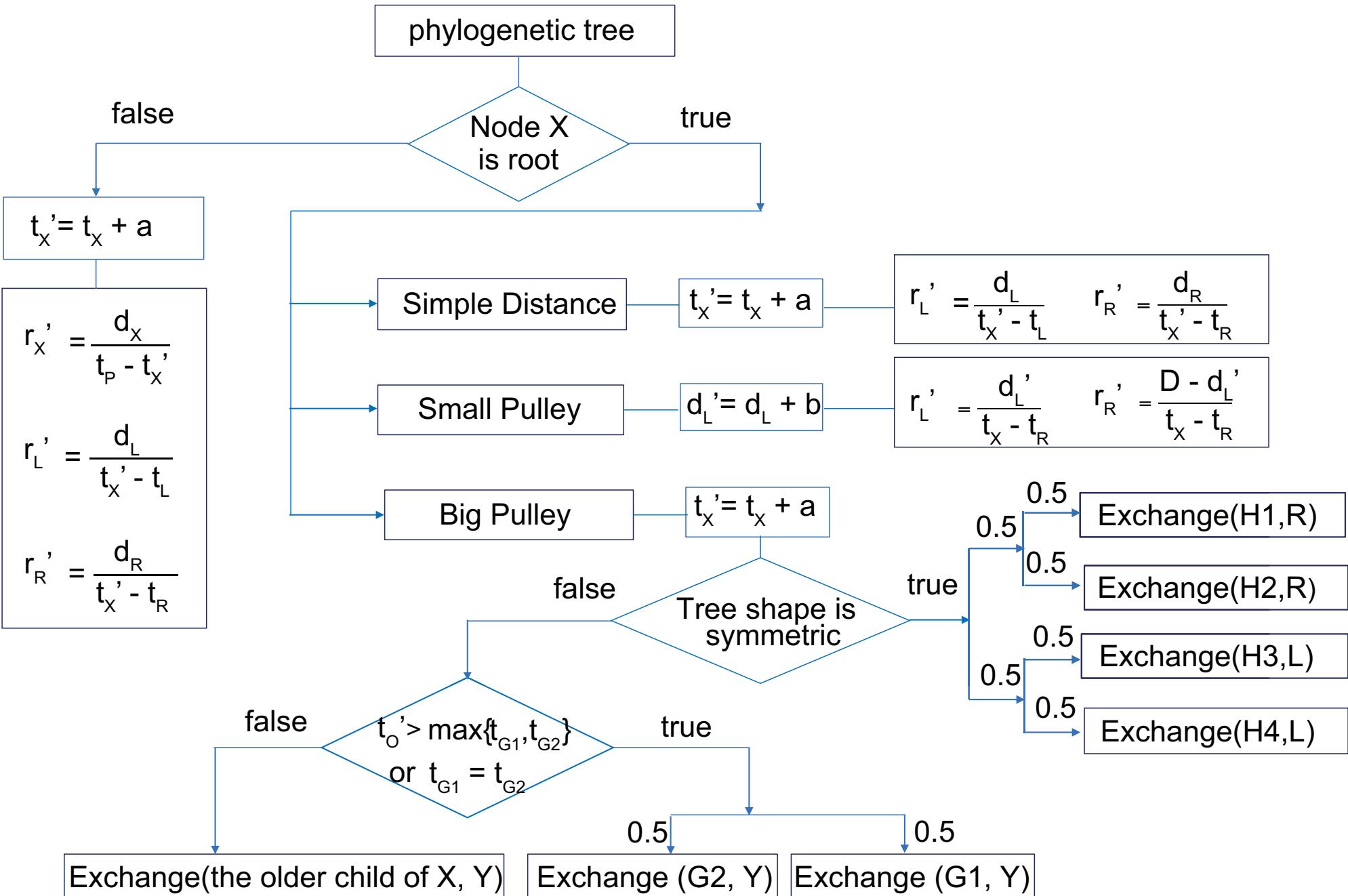
Secondly, all the branch rates are proposed based on the new $Ucldstdev'$, given the probability of original $Ucldstdev$, which is calculated using equation (39).

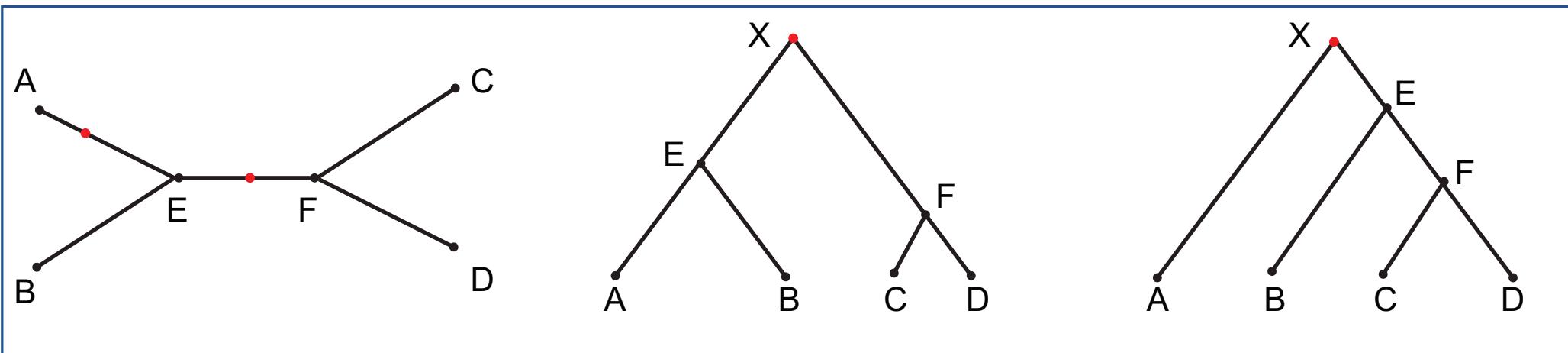
$$r'_i = \text{icdf}_{stdev'}[\text{cdf}_{stdev}(r_i)] \quad (39)$$

where the notations $cdf(\cdot)$ and $icdf(\cdot)$ represent the cumulative and inverse cumulative density function of log normal distribution.

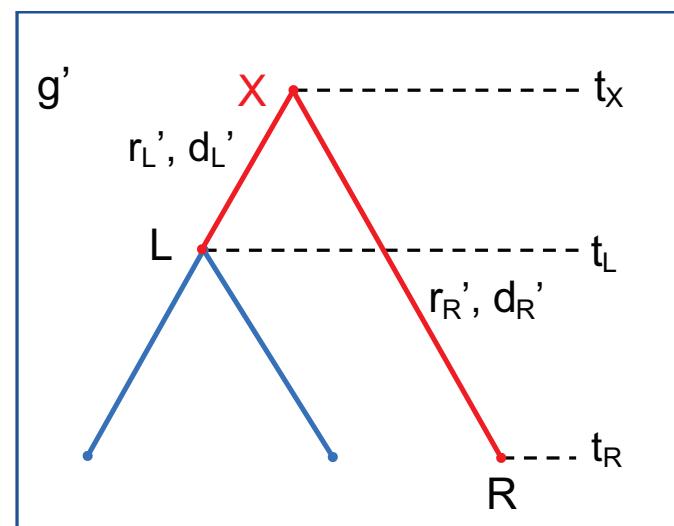
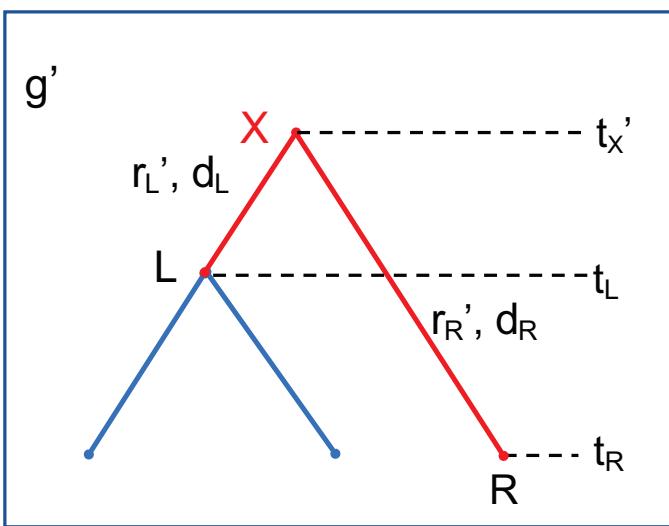
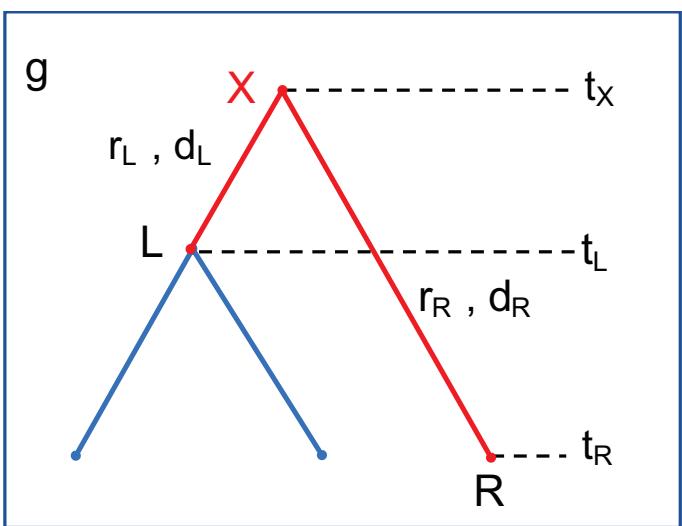
Finally, it is important to return the corrected hastings ratio, since the proposal is associated with one random variable, $Ucldstdev$ and $(2n - 1)$ branch rates. As is shown in equation (40), the ratio includes the scale operation and rates changing under the same probability.

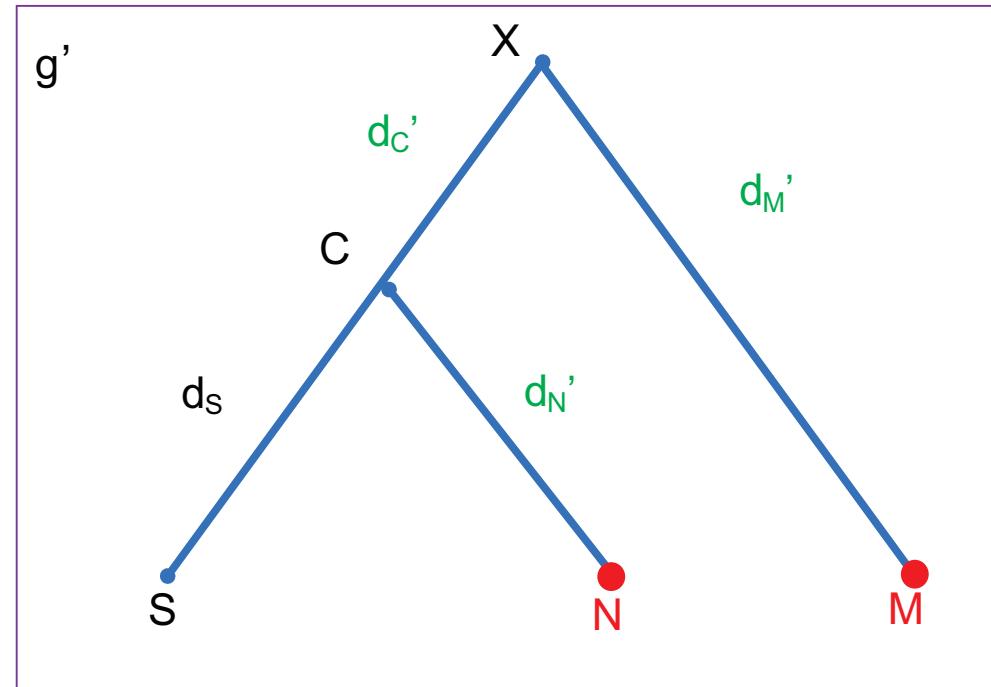
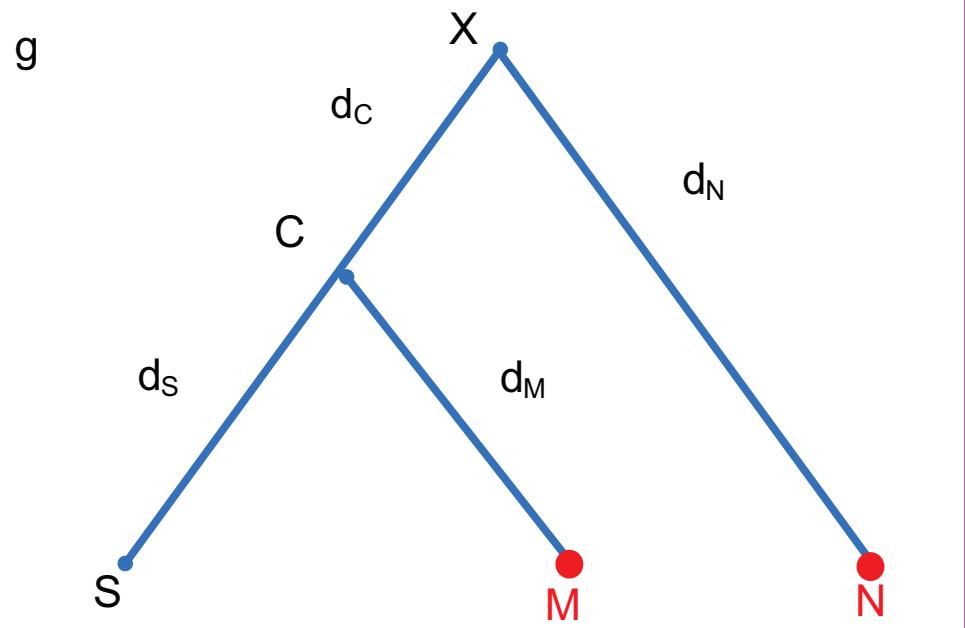
$$\mathbf{J}_{Ucldstdev} = \frac{1}{\text{scale}} \times \prod_{i=1}^{2n-1} \frac{\partial \text{icdf}_{Ucldstdev'}[\text{cdf}_{Ucldstdev}(r_i)]}{\partial r_i} \quad (40)$$

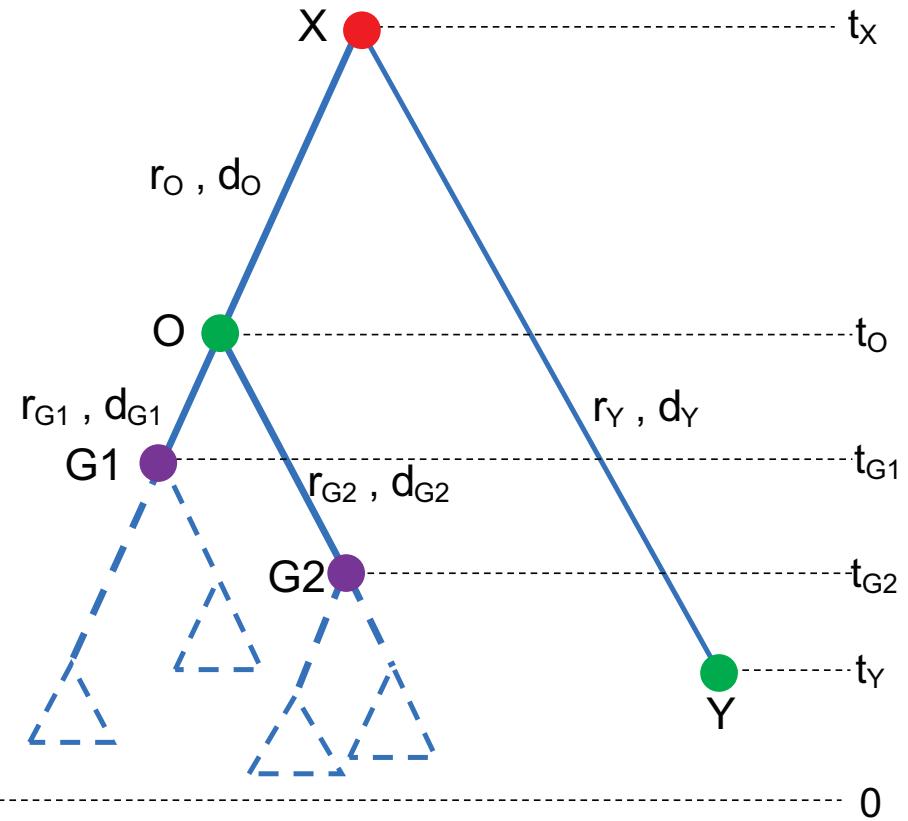
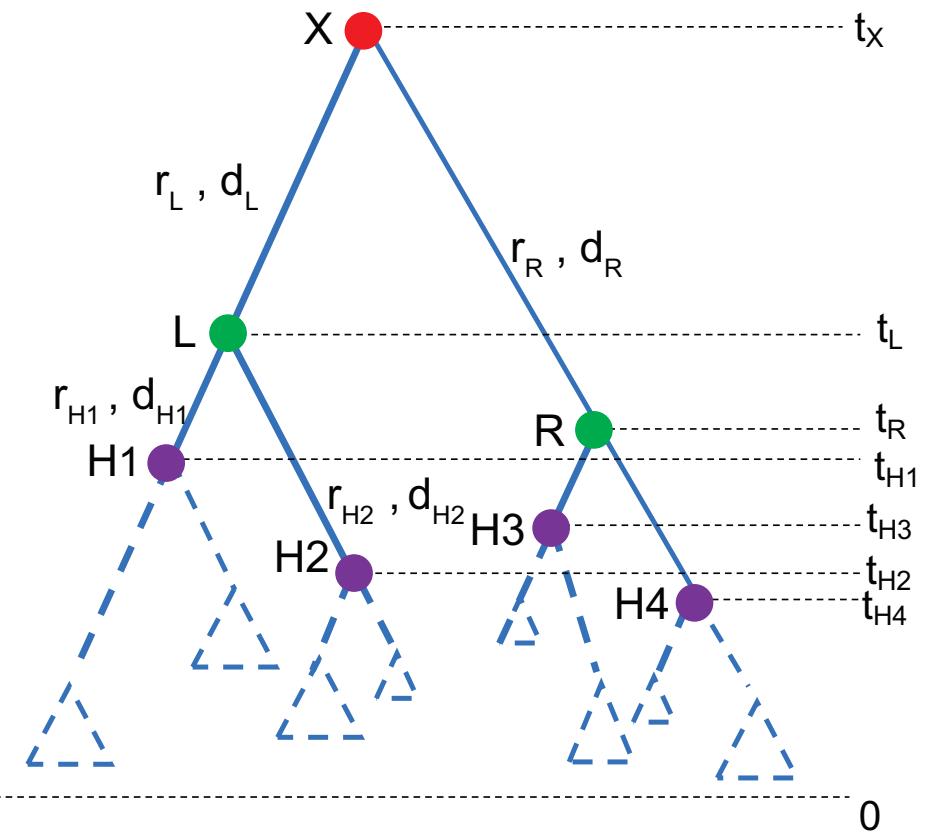




(a) An underlying unrooted tree

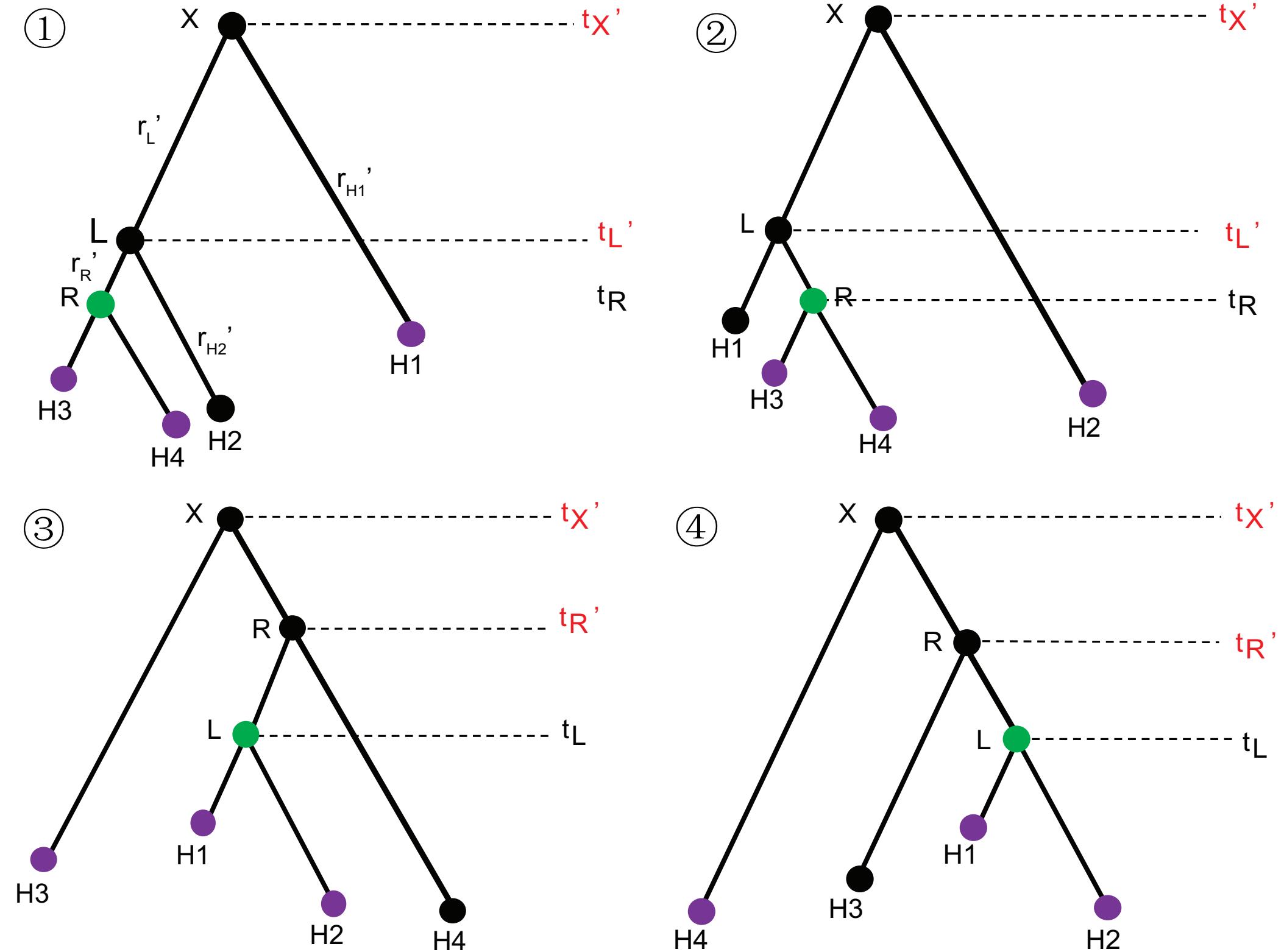


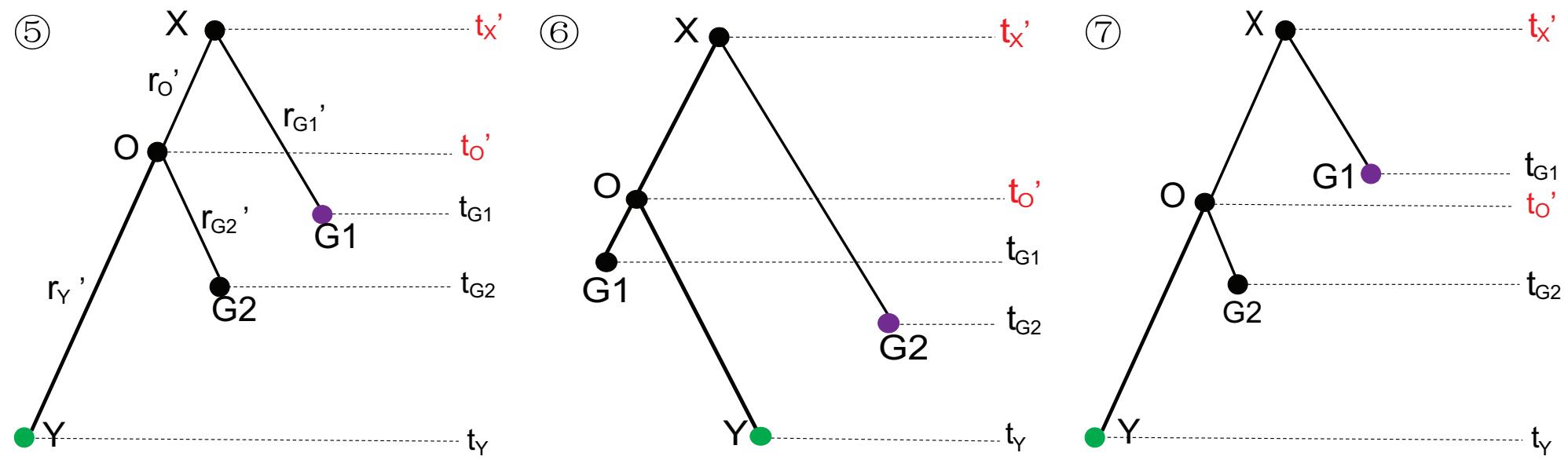


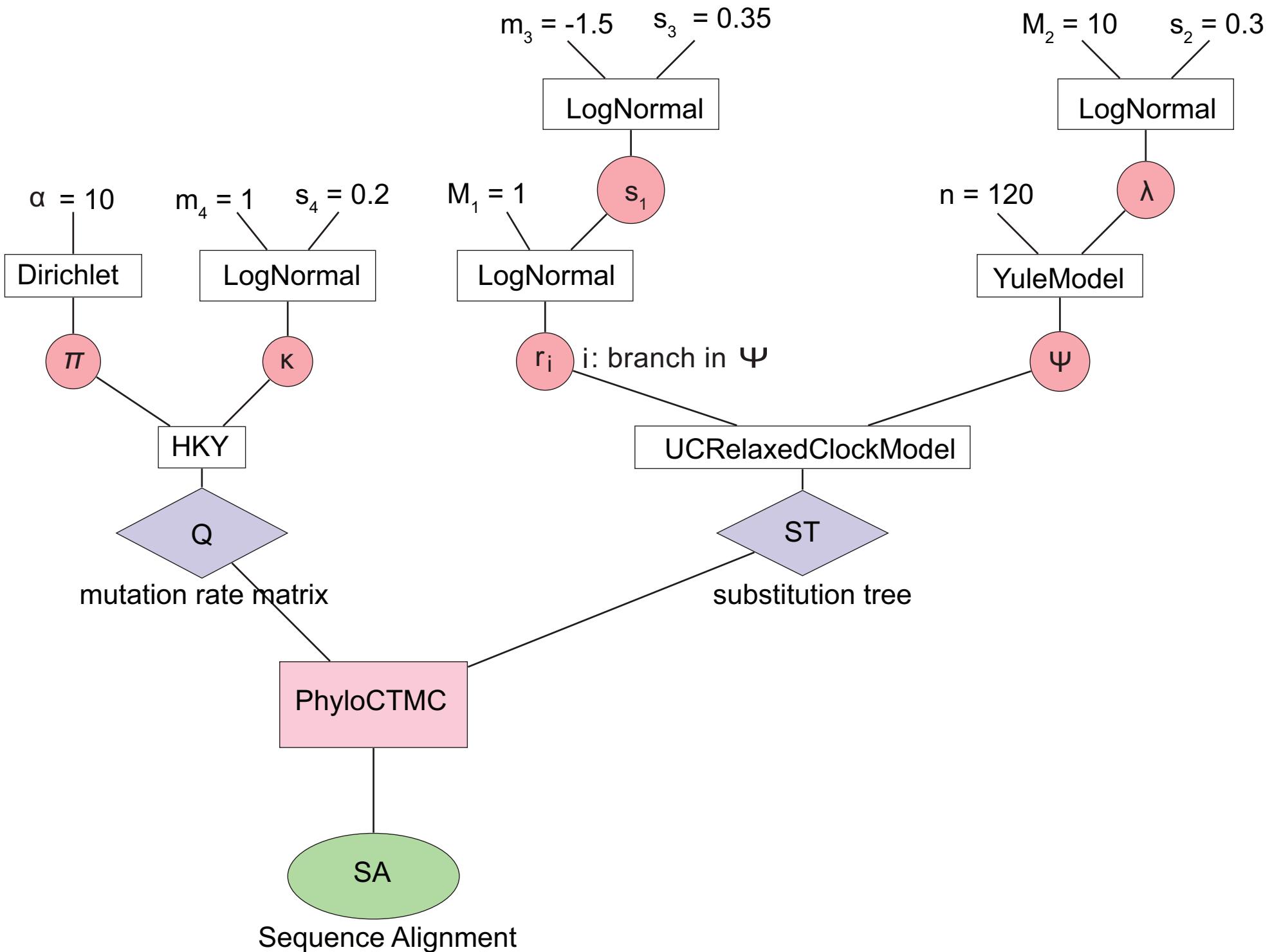


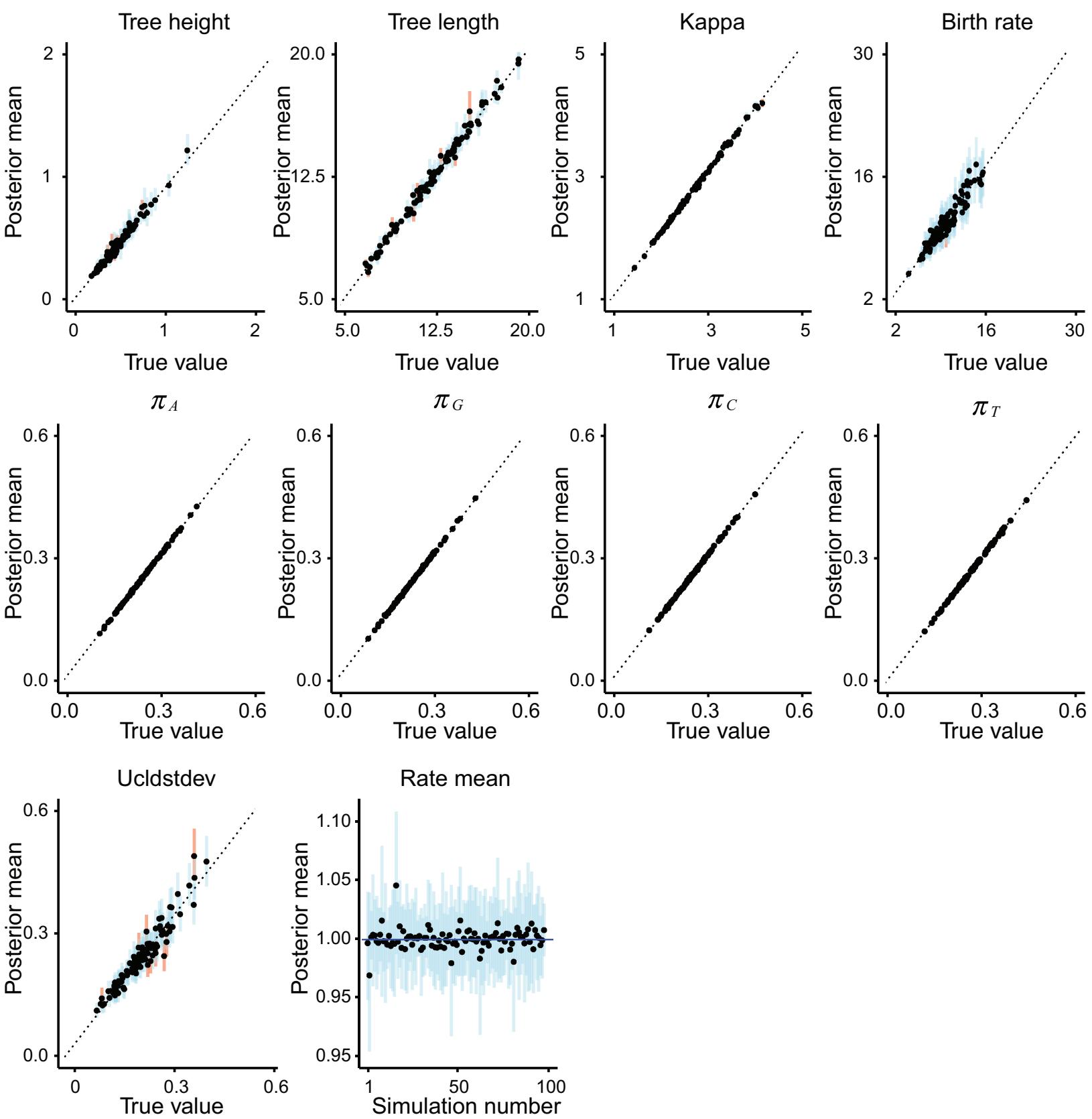
Figure

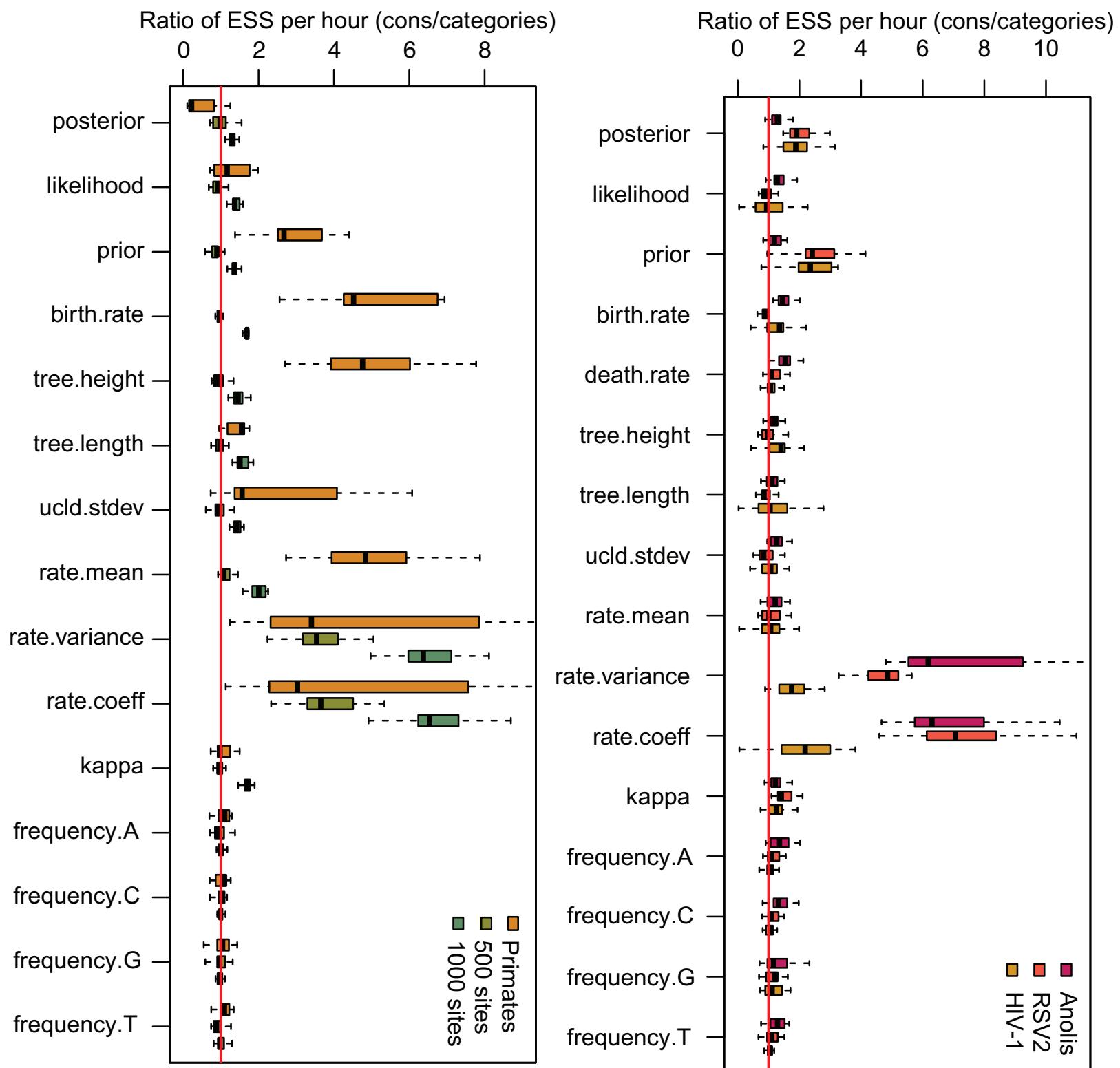
[Click here to access/download;Figure;Fig06-symmetric.eps](#)

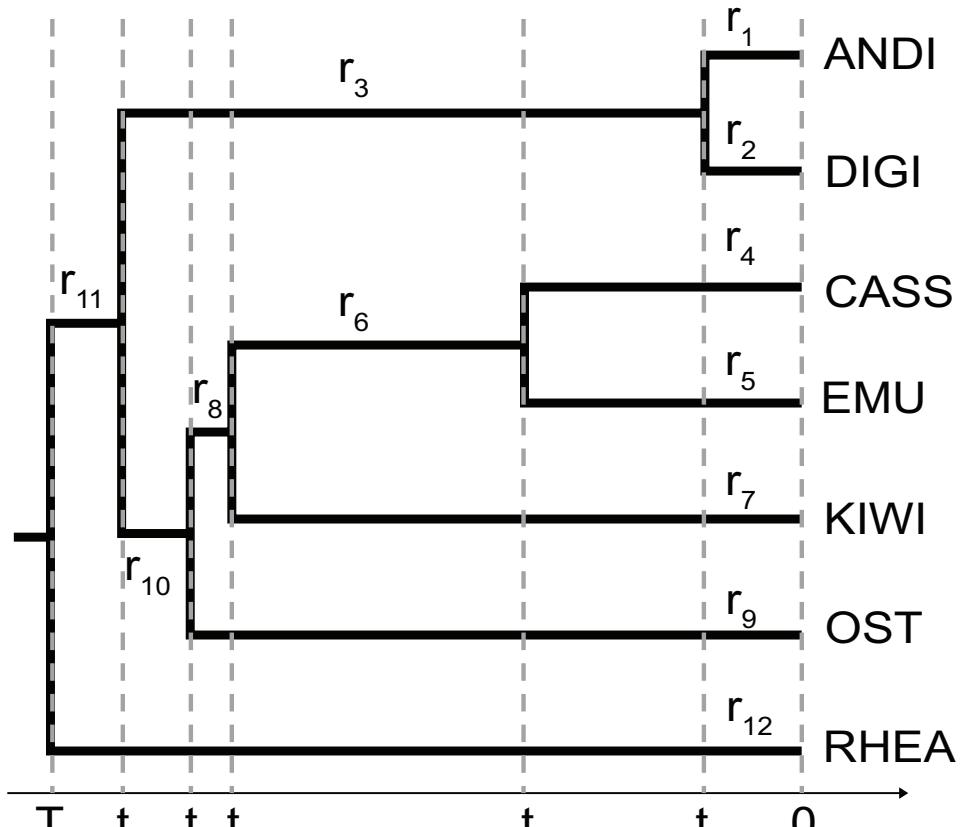




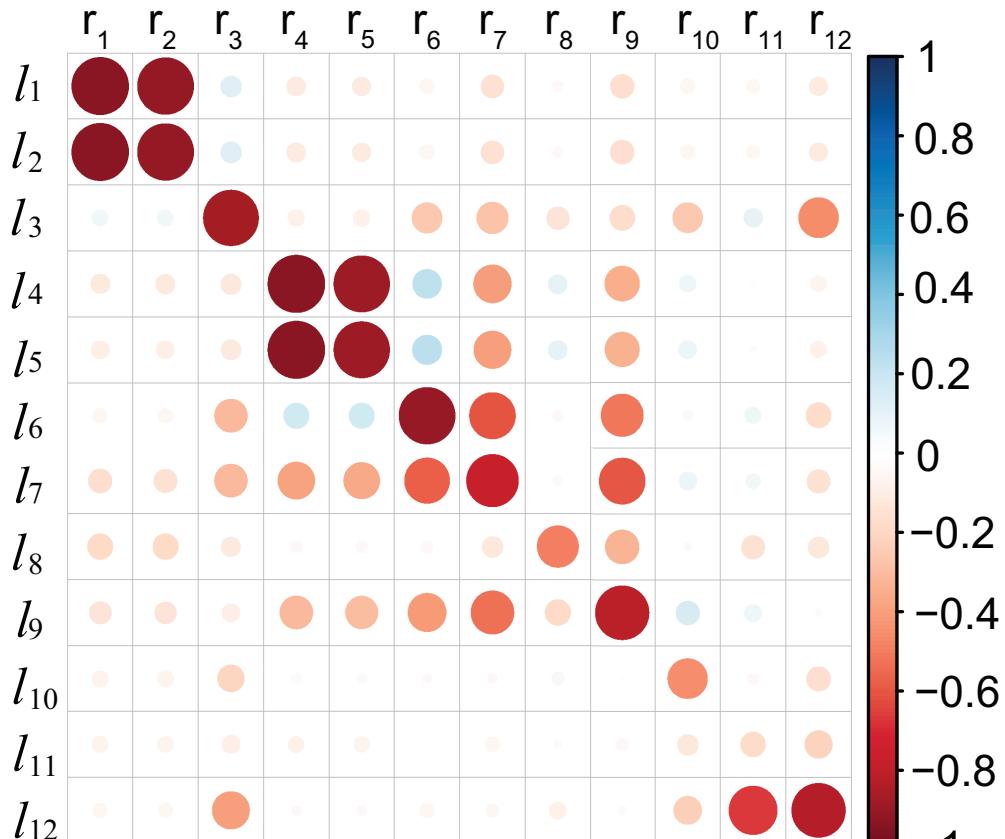




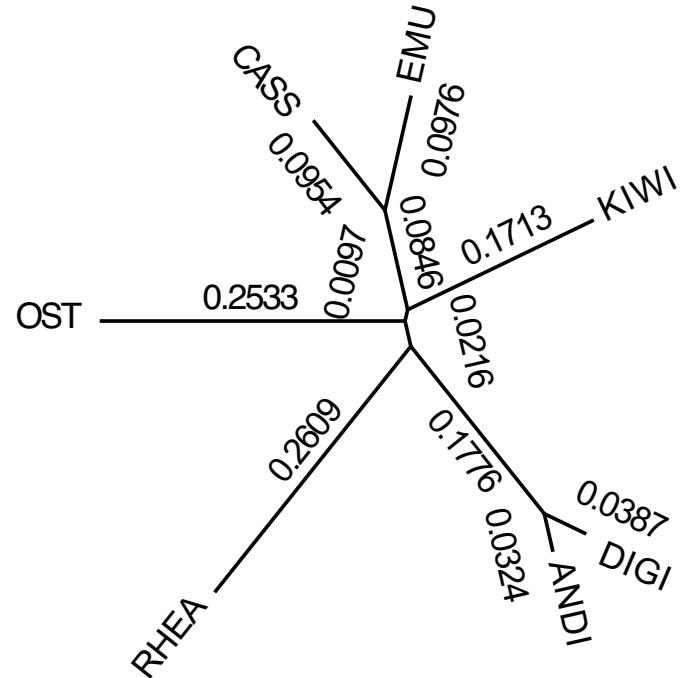




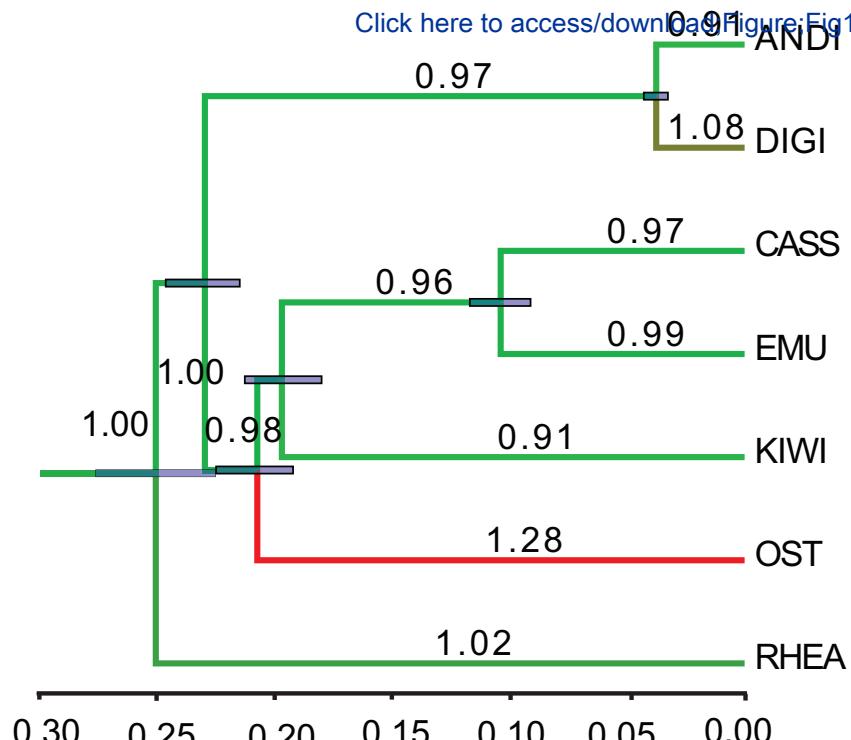
(a) Maximum clade credibility tree



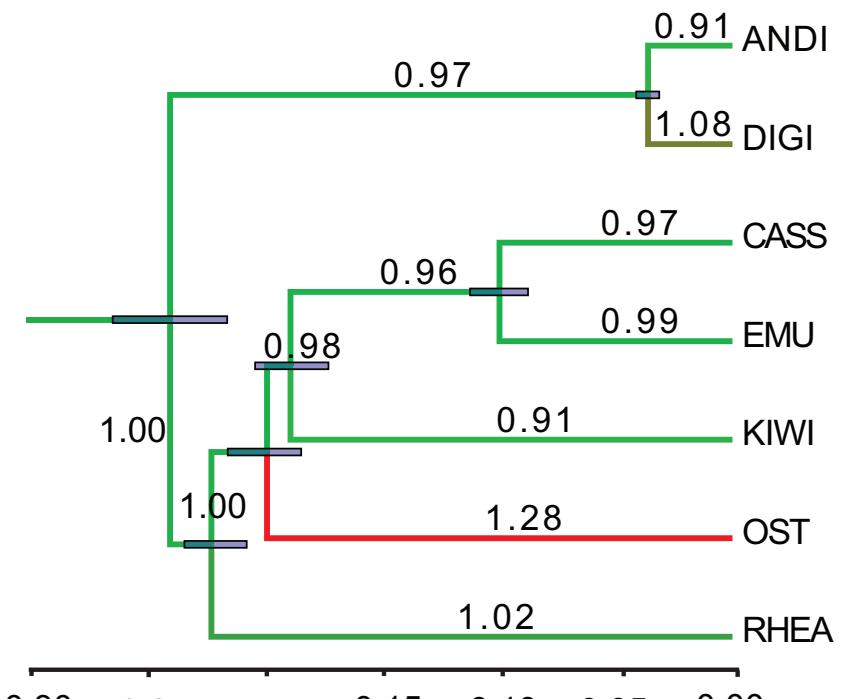
(b) Rates and branch lengths pairwise comparison



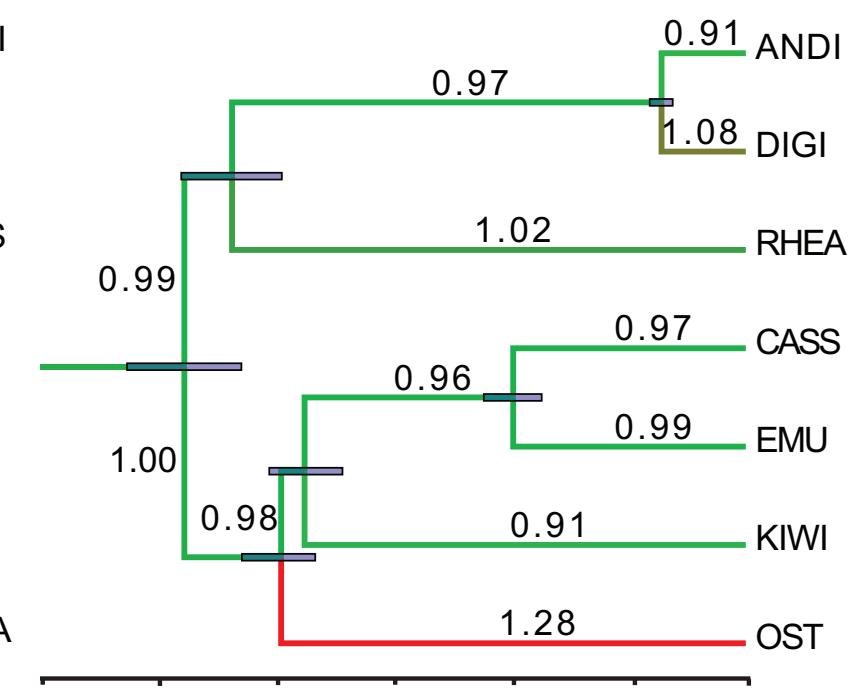
(a) Unrooted tree



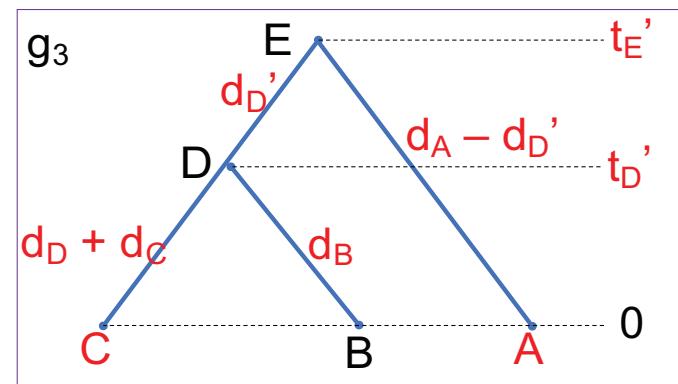
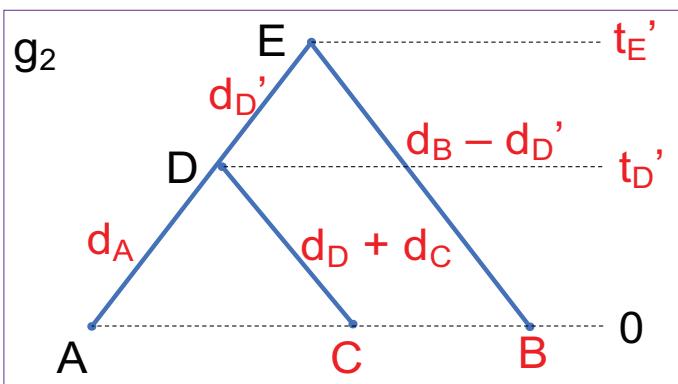
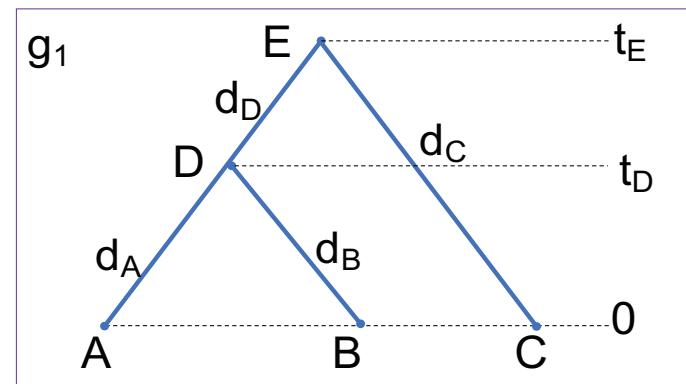
(b) The tree with 86.6% probability

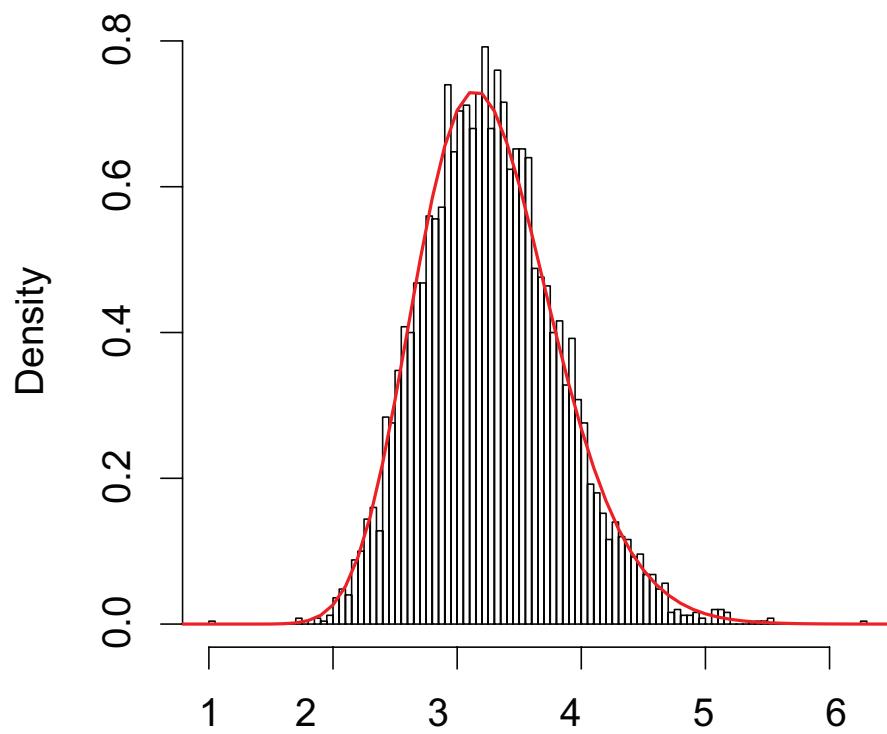


(c) The tree with 7.5% probability

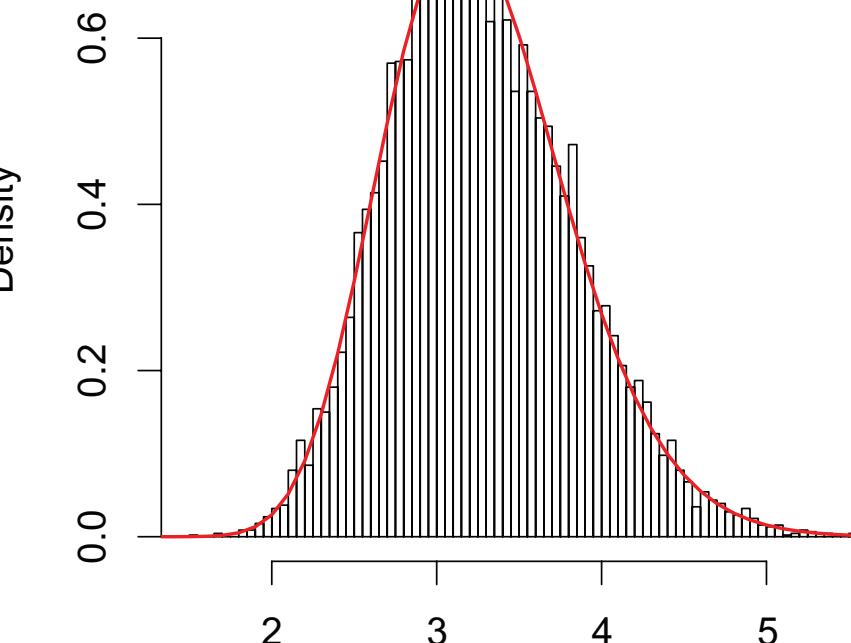


(d) The tree with 5.9% probability

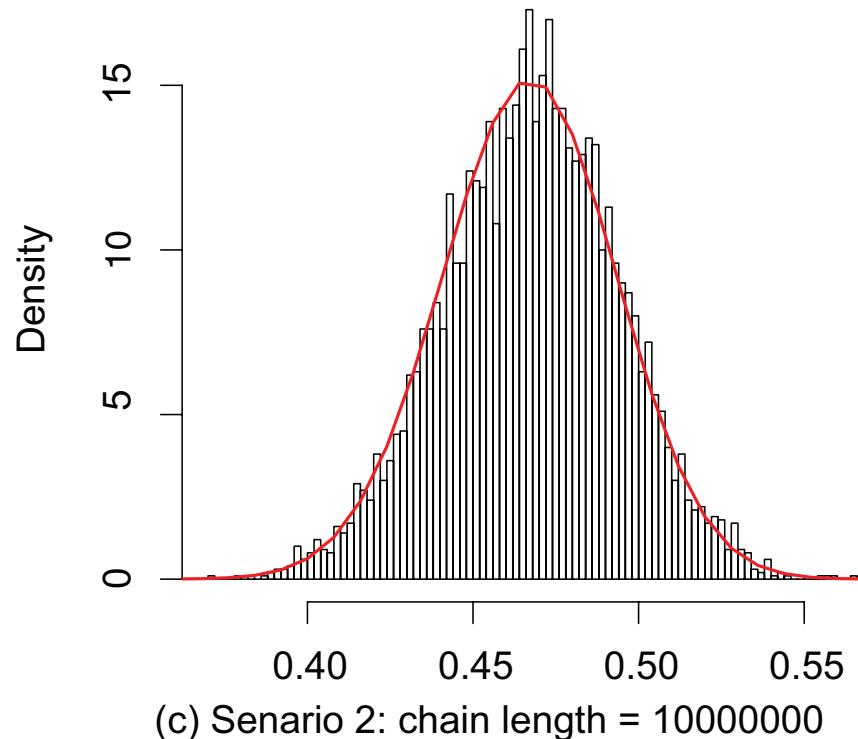




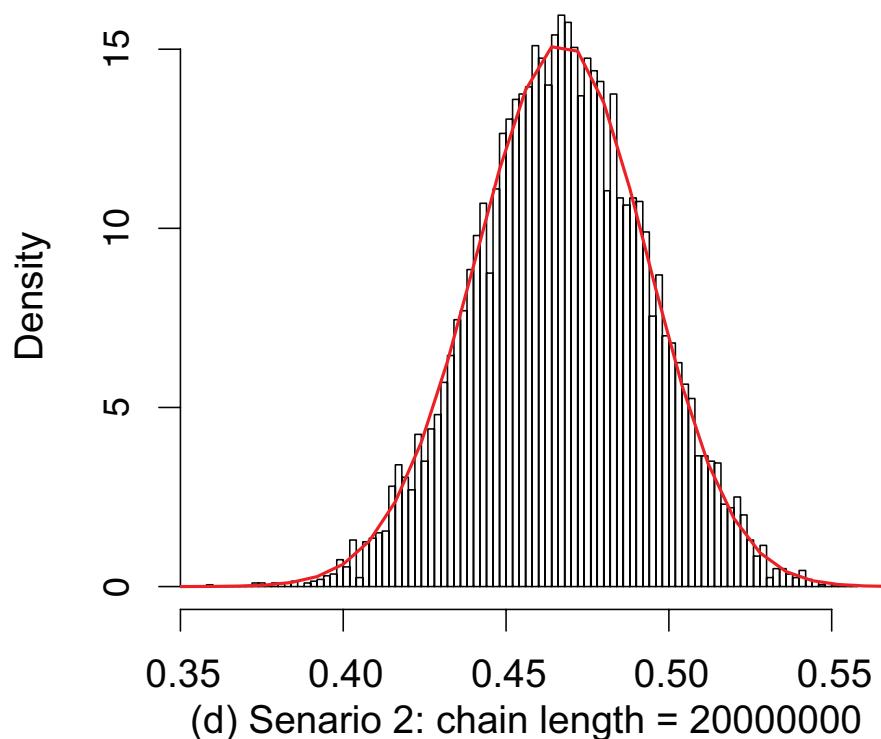
(a) Scenario 1: chain length = 10000000



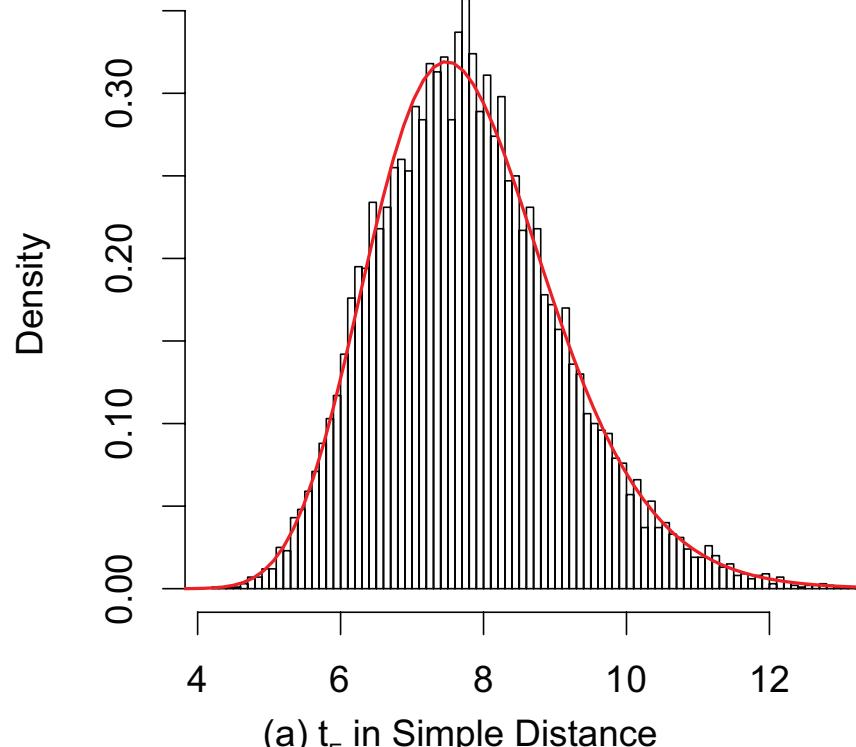
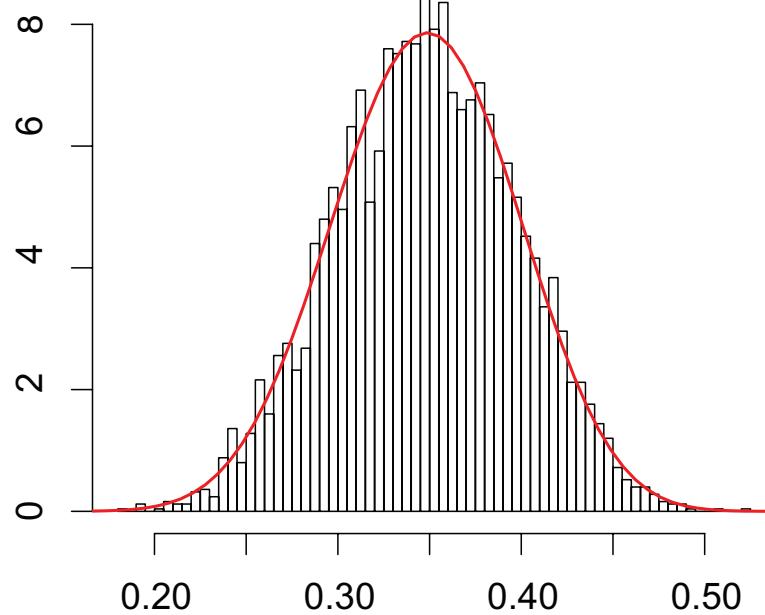
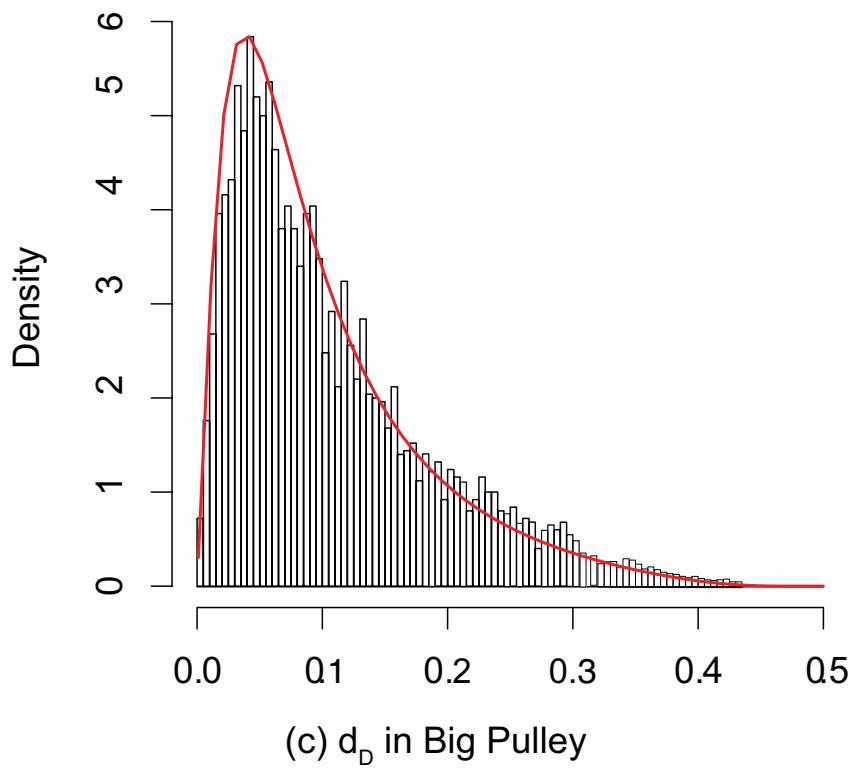
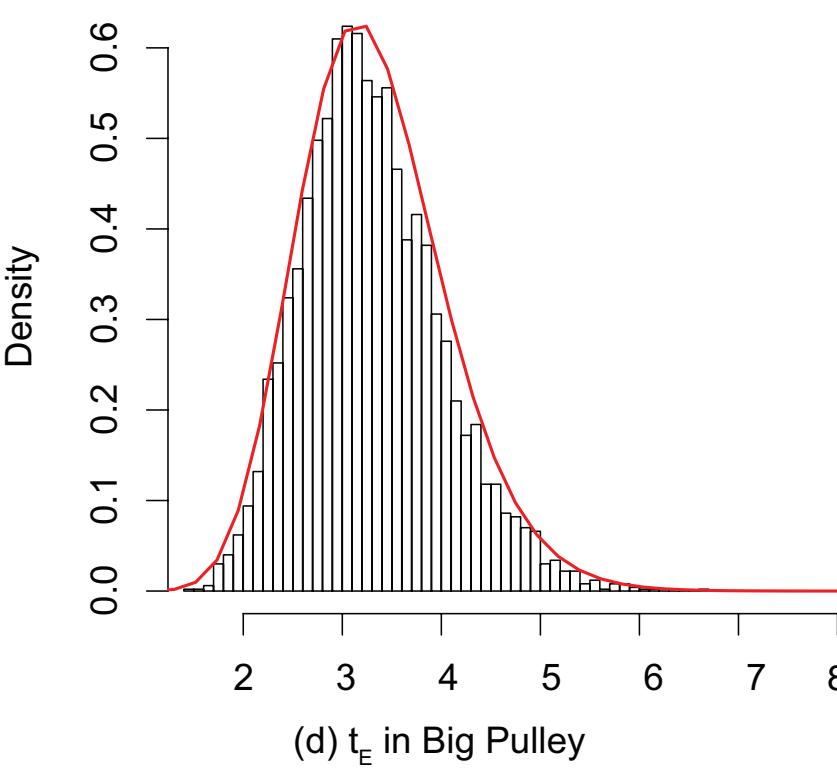
(b) Scenario 1: chain length = 20000000

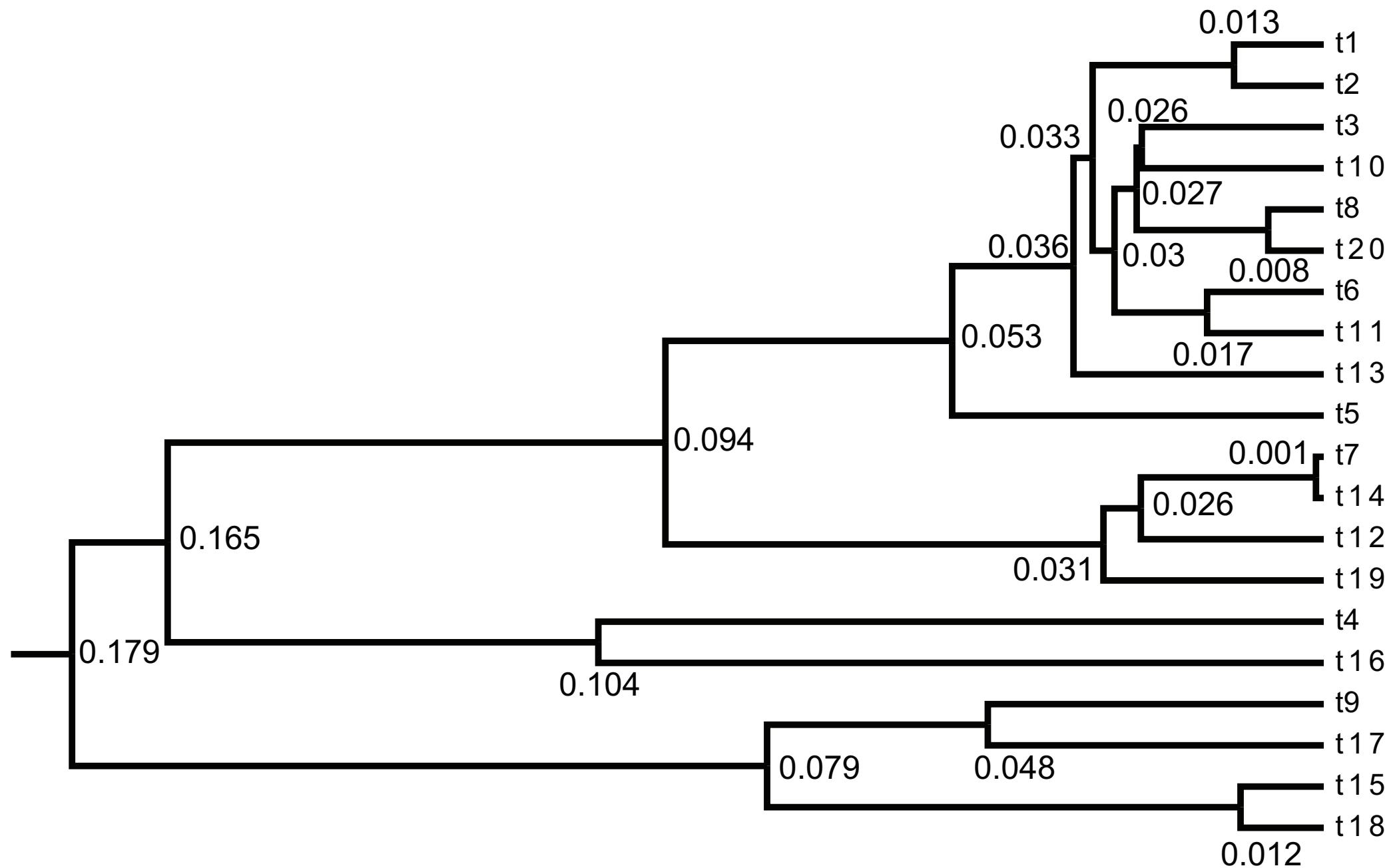


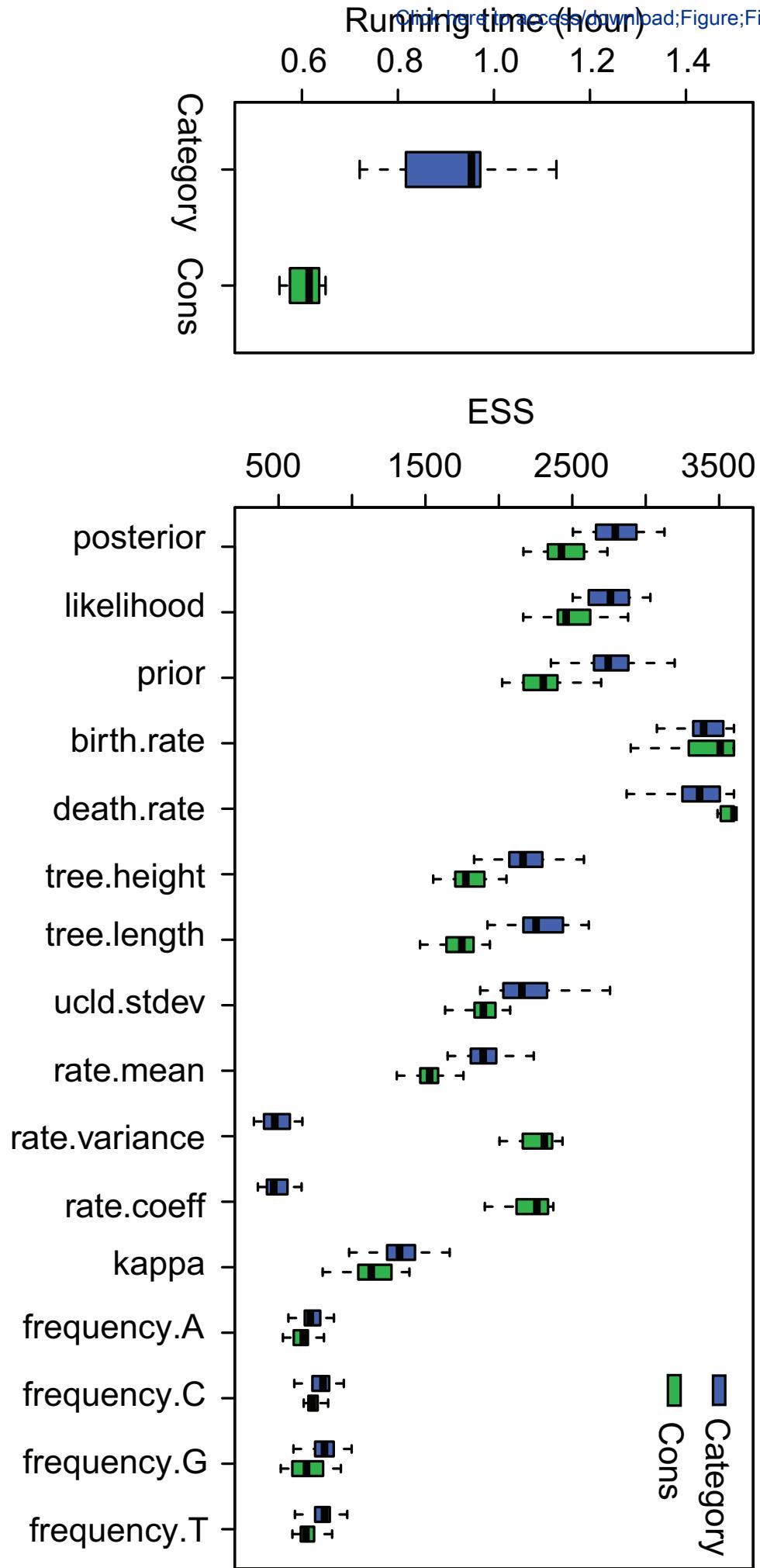
(c) Scenario 2: chain length = 10000000

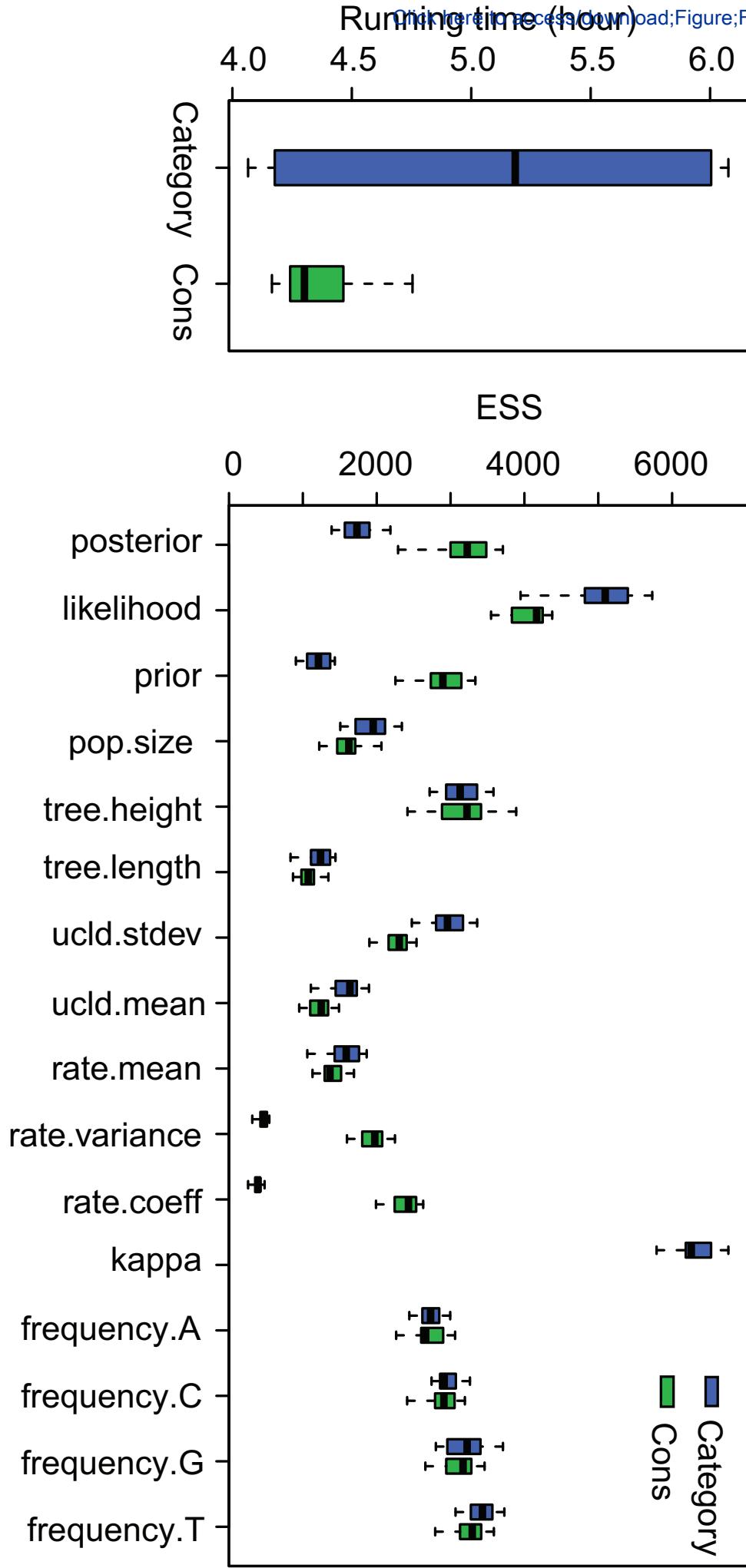


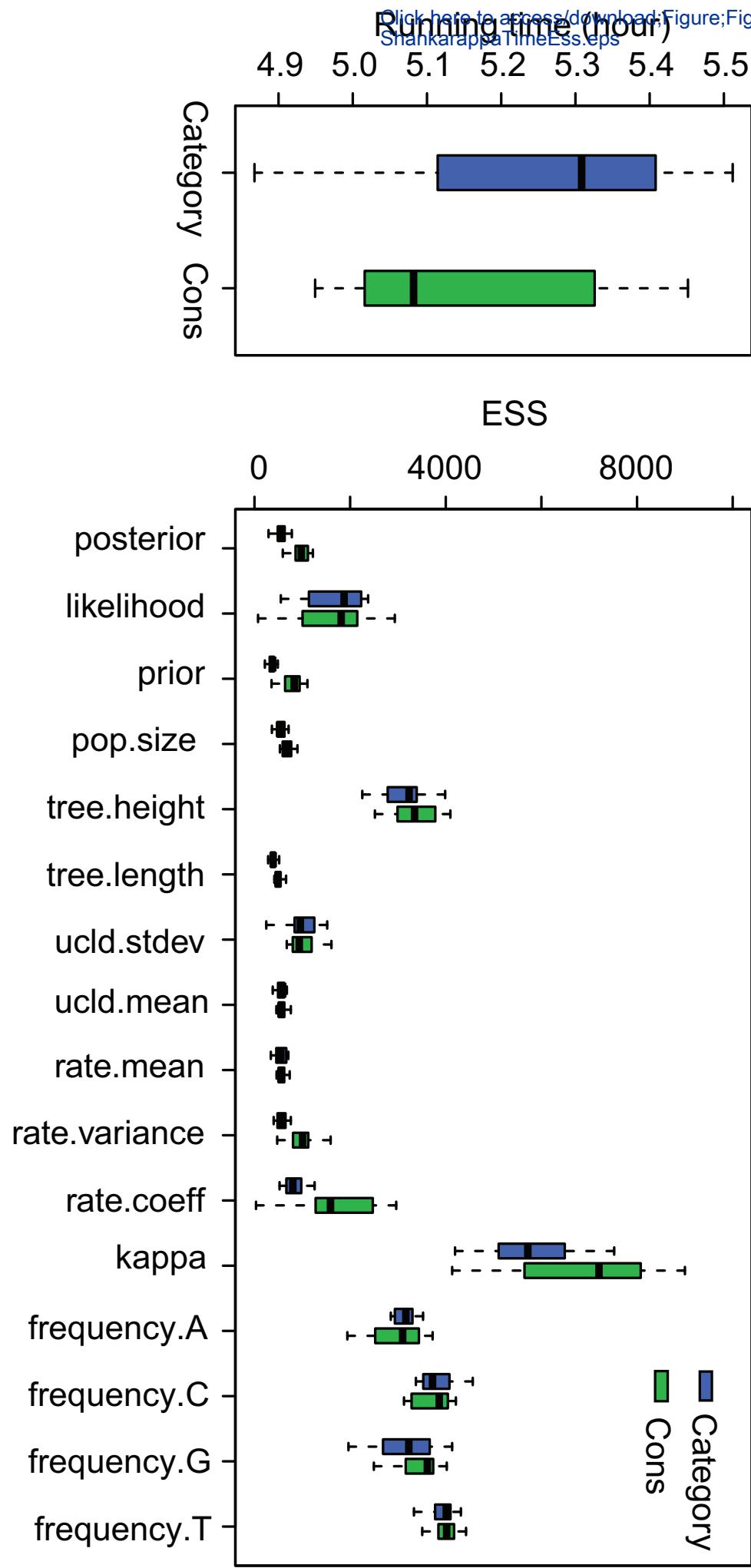
(d) Scenario 2: chain length = 20000000

(a) t_E in Simple Distance(b) d_D in Small Pulley(c) d_D in Big Pulley(d) t_E in Big Pulley



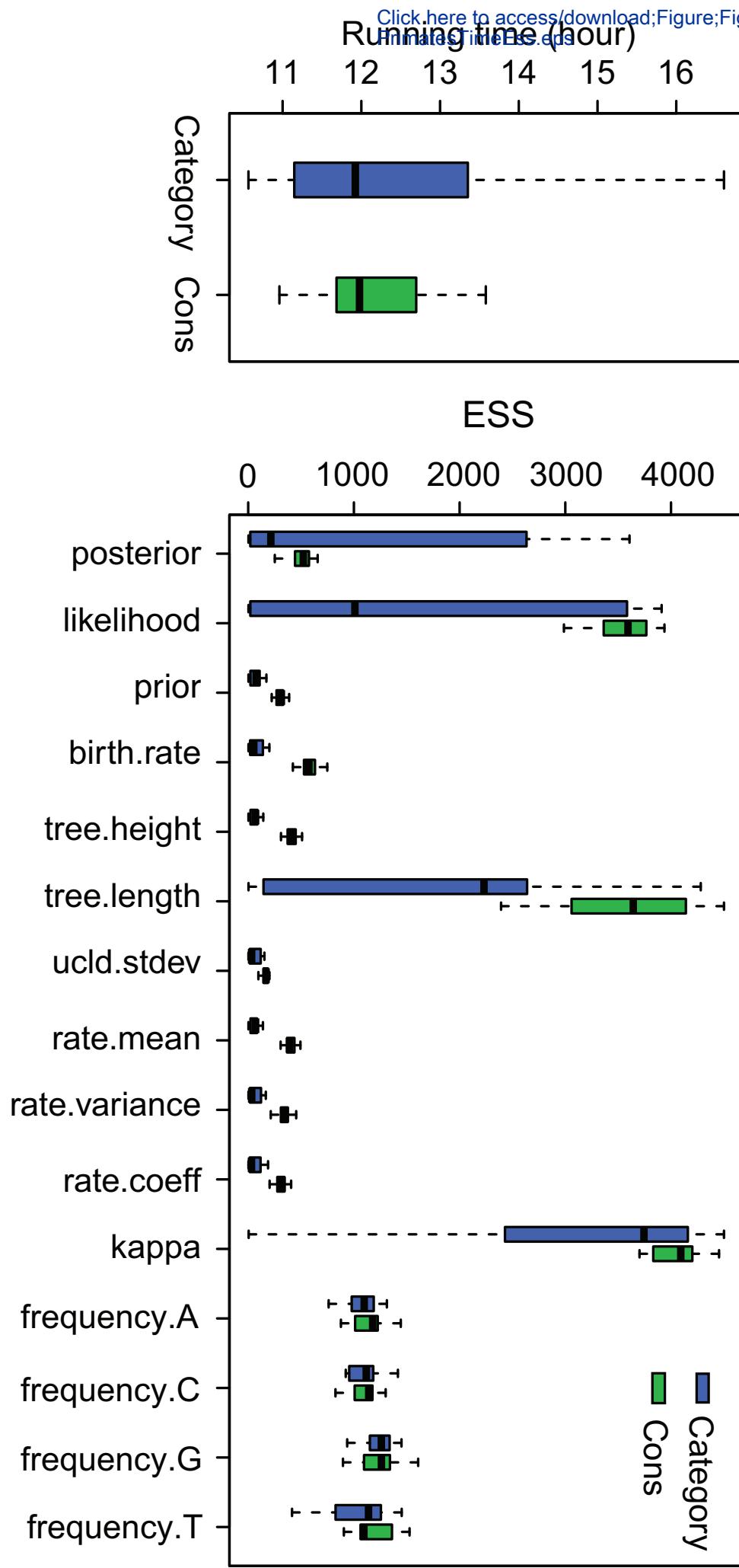




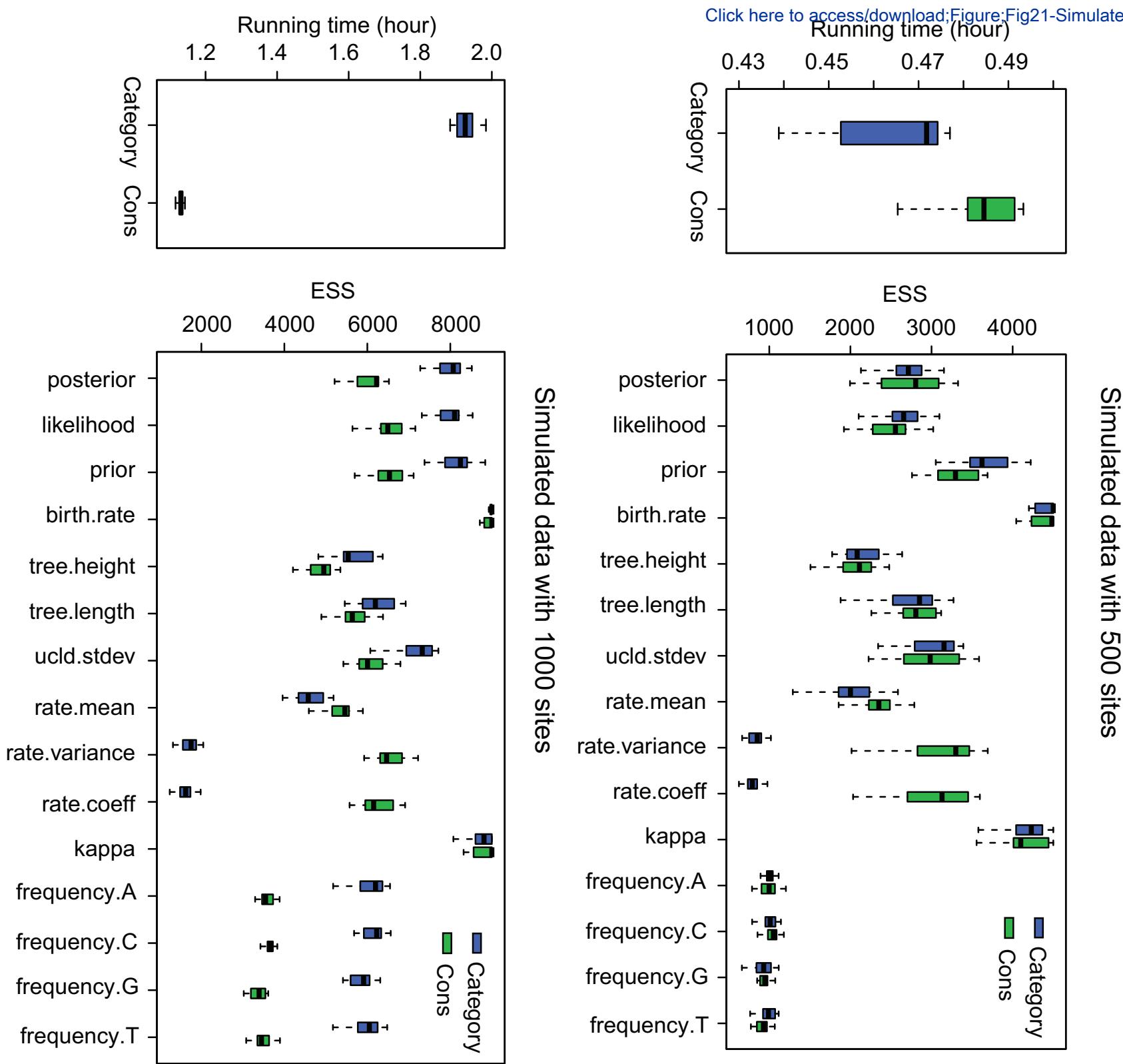


Figure

[Click here to access/download; Figure; Fig20-HmmatesTimeESS.eps](#)



Figure



Ratio of ESS per hour (cons/categories)

