METHODOLOGY ARTICLE

Improving the performance of Bayesian phylogenetic inference under relaxed clock models

Rong Zhang[1] and Alexei Drummond[2*]

*Correspondence:
alexei@cs.auckland.ac.nz
[2]School of Computer Science, University of Auckland, 38 Princes Street, 1010 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 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 and using less running time. Measured by effective samples per hour, the proposed operator is more than an order of magnitude faster than the current operators in BEAST2 on real and simulated data sets.

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

Background

Bayesian phylogenetics puts an emphasis on estimating probability distributions 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.

Historically, early implementations of Bayesian phylogenetic inference assumed a strict molecular clock where the evolutionary rates are the same at every branch [4]. This was the preferred method for estimating divergence times [3, 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 relaxed clock model has been widely applied, such as the study of Nothofagus [8] and flowering plants [9]. More specifically, as branch rates vary throughout a phylogenetic tree, it is able to identify regional gene bias in cooperation with fossil calibrations, according to Susanne's research in plant dispersal events [10]. Other studies such as accuracy evaluations [11] and performance comparisons [12] also suggested that relaxed clock models play an important role in estimating molecular rates and divergence times.

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

However the overall efficiency of the sampling process also depends strongly on the construction of the proposal mechanism. An effective proposal mechanism is proficient at exploring the posterior distribution, and can do so with fewer steps in the MCMC chain. Therefore fewer likelihood calculations are required, since each step in the chain requires a likelihood calculation.

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

In this paper, we present a novel operator that searches within a subspace of constant genetic distances. Namely, the proposed operator changes both divergence times of nodes and neighbouring branch rates so that implied genetic distances are not changed. For time-reversible substitution models the phylogenetic likelihood will also be unchanged under this operation. The proposed operator has been implemented and tested in BEAST2 [18].

Preliminaries

Bayesian MCMC

Let $\mathbf{D}$, g and $\mathbf{\Phi}$ denote the data, phylogenetic tree topology and a set of evolutionary parameters respectively. The posterior probability density can be calculated using Equation (1). It consists of prior distributions for the tree and the parameters, a phylogenetic likelihood that conveys information from data, and the posterior distribution to be inferred. These are denoted in the form of probability densities by $p(g)$, $p(\mathbf{\Phi})$, $p(\mathbf{D}|g,\mathbf{\Phi})$, $p(g,\mathbf{\Phi}|\mathbf{D})$ respectively. From a Bayesian perspective, the phylogenetic trees and the parameters are random variables described by a posterior probability distribution given the observed data $\mathbf{D}$.

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

However, due to the state space being large to explore and the marginal likelihood being infeasible to calculate, MCMC is adopted to sample the posterior distribution.

Specifically, MCMC algorithms construct a Markov chain whose stationary distribution is the posterior distribution $p(g, \mathbf{\Phi}|\mathbf{D})$, 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 $\theta = \{g, \mathbf{\Phi}\}$ from a specific proposal kernel $q(\theta'|\theta)$ and also return the Hastings-Green ratio for the proposed state transition [2, 19].

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 [20]. Similarly, scale operators multiply a subset of the current state by a scale factor [21]. They are suitable for working on a single random variable, or a single component of the model, such as a population size. Standard operators for the tree topology and divergence times include the subtree slide operator, Wilson-balding and narrow exchange operators [22, 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 $e^{\mathbf{Q}t}$ and r is modelled in the substitution-rate matrix $\mathbf{Q}$. In this way, it is able to avoid recalculations of phylogenetic likelihood which makes MCMC more efficient.

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 divergence times $\mathbf{t}$ can be obtained by the conditional distribution in Equation (2):

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

where $p(\mathbf{r}|\mathbf{\Phi})$ is the prior for rates specified in uncorrelated relaxed clock model. In the constructed Markov chain, the operator moves the original state $\theta = \{\mathbf{t}, \mathbf{r}, g, \mathbf{\Phi}\}$ by proposing a new state $\theta' = (\mathbf{t}', \mathbf{r}', g', \mathbf{\Phi}')$ state .

While the proposed operator is introduced based on uncorrelated clock models, it is admitted that it can be applied to any others relaxed models as well, 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 **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 **X** that is randomly selected among the internal nodes. Let g be the tree in the current state. The following steps propose a new node time and three rates in tree g'.

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

Step 2 Denote the nodes times of **X**, **P**, **L** and **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 **X** by $t_X{}' = t_X + a$, where a follows a Uniform distribution with a symmetric window size w, i.e. $a \sim U[-w, +w]$, 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).

$$r_X{}' = \frac{r_X \times (t_P - t_X)}{t_P - t_X{}'} \quad r_L{}' = \frac{r_L \times (t_X - t_L)}{t_X{}' - t_L} \quad r_R{}' = \frac{r_R \times (t_X - t_R)}{t_X{}' - t_R} \tag{3}$$

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 times for the special case of when **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 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 **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 and two rates in tree g' in Figure 3(c). At the same time, the genetic distances of two branches linked to the root, i.e. d_L and d_R, are kept constant constant.

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

Step 2 Propose a new node time for the root **X** by $t_X{}' = t_X + a$, where $a \sim U[-w, +w]$. Make sure that $t_X{}' > \max\{t_L, t_R\}$ holds. Otherwise, we reject the proposal.

Step 3 Propose new rates for branches on both sides of the root by using Equation (4).

$$r_L{}' = \frac{r_L \times (t_X - t_L)}{t_X{}' - t_L} \qquad r_R{}' = \frac{r_R \times (t_X - t_R)}{t_X{}' - t_R} \tag{4}$$

Step 4 Return the Green ratio α_{SD}.

Small Pulley

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

Step 1 Identify the child nodes of the root **X**, denoted by **L** and **R**. Their corresponding node times and branch rates are t_X, t_L, t_R and r_L, r_R. The implied genetic distances of the two branches linked to the root can be calculated by:

$$d_L = r_L \times (t_X - t_L) \qquad d_R = r_R \times (t_X - t_R) \tag{5}$$

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

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

$$r_L{}' = \frac{d_L{}'}{t_X - t_L} \qquad r_R{}' = \frac{D - d_L{}'}{t_X - t_R} \tag{6}$$

Step 4 Return the Green ratio α_{SP}.

Big Pulley

Big Pulley resamples the rates and times in a fixed unrooted tree. The genetic distances between the taxa are held constant, but the location of the root is readjusted.

Before describing the detailed steps, we introduce a method *Exchange* that proposes a new tree topology when it is called in Big Pulley. In Figure 4, let (i) **X** denote the root of tree g, (ii) **C** and **N** denote the two child nodes of **X**, (iii) **S** and **M** denote the two child nodes of **C**. Once the method is called by certain step in Big Pulley, the following operations will be executed.

- Call *Exchange(**M**, **N**)* to swap the two nodes by pruning and grafting, i.e. cutting **M** (**N**) at its original position and attaching it to the original position of **N** (**M**).
- Propose $d_C{}'$ by $d_C{}' = d_C + b$, where $b \sim U[-v, +v]$. Make sure that $0 < d_C{}' < D$ holds, where $D = d_C + d_N$. Otherwise, we reject the proposal.
- The distances on the other three branches, i.e. d_S, d_M and d_N, will be adjusted:

$$d_S{}' = d_S \qquad d_M{}' = d_M - d_C{}' \qquad d_N{}' = d_N + d_C \tag{7}$$

As can be seen from the above descriptions, the method *Exchange (**M**, **N**)* is actually aimed at swapping two nodes and reassigning distances on the four branches. That is to say, after using *Exchange (**M**, **N**)*, the distances d_S, d_M, d_N and d_C will be adjusted to maintain the implied genetic distances among three taxa **S**, **M** and **N**, as the tree topology changes.

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

Symmetric tree For the symmetric tree in Figure 5, the operations are illustrated in Figure 6, after which one of the four possible trees (① ② ③ ④) will be proposed.

Step 1 Identify the child nodes of the root **X**, denoted by **L** and **R**. Correspondingly, the node times are denoted by t_X, t_L, t_R. And the child nodes below them are denoted by **H1**, **H2**, **H3** and **H4**.

Step 2 Propose a new node time for the root **X** by $t_X{}' = t_X + a$, where $a \sim U[-w, +w]$.

Step 3 Propose a new node time either for **L** or **R**. And apply the method to the selected node and one of its sibling's children.

- With 0.5 probability to pick **L** and propose a new node time by $t_L{}' = t_L + a_1$, where $a_1 \sim U[-w, +w]$. Make sure that $t_R < t_L{}' < t_X{}'$ holds. Otherwise, we reject the proposal. If we don't reject then there are two options to apply the method:
  - With 0.5 probability to apply *Exchange (**H1**, **R**)* and propose tree ①
  - With 0.5 probability to apply *Exchange (**H2**, **R**)* and propose tree ②
- With 0.5 probability to pick **R** and propose a new node time by $t_R{}' = t_R + a_2$, where $a_2 \sim Uniform[-w, +w]$. Make sure that $t_L < t_R{}' < t_X{}'$ holds. Otherwise, we reject the proposal. Similarly, if we don't reject then there are two options to apply the method:
  - With 0.5 probability to apply *Exchange (**H3**, **L**)* and propose tree ③
  - With 0.5 probability to apply *Exchange (**H4**, **L**)* and propose tree ④

Step 4 Update the rates using the adjusted genetic distances divided by the proposed node times. For example, suppose we are going to propose tree ①. After the new node times for the root **X** and **L** are proposed, we apply the method by *Exchange (**H1**, **R**)*, so that four distances are adjusted, as follows:

$$d_{H1}{}' = d_{H1} - d_L{}' \quad d_{H2}{}' = d_{H2} \quad d_L{}' = d_L + b \quad d_R{}' = d_L + d_R \tag{8}$$

Finally, in this example the new rates would be updated by:

$$r_{H1}{}' = \frac{d_{H1}{}'}{t_X{}' - t_{H1}} \quad r_{H2}{}' = \frac{d_{H2}{}'}{t_L{}' - t_{H2}} \quad r_L{}' = \frac{d_L{}'}{t_X{}' - t_L{}'} \quad r_R{}' = \frac{d_R{}'}{t_L{}' - t_R} \tag{9}$$

Step 5 Return the Green ratio α_{BP}.

Asymmetric tree For an asymmetric tree such as in Figure 5 we would operate as illustrated in Figure 7, in which there are three possible trees (⑤ ⑥ ⑦).

Step 1 Identify the extant child node of the root **X**, which has two child nodes below and is denoted by **O**. The extinct child node of the root, which does not have any child nodes, is denoted by **Y**. The node times of the root **X**, **Y**, **O** and its child nodes are denoted by t_X, t_Y, t_O, t_{G1} and t_{G2} respectively.

Step 2 Propose a new node time for the root **X** by $t_X{}' = t_X + a$, where $a \sim U[-w, +w]$. Moreover, propose a new node time for **O** by $t_O{}' = t_O + a_3$, where $a_3 \sim U[-w, +w]$. To make it valid, make sure that $t_Y < t_O{}' < t_X{}'$ holds. Otherwise, we reject the proposal.

Step 3 Apply the method using **Y** and either child node of **O**, which is dependent on the value of $t_O{}'$.

- if $t_O{}'$ satisfies $t_O{}' > \max\{t_{G1}, t_{G2}\}$ or $t_{G1} = t_{G2}$, then there are two options:
  - With 0.5 Probability to apply *Exchange (**G1**, **Y**)* and propose tree ⑤
  - With 0.5 Probability to apply *Exchange (**G2**, **Y**)* and propose tree ⑥
- if $t_O{}'$ satisfies $\min\{t_{G1}, t_{G2}\} < t_O{}' < \max\{t_{G1}, t_{G2}\}$, then there is only one option: ⑦: Exchange the older child of **O** and **Y**. (For the asymmetric tree in Figure 5, we apply *Exchange (**G1**, **Y**)* and propose tree ⑦).
- if $t_O{}' < \min\{t_{G1}, t_{G2}\}$, then we reject the proposal.

Step 4 Update the rates using the adjusted genetic distances divided by the proposed node times. To give an example, assume we are going to propose tree ⑤. Firstly, $t_X{}'$ and $t_O{}'$ are proposed in *Step 3*. Then, in *Step 4*, the method *Exchange (**G1**, **Y**)* is applied, after which the four distances are adjusted as follows:

$$d_{G1}{}' = d_{G1} - d_O{}' \quad d_{G2}{}' = d_{G2} \quad d_O{}' = d_O + b \quad d_Y{}' = d_Y + d_O \tag{10}$$

And the four rates are updated as follows:

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

Step 5 Return the Green ratio α_{BP}.

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 acceptance.

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 [19], it is necessary to construct a 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 $\mathbf{OUT} = [t_X{}', r_X{}', r_L{}', r_R{}']$.

$$\mathbf{J_1} = \begin{bmatrix} \frac{\partial \mathbf{f}}{\partial t_X} & \frac{\partial \mathbf{f}}{\partial r_X} & \frac{\partial \mathbf{f}}{\partial r_L} & \frac{\partial \mathbf{f}}{\partial r_R} \end{bmatrix} = \begin{bmatrix} \frac{\partial f_1}{\partial t_X} & \frac{\partial f_1}{\partial r_X} & \frac{\partial f_1}{\partial r_L} & \frac{\partial f_1}{\partial r_R} \\ \frac{\partial f_2}{\partial t_X} & \frac{\partial f_2}{\partial r_X} & \frac{\partial f_2}{\partial r_L} & \frac{\partial f_2}{\partial r_R} \\ \frac{\partial f_3}{\partial t_X} & \frac{\partial f_3}{\partial r_X} & \frac{\partial f_3}{\partial r_L} & \frac{\partial f_3}{\partial r_R} \\ \frac{\partial f_4}{\partial t_X} & \frac{\partial f_4}{\partial r_X} & \frac{\partial f_4}{\partial r_L} & \frac{\partial f_4}{\partial r_R} \end{bmatrix}, \tag{12}$$

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:

$$\alpha_{IN} = \frac{p(-a)}{p(a)} |\mathbf{J_1}| = \frac{t_P - t_X}{t_P - t_X{}'} \times \frac{t_X - t_L}{t_X{}' - t_L} \times \frac{t_X - t_R}{t_X{}' - t_R}, \tag{13}$$

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:

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

$$\alpha_{SP} = 1, \tag{15}$$

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

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

Results

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

First, we establish correctness of the operator using a well-calibrated simulation study, which shows our operator is able to function alongside other operators. Correctness was further confirmed by sampling trees from the prior distribution i.e. without data (Refer to Appendix section 2 for more details). By comparing effective sample sizes (ESS) [24] and running times, it is demonstrated that the performance is improved when using the proposed operator. Finally, the correlation of rates and node times are discussed.

Well-calibrated simulation study

A well-calibrated simulation study is a necessary criterion to evaluate the reliability of a Bayesian model [25].

Figure 8 shows the framework used in this study, which includes the evolutionary model and the prior distributions of parameters. As is shown in the figure, the sequence alignment is simulated by a phylogenetic continuous-time Markov chain in BEAST2. It contains a substitution rate matrix given by the HKY85 [26] model and a substitution tree jointly provided by uncorrelated relaxed clock model and Yule model. More specifically, base frequencies π follow a Dirichlet distribution and the transition-transversion ratio κ follows a log-normal prior distribution. The distribution of node times is described in a Yule tree ψ with hyperparameter birth rate λ following a log-normal distribution. The rates r_i follow a log-normal distribution with mean of 1 and standard deviation $S1$ following a hyperprior distribution.

First, we independently simulated parameters and trees from the full model 100 times. The random parameters included: standard deviation of rates across branches $S1$, birth rate λ, base frequencies π and transition-transversion bias κ. Second, we simulated nucleotide alignments using the simulated parameters. In total, there are 100 data sets simulated, each with 120 taxa. Third, we used BEAST2 with Constant Distance operator to infer the tree and parameters in the 100 simulated data sets. Finally, the estimated values of the parameters were compared with the real values that were used to simulate the corresponding sequence alignment. The comparisons are shown in Figures 9.

These results show that the true values of the parameters are within the 95% highest posterior density (HPD) interval approximately 95% of the time (Table 1). This well-calibrated simulation study confirms that the Constant Distance operator can successfully provide reliable parameter estimates.

Performance comparison

To evaluate the performance of Constant Distance operator in a Bayesian phylogenetic analysis, we explored the time required to adequately sample the posterior distribution. This was achieved by examining (i) the total time taken by BEAST2 to complete the MCMC inference (running time), and (ii) the effective sample size (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 the latest BEAST2 version. We aim to demonstrate the superiority of Constant Distance operator by showing the performance of Cons configuration being better than Category configuration. In each configuration, the data set was ran 20 times with the prior distributions and all other model specifications held constant. The details of operator weights are given in Appendix 3.1.

We performed the analysis on two sets of simulated sequence alignment (See Appendix 3.2 for more details). The simulated data sets both have 20 taxa but different

sequence lengths, i.e. one data set containing 500 sites, the other containing1,000 sites. Moreover, we used four real data sets to further evaluate the performance of Constant Distance operator, including a primate data set from Ref.[27] and three data sets (anolis, RSV2 and Shankarappa) in BEAST2 [28].

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 hour. Then we compare the efficiency of two configurations by calculating the ratios of ESS per hour for simulations in the two configurations. Namely, if the ratio is larger than 1, then ESS per hour of Cons configuration is larger than that of Category configuration. As is shown in Figure 10, most ratios of the parameters are above the red line (larger than 1), which indicates that Cons configuration provides larger ESS per hour for most parameters. Although there are are several parameters sampled by Cons configuration having smaller ESS per hour in some data sets, it should be noticed that the ratio is calculated by choosing random simulations in the two configurations (See Appendix 3.3 for more details). Additionally, it is worth noting that the efficiency is improved more obviously in simulated data set having1000 sites, compared with the data set having 500 sites. This means the proposed operators sample rates and node times more efficiently if the genetic distances are more accurate. On the other hand, Table 2 lists the average running time of the data sets. It can be seen that Cons configuration finished simulations with less time in most cases. Moreover, Table 2 also shows the parameter that has the smallest ESS in Category configuration, and is compare with the corresponding ESS in Cons configuration. After calculating the ESS per hour, we conclude that Cons configuration improved the efficiency of the worst estimated parameter in Category configuration by 2.26 to 15.8.

Correlation analysis of rates and branch lengths

In this section, we conduct a pairwise comparison between rates and branch lengths in time. We used a data set from Copper et al.'s work [29]. This data set includes 7 taxa of ratites and the genome sequences have 10767 sites. After analysing the ratites data set in BEAST2 using the Constant Distance operator, we calculate Pearson coefficient and demonstrate that the proposed operators sample rates and divergence times in a correct relationship by constraining the constant distances.

The results are summarised in Figure 11. Figure 11(a) presents the topology of the maximum clade credibility tree. We utilised the programme TreeStat2 [30] 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, to a large degree, the rates are negatively correlated with branch lengths, which indicates that a larger node time will lead to a smaller rate. This is because the operators propose a branch rate r and a node time t, on condition that the distance d is constant, i.e. $d = r \times l$. The consequence is that we have a linear relationship between rate r and branch length l i.e. $r = d/l$. For example, if a larger t_1 is proposed, l_1 increases as well, but r_1 goes down. At the same time, l_8 decreases,

which causes r_8 to become larger, so that l_1 and r_8 have a positive relationship. To sum up, this dynamic change of rates and branch lengths is consistent with the mechanism of the proposed operator. Although there are some inconsistent correlations, it should be noticed that this is an average pairwise comparison in two dimensions and there are other operators sampling the rates and node times. For comparisons in higher dimensions, the results would be closer to the mechanism of the proposed operator.

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 [31], this limiting case might be approached in some data sets. As a simple test in this paper, this gives rise to an alternative approach to analysis, where a time tree, the root position and the branch rates are random variables, and the data are a set of branch lengths in units of substitution on a known unrooted tree topology.

Previous work done by Reis and Yang [14] also tried to approximate the likelihood of such an unrooted tree in Bayesian phylogenetic inference. Similar researches in [6, 32] show that these methods can account for rate changes in a relaxed clock model, but the genetic distances are not fixed, for example Stéphane Guindon used a Gibbs sampling algorithm [32]. Except Bayesian MCMC methods, other models, such as least-squares criteria [33] and maximum likelihood [34, 35], are applied to estimate substitution rates and divergence times in unrooted trees.

In this section, we investigated this approach on a fixed substitution tree reconstructed from whole mitochondrial genomes from a set of ratite species [29]. Since no uncertainty is admitted in the genetic distances and the proposed operator don't change the genetic distances, the phylogenetic likelihood is no longer needed and the unrooted tree becomes the data, rather than a multiple sequence alignment.

First of all, we used the ratites data set to construct an unrooted tree with an online program PhyML 3.0 [36, 37]. Figure 12(a) shows the unrooted tree with the genetic distances on the branches which are fixed in the subsequent relaxed clock analysis in BEAST2.

As an initial starting point, the root is randomly assigned. In this paper, we used the midpoint method to root the starting time tree. After that, according to the genetic distances among seven taxa and the topology of the rooted tree, consistent divergence times are specified and assigned to each ancestral node, so that a valid rooted time tree is obtained. Once divergence times are determined, rates on the branches are also calculated so that the products match the unrooted substitution tree.

Then we used Constant Distance operator to sample this starting tree. The resulting summary trees are shown in Figure 12(b)-(d). As can be seen, despite that there is some uncertainty in the root position, the most probable tree in Figure 12(b) is consistent with previous analyses of this data (see Figure 2 in Ref. [29]). For large data sets of long sequences, the proposed operators may prove useful to provide

faster divergence time estimates based on the assumption of known unrooted topology and branch lengths in units of expected substitutions per site.

Discussion

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

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

As data sets have increased in size the impetus to improve efficiency of Bayesian phylogenetic inference algorithms has steadily increased. Besides more effective proposal mechanisms within Metropolis-Hastings MCMC, completely novel approaches to Bayesian phylogenetics have also begun to get some attention. Variational methods are one alternative for approximating Bayesian posterior distributions [38]. These approaches make inference an optimisation problem and take advantage of tractable variational distributions that approximate the posterior distribution, thus decreasing the computational cost by avoiding high-dimensional integrals in MCMC sampling schemes. Recent work has investigated the potential for applying variational methods to phylogenetics [39, 40]. Our improved MCMC methods provide a mature performance baseline for these new approaches.

Conclusions

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

We demonstrate the correctness of this algorithm with a series of tests including a well-calibrated simulation study. Based on both simulated and real data sets, the proposed operator is much more efficient than the current algorithms implemented in BEAST2. It is able to sample the rates and times more efficiently, with performance improvements of greater than an order of magnitude increase in ESS/hour on both real and simulated data. The proposed operator is available for use as a package of BEAST2.

Abbreviations

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

Declarations

Ethics approval and consent to participate
Not applicable

Consent for publication
Not applicable

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

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

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

Authors' contributions
RZ developed the operator and was a major contributor in writing the manuscript. AJD supervised the implementation of the operator and the writing process of the manuscript. All authors read and approved the final manuscript.

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

Author details
[1]School of Computer Science, University of Auckland, 38 Princes Street, 1010 Auckland, New Zealand. [2]School of Computer Science, University of Auckland, 38 Princes Street, 1010 Auckland, New Zealand.

References

1. Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N., Teller, A.H., Teller, E.: Equation of state calculations by fast computing machines. The journal of chemical physics **21**(6), 1087–1092 (1953)
2. Hastings, W.K.: Monte carlo sampling methods using markov chains and their applications. Biometrika **57**(1), 97–109 (1970)
3. Yang, Z., Rannala, B.: Bayesian phylogenetic inference using dna sequences: a markov chain monte carlo method. Molecular biology and evolution **14**(7), 717–724 (1997)
4. Zuckerkandl, E., Pauling, L.: Evolutionary divergence and convergence in proteins, 97–166 (1965)
5. Rannala, B., Yang, Z.: Bayes estimation of species divergence times and ancestral population sizes using dna sequences from multiple loci. Genetics **164**(4), 1645–1656 (2003)
6. Thorne, J.L., Kishino, H., Painter, I.S.: Estimating the rate of evolution of the rate of molecular evolution. Mol Biol Evol **15**(12), 1647–57 (1998). doi:10.1093/oxfordjournals.molbev.a025892

7. Drummond, A.J., Ho, S.Y., Phillips, M.J., Rambaut, A.: Relaxed phylogenetics and dating with confidence. PLoS biology **4**(5), 88 (2006)
8. Knapp, M., Stöckler, K., Havell, D., Delsuc, F., Sebastiani, F., Lockhart, P.J.: Relaxed molecular clock provides evidence for long-distance dispersal of nothofagus (southern beech). PLoS Biology **3**(1), 14 (2005)
9. Smith, S.A., Beaulieu, J.M., Donoghue, M.J.: An uncorrelated relaxed-clock analysis suggests an earlier origin for flowering plants. Proceedings of the National Academy of Sciences, 201001225 (2010)
10. Renner, S.S.: Relaxed molecular clocks for dating historical plant dispersal events. Trends in plant science **10**(11), 550–558 (2005)
11. Ho, S.Y., Phillips, M.J., Drummond, A.J., Cooper, A.: Accuracy of rate estimation using relaxed-clock models with a critical focus on the early metazoan radiation. Molecular Biology and Evolution **22**(5), 1355–1363 (2005)
12. 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
13. Kobert, K., Stamatakis, A., Flouri, T.: Efficient detection of repeating sites to accelerate phylogenetic likelihood calculations. Systematic biology **66**(2), 205–217 (2017)
14. Reis, M.d., Yang, Z.: Approximate likelihood calculation on a phylogeny for bayesian estimation of divergence times. Molecular Biology and Evolution **28**(7), 2161–2172 (2011)
15. Ayres, D.L., Darling, A., Zwickl, D.J., Beerli, P., Holder, M.T., Lewis, P.O., Huelsenbeck, J.P., Ronquist, F., Swofford, D.L., Cummings, M.P., *et al.*: Beagle: an application programming interface and high-performance computing library for statistical phylogenetics. Systematic biology **61**(1), 170–173 (2011)
16. Lakner, C., Van Der Mark, P., Huelsenbeck, J.P., Larget, B., Ronquist, F.: Efficiency of markov chain monte carlo tree proposals in bayesian phylogenetics. Systematic biology **57**(1), 86–103 (2008)
17. Höhna, S., Drummond, A.J.: Guided tree topology proposals for bayesian phylogenetic inference. Syst Biol **61**(1), 1–11 (2012). doi:10.1093/sysbio/syr074
18. Bouckaert, R., Heled, J., Kühnert, D., Vaughan, T., Wu, C.-H., Xie, D., Suchard, M.A., Rambaut, A., Drummond, A.J.: Beast 2: a software platform for bayesian evolutionary analysis. PLoS computational biology **10**(4), 1003537 (2014)
19. Green, P.J.: Reversible jump markov chain monte carlo computation and bayesian model determination. Biometrika **82**(4), 711–732 (1995)
20. Suchard, M.A.: Stochastic models for horizontal gene transfer: taking a random walk through tree space. Genetics (2005)
21. Higuchi, T.: Monte carlo filter using the genetic algorithm operators. Journal of Statistical Computation and Simulation **59**(1), 1–23 (1997)
22. 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)
23. Hohna, S., Defoin-Platel, M., Drummond, A.J.: Clock-constrained tree proposal operators in bayesian phylogenetic inference. In: BioInformatics and BioEngineering, 2008. BIBE 2008. 8th IEEE International Conference On, pp. 1–7 (2008). IEEE
24. Rambaut, A., Suchard, M., Xie, D., Drummond, A.: Tracer v1. 6. 2014 (2015)
25. Dawid, A.P.: The well-calibrated bayesian. Journal of the American Statistical Association **77**(379), 605–610 (1982)
26. Hasegawa, M., Kishino, H., Yano, T.-a.: Dating of the human-ape splitting by a molecular clock of mitochondrial dna. Journal of molecular evolution **22**(2), 160–174 (1985)
27. Finstermeier, K., Zinner, D., Brameier, M., Meyer, M., Kreuz, E., Hofreiter, M., Roos, C.: A mitogenomic phylogeny of living primates. PloS one **8**(7), 69504 (2013)
28. BEAST2 Data Sets. https://github.com/CompEvol/beast2/tree/master/examples/nexus
29. 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)
30. TreeStat2. https://github.com/alexeid/TreeStat2
31. 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)
32. Guindon, S.: Bayesian estimation of divergence times from large sequence alignments. Molecular Biology and Evolution **27**(8), 1768–1781 (2010)
33. To, T.-H., Jung, M., Lycett, S., Gascuel, O.: Fast dating using least-squares criteria and algorithms. Systematic biology **65**(1), 82–97 (2015)
34. Sagulenko, P., Puller, V., Neher, R.A.: Treetime: Maximum-likelihood phylodynamic analysis. Virus evolution **4**(1), 042 (2018)
35. 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)
36. PhyML3.0: New Algorithms, Methods and Utilities. http://www.atgc-montpellier.fr/phyml/
37. 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)
38. Beal, M.J.: Variational Algorithms for Approximate Bayesian Inference, p. 281. university of London London, England (2003)
39. Zhang, C., IV, F.A.M.: Variational bayesian phylogenetic inference. In: International Conference on Learning Representations (2019). https://openreview.net/forum?id=SJVmjjR9FX
40. Dang, T., Kishino, H.: Stochastic variational inference for bayesian phylogenetics: A case of cat model. Molecular biology and evolution **36**(4), 825–833 (2019)
41. TreeTraceAnalysis. https://github.com/CompEvol/beast2/blob/master/src/beast/evolution/tree/TreeTraceAnalysis.java
42. TreeAnnotator. https://beast2.blogs.auckland.ac.nz/treeannotator/

43. Peskun, P.H.: Optimum monte-carlo sampling using markov chains. Biometrika **60**(3), 607–612 (1973)
44. Pybus, O.G., Rambaut, A.: Genie: estimating demographic history from molecular phylogenies. Bioinformatics **18**(10), 1404–1405 (2002)
45. 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
46. RateAgeBetaShift. https://github.com/revbayes/revbayes/blob/master/src/core/moves/compound/RateAgeBetaShift.cpp

Figure Legends

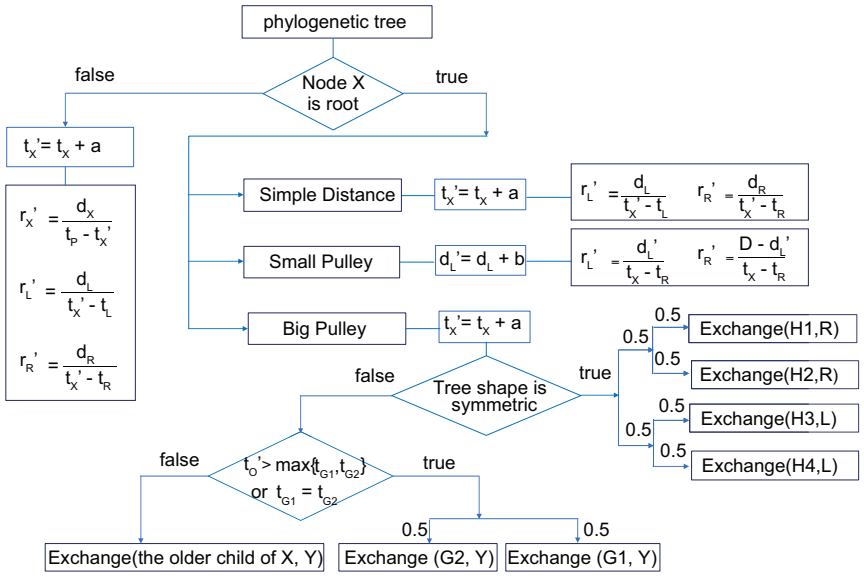


Figure 1 The flow chart of the Constant Distance operator.

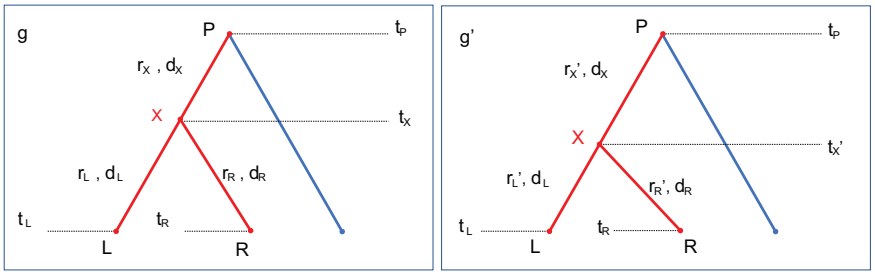


Figure 2 Illustration of the operation on an internal node. The operator proposes t_X', r_X', r_L' and r_R', during which d_L, d_R, d_X are kept constant.

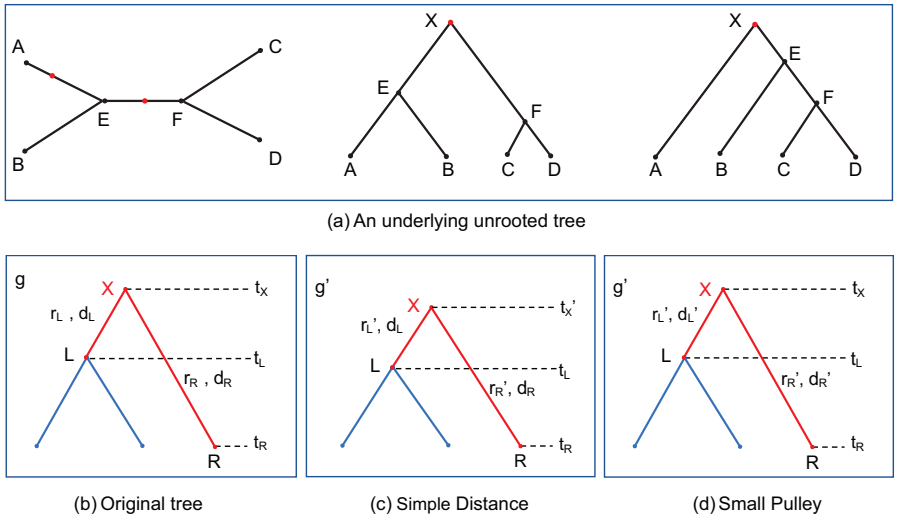


Figure 3 Illustration of operations on root. Subfigure (a) shows an example of a 4-taxa unrooted tree and two possible rooted tree for the operator to sample, during which the unrooted tree can not be changed. Based on the original tree in subfigure (b), Simple Distance proposes a node time and two rates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

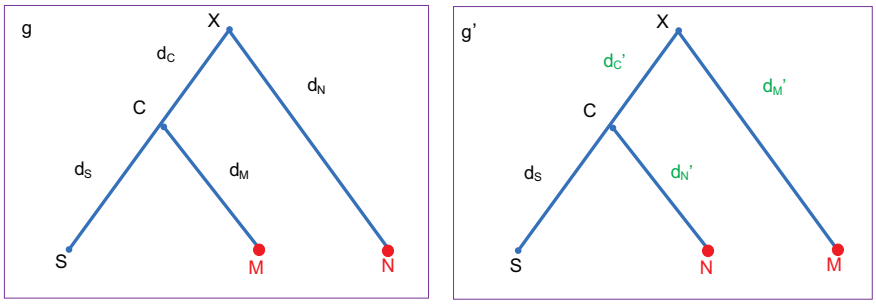


Figure 4 Illustration of Exchange (M,N) method. This method is applied to tree g and proposes g' by swapping **M** and **N**, so that the three distances are adjusted to maintain the distances among **S**, **M** and **N**. That is, $d_C{}' = d_C + b$, $d_N{}' = d_C + d_N$ and $d_M{}' = d_M - d_C{}'$,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

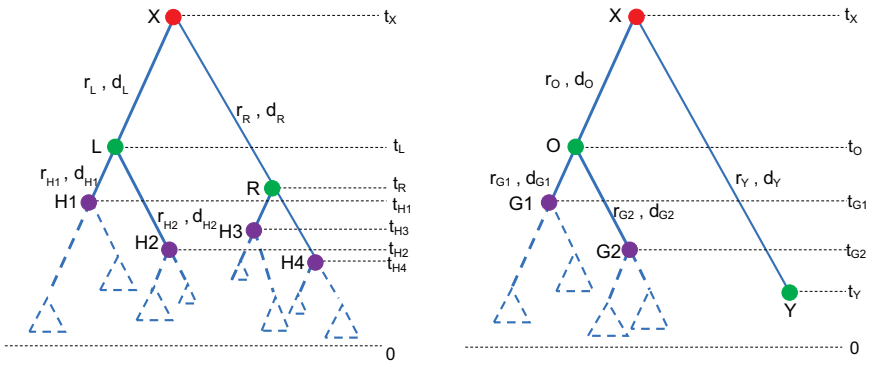


Figure 5 Two different tree shapes. The symmetric tree is on the left and the asymmetric tree is on the right. The dashed triangles represent the potential subtrees rooted at the nodes.

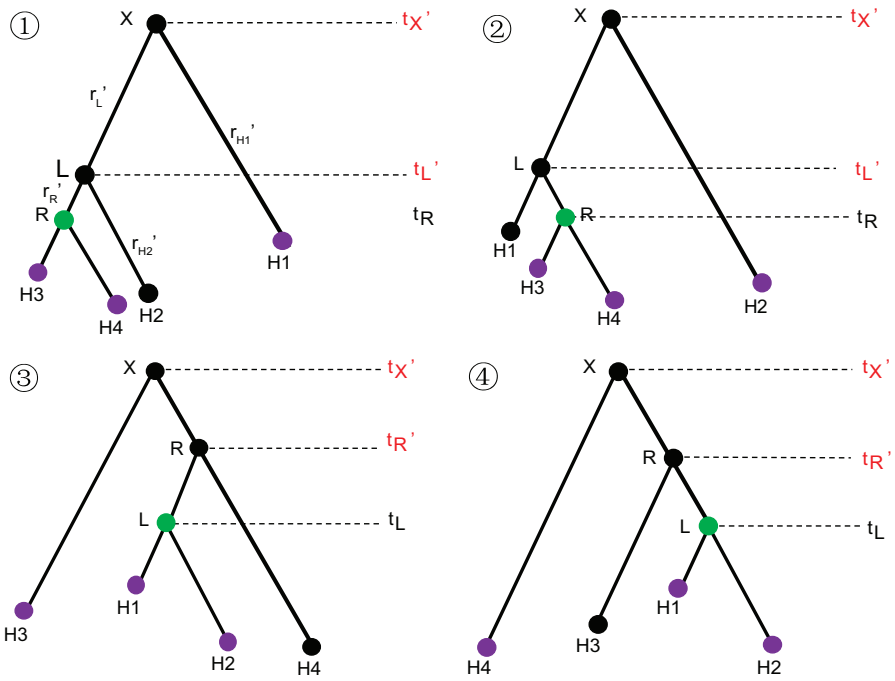


Figure 6 Illustration of operations on the symmetric tree in Figure 5. The proposed operator will propose one of the four possible trees, each with 0.25 probability.

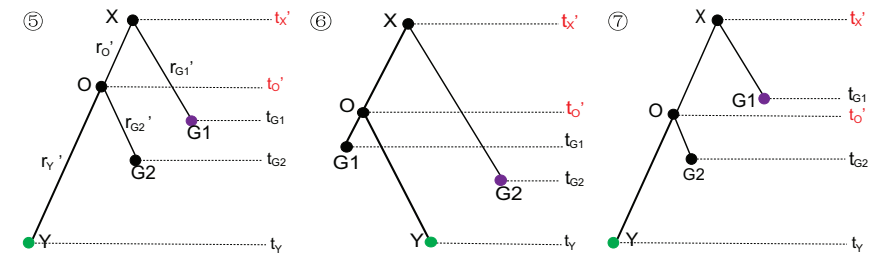


Figure 7 Illustration of operations on the asymmetric tree in Figure 5. The proposed operator will propose one of the three possible trees. If $t_o' < t_{G1}$, ⑦ has 1 probability, otherwise ⑤ and ⑥ have 0.5 probability each.

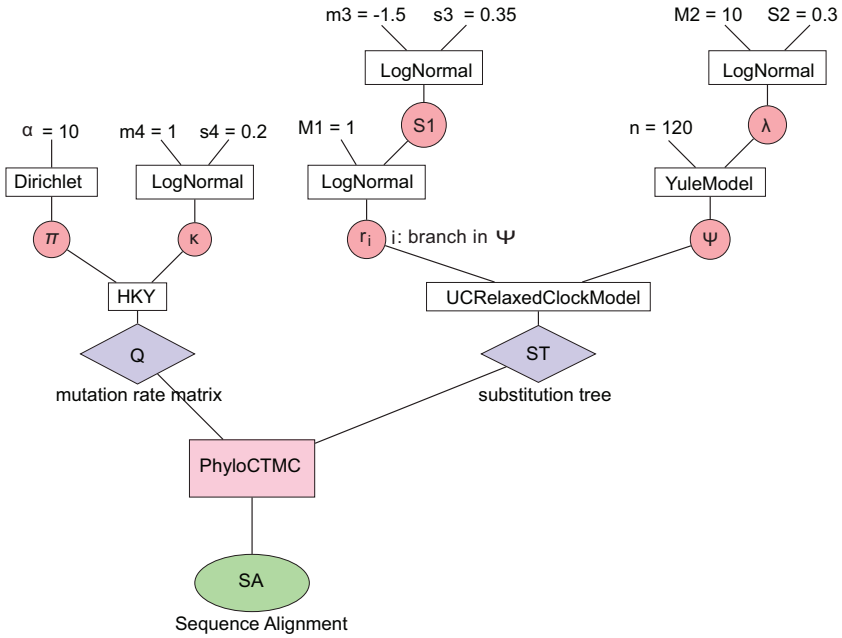


Figure 8 The models and prior distributions to simulate the sequence data. The sequence alignment (SA) is simulated through a phylogenetic continuous-time Markov Chain (PhyloCTMC) that consists of a substitution model (HKY) and an uncorrelated relaxed clock model (UCRelaxedClockModel). The random variables in HKY model construct the mutation rate matrix (Q), including base frequencies (π) 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 S1, abbreviated to UcldStdev). 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 of π, a LogNormal of κ, and a LogNormal of S1. For notations in LogNormal distributions, the uppercase letters represent the parameters in real space, and the lowercase letters represent the parameters in log space. In all the simulations, the number of taxa is fixed at 120 (n $= 120$).

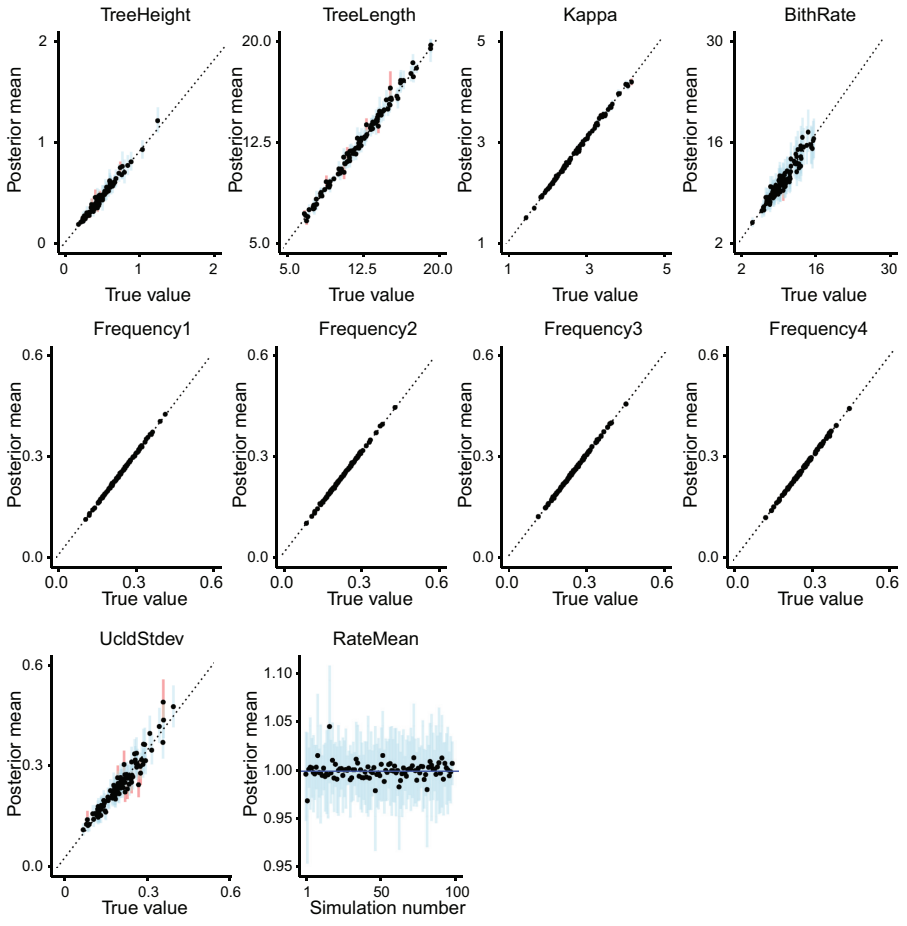


Figure 9 Comparing the sampled parameters in simulation study with 120 taxa.

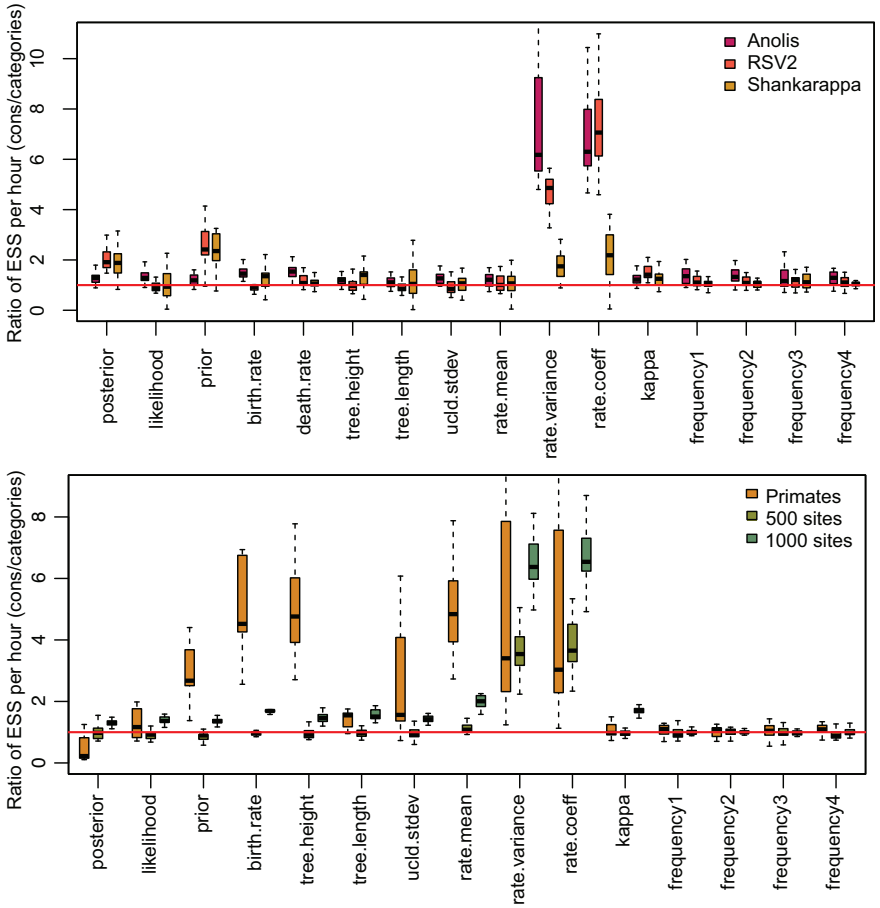
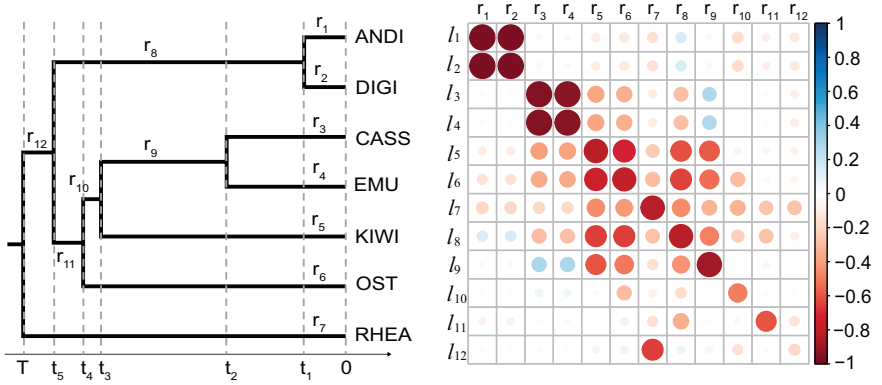


Figure 10 Comparison of ESS and running time. There are 6 data sets analysed, including 4 real data sets and 2 simulated data sets with different number of sites, as is shown in the legend. The red line represent the position where the ratio of ESS per hour is equal to 1. The horizontal axis represents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



(a) Maximum clade credibility tree

(b) Rates and branch lengths pairwise comparison

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 Pearson coefficients are calculated by converting the rates and branch lengths into log space, so that the coefficients range from -1 to 1, as is shown in the bar. Blue indicates positive correlations and red indicates negative correlations. The darker the colour is, the stronger the correlation tends to be.

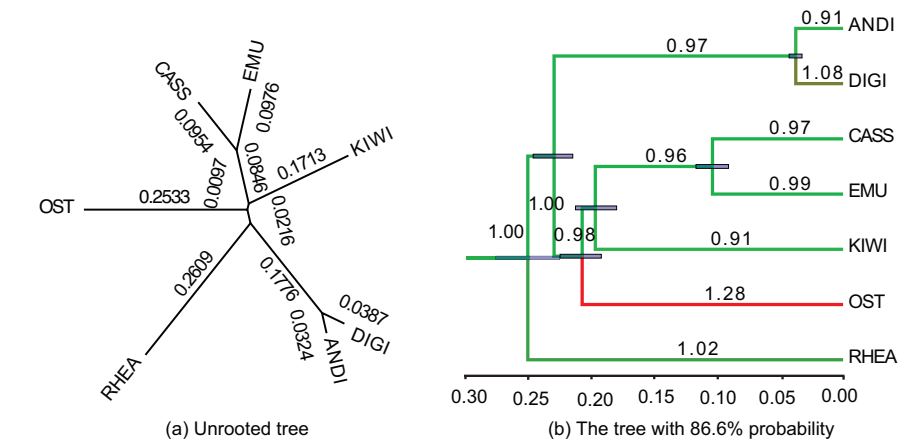


Figure 12 Illustration of sampling a fixed unrooted tree. In subfigure (a), the unrooted tree is obtained from the ratites data set [29] by a maximum likelihood method [36] and the labeled numbers represent genetic distances. The three unique tree topologies in subfigures (b) (c) and (d) are obtained from the sampled trees by using programme TreeTraceAnalysis [41]. The branch rates and node times are summarised by using programme TreeAnnotator [42]. 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.

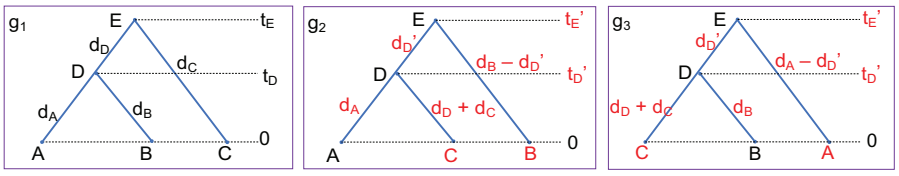


Figure 13 The illustration of sampling from prior. g_1 is set to be the original tree where an MCMC chain starts. When testing Big Pulley, the proposed operator samples the trees among g_1, g_2 and g_3.

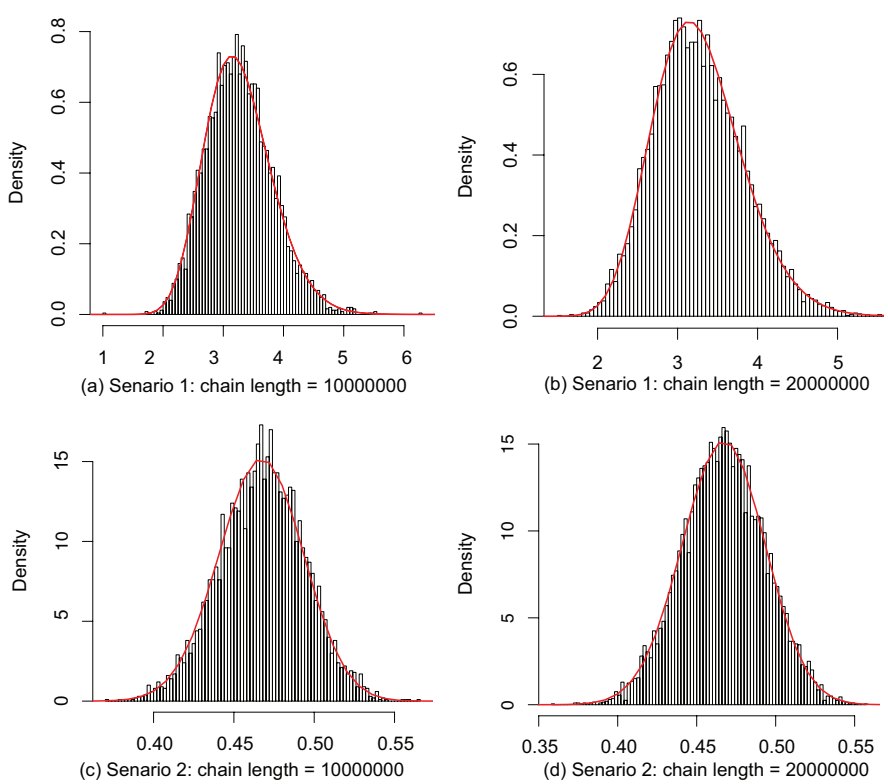


Figure 14 Sampled parameters in tests of internal nodes. The horizontal axis represents the node time of D in Figure 13. The two scenarios sample two trees with different distances specified in Table 3.

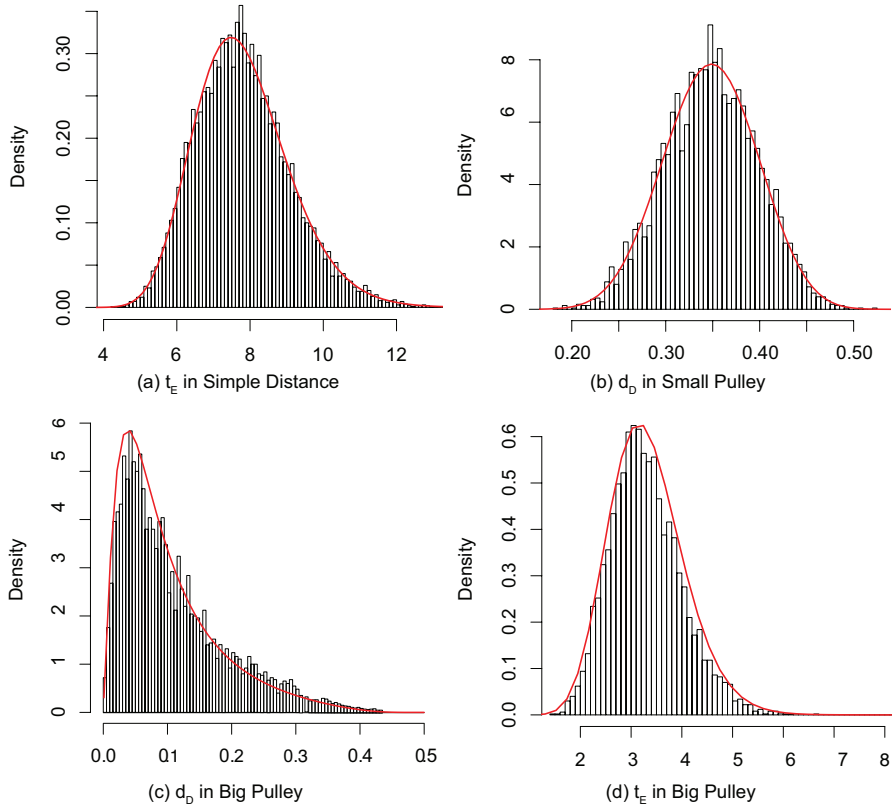


Figure 15 Sampled parameters in test of the root. For the trees in Figure 13, Simple Distance samples the root time t_E only, Small Pulley samples the distance d_D only, and Big Pulley samples t_E, t_D, d_D. To make it simple, t_E and d_D are compared.

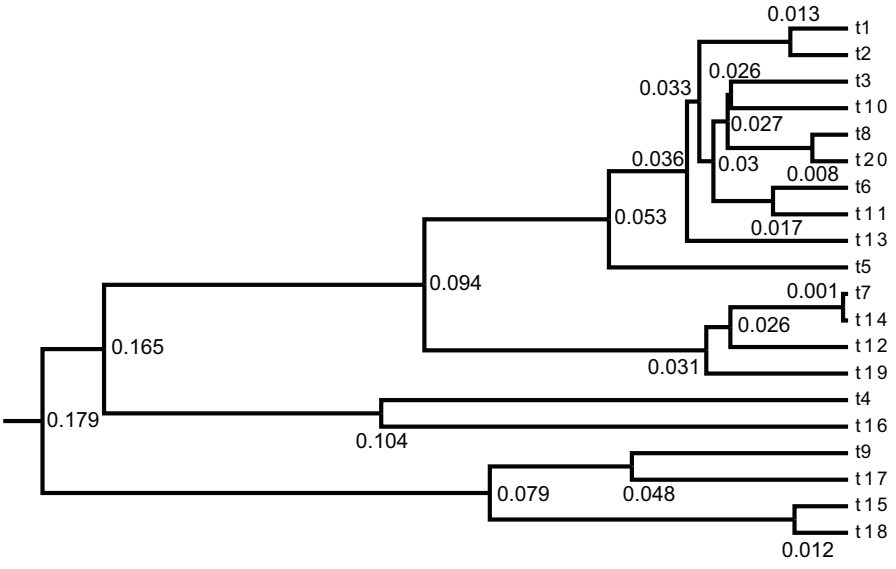


Figure 16 The tree used to simulate sequence alignment. The taxa are denoted by t1 to t20. The divergence times are drawn near the node.

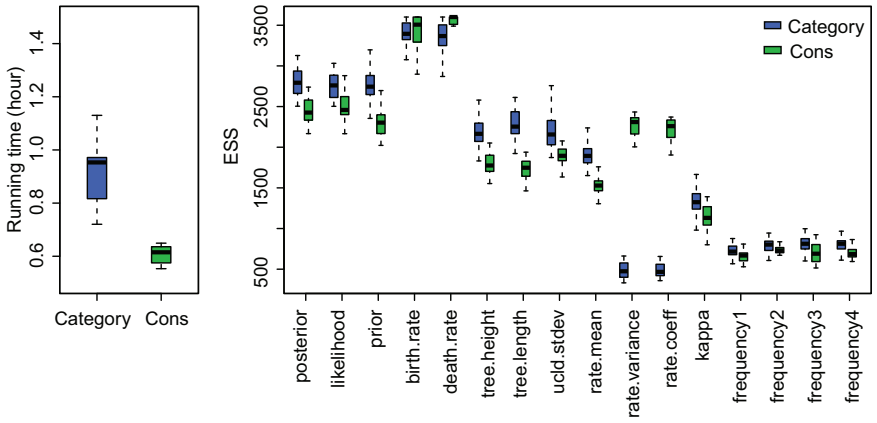


Figure 17 Running time and ESS using anolis data.

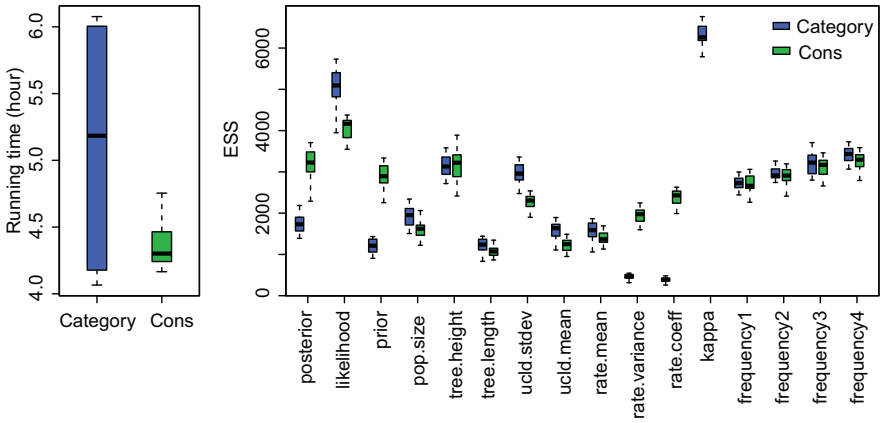


Figure 18 Running time and ESS using RSV2 data.

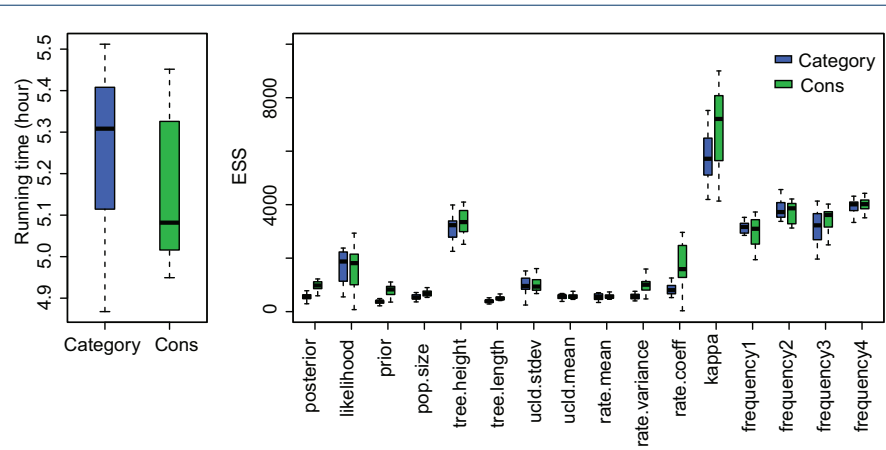


Figure 19 Running time and ESS using shankarappa data.

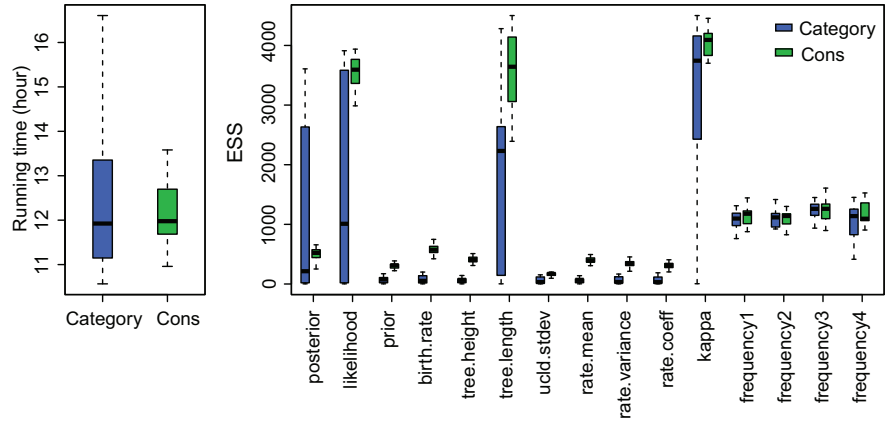


Figure 20 Running time and ESS using primates data.

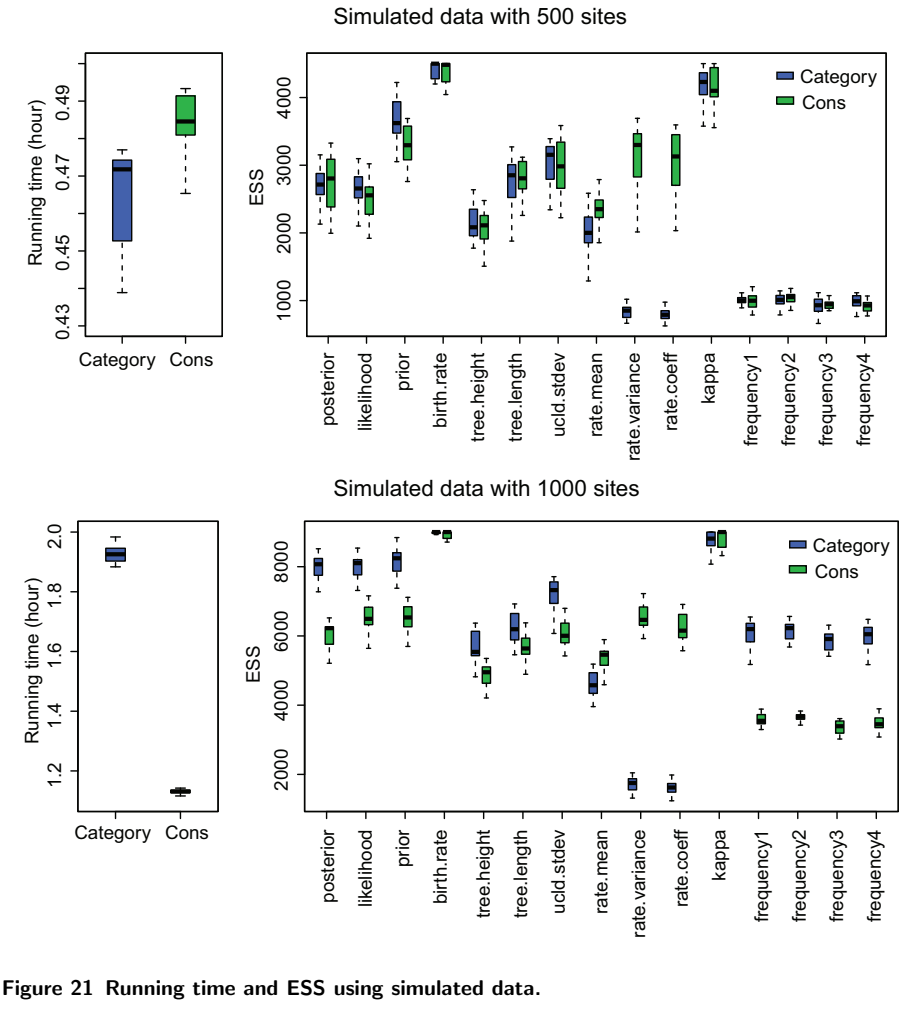


Figure 21 Running time and ESS using simulated data.

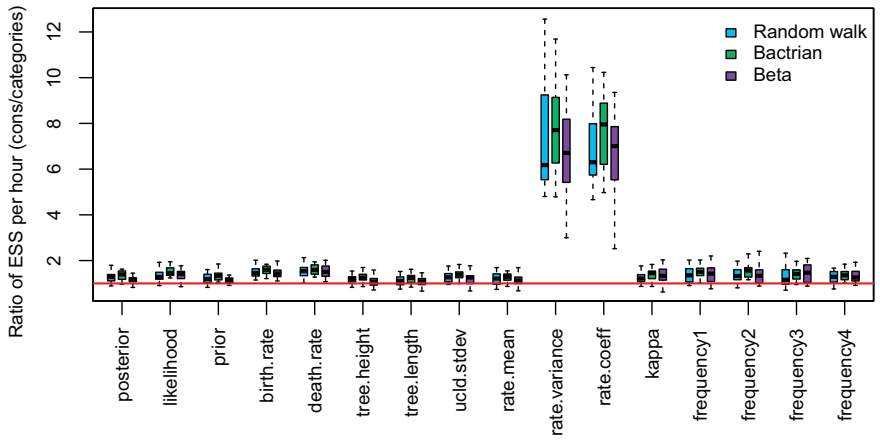


Figure 22 Efficiency comparison of proposals using Anolis data.

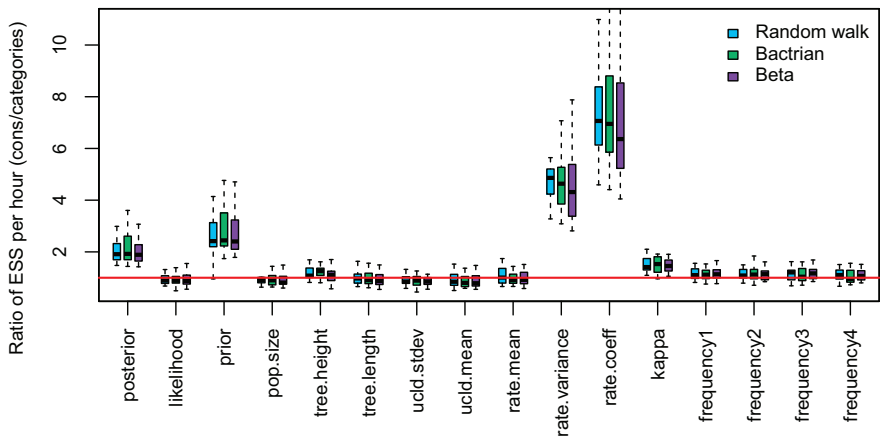


Figure 23 Efficiency comparison of proposals using RSV2 data.

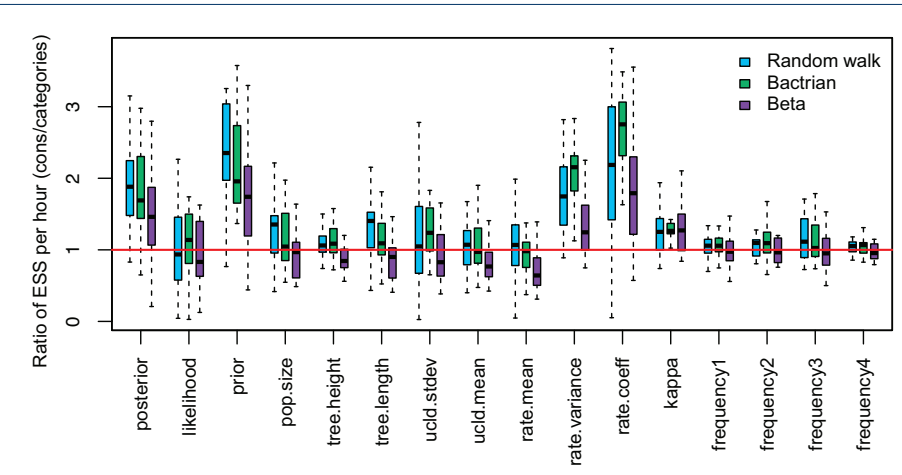


Figure 24 Efficiency comparison of proposals using Shankarappa data.

Tables

Parameters	Coverage	Parameters	Coverage
TreeHeight	89	UcldStdev	91
TreeLength	91	Frequency1	94
Kappa	97	Frequency2	96
BirthRate	99	Frequency3	95
RateMean	100	Frequency4	97

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

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

Table 2 Summary of ESS and running time

	genetic distances (fixed)				t_D	t_E	initial rates			
	d_j	d_k	d_x	d_i	initial	(fixed)	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	StdEv	Mean	Err	StdEv	
Senario 1	10000000	3.2727	8.3e-3	0.5467	3.2669	1.3e-06	0.5553	Figure 14(a)
	20000000	3.271	6.1e-3	0.5616				Figure 14(b)
Senario 2	10000000	0.4677	3.9e-04	0.0265	0.4667	3.5e-05	0.0262	Figure 14(c)
	20000000	0.4672	2.8e-04	0.0262				Figure 14(d)

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
		Mean	StdEv	Mean	StdEv	
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		Anolis		RSV2		Shankarappa		Primates	
		Cons	Category	Cons	Category	Cons	Category	Cons	Category	Cons	Category
rates times	ConstantDistance Operator	0.1919	-	0.2248	-	0.2228	-	0.2228	-	0.1850	-
	Rate Normal Operators[1]	0.1919	0.2879	0.1349	0.2698	0.1337	0.2674	0.1337	0.2674	0.1850	0.2775
	UcldStdev Scale Operator[2]	0.0288	0.0288	0.0270	0.0270	0.0267	0.0267	0.0267	0.0267	0.0278	0.0278
	UcldMean Scale Operator	-	-	-	-	0.0357	0.0089	0.0089	0.0089	-	-
	UcldMean Tree UpperDown Operator	-	-	-	-	0.0446	0.0267	0.0267	0.0267	-	-
	InternalNodeTime Scale Operator	0.0480	0.0960	0.0270	0.0270	0.0267	0.0267	0.0267	0.0267	0.0463	0.0463
	RootAge Scale Operator	0.0480	0.0480	0.0270	0.0270	0.0267	0.0267	0.0267	0.0267	0.0463	0.0463
	AllNodeTimes Uniform Operator	0.0480	0.0960	0.1799	0.2698	0.1337	0.2674	0.1783	0.2674	0.1850	0.2775
Tree	SubtreeSlide Operator	0.1440	0.1440	0.0989	0.1349	0.1070	0.1337	0.1159	0.1337	0.1388	0.1388
	NarrowExchange Operator	0.1440	0.1440	0.0989	0.1349	0.1070	0.1337	0.1159	0.1337	0.0463	0.0463
	WideExchange Operator	0.0480	0.0480	0.0270	0.0270	0.0267	0.0267	0.0267	0.0267	0.0463	0.0463
	WilsonBalding Operator	0.0480	0.0480	0.0270	0.0270	0.0267	0.0267	0.0267	0.0267	0.0463	0.0463
	BirthRate Scale Operator	0.0480	0.0480	0.0629	0.0270	-	-	-	-	0.0278	0.0278
	DeathRate Scale Operator	-	-	0.0629	0.0270	-	-	-	-	-	-
	PopulationSize Scale Operator	-	-	-	-	0.0802	0.0267	0.0624	0.0267	-	-
Substituion model	Kappa Scale Operator	0.0096	0.0096	0.0009	0.0009	0.0009	0.0009	0.0009	0.0009	0.0185	0.0185
	Frequencies DeltaExchange Operator	0.0019	0.0019	0.0009	0.0009	0.0009	0.0009	0.0009	0.0009	0.0009	0.0009

Note
1: Random walk operator and Swap operator in Cons configuration, Random walk operator, Scale operator and Swap operator in Category configuration.
2: The operator introduced in Appendix section 4 is used in Cons configuration, a Scale operator is used in Category configuration .
-: The parameter is not sampled and no operator is assigned.

Appendix

1. The Green ratio

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

Considering the subspace φ_1 on x and subspace φ_2 on x', it is assumed that there is a symmetric measure on the 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 φ. Then, Green suggested that the reversibility condition should be satisfied by detailed balance [19], as represented by Equation (17). And according to Peskun' proof, it is optimal to take Equation (18) as the acceptance probability to retain the detailed balance [43].

$$\int_A \pi(x)d_x \int_B q(x, x')\alpha(x, x')d_x = \int_B \pi(x')d_{x'} \int_A q(x', x)\alpha(x', x)d_{x'}, \tag{17}$$

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 state x' given the current state x.

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

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

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

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

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

1.1 Calculating the Green ratio for operations on internal nodes

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

$$f_1 : \mathrm{t}_X{}' = \mathrm{t}_X + \mathrm{a} \tag{20a}$$

$$f_2 : r_X{}' = \frac{r_X \times (t_P - t_X)}{t_P - t_X{}'} \tag{20b}$$

$$f_3 : r_L{}' = \frac{r_L \times (t_X - t_L)}{t_X{}' - t_L} \tag{20c}$$

$$f_4 : r_R{}' = \frac{r_R \times (t_X - t_2)}{t_X{}' - t_R} \tag{20d}$$

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

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

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

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).

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

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

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

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

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 $|\mathbf{J}_3| = 1$.

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

1.3 Calculating the Green ratio for Big Pulley

Two new node times are proposed in Big Pulley. One is the root time (Equation (28a)), the other is the node time 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 node time proposed, as is seen in Equation (28b). In addition, the distances are adjusted by the method *Exchange (***M**, **N***)*, dependent on which nodes are chosen. As a result, the four rates are proposed, as is shown in Equation (28c)∼Equation (28f)

$$ {t_X}' = t_X + a \tag{28a} $$

$$ {t_C}' = t_C + a_{1,2,3} \tag{28b} $$

$$ {r_C}' = \frac{r_C \times (t_X - t_C) + b}{t'_X - {t_C}'} \tag{28c} $$

$$ {r_S}' = \frac{r_2 \times (t_C - t_S)}{t'_C - t_S} \tag{28d} $$

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

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

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

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

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

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

Algorithm 1 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

2. Sampling from the prior

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

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

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

In addition, a Coalescent model [44] with constant population size ($N = 0.3$) is used to describe the tree prior. Hence, for the tree in Figure 13, the probability of node times is calculated by Equation (32).

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

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

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

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

2.1 Test the operator on internal nodes

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

move between node A and E. Besides, to make sure that the result is robust, two different MCMC chain lengths are performed in each scenario, i.e. 10 million and 20 million.
The mean, mean error and the standard deviation of the MCMC samples are summarised in Table 4. Besides, according to Equation (33), the actual joint distribution is obtained by using Equation (34), and is used to evaluate the results, which is also included in Table 4. Moreover, the histograms of MCMC samples that indicate the sampled distributions, as well as the curves of the numerical integration of Equation (34), are shown in Figure 14. From Table 4 and Figure 14, it can be seen that the red curves well fit the black histograms, and the mean values and standard deviations are consistent, which makes it safe to conclude that the proposed operator samples the internal node correctly.

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

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).

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

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

2.2.2 Using Small Pulley Although both d_x and d_i are changed during the sampling process when using Small Pulley, the sum of d_D and d_C are kept 0.67 in this test, as the initial setting shown in Table 5. To make it simple, only d_D is compared.
Then, based on Equation (33), the exact distribution of d_i can be obtained by Equation (36), which is compared with the sampled distribution in Table 6 and Figure 15(b). Even though there exist some errors, the sampled parameters can be considered to follow the same distribution. So the Small Pulley is also able to provide correct samples.

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

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).

$$\begin{aligned} p(r,t) = & \int_{t_E=0}^{+\infty} \int_{t_D=0}^{t_E} \int_{d_D=0}^{0.5} p(t_E,t_D) \times p(\frac{0.5}{t_D}) \\ & \times p(\frac{0.5}{t_D}) \times p(\frac{d_D}{t_E - t_D}) \times p(\frac{0.5 - d_D}{t_E}) d_{d_D} d_{t_D} d_{t_E} \end{aligned} \tag{37}$$

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 [45] and a Beta proposal [46], assign a specific distribution on the new node time so that it is more probable to move to a certain height on the branch, either far away from or close to its original position. This section applied Random walk proposal (the operators in this paper), Bactrian proposal and Beta proposal to the three data sets, and the results are compared to those using Category configuration.
The comparisons are shown in Figure 22, Figure 23 and Figure 24. It is indicated that Beta proposal achieved worst performance in the three analysed data sets. The performance of the Constant Distance operator (Random walk) and Bactrian proposal varies depending on different data sets. However, these two proposal methods are both more efficient that the Category configuration. Therefore, it still needs further investigation to demonstrate the effectiveness of different proposals.

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

$$Ucldstdev' = Ucldstdev \times \text{scale} \tag{38}$$

where $\text{scale} = \text{Factor} + \left[\xi \times (\frac{1}{\text{Factor}} - \text{Factor})\right]$ and ξ is a random variable from a $Uniform(0, 1)$, Factor is a user-defined parameter to specify how bold the proposal is.
Secondly, all the branch rates are proposed based on the new $Ucldstdev'$, given the probability of original $Ucldstdev$, which is calculated using Equation (39).

$$r_i' = icdf_{stdev'}\left[cdf_{stdev}(r_i)\right] \tag{39}$$

where the notations $cdf(\cdot)$ and $icdf(\cdot)$ represent the cumulative and inverse cumulative density function of log normal distribution.
Finally, it is important to return the corrected hastings ratio, since the proposal is associated with one random variable, $Ucldstdev$ and $(2n - 1)$ branch rates. As is shown in Equation (40), the ratio includes the scale operation and rates changing under the same probability.

$$\mathbf{J}_{Ucldstdev} = \frac{1}{scale} \times \prod_{i=1}^{2n-1} \frac{\partial icdf_{Ucldstdev'}[cdf_{Ucldstdev}(r_i)]}{\partial r_i} \tag{40}$$