

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

Manuscript Number:	EVOB-D-19-00107R2	
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. 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 more efficient than the current operators in BEAST2. Especially for large data sets, the improvement is up to half an order of magnitude.</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</p>	

below.

Response to reviewer #1

General comments

Some remaining typos

p11, l60 don't change -> doesn't change

p14, l37 cateogry -> category

Table 2 simualted -> simulated

Table 4, figure 14 senario -> scenario

Author's Response:

Thank you very much for pointing out the typos.

In the latest manuscript, the mistakes have been corrected.

Response to reviewer #2

General comments

1. The terminology around taxa with sampling times in the past is confusing.

(1) In algorithm 2, the use of extant/extinct does not match any common usage. The definitions on page 7 line 33 for Y and O are clearer and avoid this confusion. I suggest removing extant/extinct entirely.

(2) It seems that the authors are using "sampled ancestor" to mean "a tip whose sampling time is in the past," or more compactly "a heterochronous tip." This is not the standard usage of sampled ancestor, which generally means "a sampled taxon who has descendants which are also sampled taxa."

(3) On page 6 (l 24-25), it is stated that a sampled ancestor has no descendants (is of degree 1, having only a parent). This suggests that "sampled ancestor" is being used to mean "heterochronous tip."

(4) Does the "Big Pulley" algorithm apply if O is a node of degree 2 (having a parent and a single child)? It seems like it should, but I do not know if it would fall under the symmetric or asymmetric case. The authors are free to leave this case to future work, so long as the terminology surrounding what O represents is made clearer.

Author's Response:

Thank you for your professional comments and suggestions.

(1) We have removed the notation of "extant/extinct" in the revised manuscript. The two child nodes of the root are denoted by O (having two child nodes) and Y (having no child nodes).

(2) – (3) We have removed the usage of "sampled ancestor" in the revised manuscript. In the trees where the Big Pulley operator works on, node Y means a heterochronous tip.

(4) In the revised manuscript, we have made it clear that O is the node with two child nodes. The situation where O has a parent and a single child node will be handled in our future work.

2. Figure 10 and Table 7 appear contradictory about the models used to analyze RSV2 and HIV-1. In the table it is stated that there is an operator on the population size, but in the figure there are efficiencies listed for birth and death rates instead.

Author's Response:

Thank you for pointing out this mistake.

In the latest manuscript, we have corrected the inconsistent parameters in Figure 10 and Table 7. To be specific, the Anolis data set has a birth-death tree prior in our model and two corresponding parameters (birth rate and death rate) are sampled in the analysis. For RSV2 and HIV-1 data sets, we used coalescent model as the tree prior, where the parameter population size (pop.size) is sampled. Moreover, we also sampled clock mean (ucl.d.mean) for RSV2 and HIV-1 because we specified sampled dates at the tips. Therefore, in Figure 7, birth.rate and death.rate are compared in Anolis data set, pop.size and ucl.d.mean are compared in RSV2 and HIV-1. The rest of parameters are the same in the three data sets analysis.

Typos and other minor comments:

1. The comparison of ESS for the clock standard deviation between "cons" and "categories" doesn't seem exactly fair, since there is an entirely new operator for the standard deviation of the clock in "cons." -It may be useful to mention the "nocons" tests were run in previous analyses. This would assuage any concerns that the difference between "categories" and "cons" is about the change from discretized to continuous branch rates, rather than about the use of the constant distance operator.

	<p>Author's Response: Thank you for your professional comments In the revised manuscript, we have added a new subsection to discuss the "NoCons" configuration in Appendix Section 4. The details are as follows. In the comparison of ESS for the clock standard deviation (denoted by ucld.stdev in Figure 10) we specified a normal scale operator in "Category" configuration. In "Cons" configuration, the UcldstddevScaleOperator is used to sample the clock standard deviation of continuous rates. To avoid the concern that the difference between "Category" and "Cons" is a result of how rates are parameterised (i.e. discrete or continuous), we set another configuration where continuous rates are sampled without using the ConstantDistance operator (denoted by "NoCons" configuration). The weights of the operators in "NoCons" are the same as those in "Category" which is detailed in Table 7. We ran the analysis using the three real data sets (Anolis, RSV2 and HIV-1) and the comparison of ESS per hour between "Category", "Cons" and "NoCons" is summarised in Figure 25. The figure shows ESS per hour in log_10 space of ucld.stdev in 20 independent MCMC chains. As can be seen, "Cons" configuration gives similar performance, comparing with "Category". This indicates UcldstddevScaleOperator works properly on continuous rates. Moreover, ESS per hour is much larger in "Cons" than in "NoCons", where both continuous rates are sampled. Therefore, the proposed operators contribute to the improved performance. However, we noticed that the rate parameterisation does have some mixing issues in MCMC chains. In the future, we will further investigate how to parameterise branch rates to get better performance when using the proposed operators.</p> <p>2. Table 7: "Substitution model"</p> <p>Author's Response: In the revised manuscript, it has been replaced by "Substitution model".</p> <p>3. p26 l28: the simulations are not definitive proof, but rather a convincing demonstration</p> <p>Author's Response: In the revised manuscript, the inappropriate statement "Thus, Simple Distance is proved to be correct." is replaced by "Thus, Simple Distance samples the root time and two branch rates correctly."</p> <p>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.</p>
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
1 Zhang and Drummond

2
3
4
5
6
7

METHODOLOGY ARTICLE

Improving the performance of Bayesian phylogenetic inference under relaxed clock models

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

*Correspondence:
rzh419@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 more efficient than the current operators in BEAST2. Especially for large data sets, the improvement is up to half an order of magnitude.

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

10 Bayesian phylogenetic inference via MCMC is computationally intensive for large
11 data sets. Two approaches to improve efficiency are (i) by making faster likelihood
12 calculations, and (ii) by incorporating more effective proposal kernels. Calculating
13 the phylogenetic likelihood is computationally expensive. Hence, researchers have
14 tried many ways to tackle the computation burden in the likelihood calculations,
15 such as detection of repeating sites [12], approximate methods (e.g. [13, 14]) and
16 the use of parallelisation strategies (e.g. BEAGLE [15]).
17

18 However the overall efficiency of the sampling process also depends strongly on
19 the construction of the proposal mechanism. An effective proposal mechanism is
20 proficient at exploring the posterior distribution, and can do so with fewer steps
21 in the MCMC chain. Therefore fewer likelihood calculations are required, since
22 each step in the chain that changes the tree or substitution parameters requires a
23 likelihood calculation.
24

25 A major limitation in Bayesian MCMC analysis of phylogeny lies in the efficiency
26 with which operators sample the tree space [16, 17]. Fast and reliable estimation
27 is dependent on a good mixture of operators, since the posterior distribution often
28 exhibits correlations between the tree and other random variables.
29

30 In this paper, we present a novel operator that works alongside standard operators
31 by proposing moves within a subspace of constant genetic distances. Namely, the
32 proposed operator changes both divergence times of nodes and neighbouring branch
33 rates so that the implied genetic distances are not changed. For time-reversible
34 substitution models the phylogenetic likelihood will also be unchanged under this
35 operation. The proposed operator has been implemented and tested in BEAST2
36 [18].
37

38 Preliminaries 39

40 Bayesian MCMC 41

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

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
58

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].

Methods

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.

Operations on internal nodes

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}' .

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

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.

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.

Step 4 Propose new rates by using equation 3.

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

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

Operations on the root

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.

Simple Distance

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

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

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

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

16 *Step 4* Return the Green ratio α_{SD} .

17 Small Pulley

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

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

$$32 \quad d_L = r_L \times (t_X - t_L) \quad d_R = r_R \times (t_X - t_R) \quad (5)$$

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

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

$$42 \quad r_L' = \frac{d_L'}{t_X - t_L} \quad r_R' = \frac{D - d_L'}{t_X - t_R} \quad (6)$$

45 *Step 4* Return the Green ratio α_{SP} .

46 Big Pulley

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

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

- 54 • Swap the two nodes by pruning and regrafting, i.e. cutting \mathbf{M} (\mathbf{N}) at its
55 original position and attaching it to the original position of \mathbf{N} (\mathbf{M}).
- 56 • Propose $d_C' \leftarrow d_C + b$, where $b \sim \text{Uniform}[-v, +v]$. Make sure that $0 < d_C' <$
57 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 $d_S' = d_S \quad d_M' = d_M - d_C' \quad d_N' = d_N + d_C$ (7)
11
12

13 As can be seen from the above descriptions, the method *Exchange(M, 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 Additionally, operations in Big Pulley vary depending on the shape of phylogenetic
18 tree. In Figure 5, a symmetric tree is shown on the left, in which both the child
19 nodes of the root have two child nodes, i.e. **L** having children **H1**, **H2** and **R** having
20 children **H3**, **H4**. But in the asymmetric tree on the right, only one of the child
21 nodes of the root has two child nodes below it, i.e. **O** having children **G1**, **G2**. But
22 the other child node **Y** doesn't have any child node, which is a heterochronous tip.
23 The corresponding operations are detailed in the following two parts.
24
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 **Algorithm 1** Proposal for symmetric trees in Big pulley

34 {Step 1: Identify the two child nodes of the root **X**, denoted by **L** and **R**. Correspondingly, the node
35 times are denoted by t_X , t_L , t_R . The child nodes below them are denoted by **H1**, **H2**, **H3** and **H4**.}
36

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
 {Step 4: Update the rates on the corresponding branches.}
 {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 Finally, in this example the new rates would be updated by:
 15
16

$$17 \quad 18 \quad 19 \quad r_{H1}' = \frac{d_{H1}'}{t_{X'} - t_{H1}} \quad r_{H2}' = \frac{d_{H2}'}{t_{L'} - t_{H2}} \quad r_L' = \frac{d_L'}{t_{X'} - t_{L'}} \quad r_R' = \frac{d_R'}{t_{L'} - t_R} \quad (9)$$

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

Algorithm 2 Proposal for asymmetric trees in Big pulley

25 {Step 1: Identify the child node of the root **X** that has two child nodes below, which is denoted by **O**.
 26 The other child node of the root, which does not have any child nodes and means a heterochronous
 27 tip, is denoted by **Y**. The node times of the root **X**, **Y**, **O** and its child nodes are denoted by t_X , t_Y ,
 28 t_O , t_{G1} and t_{G2} respectively.}
 29 Let **X** be the root of the tree.
 30 Let **O** be the child of **X** that has children, and let **Y** be the child of **X** that does not have children.
 31 {Step 2: Propose a new node time for the root **X**.}
 32 $a \sim \text{Uniform}[-w, +w]$
 $t_{X'} \leftarrow t_X + a$
 33 {Step 3: Propose a new node time for the node **O**.}
 34 $a_3 \sim \text{Uniform}[-w, +w]$
 $t_{O'} \leftarrow t_O + a_3$
 35 **if** $t_{O'} < t_Y$ or $t_{O'} > t_{X'}$ **then**
 Reject the proposal.
 36 **end if**
 37 {Step 4: Adjust the distances according to the tree corresponding topologies.}
 38 **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
 39 **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).
 40 **else if** $t_{O'} < \min\{t_{G1}, t_{G2}\}$ **then**
 Reject the proposal.
 41 **end if**
 42 {Step 4: Update the rates on the corresponding branches.}
 43 {Step 5: Return the Green ratio α_{BP} .}

44 To give an example, assume we are going to propose tree ⑤. Firstly, $t_{X'}$ and $t_{O'}$
 45 are proposed in *Step 3* and *Step 4*. Then, in *Step 4*, the method *Exchange* (**G1**, **Y**)
 46 is applied, after which the four distances are adjusted as follows:
 47
48

$$49 \quad 50 \quad 51 \quad 52 \quad d_{G1}' = d_{G1} - d_{O'} \quad d_{G2}' = d_{G2} \quad d_{O'} = d_O + b \quad d_Y' = d_Y + d_O \quad (10)$$

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

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

45 These results show that the true values of the parameters are within the 95%
46 highest posterior density (HPD) interval approximately 95% of the time (Table 1).
47 This well-calibrated simulation study formed part of the validation of our imple-
48 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. Each setting is benchmarked using an Intel(R) Xeon(R) Gold 6138 CPU (2.00 GHz).

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, the efficiency varies in different data sets and also depends on what model is used in the analysis. For Anolis data set (29 taxa and 1456 sites), Category configuration performs better than Cons configuration, since most ratios of the parameters are slightly below the red line (which means smaller than 1). Moreover, for RSV2 (129 taxa and 629 sites) and HIV-1 data sets (117 taxa and 663 sites), some ratios of the parameters, “posterior” and “prior” in particular, are above the red line (larger than 1), which indicates that Cons configuration provides larger ESS per hour. Although there are several parameters sampled by Cons configuration having smaller ESS per hour, 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 indicates that the proposed operators behave better when sequence length is long. More specifically, in Primates data set (87 taxa and 19220 sites), the longer molecular sequence provides more accurate genetic distances, which leads to peaked likelihood distributions. In this circumstance, the proposed operators sample rates and node times that fit the constant genetic distances more efficiently.

Table 2 lists the average running time of 20 simulations for each data set. It can be seen that Cons configuration finished simulations with a little bit more time in most cases. This is because the continuous rates have to be adjusted for a new clock

standard deviation (See Appendix Section 4 for more details). Moreover, Table 2 also shows the parameter that has the smallest ESS in Category configuration, and is compare with the corresponding ESS in Cons configuration. Although the improvement in ESS is not obvious for both simulate data sets, it is noticed that ESS of the parameters are much larger in Cons configuration for all the real data sets. 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 1.55 to 8.53.

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.

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

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

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

Sampling a fixed unrooted tree

A limiting case for the relaxed molecular clock model (and one exploited in some of our validation tests) occurs for long sequences, when the branch lengths of the unrooted tree, in units of expected substitutions per site, become known without error. With full length genomes now available, although inferring trees from genomes involves complexities and assumptions such as a good partition scheme [35], this limiting case might be approached in some data sets. As a simple test in this paper,

6 this gives rise to an alternative approach to analysis, where divergence times, a root
7 position and the branch rates are random variables, and the data are a set of branch
8 lengths in units of substitution on a known unrooted tree topology.

9 Previous work done by Reis and Yang [14] also tried to approximate the likelihood
10 of such an unrooted tree in Bayesian phylogenetic inference. Similar researches in
11 [6, 13] show that these methods can account for rate changes in a relaxed clock
12 model, but the genetic distances are not fixed, for example Stéphane Guindon used
13 a Gibbs sampling algorithm [13]. Outside of the Bayesian MCMC formalism, least-
14 squares criteria [36] and maximum likelihood [37, 38], can also be applied to estimate
15 substitution rates and divergence times in unrooted trees.

16 In this section, we investigated this approach on a fixed substitution tree recon-
17 structed from whole mitochondrial genomes from a set of ratite species [33]. Since no
18 uncertainty is admitted in the genetic distances and the proposed operator doesn't
19 change the genetic distances, the phylogenetic likelihood is no longer needed and
20 the unrooted tree becomes the data, rather than a multiple sequence alignment.

21 First of all, we used the ratites data set to construct an unrooted tree with PhyML
22 3.0 [39, 40]. Figure 12(a) shows the unrooted tree with the genetic distances on the
23 branches which are fixed in the subsequent relaxed clock analysis in BEAST2.

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

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

45 Discussion

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

56 The key idea of the presented operator (to maintain the genetic distances) shows
57 a novel direction for more efficient proposals in Bayesian phylogenetic MCMC. For
58

6 example, the operations on the internal nodes, in the current study, involve one
7 random internal node, one node time and three branch rates. If two or more nodes
8 are selected, then more associated rates and node times can be sampled in one
9 proposal, which may achieve even better efficiency. Another possible approach is
10 to make small changes to the genetic distances as well. To minimise the number
11 of changes to genetic distances, a two-dimensional random draw will be used to
12 change four parameters (one divergence time and one rate changed directly, the
13 other two rates derived so as to minimise changes to genetic distances). What's
14 more, it should be pointed out that Small Pulley and Big Pulley can only be applied
15 to reversible continuous-time Markov chain models where unrooted trees can be
16 used in inference, because these operators require the underlying unrooted tree to
17 be unchanged. Future work could elaborate a larger class of operators along these
18 lines.

19 As data sets have increased in size the impetus to improve efficiency of Bayesian
20 phylogenetic inference algorithms has steadily increased. Besides more effective pro-
21 posal mechanisms within Metropolis-Hastings MCMC, completely novel approaches
22 to Bayesian phylogenetics have also begun to get some attention. Variational meth-
23 ods are one alternative for approximating Bayesian posterior distributions [41].
24 These approaches make inference an optimisation problem and take advantage of
25 tractable variational distributions that approximate the posterior distribution, thus
26 decreasing the computational cost by avoiding high-dimensional integrals in MCMC
27 sampling schemes. Recent work has investigated the potential for applying varia-
28 tional methods to phylogenetics [42, 43]. Our improved MCMC methods provide a
29 performance baseline for these new approaches.

30 Conclusions

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

35 We demonstrate the correctness of this algorithm with a series of tests including
36 a well-calibrated simulation study. Based on both simulated and real data sets,
37 the proposed operator is more efficient than current algorithms implemented in
38 BEAST2 for datasets with long sequences. This is a desirable property because
39 efficiency is most important for larger datasets. The proposed operator is available
40 for use as a package of BEAST2.

41 Abbreviations

42 MCMC	Markov chain Monte Carlo
43 ESS	effective sample sizes
44 HPD	highest posterior density
45 Category	Using the current operators in BEAST2 to sample discrete rate categories
46 Cons	Using the Constant Distance operator to sample continuous rates specified by an uncorrelated related clock model

47 Declarations

48 Ethics approval and consent to participate
49 Not applicable

6 Consent for publication
7 Not applicable

8 Availability of data and material

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

11 Competing interests
12 The authors declare that they have no competing interests.

13 Funding

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

17 Authors' contributions

18 RZ developed the operator and was a major contributor in writing the manuscript. AJD supervised the
19 implementation of the operator and the writing process of the manuscript. All authors read and approved the final
20 manuscript.

21 Acknowledgements

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

27 Author details

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

30 References

- 31 1. Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N., Teller, A.H., Teller, E.: Equation of state calculations by
32 fast computing machines. *The journal of chemical physics* **21**(6), 1087–1092 (1953)
- 33 2. Hastings, W.K.: Monte carlo sampling methods using markov chains and their applications. *Biometrika* **57**(1),
34 97–109 (1970)
- 35 3. Zuckerkandl, E., Pauling, L.: Evolutionary divergence and convergence in proteins, 97–166 (1965)
- 36 4. Yang, Z., Rannala, B.: Bayesian phylogenetic inference using dna sequences: a markov chain monte carlo
37 method. *Molecular biology and evolution* **14**(7), 717–724 (1997)
- 38 5. Rannala, B., Yang, Z.: Bayes estimation of species divergence times and ancestral population sizes using dna
39 sequences from multiple loci. *Genetics* **164**(4), 1645–1656 (2003)
- 40 6. Thorne, J.L., Kishino, H., Painter, I.S.: Estimating the rate of evolution of the rate of molecular evolution. *Mol
41 Biol Evol* **15**(12), 1647–57 (1998). doi:[10.1093/oxfordjournals.molbev.a025892](https://doi.org/10.1093/oxfordjournals.molbev.a025892)
- 42 7. Drummond, A.J., Ho, S.Y., Phillips, M.J., Rambaut, A.: Relaxed phylogenetics and dating with confidence.
PLoS biology **4**(5), 88 (2006)
- 43 8. Knapp, M., Stöckler, K., Havell, D., Delsuc, F., Sebastiani, F., Lockhart, P.J.: Relaxed molecular clock provides
44 evidence for long-distance dispersal of nothofagus (southern beech). *PLoS Biology* **3**(1), 14 (2005)
- 45 9. Smith, S.A., Beaulieu, J.M., Donoghue, M.J.: An uncorrelated relaxed-clock analysis suggests an earlier origin
46 for flowering plants. *Proceedings of the National Academy of Sciences*, 201001225 (2010)
- 47 10. Ho, S.Y., Phillips, M.J., Drummond, A.J., Cooper, A.: Accuracy of rate estimation using relaxed-clock models
48 with a critical focus on the early metazoan radiation. *Molecular Biology and Evolution* **22**(5), 1355–1363
49 (2005)
- 50 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)
- 51 12. Kober, K., Stamatakis, A., Flouri, T.: Efficient detection of repeating sites to accelerate phylogenetic
52 likelihood calculations. *Systematic biology* **66**(2), 205–217 (2017)
- 53 13. Guindon, S.: Bayesian estimation of divergence times from large sequence alignments. *Molecular Biology and
54 Evolution* **27**(8), 1768–1781 (2010)
- 55 14. Reis, M.d., Yang, Z.: Approximate likelihood calculation on a phylogeny for bayesian estimation of divergence
56 times. *Molecular Biology and Evolution* **28**(7), 2161–2172 (2011)
- 57 15. Ayres, D.L., Darling, A., Zwickl, D.J., Beerli, P., Holder, M.T., Lewis, P.O., Huelsenbeck, J.P., Ronquist, F.,
58 Swofford, D.L., Cummings, M.P., et al.: Beagle: an application programming interface and high-performance
59 computing library for statistical phylogenetics. *Systematic biology* **61**(1), 170–173 (2011)
- 60 16. Lakner, C., Van Der Mark, P., Huelsenbeck, J.P., Larget, B., Ronquist, F.: Efficiency of markov chain monte
61 carlo tree proposals in bayesian phylogenetics. *Systematic biology* **57**(1), 86–103 (2008)
- 62 17. Höhna, S., Drummond, A.J.: Guided tree topology proposals for bayesian phylogenetic inference. *Syst Biol*
63 **61**(1), 1–11 (2012). doi:[10.1093/sysbio/syr074](https://doi.org/10.1093/sysbio/syr074)
- 64 18. Bouckaert, R., Heled, J., Kühnert, D., Vaughan, T., Wu, C.-H., Xie, D., Suchard, M.A., Rambaut, A.,
65 Drummond, A.J.: Beast 2: a software platform for bayesian evolutionary analysis. *PLoS computational biology*
10(4), 1003537 (2014)

- 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
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. Hohna, S., Defoin-Plate, 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)
 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/alexied/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., Rodríguez, 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-taxon 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).

[width=12cm]Fig04-exchangemethod.eps

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

[width=12cm]Fig05-treechange.eps

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.

[width=12cm]Fig06-symmetric.eps

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.

[width=12cm]Fig07-asymmetric.eps

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.

[width=12cm]Fig08-modelvalidation.eps

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 ($n = 120$).

[width=12cm]Fig09-calibrated.eps

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.

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.

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 .

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.

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.

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.

Figure 17 Running time and ESS using Anolis data.

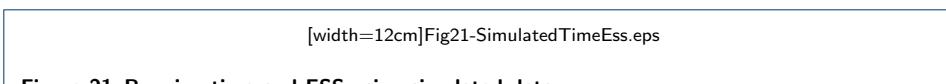
Figure 18 Running time and ESS using RSV2 data.

[width=12cm]Fig19-ShankarappaTimeEss.eps

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

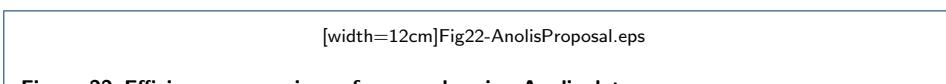
[width=12cm]Fig20-PrimatesTimeEss.eps

Figure 20 Running time and ESS using primates data.



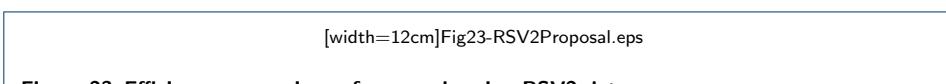
[width=12cm]Fig21-SimulatedTimeEss.eps

Figure 21 Running time and ESS using simulated data.



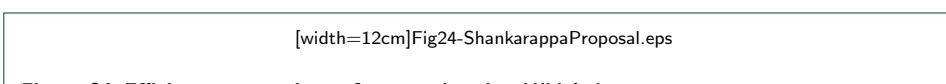
[width=12cm]Fig22-AnolisProposal.eps

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.



[width=12cm]Fig25-UcllScalar.eps

Figure 25 Efficiency comparison of clock standard deviation.

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

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

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

Data	Configuration	Average running time (hour)	Parameter	ESS
Anolis	Category Cons	0.3788 0.4212	frequency.A	698.95 750.53
RSV2	Category Cons	2.0509 2.6742	prior	1231.32 3409.88
HIV-1	Category Cons	2.6040 3.2680	prior	387.83 753.48
Primates	Category Cons	31.4059 21.6584	rate.mean	71.79 422.24
Simulated data with 500 sites	Category Cons	0.0728 0.0834	frequency.G	819.39 837.50
Simulated data with 1000 sites	Category Cons	0.4403 0.4863	frequency.G	2760.41 2961.67

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

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	
Scenario 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				
Scenario 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				

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

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	
Simple Distance	t_E	7.8081	1.2884	7.8187	1.2992	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)

Table 6 Results of sampling the root

Table 7 Operator weights in MCMC chains

Operator class	Name	Simulated data Cons Category	Anolis Category	RSV2 Cons Category	HIV-1 Cons Category	Primates Cons Category
rates times	ConstantDistance Operator	0.2170	-	0.2228	-	0.2402
	Rate Normal Operators ¹	0.1302	0.2604	0.1349	0.2674	0.1441
	UcidStdDev Scale Operator ²	0.0260	0.0260	0.0270	0.0267	0.0288
	UcidMean Scale Operator	-	-	0.0089	0.0089	-
	UcidMean Tree UpperDown Operator	-	-	0.0267	0.0267	-
	InternalNode Time Scale Operator	0.0174	0.0260	0.0180	0.0270	0.0178
	RootAge Scale Operator	0.0174	0.0260	0.0180	0.0270	0.0178
	AllNodeTimes Uniform Operator	0.1910	0.2604	0.1978	0.2698	0.1961
	SubtreeSlide Operator	0.1302	0.1302	0.0989	0.0989	0.1337
	NarrowExchange Operator	0.1302	0.1302	0.0989	0.0989	0.1337
Tree	WideExchange Operator	0.0434	0.0434	0.0270	0.0270	0.0267
	WilsonBalding Operator	0.0434	0.0434	0.0270	0.0270	0.0267
	BirthRate Scale Operator	0.0434	0.0434	0.0629	0.0629	-
	DeathRate Scale Operator	-	-	0.0629	0.0629	-
	PopulationSize Scale Operator	-	-	-	0.0267	0.0267
	Kappa Scale Operator	0.0087	0.0087	0.0009	0.0009	0.0009
	Frequencies DeltaExchange Operator	0.0017	0.0017	0.0009	0.0009	0.0010

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

7 **1. The Green ratio**

8 When developing an operator for MCMC, the proposal function must be reversible. In other words, the probability
9 that the operator propose a new state from the current state is required to be equal to the probability that the
10 proposed state goes back to current state. To be specific, let $\pi(x)$ be the target probability distribution and
11 $p(x, x')$ be the transition kernel in the continuous Markov chain. The reversibility condition requires that
12 $\pi(x)p(x, x') = \pi(x')p(x', x)$. And an operator provides a proposal $q(x, x')$ with some probability $\alpha(x, x')$ that
13 the proposal is accepted. Thus, the reversibility condition is rewritten as
14 $\pi(x)q(x, x')\alpha(x, x') = \pi(x')q(x', x)\alpha(x', x)$.

15 Considering the subspace φ_1 on x and subspace φ_2 on x' , it is assumed that there is a symmetric measure on the
16 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
17 φ . Then, Green suggested that the reversibility condition should be satisfied by detailed balance [20], as represented
18 by equation (17). And according to Peskun' proof, it is optimal to take equation (18) as the acceptance probability
19 to retain the detailed balance [46].

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

20 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
21 state x' given the current state x .

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

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

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

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

26 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
27 as the Green ratio, and \mathbf{J} ensures that the proposal have a symmetric measure on each subspace in state x and x' .

28 **1.1 Calculating the Green ratio for operations on internal nodes**

29 The Constant Distance Operator firstly proposes a new time for the randomly selected internal node (equation
30 (20a)), and then proposes three rates by the original distances and new node times(equation (20b)~equation
31 (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 $|\mathbf{J}_3| = 1$.

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

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).

$$\mathbf{J}_4 = \begin{bmatrix} \frac{\partial t_X'}{\partial t_X} & \frac{\partial t_X'}{\partial t_C} & \frac{\partial t_X'}{\partial r_S} & \frac{\partial t_X'}{\partial r_M} & \frac{\partial t_X'}{\partial r_N} & \frac{\partial t_X'}{\partial r_{N2}} \\ \frac{\partial t_C'}{\partial t_X} & \frac{\partial t_C'}{\partial t_C} & \frac{\partial t_C'}{\partial r_S} & \frac{\partial t_C'}{\partial r_M} & \frac{\partial t_C'}{\partial r_N} & \frac{\partial t_C'}{\partial r_{N2}} \\ \frac{\partial t_X}{\partial r_S} & \frac{\partial r_C}{\partial t_C} & \frac{\partial r_C}{\partial r_S} & \frac{\partial r_C}{\partial r_M} & \frac{\partial r_C}{\partial r_N} & \frac{\partial r_C}{\partial r_{N2}} \\ \frac{\partial r_S}{\partial t_X} & \frac{\partial r_S}{\partial t_C} & \frac{\partial r_S}{\partial r_S} & \frac{\partial r_S}{\partial r_M} & \frac{\partial r_S}{\partial r_N} & \frac{\partial r_S}{\partial r_{N2}} \\ \frac{\partial r_M}{\partial t_X} & \frac{\partial r_M}{\partial t_C} & \frac{\partial r_M}{\partial r_S} & \frac{\partial r_M}{\partial r_M} & \frac{\partial r_M}{\partial r_N} & \frac{\partial r_M}{\partial r_{N2}} \\ \frac{\partial r_N}{\partial t_X} & \frac{\partial r_N}{\partial t_C} & \frac{\partial r_N}{\partial r_S} & \frac{\partial r_N}{\partial r_M} & \frac{\partial r_N}{\partial r_N} & \frac{\partial r_N}{\partial r_{N2}} \end{bmatrix} \quad (29)$$

$$= \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ \frac{r_C}{t_X' - t_C'} & \frac{-r_C}{t_X' - t_C'} & \frac{t_X' - t_C}{t_X' - t_C'} & 0 & 0 & 0 \\ 0 & \frac{r_S}{t' - t_S} & 0 & \frac{t_C - t_S}{t_C' - t_S} & 0 & 0 \\ \frac{-r_C}{t_X' - t_M} & \frac{r_{N1} + r_C}{t_X' - t_M} & \frac{-(t_X - t_C)}{t_X' - t_M} & 0 & \frac{t_C - t_M}{t_X' - t_M} & 0 \\ \frac{r_C + r_S}{t_C' - t_N} & \frac{-(r_C + r_S)}{t_C' - t_N} & \frac{t_X - t_C}{t_C' - t_N} & 0 & 0 & \frac{t_X - t_N}{t_C' - t_N} \end{bmatrix}$$

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

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

```

Original tree is symmetric:
if the node that has been exchanged with L or R has child nodes then
     $\alpha = \beta = 0.25$ 
else if  $t_R > t_L$  then
     $\alpha = 1, \beta = 0.5$ 
else if  $t_R < t_L$  then
     $\alpha = 0.5, \beta = 1$ 
else if  $t_R = t_L$  then
     $\alpha = \beta = 1$ 
end if
if Proposed tree belongs to ① or ② then
    Return  $\mu = \frac{\alpha}{0.25}$ 
end if
if Proposed tree belongs to ③ or ④ then
    Return  $\mu = \frac{\beta}{0.25}$ 
end if

Original tree is asymmetric:
if the node that has been exchanged with O has child nodes then
     $\gamma = 0.25$ 
else
     $\gamma = 0.5$ 
end if
if Proposed tree belongs to ⑤ or ⑥ then
    Return  $\mu = \frac{\gamma}{0.5}$ 
end if
if Proposed tree belongs to ⑦ then
    Return  $\mu = \frac{0.25}{1}$ 
end if

```

30
31
32 2. Sampling from the prior

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

38 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
39 in the experiments in this section. In the figure, g_1 is set as the initial tree. Firstly, a LogNormal distribution is used
40 as the rate prior in the uncorrelated relaxed clock model, given by equation (31).

$$41 \quad r = \{r_A \quad r_B \quad r_C \quad r_D\} \sim \text{LogNormal}(m = -3, s = 0.25) \quad (31)$$

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

$$44 \quad p(t = \{t_E, t_D\}) = \left(\frac{1}{N} \times e^{-\frac{1}{N}(t_E - t_D)}\right) \times \left(\frac{1}{N} \times e^{-\frac{3}{N}t_D}\right) \quad (32)$$

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

$$48 \quad p(r, t) = p(t_E, t_D) \times p(r_D) \times p(r_A) \times p(r_B) \times p(r_C) \\ 49 \quad = p(t_E, t_D) \times p\left(\frac{d_D}{t_E - t_D}\right) \times p\left(\frac{d_A}{t_D - t_A}\right) \times p\left(\frac{d_B}{t_D - t_B}\right) \times p\left(\frac{d_C}{t_E - t_C}\right), \quad (33)$$

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

53 2.1 Test the operator on internal nodes

54 The genetic distances, node times and rates for g_1 in Figure 13 are given in Table 3. To test roundly, two scenarios
55 are designed. In each scenario, the genetic distances are fixed, the node time t_D starts from the initial value and will
56 be changed by the proposed operator during the sampling process. Essentially, the proposed operator makes node D
57

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 samples the root time and two branch rates correctly.

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 achieved similar performance in RSV2 data set, while Bactrian proposal provided larger ESS per hour for most parameters in HIV-1 data set. Therefore, it still needs further investigation to demonstrate the effectiveness of different proposals when analysing various data sets. Our current implementation of the operators enables users to specify which proposal style will be used in Beast2 analysis.

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 $\text{Uniform}(0, 1)$, 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).

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

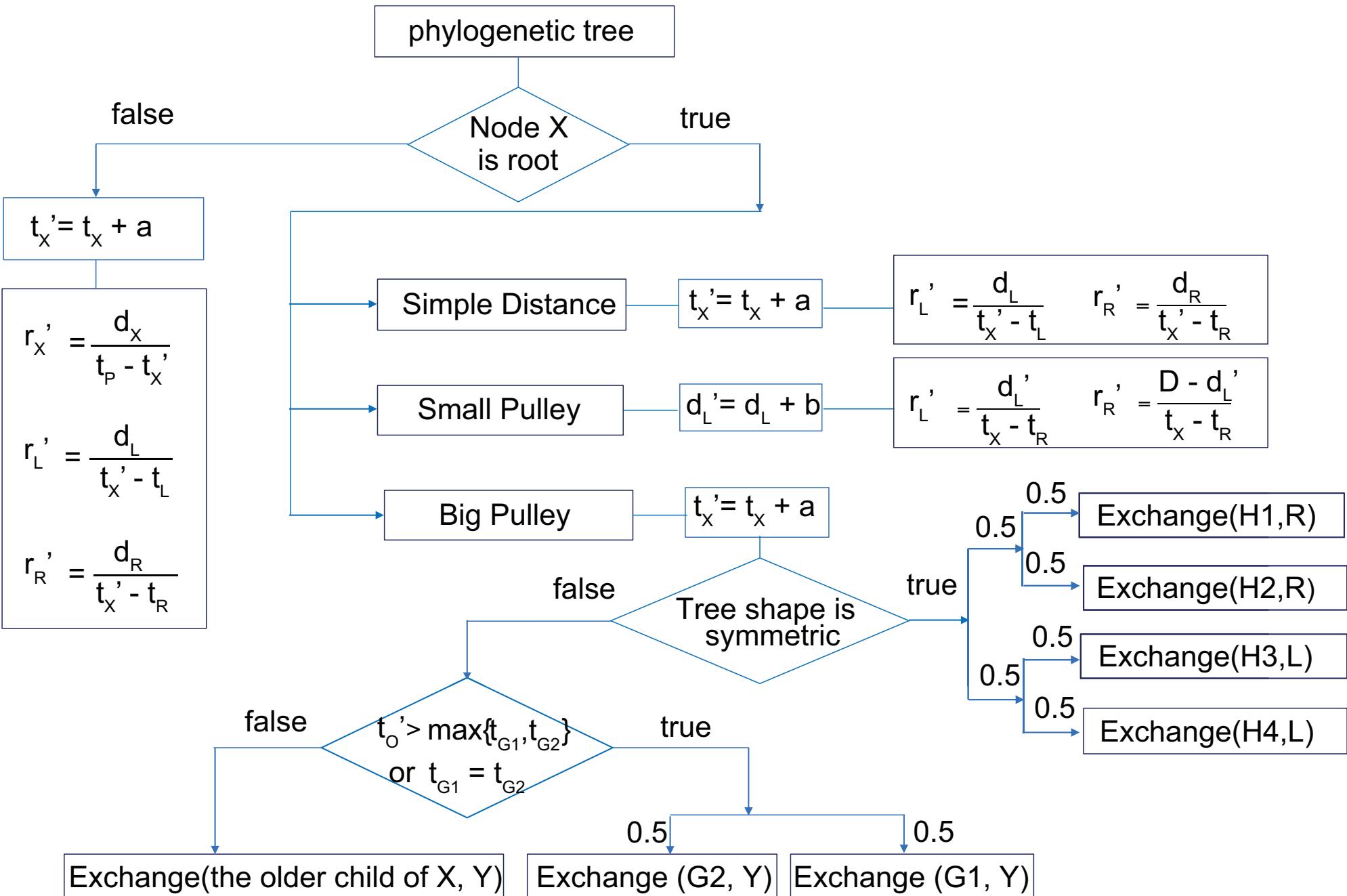
where the notations $\text{cdf}(\cdot)$ and $\text{icdf}(\cdot)$ represent the cumulative and inverse cumulative density function of log normal distribution. Because of the calculation of $\text{cdf}(\cdot)$ and $\text{icdf}(\cdot)$ for each branch rates, the "Cons" configuration requires more running time than "category", as is discusses in "Performance comparison" section. However, it is acceptable as ConstantDistance operator gives larger ESS.

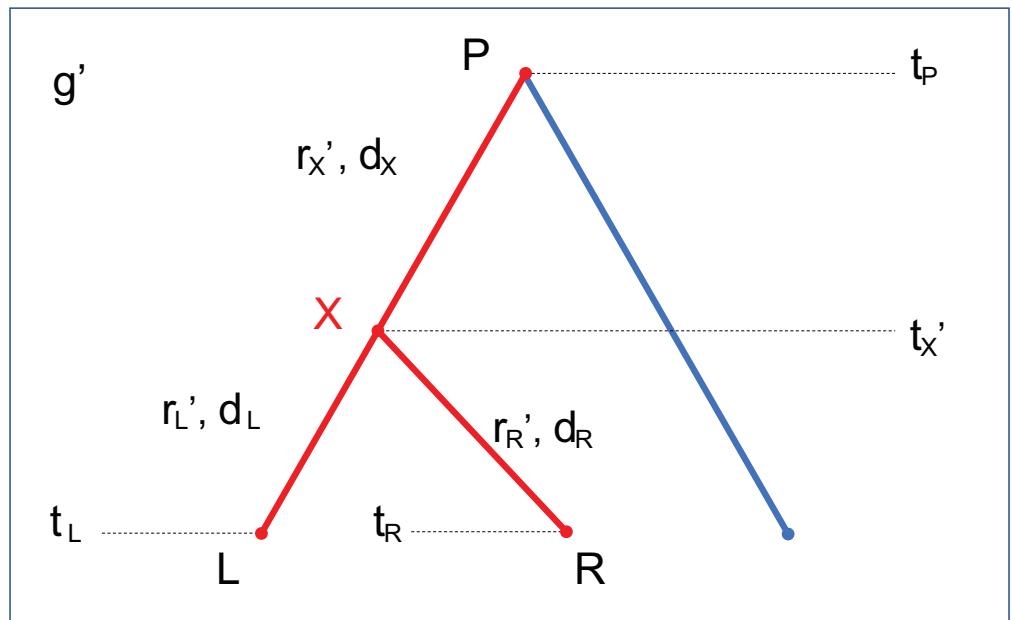
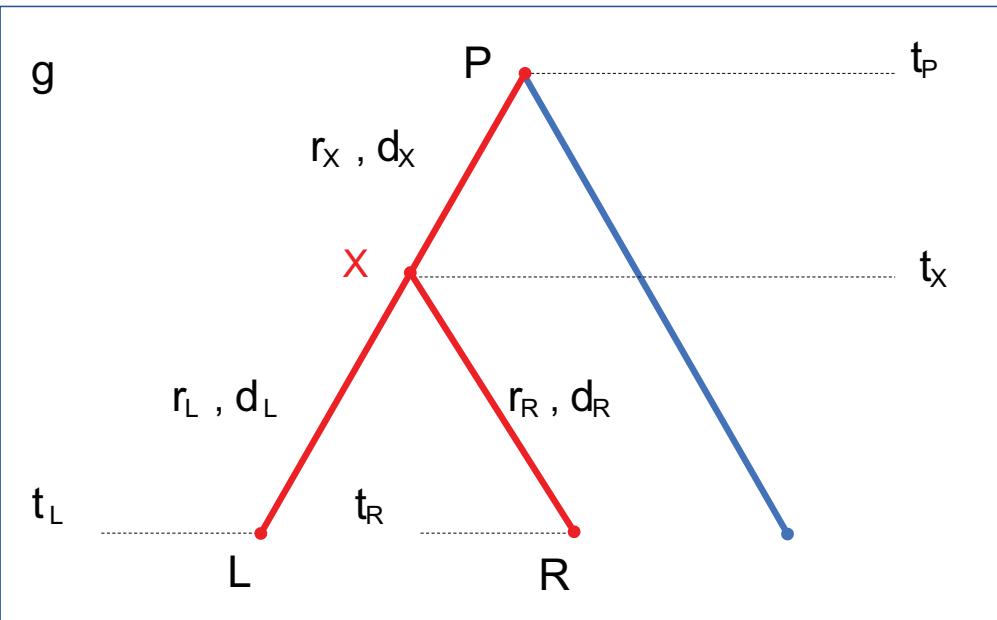
Finally, it is important to return the corrected hastings ratio, since the proposal is associated with one random variable, $Ucldstdev$ and $(2n - 1)$ branch rates. As is shown in equation (40), the ratio includes the scale operation and rates changing under the same probability.

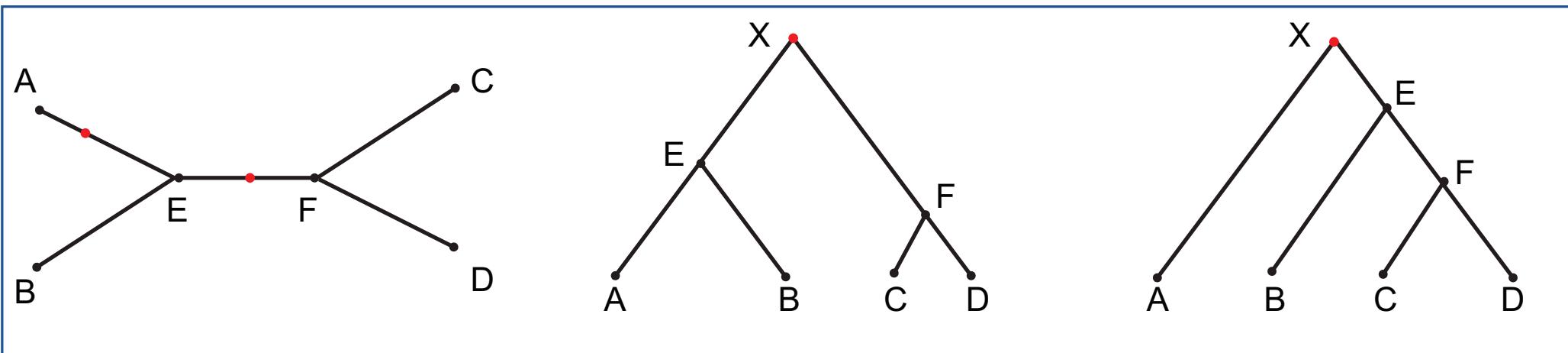
$$\mathbf{J}_{Ucldstdev} = \frac{1}{\text{scale}} \times \prod_{i=1}^{2n-1} \frac{\partial \text{icdf}_{Ucldstdev'}[\text{cdf}_{Ucldstdev}(r_i)]}{\partial r_i} \quad (40)$$

In the comparison of ESS for the clock standard deviation (denoted by ucl.stdev in Figure 10), we specified a normal scale operator in "Category" configuration. In "Cons" configuration, the UcldstdevScaleOperator is used to sample the clock standard deviation of continuous rates. To avoid the concern that the difference between "Category" and "Cons" is a result of how rates are parameterised (i.e. discrete or continuous), we set another configuration where continuous rates are sampled without using the ConstantDistance operator (denoted by "NoCons" configuration). The weights of the operators in "NoCons" are the same as those in "Category" which is detailed in Table 7. We ran the analysis using the three real data sets (Anolis, RSV2 and HIV-1) and the comparison of ESS per hour between "Category", "Cons" and "NoCons" is summarised in Figure 25. The figure shows ESS per hour in \log_{10} space of ucl.stdev in 20 independent MCMC chains. As can be seen, "Cons" configuration gives similar performance, comparing with "Category". This indicates UcldstdevScaleOperator works properly on continuous rates. Moreover, ESS per hour is much larger in "Cons" than in "NoCons", where both continuous rates

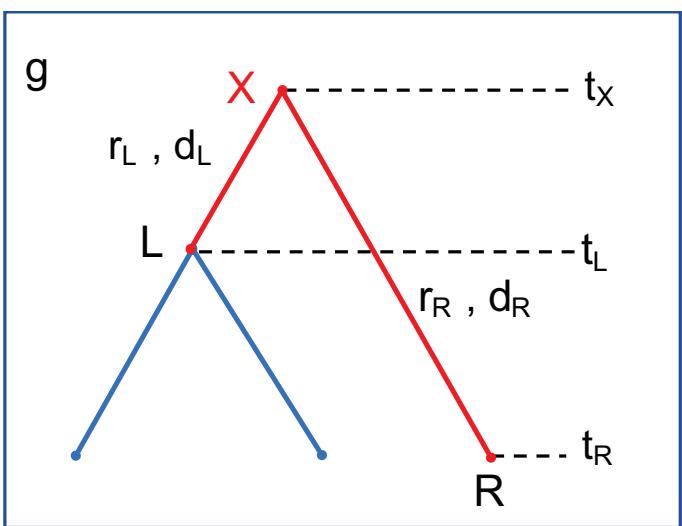
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
are sampled. Therefore, the proposed operators contribute to the improved performance. However, we noticed that the rate parameterisation does have some mixing issues in MCMC chains. In the future, we will further investigate how to parameterise branch rates to get better performance when using the proposed operators.



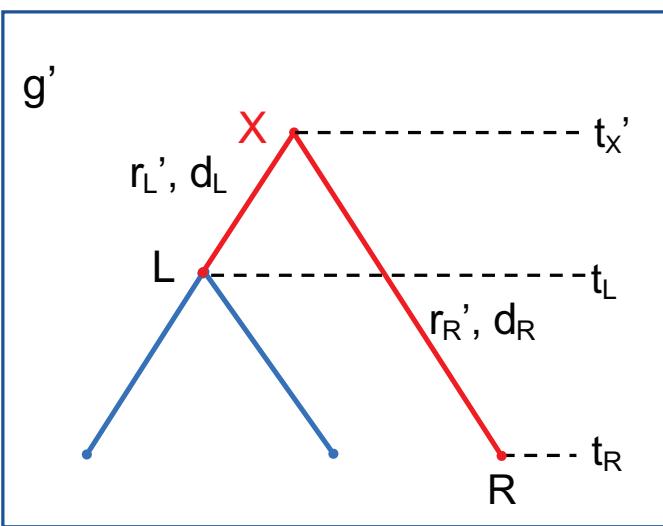




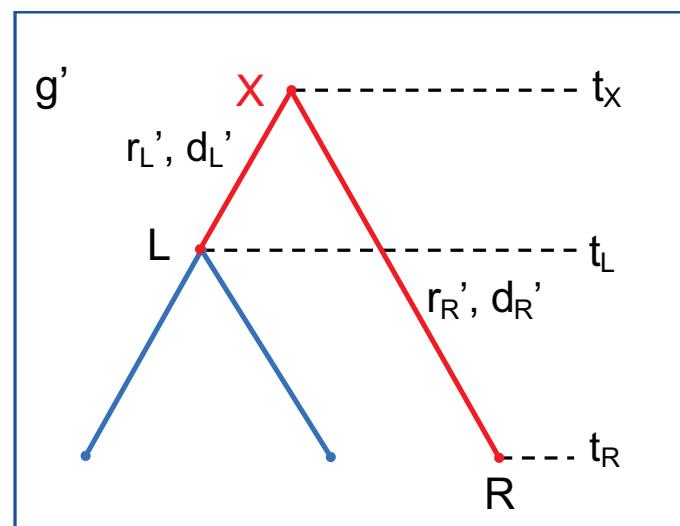
(a) An underlying unrooted tree



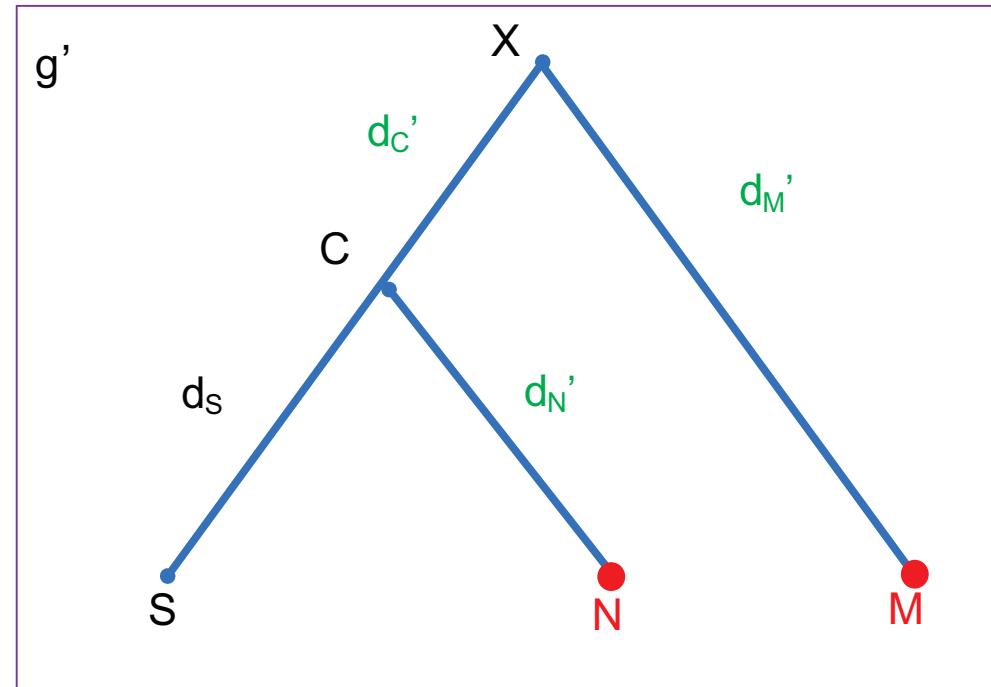
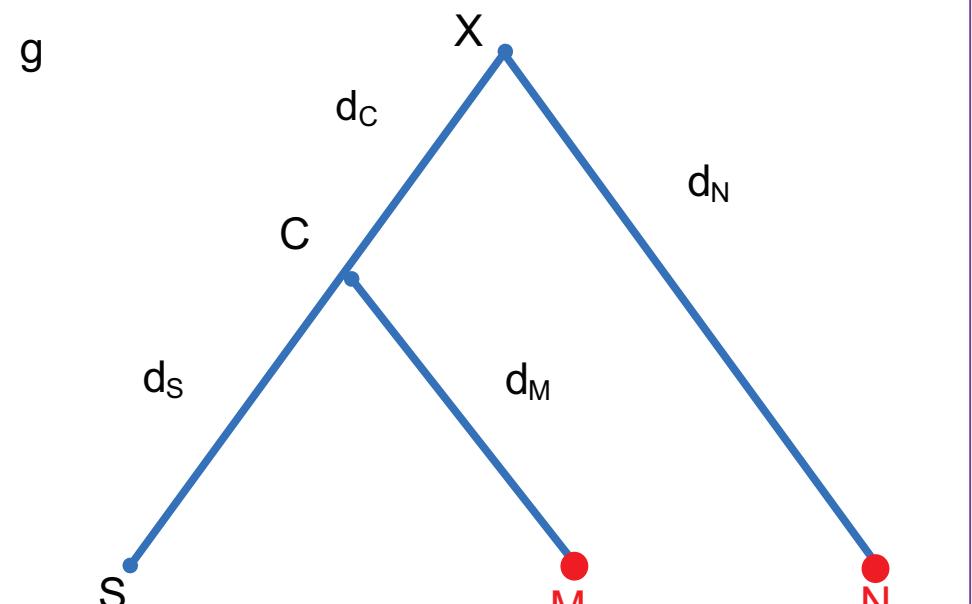
(b) Original tree

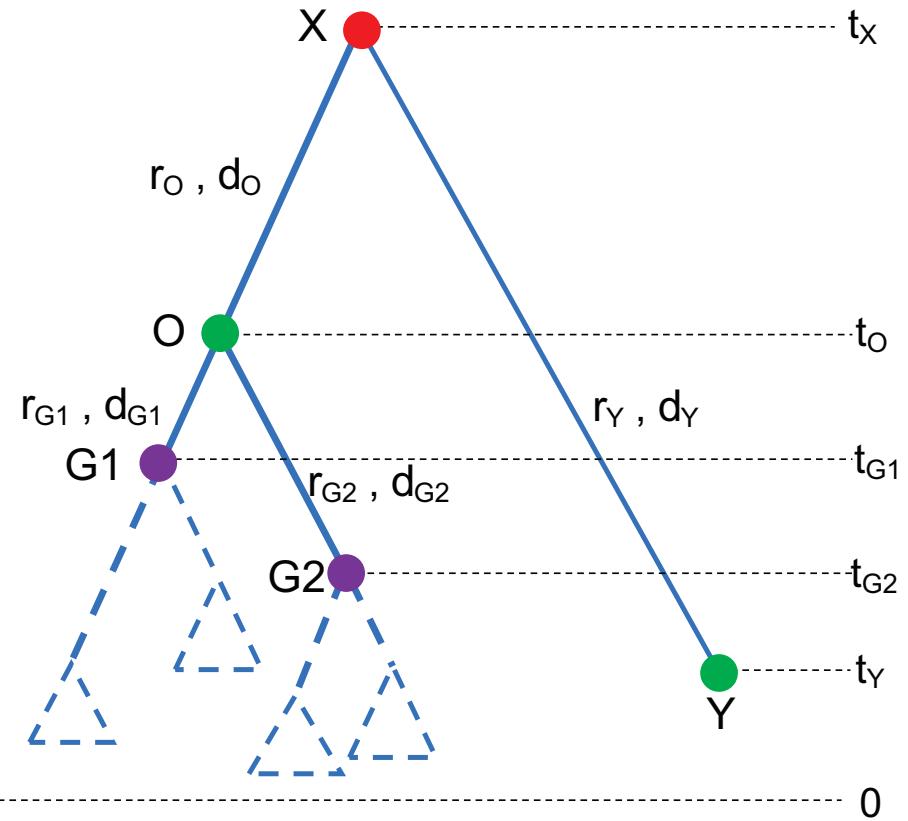
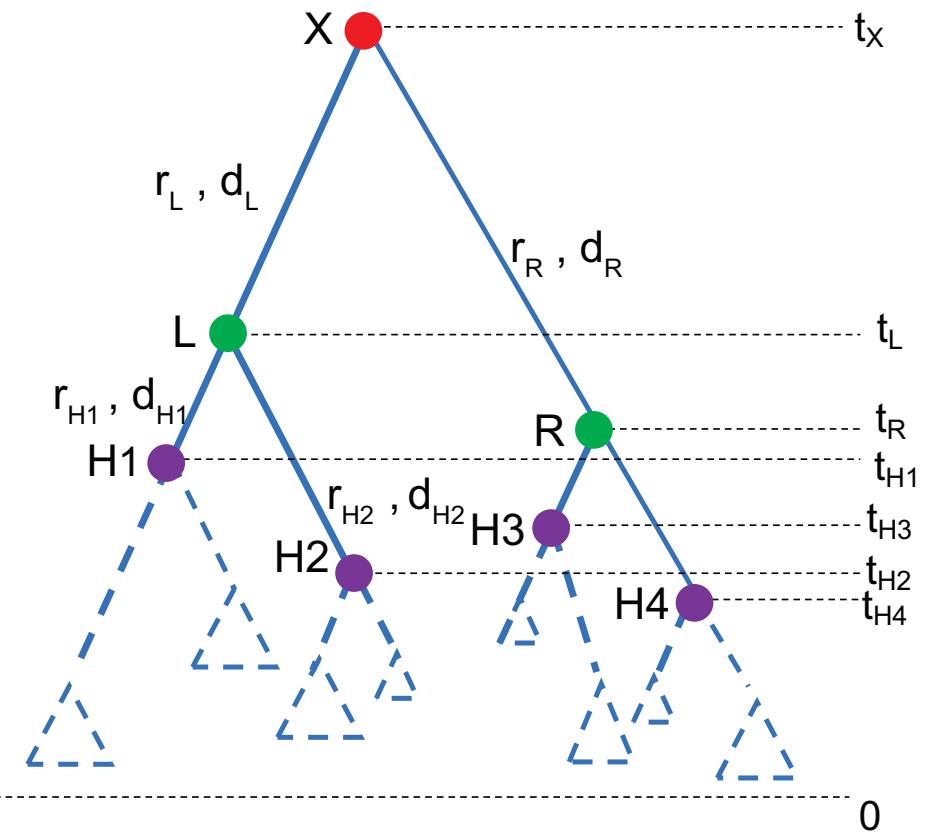


(c) Simple Distance



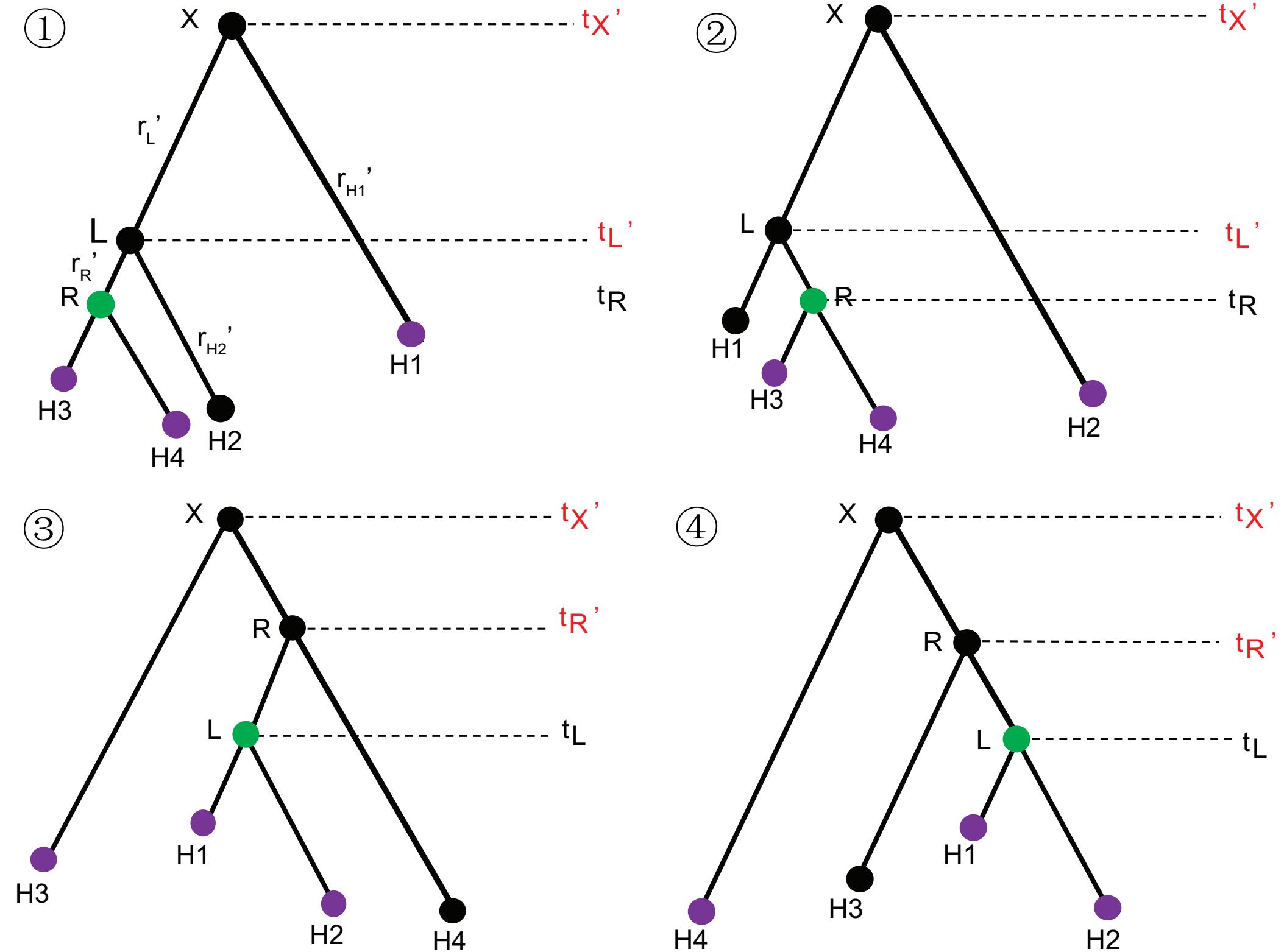
(d) Small Pulley

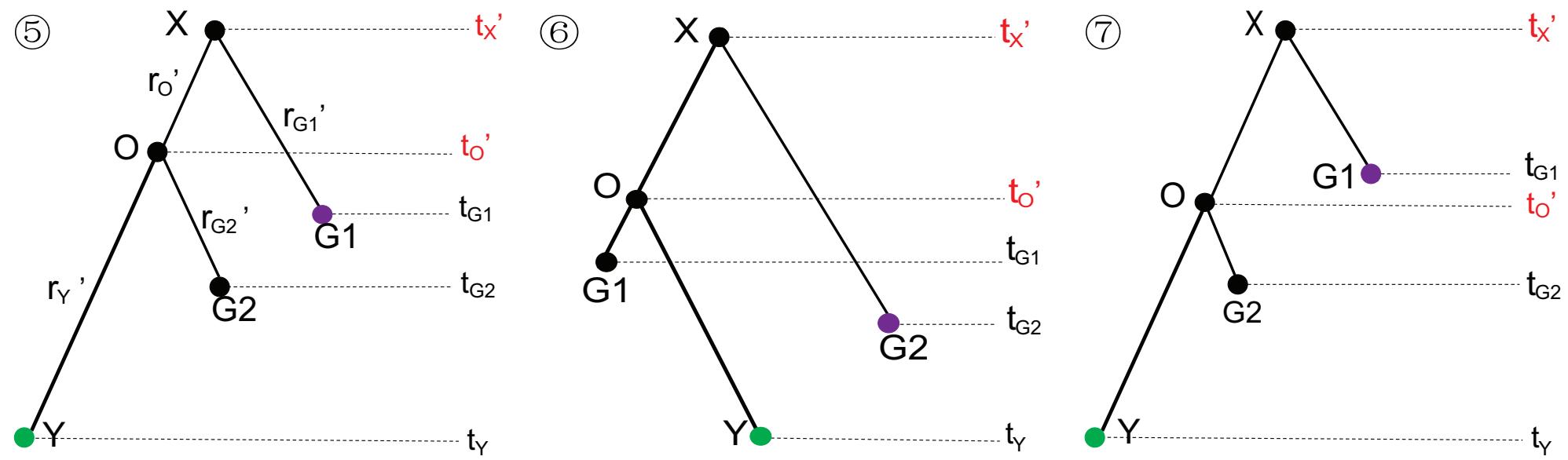


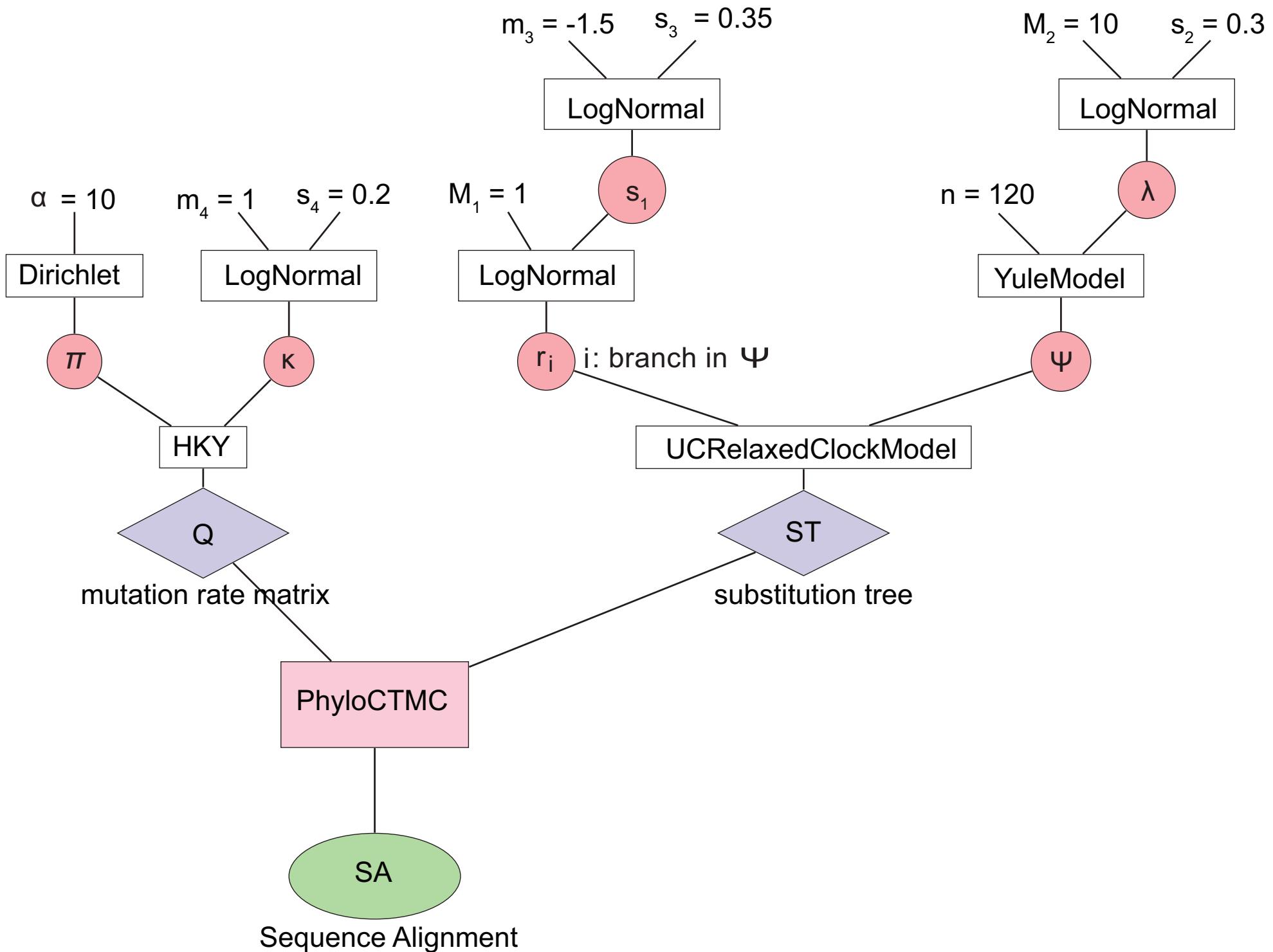


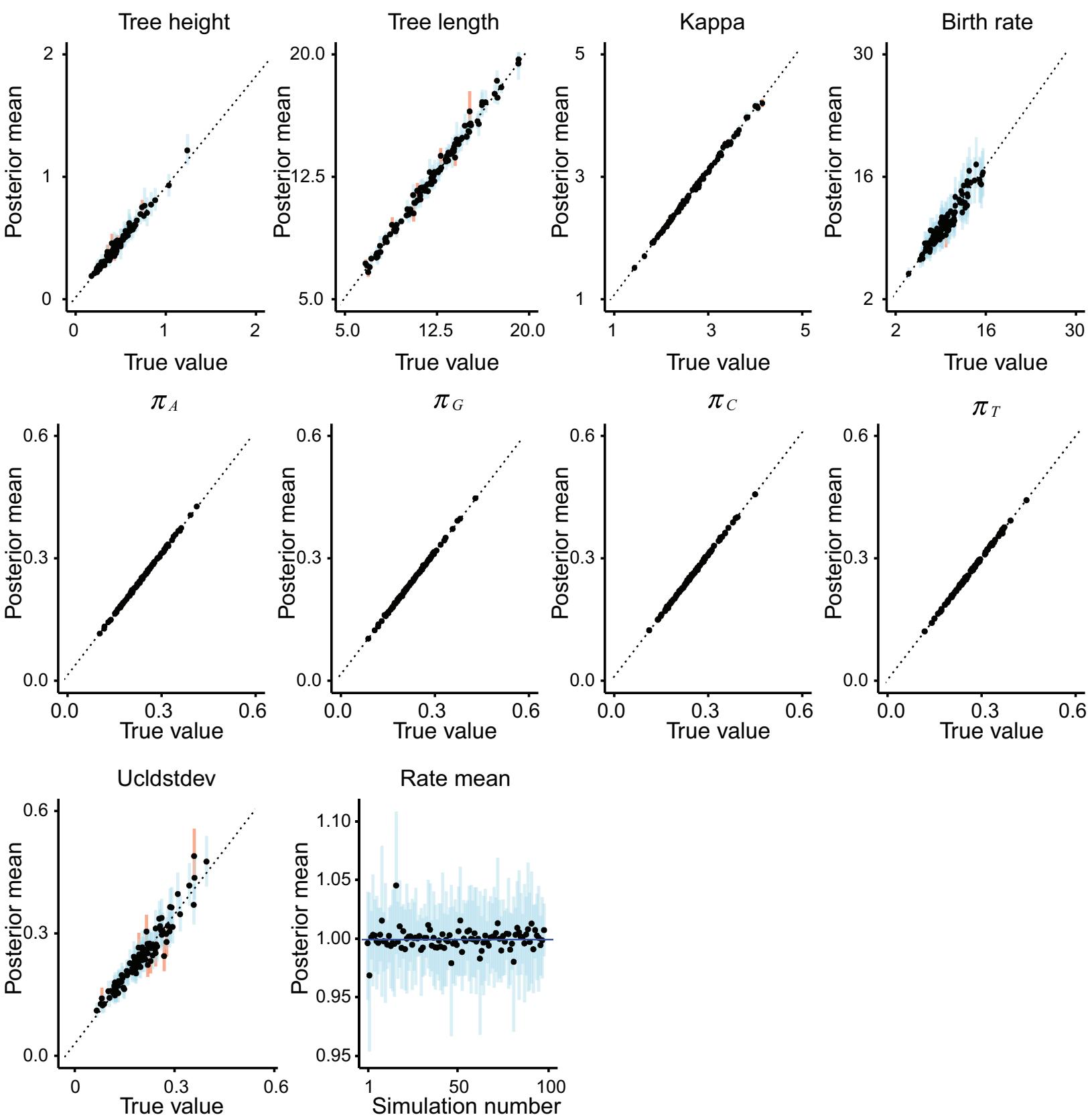
Figure

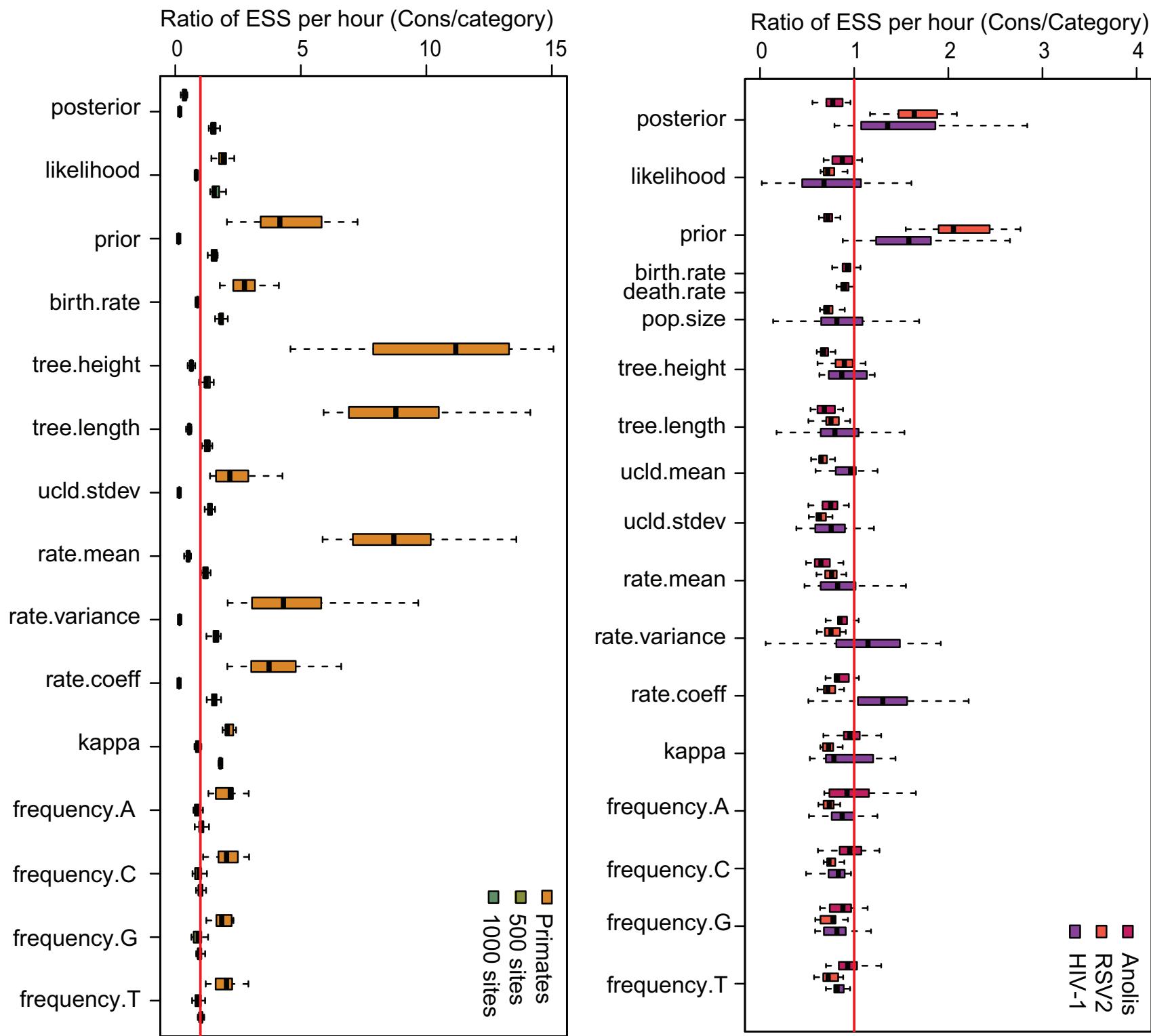
[Click here to access/download;Figure;Fig06-symmetric.eps](#)

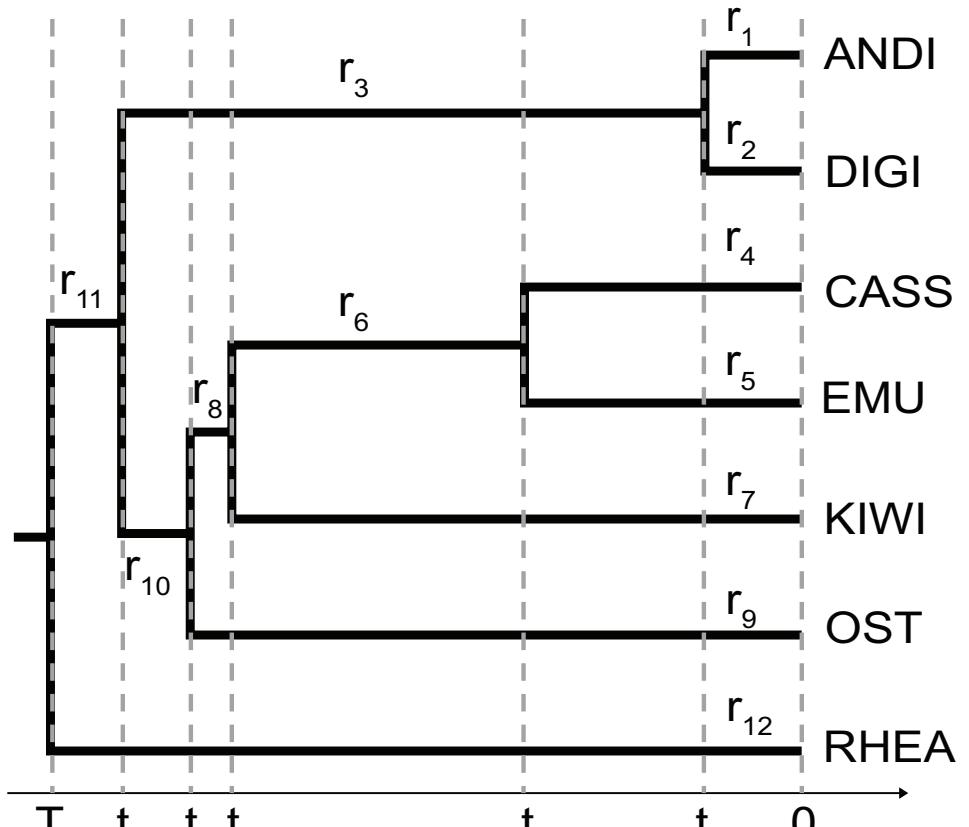




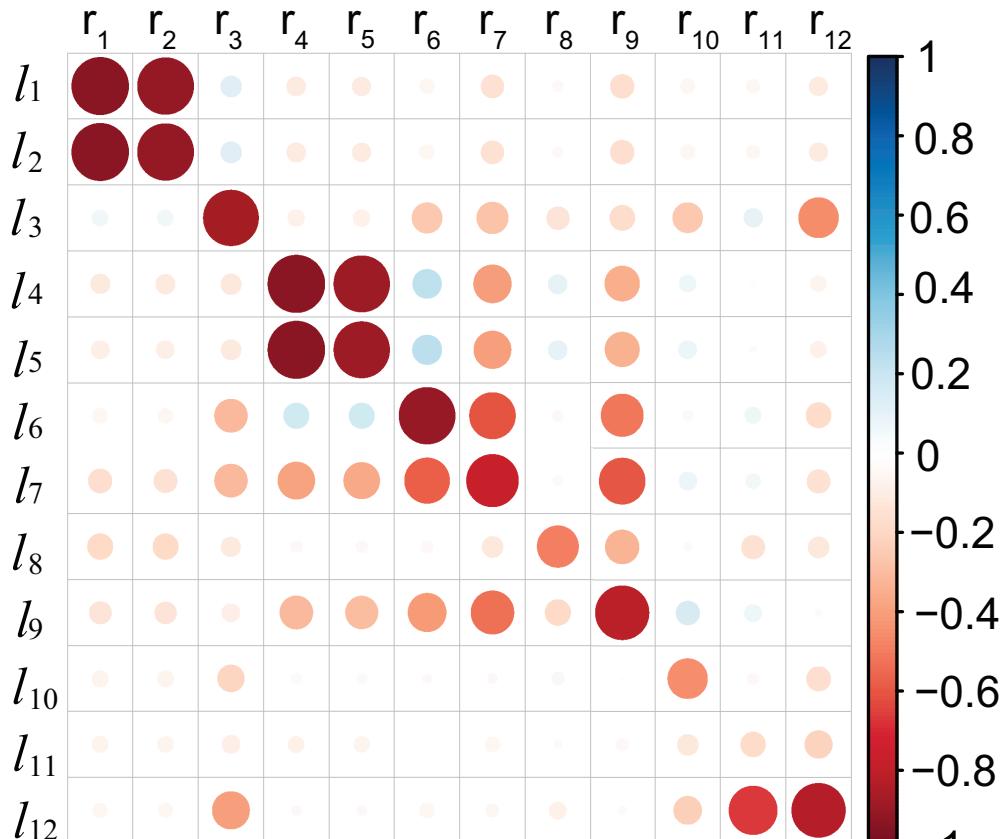




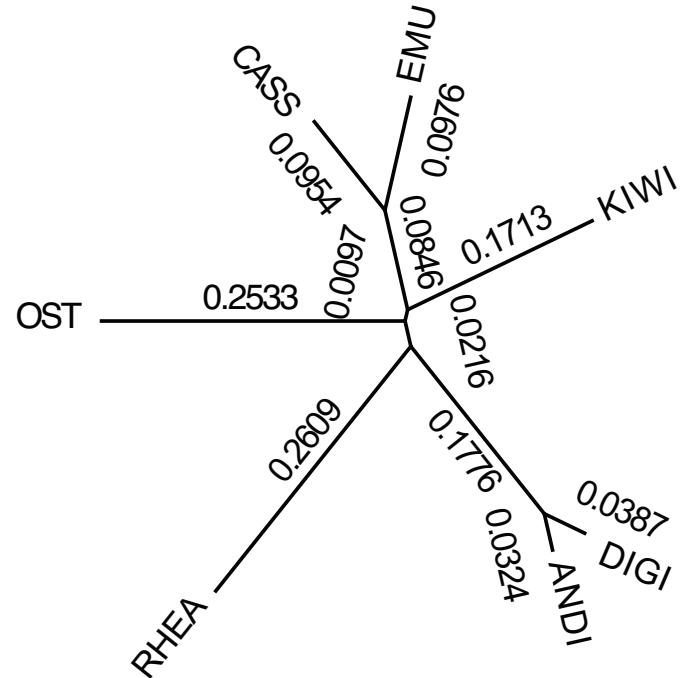




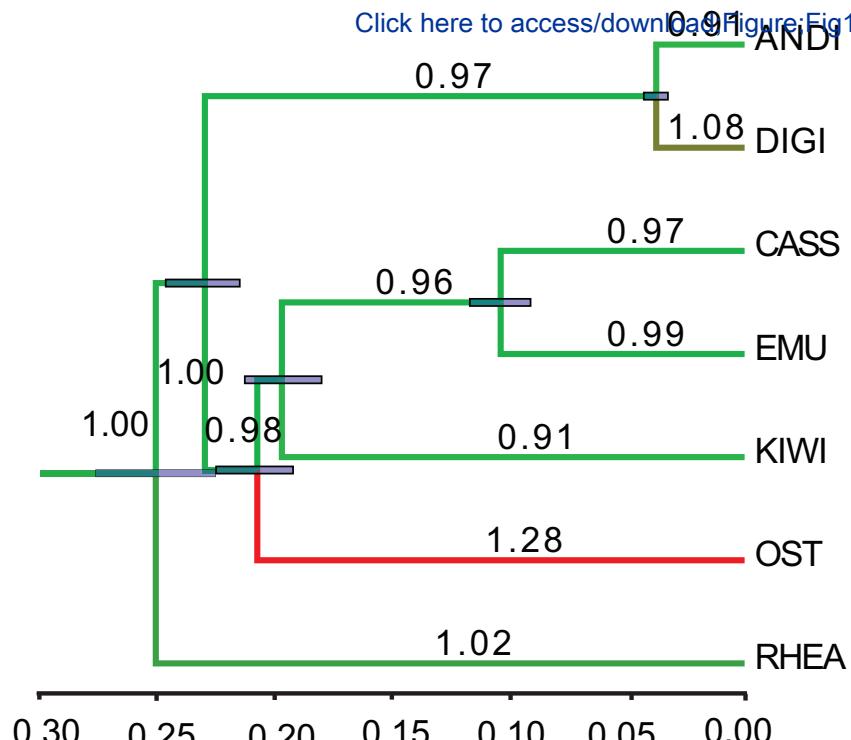
(a) Maximum clade credibility tree



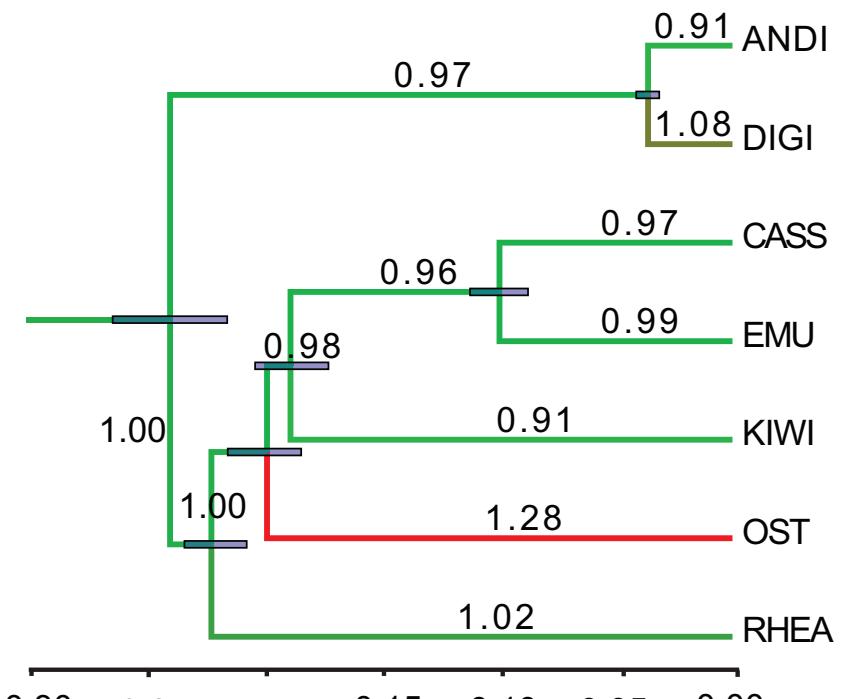
(b) Rates and branch lengths pairwise comparison



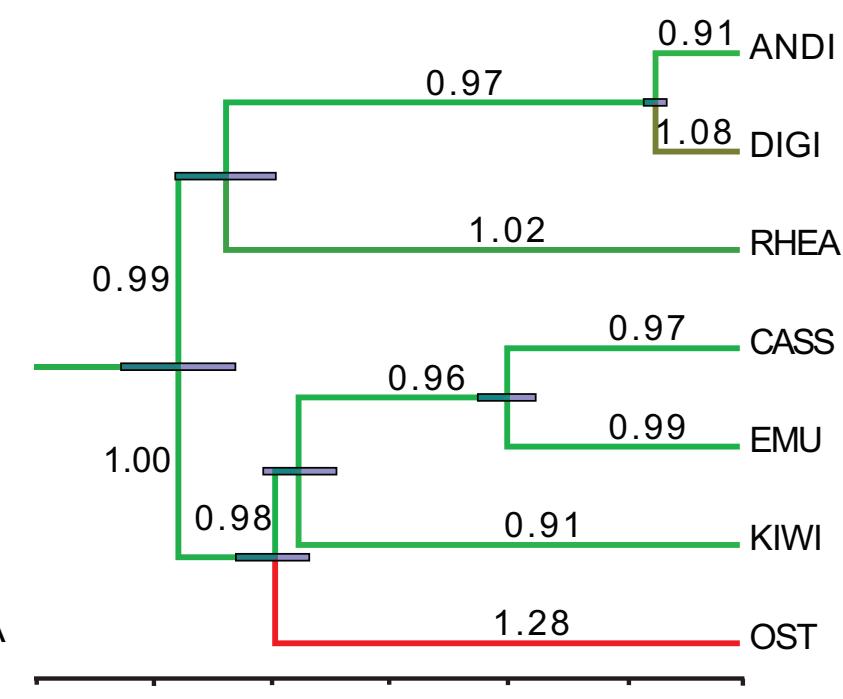
(a) Unrooted tree



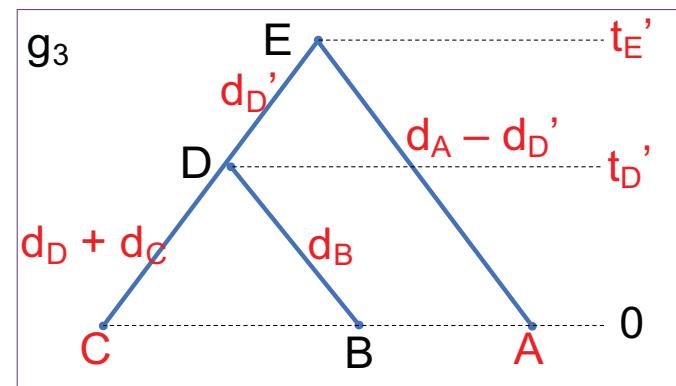
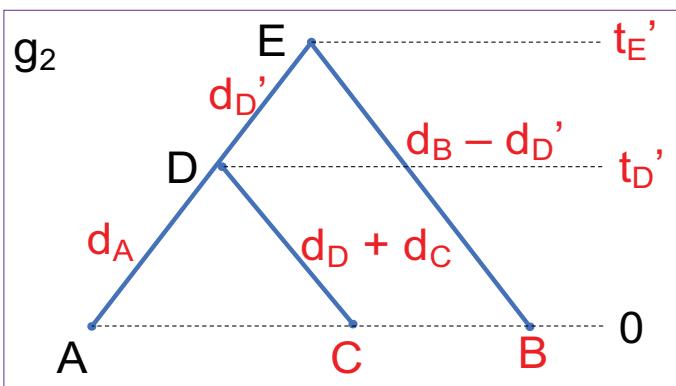
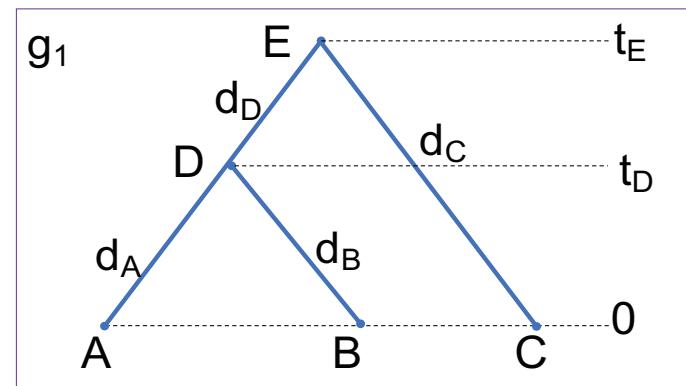
(b) The tree with 86.6% probability

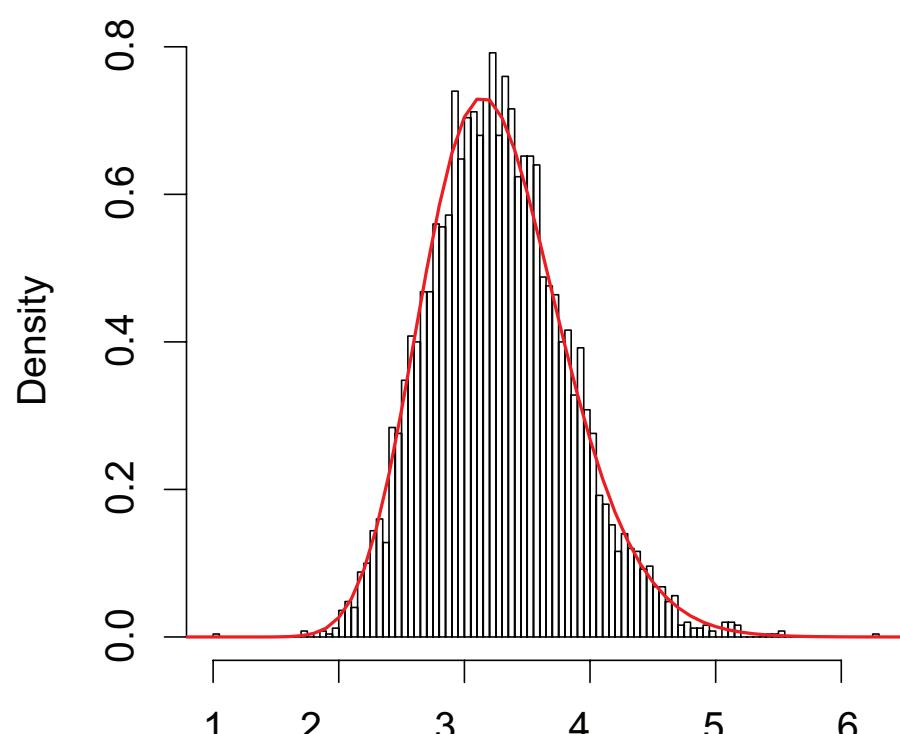


(c) The tree with 7.5% probability

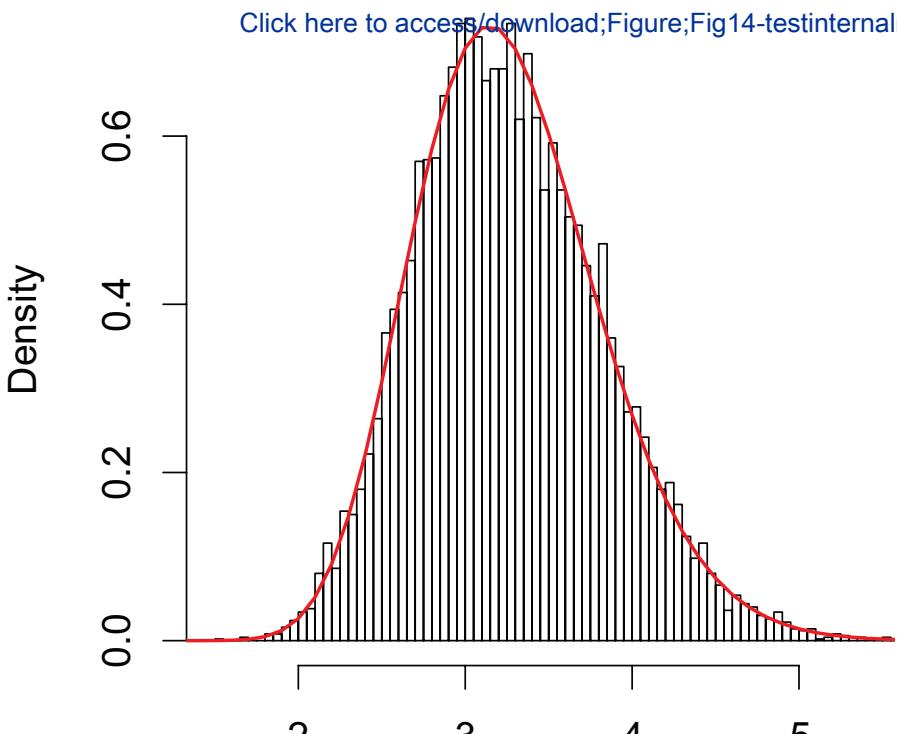


(d) The tree with 5.9% probability

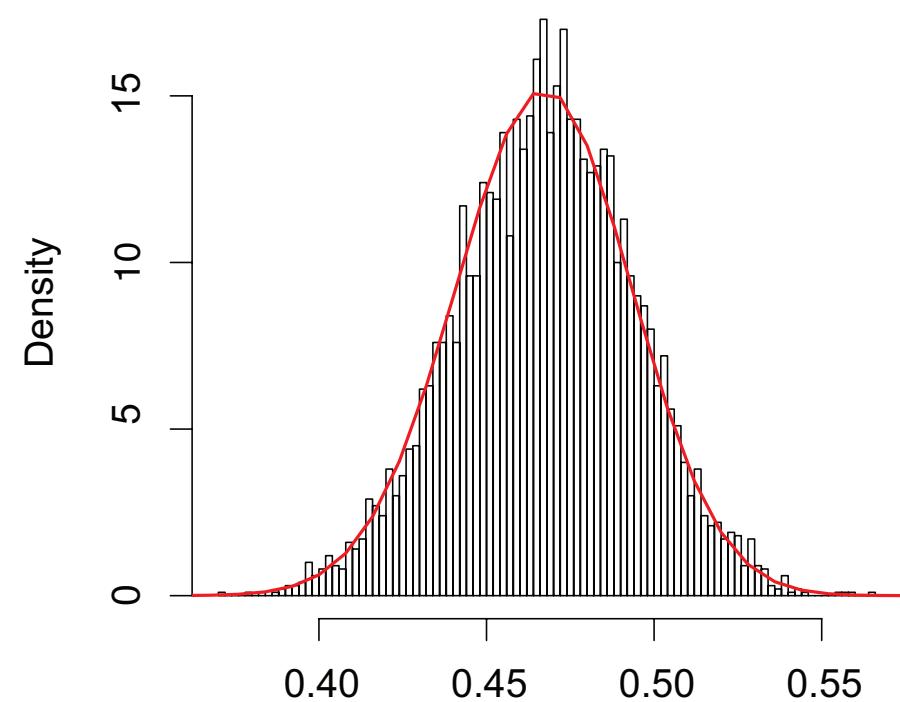




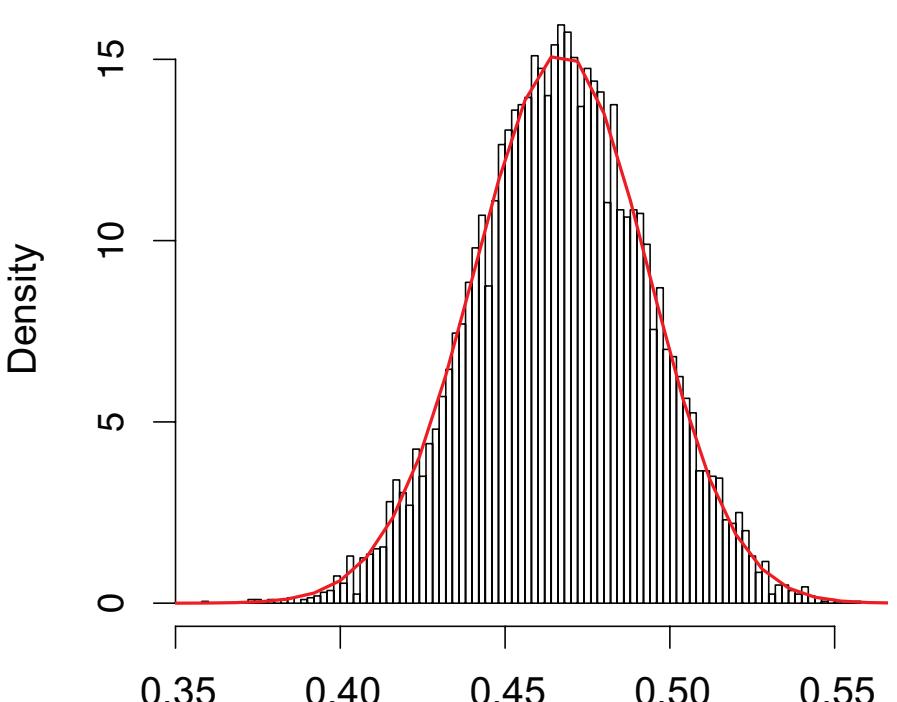
(a) Scenario 1: chain length = 100000000



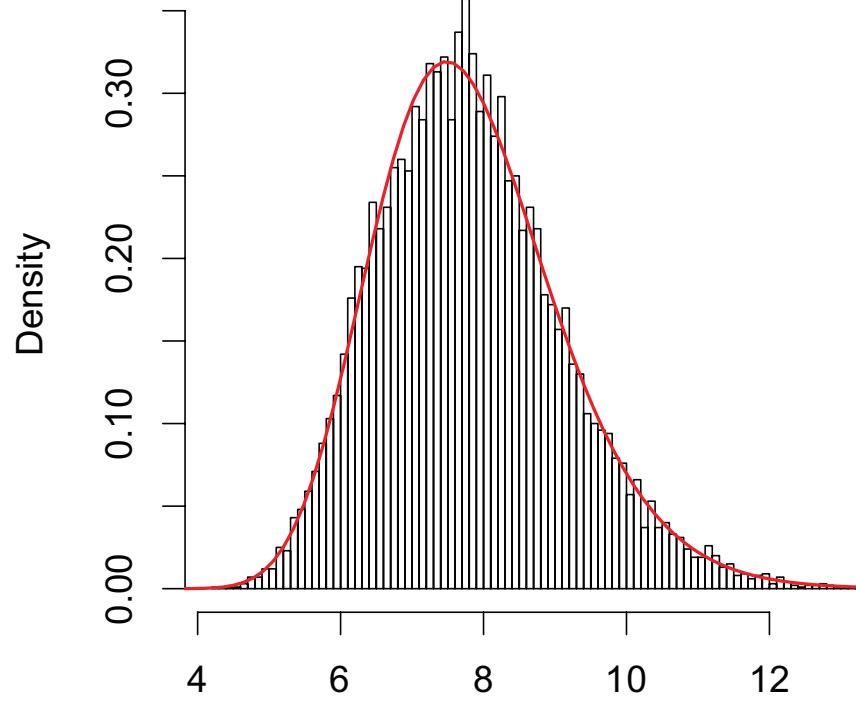
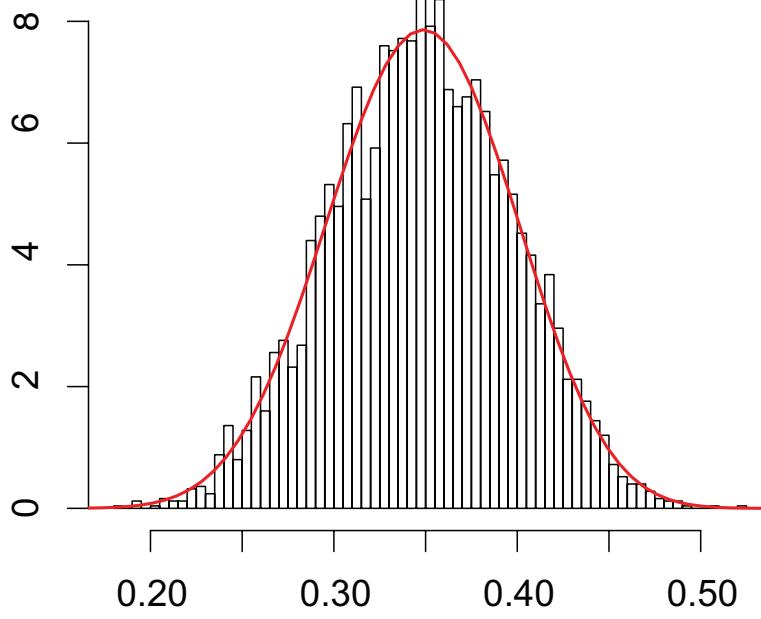
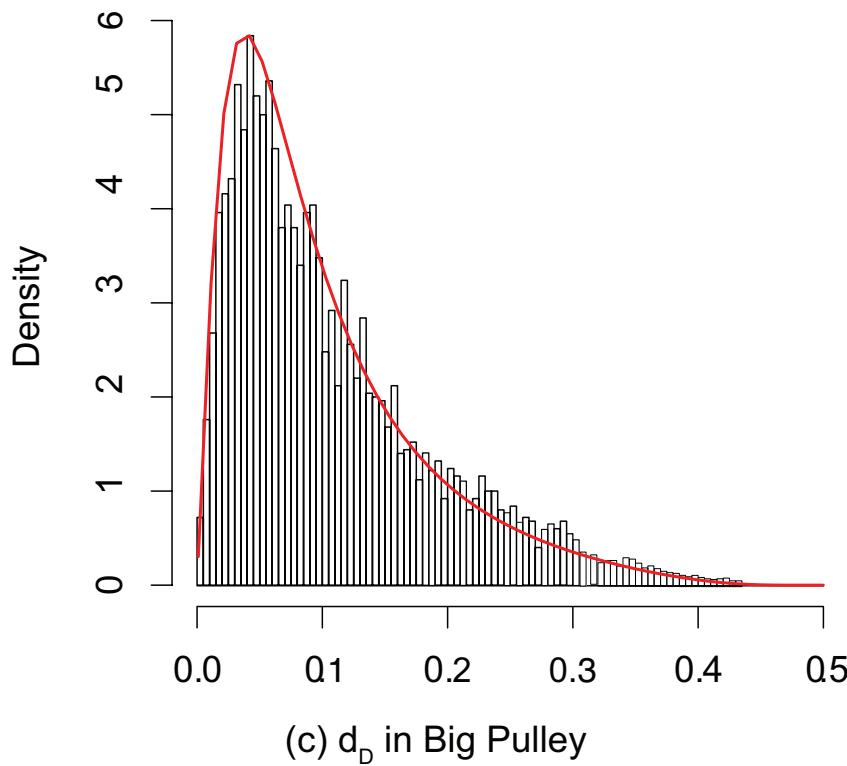
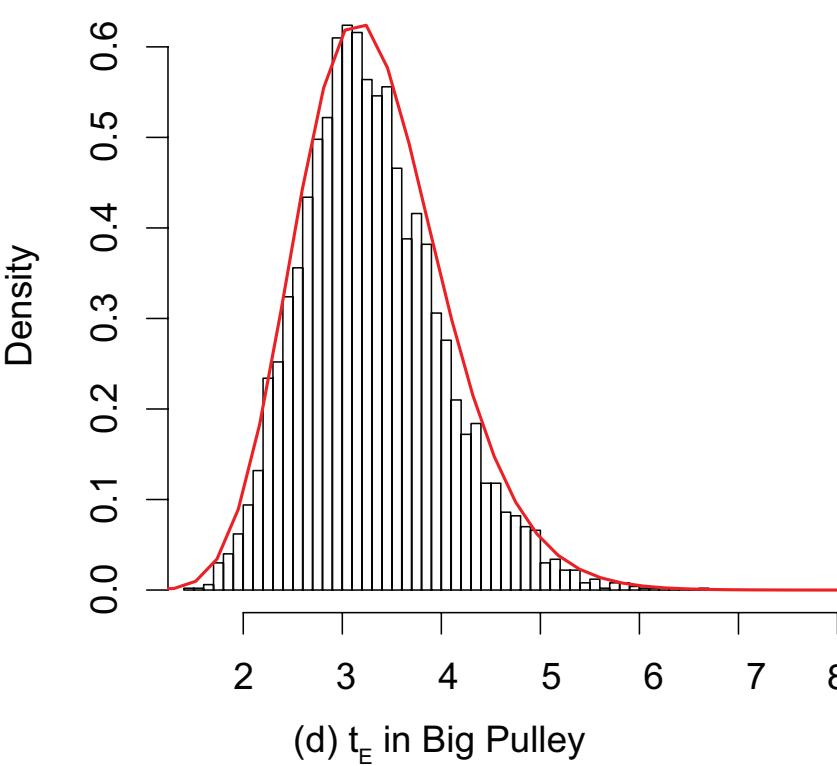
(b) Scenario 1: chain length = 200000000

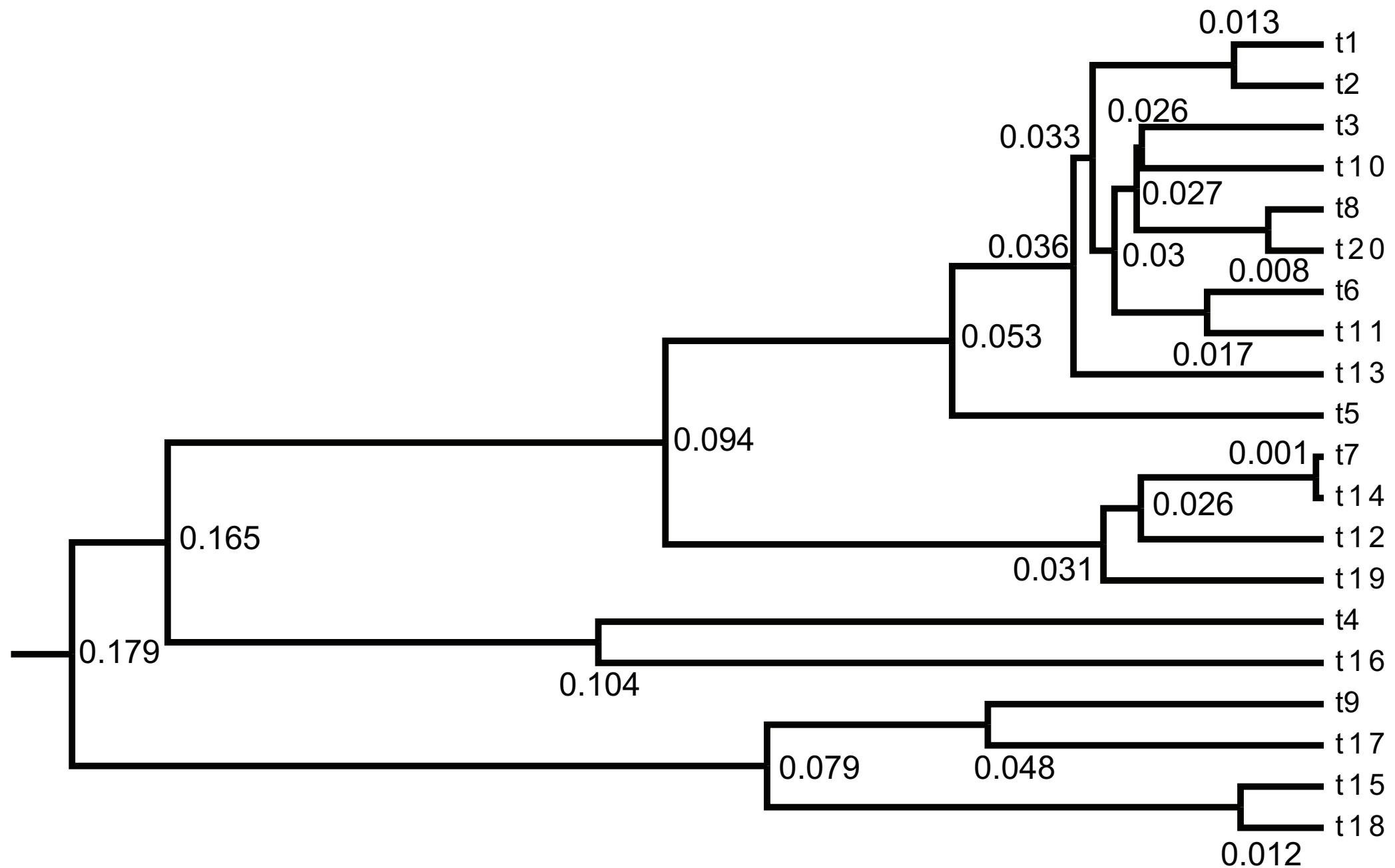


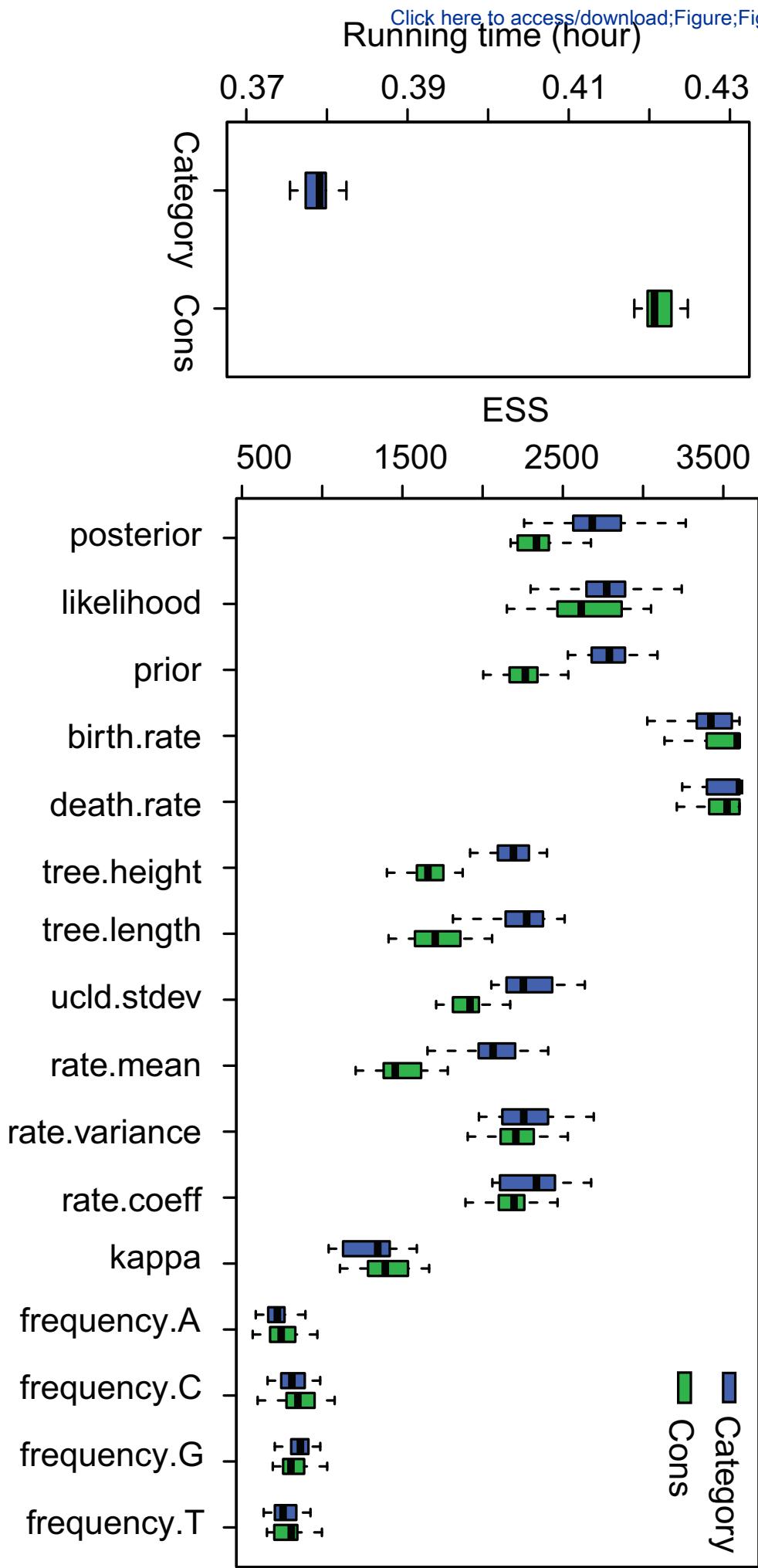
(c) Scenario 2: chain length = 100000000

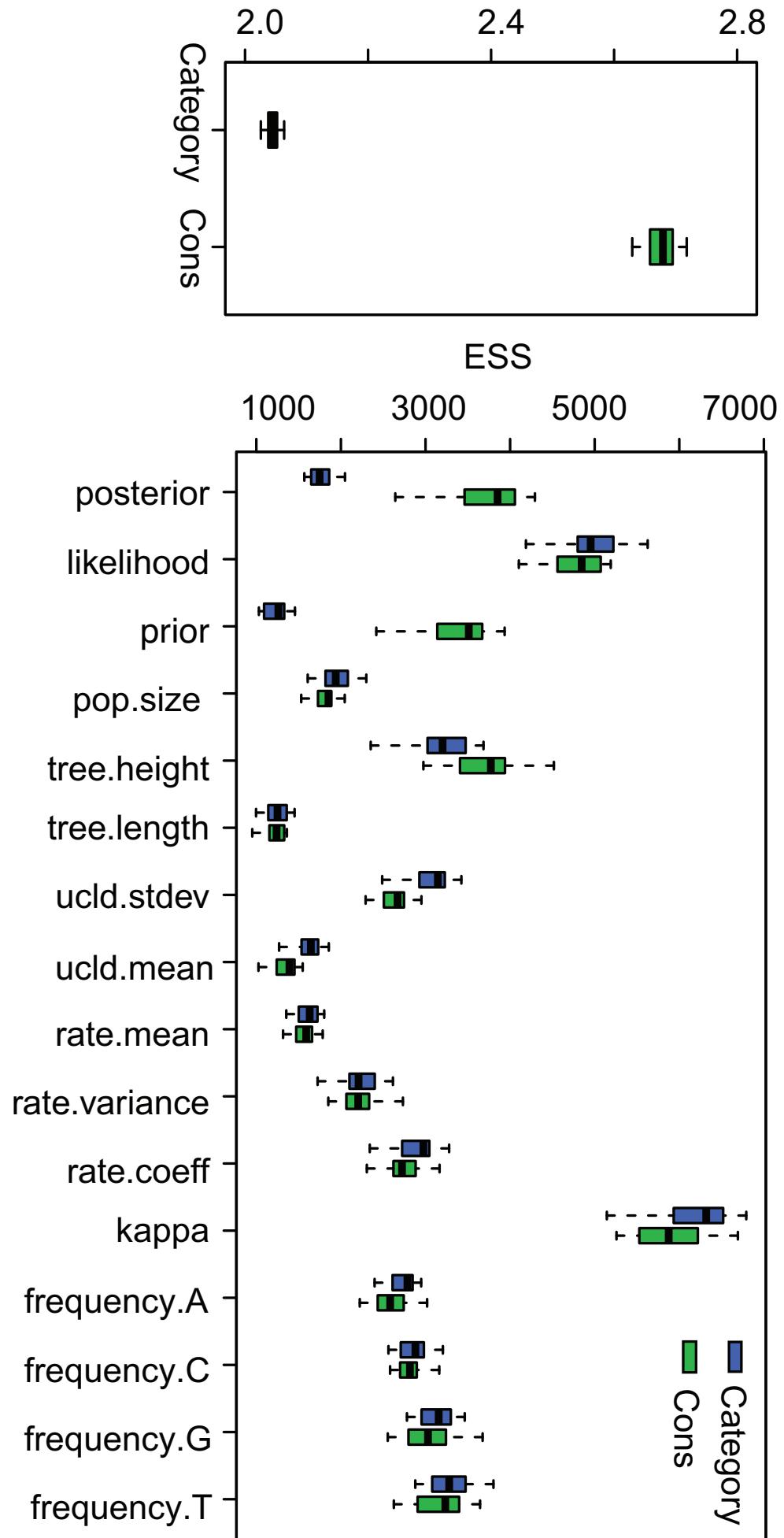


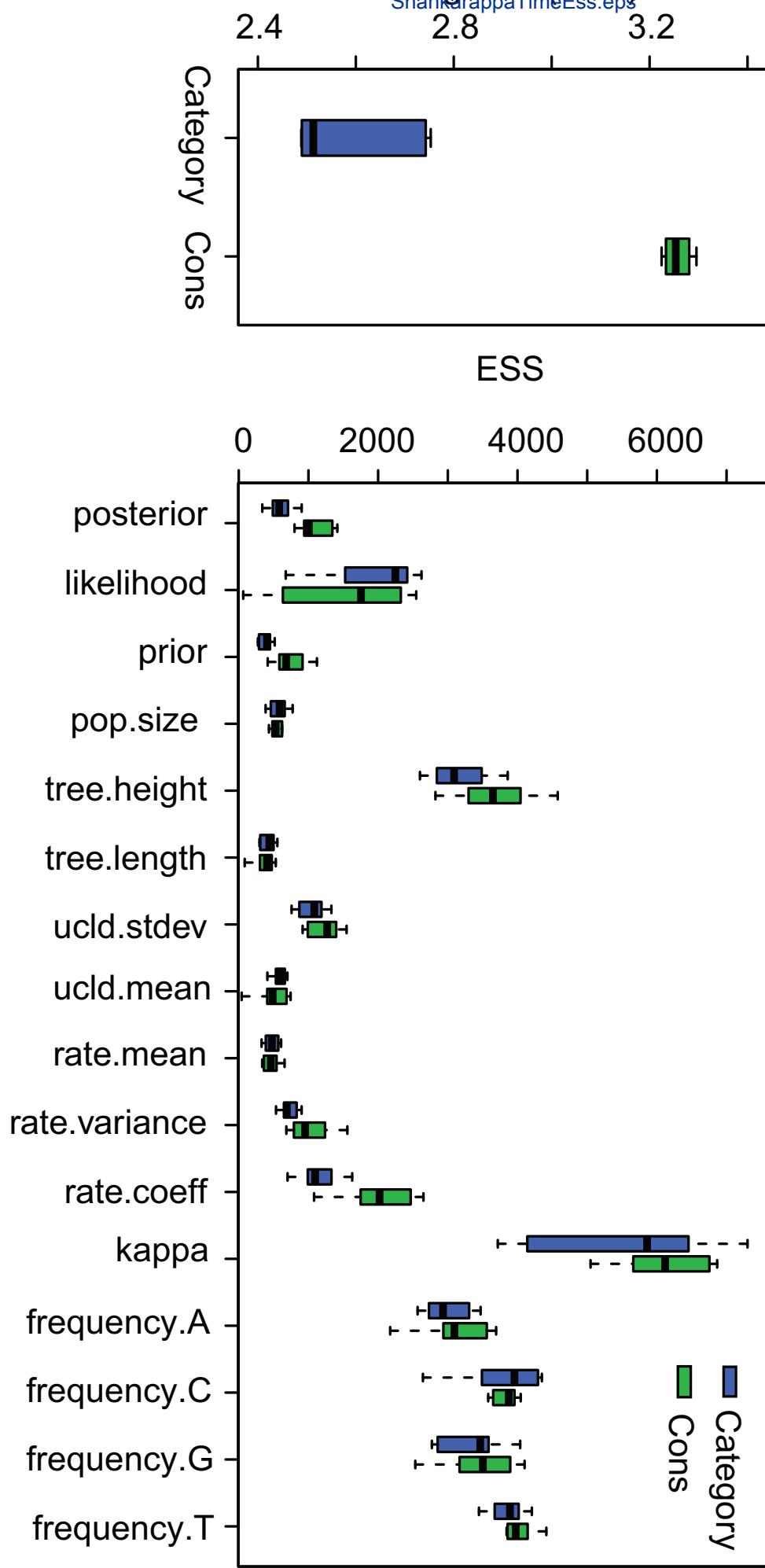
(d) Scenario 2: chain length = 200000000

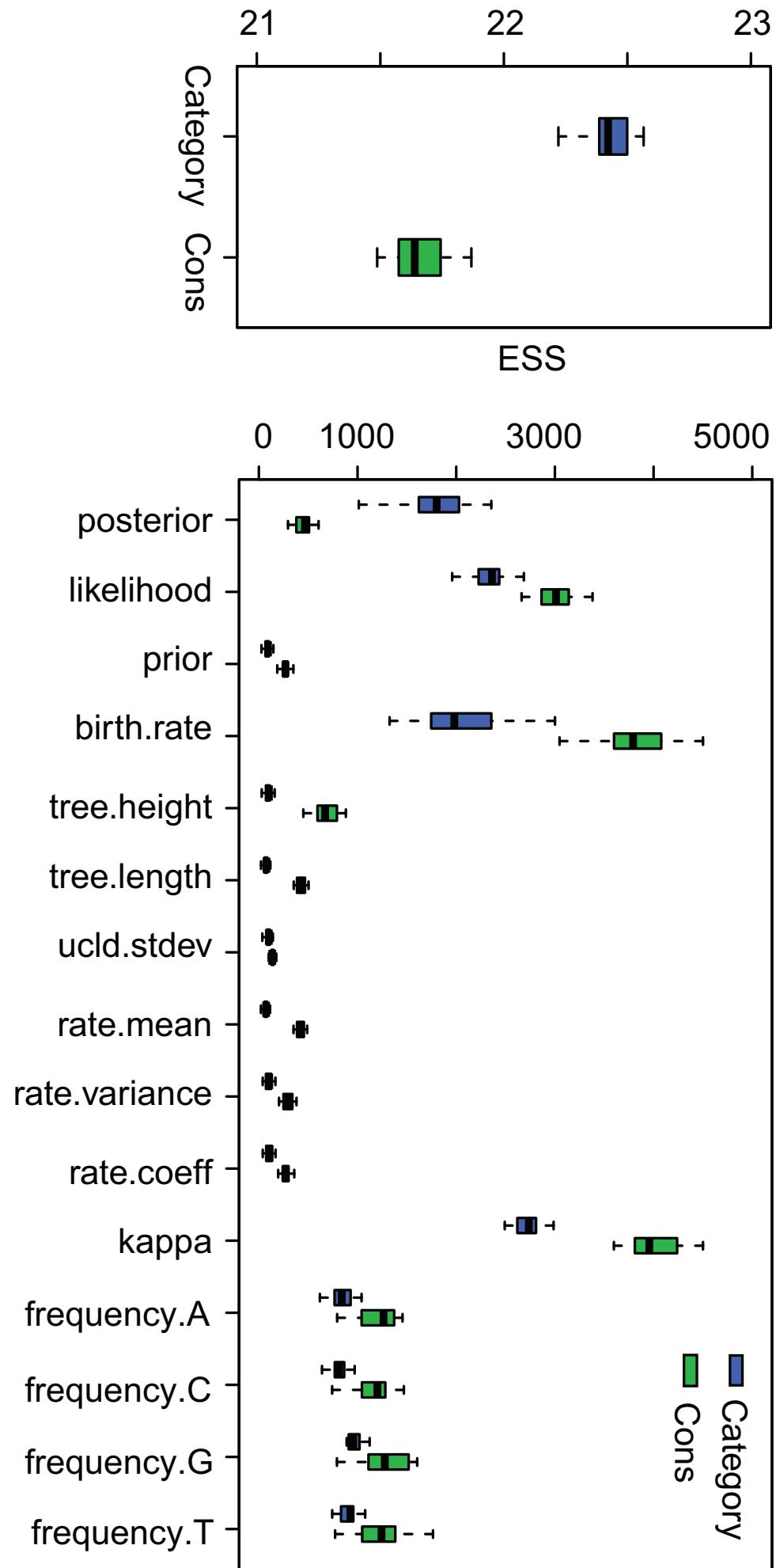
(a) t_E in Simple Distance(b) d_D in Small Pulley(c) d_D in Big Pulley(d) t_E in Big Pulley

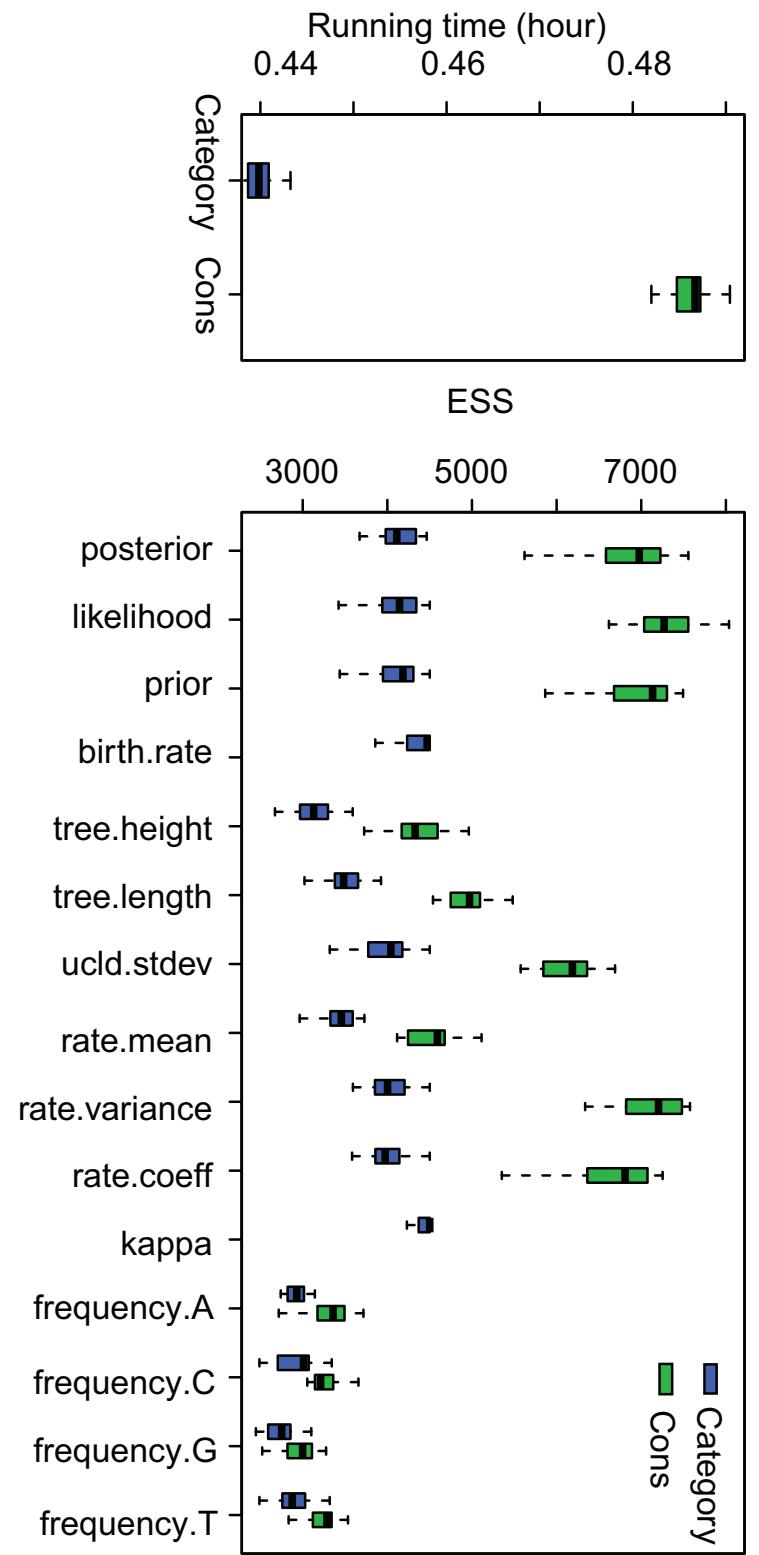




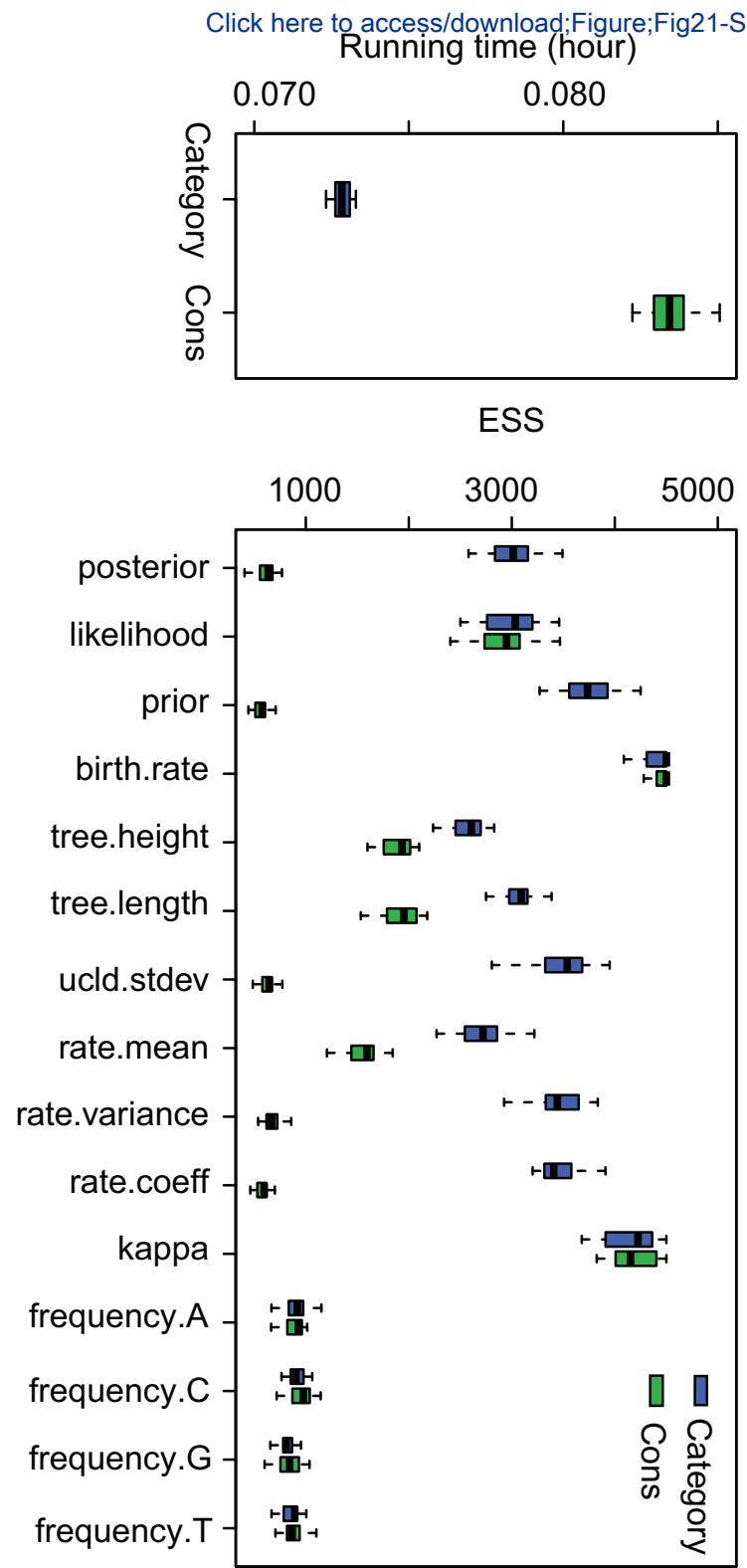






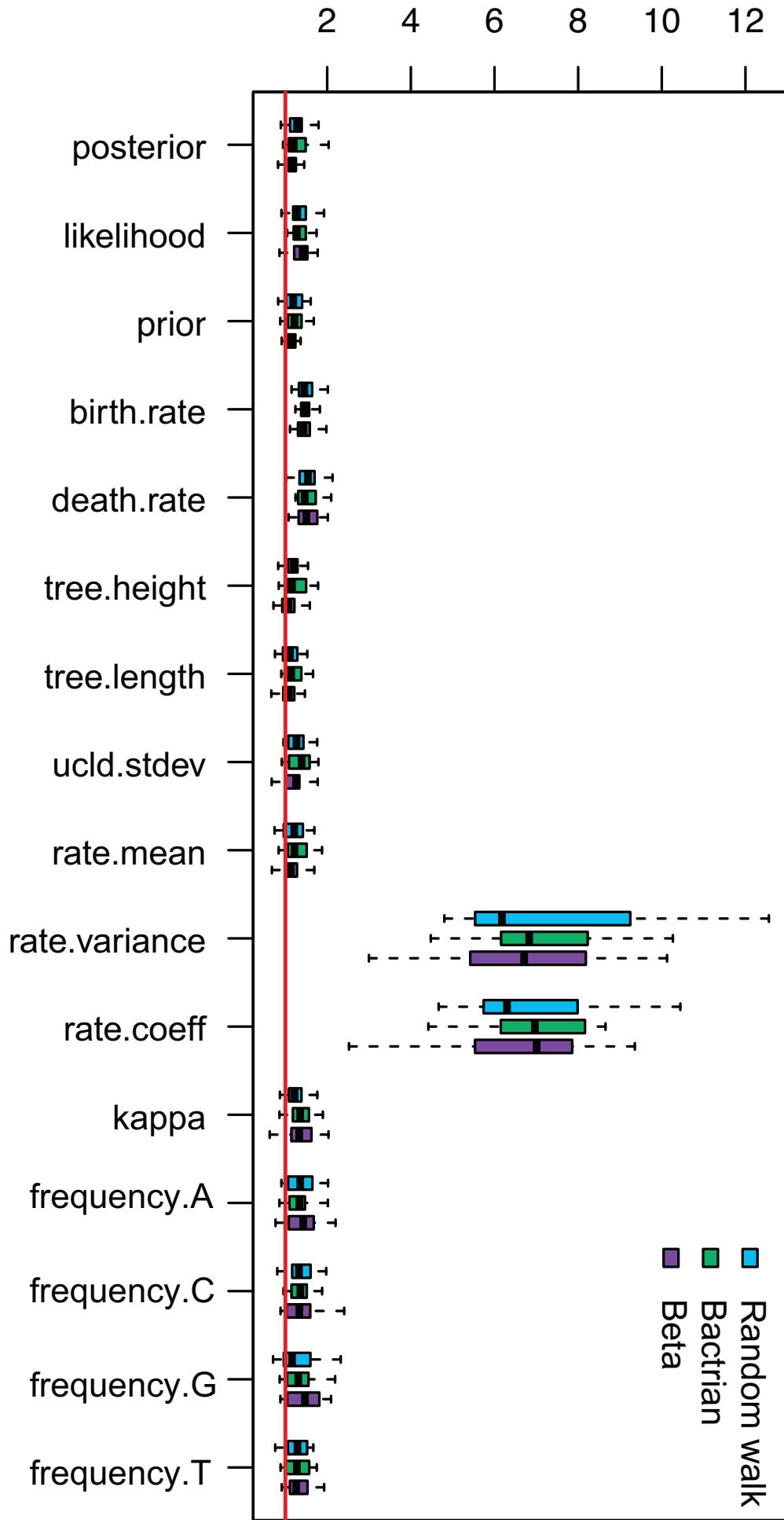


Simulated data with 1000 sites

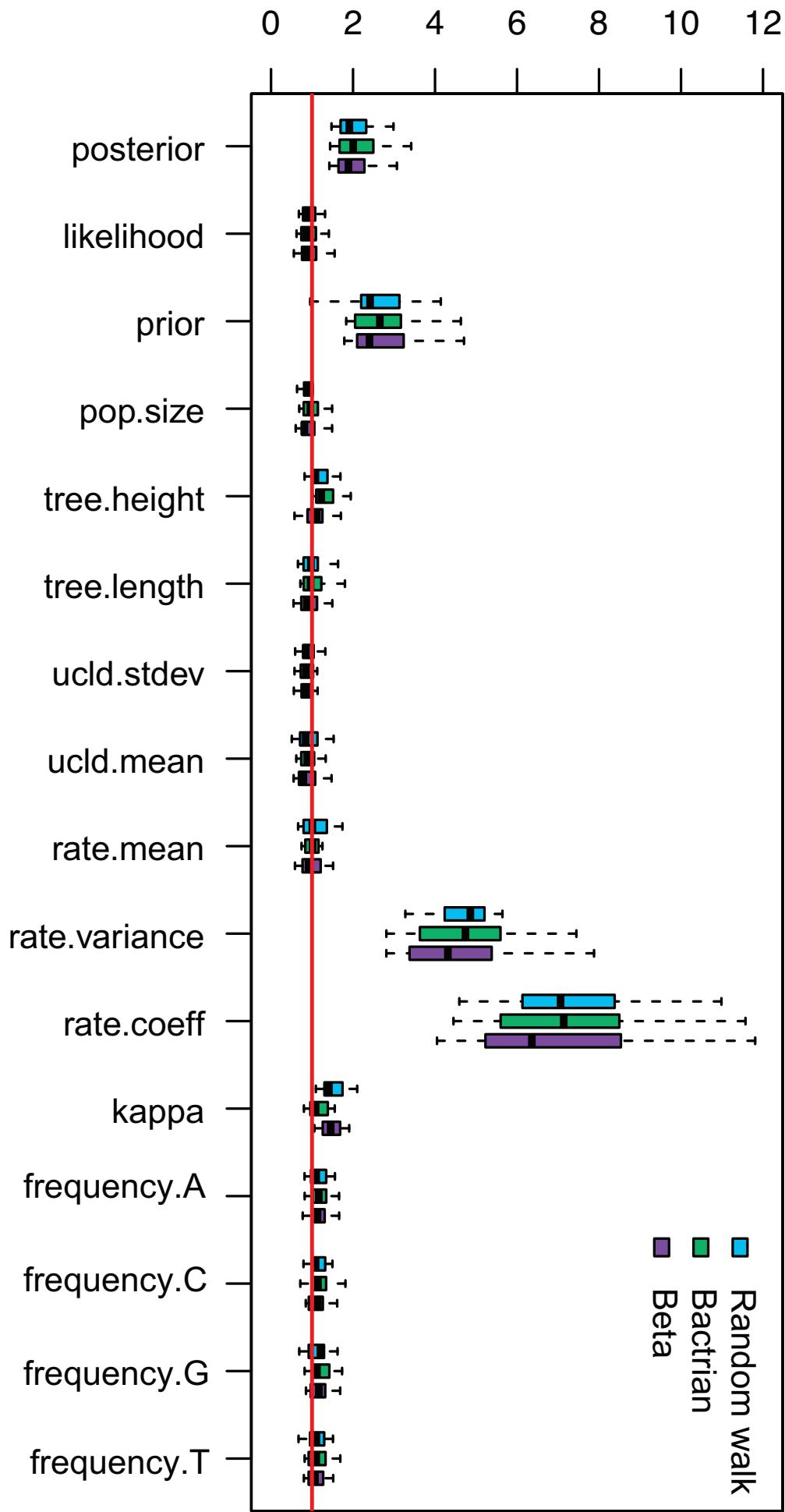


Simulated data with 500 sites

Ratio of ESS per hour (cons/categories)



Ratio of ESS per hour (cons/categories)



Ratio of ESS per hour (cons/categories)

