RESEARCH

Improving the performance of Bayesian phylogenetic inference under relaxed clock models

Rong Zhang and Alexei Drummond*

*Correspondence:
alexei@cs.auckland.ac.nz
School of Computer Science,
University of Auckland, Princes
Street, 1010 Auckland, New
Zealand
Full list of author information is
available at the end of the article

Abstract

Bayesian MCMC has become a common approach for phylogenetic inference. This paper develops a new operator to improve the efficiency of Bayesian phylogenetic inference for models that include a per-branch rate parameter. In an MCMC algorithm, the presented operator 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. It is noticed 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. 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; Operator; Genetic distances; Divergence times; Evolutionary rates

Introduction

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 [3] 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]. By allowing rates to vary across lineages, it is considered that better estimates of divergence times can be obtained [10, 11, 12].

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].

Prelimiaries

Bayesian MCMC

Let D denote the data, and let g and Φ denote the phylogenetic tree 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 $p(g)$, $p(\Phi)$, $\Pr(D|g,\Phi)$, $p(g,\Phi|D)$ respectively. From a Bayesian perspective, the phylogenetic trees and the parameters are random variables described by a posterior probability distribution given the observed data D.

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

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 $p(g,\Phi|D)$, in such a way that the computation of the marginal likelihood $p(D)$ 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, 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].

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, the proposed operator is based on an uncorrelated relaxed clock model, in which the rates vary from branch to branch, 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 r and divergence times t can be obtained by the conditional distribution in Equation (2):

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

where $p(r|\Phi)$ is the prior for rates specified in uncorrelated relaxed clock model. The proposed operator is to sample the state (t, r, Φ) in the constructed Markov chain.

Methodology: the proposed operator

In this section, we define the Constant Distance Operator. Figure 1 illustrates the flow chart of the proposed operator. In a phylogenetic tree, the node to operate on is denoted by **X**. The proposed 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 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 $\mathbf{X}$ 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.

Simple Distance

Figure 3 shows the trees that are rooted at the node $\mathbf{X}$. The original tree in the current state is shown in Figure 3(a), which is denoted by g. Similar to the operations on internal nodes, we will use the following steps to propose a new tree g', while keeping the genetic distances of two branches linked to the root d_L and d_R constant at the same time. These steps are illustrated in Figure 3(b).

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.

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

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

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

Step 4 Return the Green ratio α_{SD}.

Small Pulley

In contrast to Simple Distance, Small Pulley proposes a new genetic distance of a branch on one side of the root. As is illustrated in Figure 3, a new tree g' is proposed based on the original tree g. In order to maintain the total genetic distance d_L and d_R of the two branches linked to the root, after ${d_L}'$ is proposed, d_R will be adjusted simultaneously. The detailed process includes the following 4 steps.

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. The implied genetic distances of the two branches linked to the root can be calculated by:

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

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, where $D = d_L + d_R$. Otherwise, we reject the proposal.

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

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

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.

Firstly, a method called *Exchange* is designed to propose a new tree topology. Let **R** denote the root of tree g, let **C** and **N** denote the two child nodes of **R**, and let **S** and **M** denote the two child nodes of **C** (Figure 4). The following operations will be performed to propose a new tree g'.

- 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.

Secondly, before applying this method in Big Pulley, there are two different tree shapes to take into consideration. In Figure 5, a symmetric tree is shown on the left, in which both the child nodes of the root have child nodes. But in the asymmetric tree on the right, only one of the child nodes of the root has child nodes below it, and the other child node is a leaf node. 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 using **R** and either child node of **L**.

- 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:

$$ {d_1}' = d_1 - {d_L}' \qquad {d_2}' = d_2 \qquad {d_L}' = d_L + b \qquad {d_R}' = d_L + d_R \tag{8} $$

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

$$ {r_1}' = \frac{{d_1}'}{{t_X}' - t_1} \qquad {r_2}' = \frac{{d_2}'}{{t_j}' - t_2} \qquad {r_L}' = \frac{{d_L}'}{{t_X}' - {t_L}'} \qquad {r_R}' = \frac{{d_R}'}{{t_L}' - t_R} \tag{9} $$

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 older child of the root **X**, denoted by **O**, and the younger child of the root is denoted by **Y**. The node times of the root **X**, **O** and its child nodes are denoted by t_X, 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 ${t_O}' < {t_X}'$ 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).

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:

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

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 ‘The Green ratio’ in the Appendix section). 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 the Appendix.

Experimental results and analysis

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 subsection Sampling from the prior in Appendix). 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 11 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 $S1$, speciation rate λ, base frequencies π and transition-transversion bias κ. Second, we simulated nucleotide alignments using the simulated parameters. To make this study more robust, two groups of sequence data were simulated, one group with $n = 20$ taxa and the other with $n = 120$ taxa. Each group contained 100 simulated sequence alignments with 10,000 sites. Third, we used BEAST2 with Constant

Distance operator to infer the tree and parameters in each of the 200 simulated 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 of the two groups are shown in Figures 12 and 13.

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 5) for both the small ($n = 20$) and the large ($n = 120$) trees. 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 simulation, and ii) the effective sample size (ESS) of $S3$ - the standard deviation of the relaxed clock. 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 three tree operator configurations. 1) Using the current operators in BEAST2 to sample discrete rate categories (categories). 2) Using the ConstantDistance operator to sample continuous rates specified by an uncorrelated related clock model (cons). 3) Using standard BEAST2 operators and the same clock model as cons (nocons). This analysis was performed 100 times on each dataset for each configuration, with the prior distributions and all other model specifications held constant.

Performance on simulated data

The framework described in Figure 11 was used to simulate two groups of sequence alignment: one with 20 taxa and one with 120 taxa. In each group, three data sets with different sequence lengths were simulated: containing 5,000 sites (short), 10,000 sites (medium), and 20,000 sites (long). Each of the 6 data sets were run 100 times in BEAST2, on each of the 3 configurations.

The ESS and running time are summarised in Figure 14 and Table 6. These results show that for all 6 simulated datasets, our proposed operator achieved a larger ESS of $S3$ per chain, compared with the other two configurations. Furthermore, the total running time under cons is on a similar timescale to the other configurations.

We calculated the ESS per hour for each configuration, averaged across the 6 simulated datasets (Table 6). These results show that our proposed operator achieved an ESS per hour 21.4 times larger than categories and 4.1 times larger than nocons.

Performance on real data

We used a real data set with 83 primates and 1234 sites [27] to further evaluate the performance of Constant Distance operator. Similar to the process of simulated data sets, this analysis was performed in three configurations, i.e. categories, cons and nocons. Each configuration was run 100 times on this dataset.

These results (Figure 15) show that, on average, the configuration using Constant Distance operator achieves a larger ESS of S3 and in less running time. Overall, Constant Distance operator improved the performance by providing 18.3 and 10.8 times larger ESS per hour, compared with categories and nocons respectively.

Correlation analysis of rates and node times

After analysing the ratites data set (with 7 taxa and 10767 sites [28]) in BEAST2 using the Constant Distance operator, the sampled trees in the output tree file are filtered by the shared common ancestor of each taxa, which is implemented in a program called TreeStat2 [29]. Then, we conducted a pairwise comparison between each rate and node time in the filtered trees in order to see how they are correlated.

The results are shown in Figure 16. The tree in Figure 16(a) is used to filter the sampled trees to ensure that all trees examined have the same topology. Afterwards, a pairwise comparison was made by using 12 branch rates and 5 node times. As can be seen from upper right of Figure 16(b), to a large degree, the rates are negatively correlated with the node times. This indicates that when a larger node time is proposed, the corresponding rates become smaller. For example, $t1$ has a negative correlation with $r1$ and $r2$, but has positive correlation with $r8$. That is to say, with the increase of $t1$, the rate $r8$ will also increase, but $r1$ and $r2$ will decrease. In addition, in the upper left of Figure 16(b), most rates on two adjacent branches have positive correlations, but have a negative correlation with the rate above them. Take $r6$ and $r7$ as an example. They have a positive correlation with each other, but are negatively related to $r12$. This means a larger $r6$ incurs a larger $r7$ and a smaller $r12$. To sum up, this dynamic change of rates and node times 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. 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, becomes known without error. With full length genomes now available, this limiting case might be approached in some data sets. 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.

We investigated this approach on a fixed substitution tree reconstructed from whole mitochondrial genomes from a set of ratite species. Since no uncertainty is admitted in 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 [30, 31]. Figure 17(a) shows the unrooted tree with the genetic distances on the branches which are fixed in a subsequent relaxed clock analysis in BEAST2.

As an initial starting point, the midpoint method is adopted to root the starting time tree. After that, based on the genetic distances among seven taxa and the topology of the rooted tree, consistent divergence times are 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 determined so that the products match the unrooted substitution tree.

The resulting summary tree (Figure 17(b)) is consistent with previous analyses of this data [28] (Figure 17(c)). For large data sets of long sequences, this method 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.

Conclusion

As data sets continue to increase in size, the need for computational efficiency of Bayesian phylogenetic analyses is also increasing. 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 in the core code of BEAST2.

Competing interests

The authors declare that they have no competing interests.

Acknowledgements

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

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. In: Evolving Genes and Proteins, pp. 97–166. Elsevier, ??? (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. Ho, S.Y., Phillips, M.J., Drummond, A.J., Cooper, A.: Accuracy of rate estimation using relaxed-clock models with a critical focus on the early metazoan radiation. Molecular Biology and Evolution **22**(5), 1355–1363 (2005)

11. Renner, S.S.: Relaxed molecular clocks for dating historical plant dispersal events. Trends in plant science **10**(11), 550–558 (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)
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. 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)
29. TreeStat2. https://github.com/alexeid/TreeStat2
30. PhyML3.0: New Algorithms, Methods and Utilities. http://www.atgc-montpellier.fr/phyml/
31. 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)
32. Peskun, P.H.: Optimum monte-carlo sampling using markov chains. Biometrika **60**(3), 607–612 (1973)
33. Pybus, O.G., Rambaut, A.: Genie: estimating demographic history from molecular phylogenies. Bioinformatics **18**(10), 1404–1405 (2002)

Figures

[width=12cm]flowchart.eps

Figure 1 The flow chart of the Constant Distance operator.

[width=12cm]internalnodes.eps

Figure 2 Illustration of the operation on an internal node. The operator proposes a tree ${g_{in}}'$ based on tree g_{in}, during which d_i, d_j, d_k are kept constant.

[width=12cm]rootstrategy.eps

Figure 3 Illustration of Simple Distance and Small Pulley sub-moves. Simple Distance proposes ${g_{r1}}'$ and keeps d_i, d_x constant. Small Pulley proposes ${g_{r2}}'$ and $d_i + d_x$ remains constant.

[width=12cm]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]treeshape.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]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]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]bigpulleyExp.eps

Figure 8 The illustration of sampling from prior. g_1 is set to be the original tree where an MCMC chain starts. When testing Big Pulley, the proposed operator samples the trees among g_1, g_2 and g_3.

[width=0.45]Fig1	[width=0.45]Fig2	[width=0.45]Fig7	[width=0.45]Fig8
(a) Senario 1: chain length = 10000000	(b) Senario 1: chain length = 20000000	(c) Senario 2: chain length = 10000000	(d) Senario 2: chain length = 20000000

Figure 9 Sampled parameters in tests of internal nodes. The horizontal axis represents the node time of D in Figure 8. The two scenarios sample two trees with different distances specified in Table 1.

[width=0.45h]RplotSD	[width=0.45]RplotSP2	[width=0.45]bigd	[width=0.45]bigT
(a) t_E in Simple Distance	(b) d_D in Small Pulley	(c) d_D in Big pulley	(d) t_E in Big pulley

Figure 10 Sampled parameters in test of the root. For the trees in Figure 8, Simple Distance samples the root time t_E only, Small Pulley samples the distance d_D only, and Big Pulley samples t_E, t_D, d_D. To make it simple, t_E and d_D are compared.

[width=12cm]ModelValidation.eps

Figure 11 The models and prior distributions to simulate the sequence data. The sequence alignment is simulated through a phylogenetic continuous-time Markov Chain that consists of a substitution model (HKY) and an uncorrelated relaxed clock model. The random variables in HKY model construct the mutation rate matrix, including frequency and ratio. The rates and time trees specified by a Yule model construct the substitution tree. The standard deviation of rates and the birth rate in Yule model are both random variables following LogNormal distributions.

[width=12cm]SmallTree.eps

Figure 12 Comparing the sampled parameters in simulation study with 20 taxa. The red crossings represent the coordinates of real and estimated value. The vertical lines with circles show the 95% HPD of the sampled parameters. The blue lines indicate that the estimated value is equal to the real value. The red line represents the fixed value of rate mean.

[width=12cm]LargeTree.eps

Figure 13 Comparing the sampled parameters in simulation study with 120 taxa.

[width=12cm]Efficiency.eps

Figure 14 Comparison of ESS and running time using simulated data. The term long, medium and short represent the length of sequence with 20 thousand, 10 thousand and 5 thousand respectively.

[width=0.45]primates2.eps [width=0.45]primates1.eps

Figure 15 Comparison of ESS and running time using primates data. The orange curve links the mean values of each boxplot and the values are shown on the left of the boxplot.

[width=0.5]correlationtree [width=0.4]rateandtime.eps

(a) Clades (b) Pairwise compare

Figure 16 Correlation between rates and node times in the ratites tree. The rates and node times are in correspondence with the notations in Figure 17(b). Blue indicates the positive relation and red indicates the negative relation. The darker the colour is, the stronger the relation tends to be.

[width=0.35]unrootedtree [width=0.4]initialtree [width=0.4]summary3 [width=0.45]ratites

(a) Unrooted tree (b) Rooted time tree (c) Summary tree of sampled trees (d) Ratite phylogeography in Ref. [28]

Figure 17 Illustration of sampling a fixed unrooted tree. The unrooted tree is obtained from the ratites data set. The rooted time tree is the initial state in MCMC, with time on the nodes and rates on the branches.

Tables

	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 1 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 9(a)
	20000000	3.271	6.1e-3	0.5616				Figure 9(b)
Senario 2	10000000	0.4677	3.9e-04	0.0265	0.4667	3.5e-05	0.0262	Figure 9(c)
	20000000	0.4672	2.8e-04	0.0262				Figure 9(d)

Table 2 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 3 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 10(a)
Small Pulley	d_i	0.3480	0.0492	0.3476	0.0494	Figure 10(b)
Big Pulley	d_i	0.1016	0.0766	0.0960	0.0760	Figure 10(c)
	t_E	3.3017	0.6908	3.3095	0.6912	Figure 10(d)

Table 4 Results of sampling the root

	BirthRate	TreeHeight	RateMean	UcldStdev	Kappa	Frequency
20 taxa	93	98	100	95	100	100
120 taxa	100	98	85	94	100	100

Table 5 Number of real values lying in the 95% HPD in Figure 12 and Figure 13

		ESS of of analysis		Running time(second)		ESS per hour	
	Length	20 taxa	120 taxa	20 taxa	120 taxa	20 taxa	120 taxa
categories	20000	13	4	18635	44155	2.47	0.36
	10000	58	6	8660	36406	24.11	0.62
	5000	171	8	4690	15956	131.09	1.89
Average		81	6	10662	32172	27.19	0.71
cons	20000	616	147	20207	35406	109.72	14.92
	10000	646	161	8967	25589	259.24	22.68
	5000	993	186	4581	12487	780.50	53.70
Average		752	165	11252	24494	240.24	24.21
nocons	20000	86	63	19344	38245	16.09	5.91
	10000	153	20	8361	30521	65.83	2.31
	5000	296	48	4499	12940	237.04	13.22
Average		179	43	10735	27236	59.87	5.72

Table 6 Summary of ESS and running time using simulated data

Appendix

0.1 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 the detailed balance [19], as is 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 [32].

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

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

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).

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

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

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 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 number of distances of both branches linked to the root is 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 $|\mathbf{J}_3| = 1$.

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

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).

$$\mathbf{J}_4 = \begin{bmatrix} \frac{\partial t_X{}'}{\partial t_X} & \frac{\partial t_X{}'}{\partial t_C} & \frac{\partial t_X{}'}{\partial r_C} & \frac{\partial t_X{}'}{\partial r_S} & \frac{\partial t_X{}'}{\partial r_M} & \frac{\partial t_X{}'}{\partial r_{N2}} \\ \frac{\partial t_C{}'}{\partial t_X} & \frac{\partial t_C{}'}{\partial t_C} & \frac{\partial t_C{}'}{\partial r_C} & \frac{\partial t_C{}'}{\partial r_S} & \frac{\partial t_C{}'}{\partial r_M} & \frac{\partial t_C{}'}{\partial r_{N2}} \\ \frac{\partial r_C{}'}{\partial t_X} & \frac{\partial r_C{}'}{\partial t_C} & \frac{\partial r_C{}'}{\partial r_C} & \frac{\partial r_C{}'}{\partial r_S} & \frac{\partial r_C{}'}{\partial r_M} & \frac{\partial r_C{}'}{\partial r_{N2}} \\ \frac{\partial r_S{}'}{\partial t_X} & \frac{\partial r_S{}'}{\partial t_C} & \frac{\partial r_S{}'}{\partial r_C} & \frac{\partial r_S{}'}{\partial r_S} & \frac{\partial r_S{}'}{\partial r_M} & \frac{\partial r_S{}'}{\partial r_{N2}} \\ \frac{\partial r_M{}'}{\partial t_X} & \frac{\partial r_M{}'}{\partial t_C} & \frac{\partial r_M{}'}{\partial r_C} & \frac{\partial r_M{}'}{\partial r_S} & \frac{\partial r_M{}'}{\partial r_M} & \frac{\partial r_M{}'}{\partial r_N} \\ \frac{\partial t_N{}'}{\partial t_X} & \frac{\partial t_N{}'}{\partial t_C} & \frac{\partial t_N{}'}{\partial r_C} & \frac{\partial t_N{}'}{\partial r_S} & \frac{\partial t_N{}'}{\partial r_N} & \frac{\partial t_N{}'}{\partial r_N} \end{bmatrix}$$
$$= \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ \frac{r_C}{t_X{}'-t_C{}'} & \frac{-r_C}{t_X{}'-t_C{}'} & \frac{t_X{}'-t_C}{t_X{}'-t_C{}'} & 0 & 0 & 0 \\ 0 & \frac{r_S}{t'-t_S} & 0 & \frac{t_C-t_S}{t_C{}'-t_S} & 0 & 0 \\ \frac{-r_C}{t_X{}'-t_M} & \frac{r_{N1}+r_C}{t_X{}'-t_M} & \frac{-(t_X-t_C)}{t_X{}'-t_M} & 0 & \frac{t_C-t_M}{t_X{}'-t_M} & 0 \\ \frac{r_C+r_S}{t_C{}'-t_N} & \frac{-(r_C+r_S)}{t_C{}'-t_N} & \frac{t_X-t_C}{t_C{}'-t_N} & 0 & 0 & \frac{t_X-t_N}{t_C{}'-t_N} \end{bmatrix} \tag{29}$$

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

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

```
if the node that has been exchanged with dau or dau 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 the node that has been exchanged with O has child nodes then
    $\gamma = 0.25$
else
    $\gamma = 0.5$
end if
for ① ② do
    Return $\mu = \frac{\alpha}{0.25}$
end for
for ③ ④ do
    Return $\mu = \frac{\beta}{0.25}$
end for
for ⑤ ⑥ do
    Return $\mu = \frac{\gamma}{0.5}$
end for
for ⑦ do
    Return $\mu = \frac{0.25}{1}$
end for
```

0.2 Sampling from the prior

In Figure 8, a tree with three taxa A, B and C (plus one internal node D, and root E) is used as a small example in this experiment. 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).

$$r = \{r_A \quad r_B \quad r_C \quad r_D\} \sim LogNormal(M = -3, S = 0.5) \tag{31}$$

In addition, a Coalescent model [33] with constant population size ($N = 0.3$) is used to describe the tree prior. Hence, for the tree in Figure 8, 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.

Test the operator on internal nodes

The genetic distances, node times and rates for g_1 in Figure 8 are given in Table 1. 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 2. 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 2. 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 9. From Table 2 and Figure 9, 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.

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

Test the operator on root

Still starting from g_1 in Figure 8, the initial settings for testing the root are given in Table 3. And the three strategies are tested separately in the following parts.

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 4 and Figure 10(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.

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 3. 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 4 and Figure 10(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.

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

Using Big Pulley For g_1 in Figure 8, 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 8. So the samples are repeatedly drawn from the 3 trees. Besides, according to the initial settings in Table 3, 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 4. The histograms of samples and numerical curves of d_D and t_E are pictured in Figure 10(c) and Figure 10(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.