

## Response to the editor and reviewer

We greatly appreciate the editor and the reviewer for the efforts and the valuable suggestions and hope that deficiencies pointed out in the original submission are overcome in the revised version. Our responses of the Referee's Report are given below.

---

### Major Issues

1. *More background on the operator is required.*

(1) *It is necessary to explain why preserving genetic distances is the goal of the operator. That is, to point out that the transition probability matrix for a branch is  $\exp(Qrt)$  so holding  $d = rt$  constant does not change the likelihood along that branch, and thus requires no re-computation of any partial likelihoods, speeding up MCMC.*

(2) *Similarly, a brief introduction to the notion of an underlying unrooted phylogenetic tree would be useful for understanding the Pulley operators.*

#### Author's Response:

Thanks for your comment.

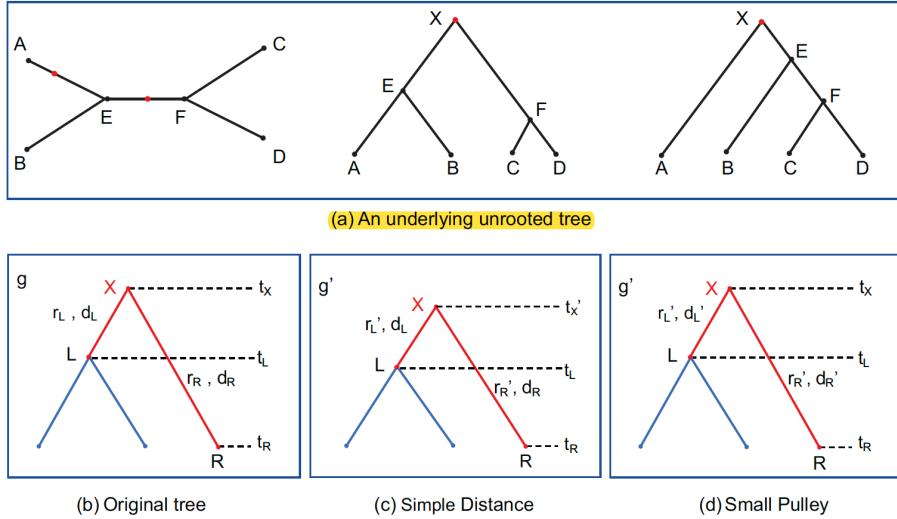
In the revised manuscript, we have added the necessary backgrounds you mentioned. The details are shown as follows.

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

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  $e^{Qt}$  and  $r$  is modelled in the substitution-rate matrix  $Q$ . In this way, it is able to avoid recalculations of phylogenetic likelihood which makes MCMC more efficient.

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

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.



**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).

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

## Author's Response:

Thanks for pointing out this detail.

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

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

3. The description of the asymmetric case in Big Pulley appears to assume that the younger child is a tip, but this is only a given if the tree has no heterochronous samples (which are increasingly common in real datasets). At a quick glance, it appears that the move could still work in this case but would require  $t\_Y < t\_O' < t\_X'$  and not just  $t\_O' < t\_X'$ .

## **Author's Response:**

Thanks for your feedback.

After careful thoughts, we confirm that it is not necessary to assume the younger child node to be a tip. To make it clear, we made a statement that node **O** refers to the node having child

nodes and node **Y** refers to the node having no child nodes, in the revised manuscript. And we use the term “extant” and “extinct” to describe node **O** and **Y** respectively, instead of “older” and “younger”.

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

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

The details are as follows.

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.

*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.

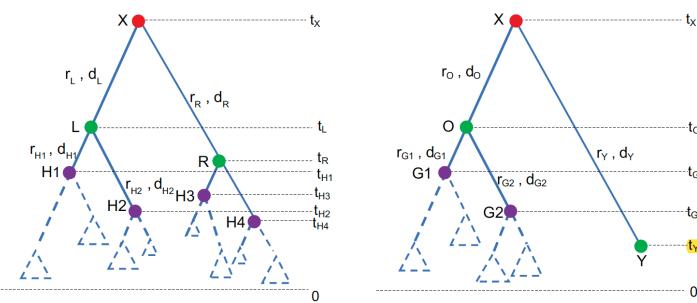


Figure 5 Two different tree shapes. The symmetric tree is on the left and the asymmetric tree is on the right. The dashed triangles represent the potential subtrees rooted at the nodes.

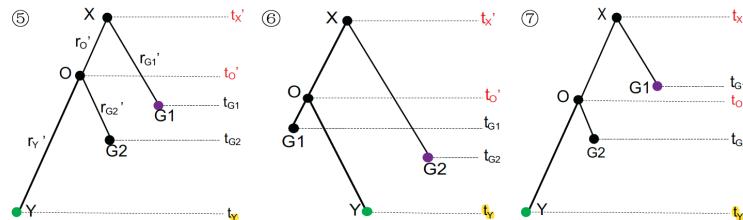


Figure 7 Illustration of operations on the asymmetric tree in Figure 5. The proposed operator will propose one of the three possible trees. If  $t_{O'} < t_{G1}$ , ⑦ has 1 probability, otherwise ⑤ and ⑥ have 0.5 probability each.

**4. Additional information is required about the simulation study.**

- (1) What priors were used for inference? Especially important is the prior on the root age.
- (2) What other operators were used on the tree and the branch rates? These are the only operators that can change the underlying unrooted phylogeny, which makes them crucial to performance.

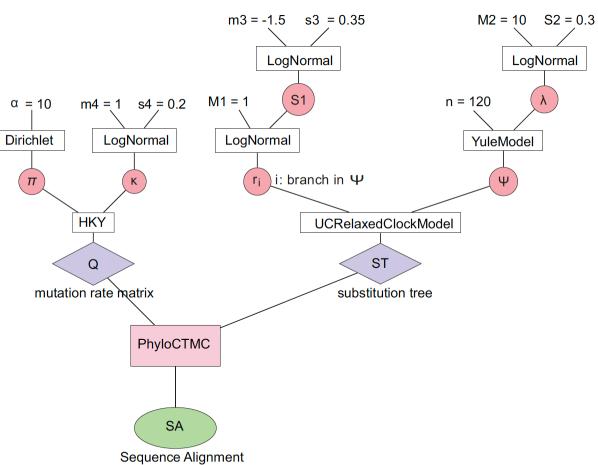
**Author's Response:**

Thanks for your professional questions.

(1) The priors used in the well-calibrated simulation study are basically presented in the framework in Figure 8. To be more specific, the priors include: (a) a Yule model tree prior where the birth rate has a LogNormal( $M2=10, S2=0.3$ ) distribution as prior, (b) base frequency having a Dirichlet ( $\alpha=10$ ) distribution as prior, (c) kappa having a LogNormal( $m4=1.0, s4=0.2$ ) distribution as prior, (d) branch rates having a LogNormal ( $M1=1, S1$ ) distribution as prior, and (e) standard deviation of rates prior having a LogNormal( $m3=-1.5, s3=0.35$ ) as hyper prior.

To answer your question, the root age ( $t_{Root}$ ) is the sum of interval lengths between two speciation events ( $\tau_i$ ). Under a Yule model,  $\tau_i$  follows an Exponential distribution with rate  $i\lambda$ , where the birth rate  $\lambda$  has a LogNormal( $M2=10, S2=0.3$ ) prior distribution. Therefore, the prior distribution of the root age can be represented by

$$t_{Root} = \sum_{i=1}^N \tau_i, \text{ where } \tau_i \sim \text{Exponential}(i\lambda) \text{ and } \lambda \sim \text{LogNormal}(M2=10, S2=0.3).$$



**Figure 8** The models and prior distributions to simulate the sequence data. The sequence alignment (SA) is simulated through a phylogenetic continuous-time Markov Chain (PhyloCTMC) that consists of a substitution model (HKY) and an uncorrelated relaxed clock model (UCRelaxedClockModel). The random variables in HKY model construct the mutation rate matrix ( $Q$ ), including base frequencies ( $\pi$ ) and kappa ( $\kappa$ ). The time trees ( $w$ ) and branch rates ( $r_i$  for each branch  $i$  in  $w$ ) 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 UclStdDev). And the time trees have a Yule model prior with birth rate ( $\lambda$ ) having a LogNormal prior. The other prior distributions include a Dirichlet distributions of  $\pi$ , a LogNormal of  $\kappa$ , 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 ( $n = 120$ ).

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

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

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

([https://github.com/Rong419/OperatorPaper/validation/calibrated/cal\\_val\\_120\\_template.xml](https://github.com/Rong419/OperatorPaper/validation/calibrated/cal_val_120_template.xml))

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

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

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

#### **Author's Response:**

Thank you for your valuable advice.

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

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

## **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.

## 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 tree 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.

### 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.

**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 <sup>1</sup>	0.1919	0.2879	0.1349	0.2698	0.1337	0.2674	0.1337	0.2674	0.1850	0.2775
	Ucldstdev Scale Operator <sup>2</sup>	0.0288	0.0288	0.0270	0.0270	0.0267	0.0267	0.0267	0.0267	0.0278	0.0278
	UcldmMean Scale Operator	-	-	-	-	0.0357	0.0089	0.0089	0.0089	-	-
	UcldmMean TreeUpperDown 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
Tree	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
	SubtreeSlide Operator	0.1440	0.1440	0.0989	0.1349	0.1070	0.1337	0.1159	0.1337	0.1388	0.1388
Substitution model	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	-	-
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.

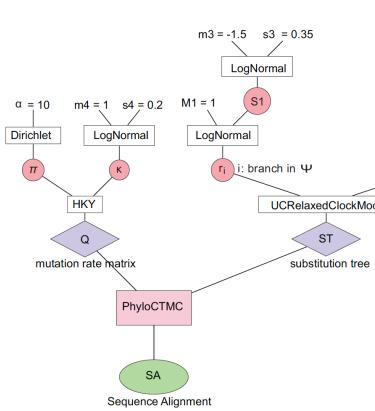
2: The operator introduced in Appendix (UcldstdevScaleOperator) is used in Cons configuration, a Scale operator is used in Category configuration .

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

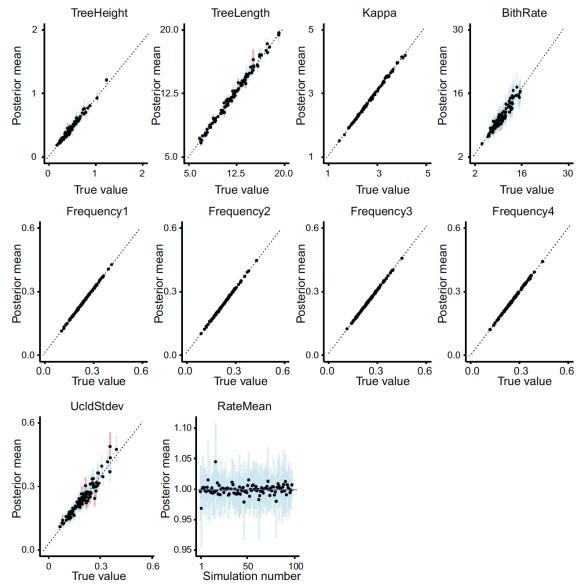
6. Figures 12 and 13 appear to be completely identical, it would appear that the 20-taxon figure was duplicated.

### Author's Response:

We have deleted the redundant figure of the 20-taxon results, and the revised manuscript only shows the results of 120-taxon data set. The details are shown as follows:



**Figure 8** The models and prior distributions to simulate the sequence data. The sequence alignment (SA) is simulated through a phylogenetic continuous-time Markov Chain (PhyloCTMC) that consists of a substitution model (HKY) and an uncorrelated relaxed clock model (UCRelaxedClockModel). The random variables in HKY model construct the mutation rate matrix ( $Q$ ), including base frequencies ( $\pi$ ) and kappa ( $\kappa$ ). The time trees ( $\psi$ ) and branch rates ( $r_i$  for each branch  $i$  in  $\psi$ ) 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 UclStdDev). And the time trees have a Yule model prior with birth rate ( $\lambda$ ) having a LogNormal prior. The other prior distributions include a Dirichlet distributions of  $\pi$ , a LogNormal of  $\kappa$ , 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.

## Minor Issues

7. In the preliminaries, there are some issues with switching between parameterizations in terms of node times,  $t$ , and in terms of the tree,  $g$ .

(1) The change from  $Pr(g)$  in equation 1 to  $Pr(t | \Phi)$  in equation 2 is a bit jarring and equation 2 is less general.  $Pr(t | \Phi)$  assumes independence between tree topology and divergence times, which is not always the case (for example the model of Barido-Sottani et al. (2018)).

(2) Page 3, lines 52-54 refer to proposing a tree  $g'$ , whereas page 3 line 38 states the operator works on times.

(3) Readers will have an easier time if one parameterization is used consistently. I personally see no strong argument in favor of  $Pr(t | \Phi)$ ,  $Pr(g | \Phi)$  still allows the use of the vector of node times,  $t$ .

### Author's Response:

Thank you for helping us find the issues.

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

(1) We have modified Equation 2 and make it more appropriate.

Firstly, we introduce the notations of probability density in Equation 1.

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

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

Then, Equation 2 is written by using the forms of conditional probability.

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):

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

where  $p(\mathbf{r}|\Phi)$  is the prior for rates specified in uncorrelated relaxed clock model. In (2) To make it clear, we claim that  $g$  represents the tree topology in current state, and  $g'$  represents the tree topology in proposal state. For operations that do not change the tree topology, there exists  $g = g'$ . In the revised manuscript, we have rewritten the statement that the operations on internal node proposes one node time and three branch rates only, the tree topology remains the same. Moreover, we have also clearly stated what is exactly proposed by the operator in the rest of manuscript.

#### Operations on internal nodes

Figure 2 represents the tree (or subtree) with the node  $\mathbf{X}$  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  $g'$ .

#### Simple Distance

Figure 3 (b), (c) and (d) show the trees that are rooted at the node  $\mathbf{X}$ . The original tree  $g$  in the current state is shown in Figure 3(b). Similar to the operations on internal nodes, we will use the following steps to propose a new root time and two rates in tree  $g'$  in Figure 3(c). At the same time, the genetic distances of two

### *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

(3) To make our manuscript more readable, we have used parameters  $\mathbf{r}$ ,  $\mathbf{t}$ ,  $g$  to represent the branch rates, divergence times and tree topology in the original state respectively. And parameters  $\mathbf{r}'$ ,  $\mathbf{t}'$ ,  $g'$  are used to represent the branch rates, divergence times and tree topology in proposed state respectively. In the revised manuscript, we have also used the bold style to represent the vector of parameters, for instance, the vector of all divergence times is denoted by  $\mathbf{t}$ .

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

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

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

#### **Author's Response:**

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

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

The details are shown as follows.

### *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.

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

(1) Explaining Exchange() before the moves is important, but the sentence "Firstly, a method called Exchange is designed to propose a new tree topology" is confusing when in fact calling Exchange() is step 3.

(2) The description of symmetric tree step 3 (page 6 lines 6-7) is confusing, as 50% of the time we will apply the method to L and either child of R.

(3) In equation 8, presumably  $d_1$  is  $d_{H1}$ , but this is not stated. Equation 10 uses  $d_{G1}$  instead of  $d_1$ , which seems more clear.

### Author's Response:

Thank you for your comment and suggestions.

(1) In the revised manuscript, we have modified the descriptions when introducing the Exchange() method.

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 (1) X

(2) We have eliminated the confusing description and made it clear that the Exchange() method will be applied to the selected node and one of its sibling's child nodes.

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.

(3) In the revised manuscript, we have replaced the unclear notations of distances " $d_1, d_2$ " with " $d_{H1}, d_{H2}$ ", so that it is explicit to understand the notations in Equation 8.

posed 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:

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

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

$$r_{H1}' = \frac{d_{H1}'}{t_{X'} - t_{H1}} \quad r_{H2}' = \frac{d_{H2}'}{t_{L'} - t_{H2}} \quad r_{L'} = \frac{d_{L}'}{t_{X'} - t_{L'}} \quad r_{R'} = \frac{d_{R}'}{t_{L'} - t_{R'}} \quad (9)$$

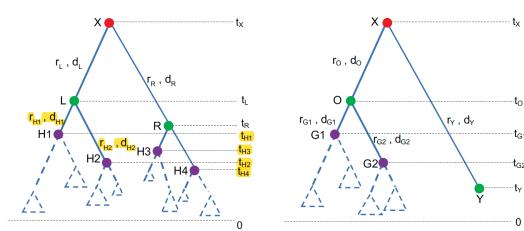


Figure 5 Two different tree shapes. The symmetric tree is on the left and the asymmetric tree is on the right. The dashed triangles represent the potential subtrees rooted at the nodes.

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

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

(2) The current comparison scheme is difficult to interpret. The rate-to-rate and age-to-age correlations do not seem to be important, but take up more of the figure than the important comparisons. It would be simpler to directly compare branch lengths to the rates of those branches, perhaps by taking the Pearson correlation coefficient of length and rate across the

*posterior. Branches could be matched across trees much as they currently are. The results could be presented as a histogram or a heatmap as is currently done.*

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

**Author's Response:**

Thank you for your professional comment.

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

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

The details are as follows:

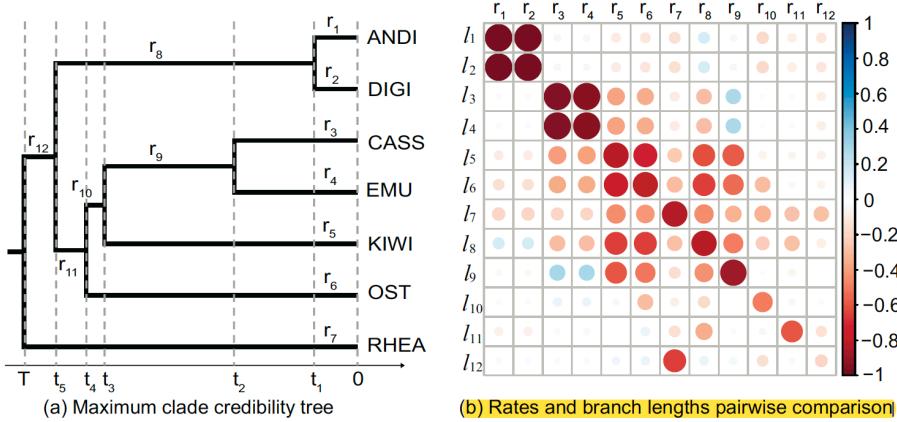
**Correlation analysis of rates and branch lengths**

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.  $d = r \times t$ . The consequence is that we have a negative relationship between rate  $r$  and branch length  $t$  i.e.  $r = d/t$ . For example, if a larger  $t_1$  is proposed,  $t_1$  increases as well, but  $r_1$  goes down. This indicates the negative

correlation between  $r_1$  and  $l_1$ . 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.



**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 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.

(3) We have corrected our original statement. In the revised manuscript, we have cited the referred work to show that there exist details in inferring trees from genomes. Nevertheless, this paper uses this approach as a simple test to demonstrate that the operators are useful in sampling rates and divergence times in relaxed clock models.

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.

31. Lanfear, R., Calcott, B., Ho, S.Y., Guindon, S.: Partition\_nder: combined selection of partitioning schemes and substitution models for phylogenetic analyses. Molecular biology and evolution 29(6), 1695{1701 (2012)

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

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

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

### **Author's Response:**

Thank you for your comment.

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

---

#### **Algorithm 1** Calculation of $\mu$ for Big pulley

---

```
1: Original tree is symmetric:  
2: if the node that has been exchanged with L or R has child nodes then  
3:    $\alpha = \beta = 0.25$   
4: else if  $t_R > t_L$  then  
5:    $\alpha = 1, \beta = 0.5$   
6: else if  $t_R < t_L$  then  
7:    $\alpha = 0.5, \beta = 1$   
8: else if  $t_R = t_L$  then  
9:    $\alpha = \beta = 1$   
10: end if  
11: if Proposed tree belongs to ① or ② then  
12:   Return  $\mu = \frac{\alpha}{0.25}$   
13: end if  
14: if Proposed tree belongs to ③ or ④ then  
15:   Return  $\mu = \frac{\beta}{0.25}$   
16: end if  
17:  
18: Original tree is asymmetric:  
19: if the node that has been exchanged with O has child nodes then  
20:    $\gamma = 0.25$   
21: else  
22:    $\gamma = 0.5$   
23: end if  
24: if Proposed tree belongs to ⑤ or ⑥ then  
25:   Return  $\mu = \frac{\gamma}{0.5}$   
26: end if  
27: if Proposed tree belongs to ⑦ then  
28:   Return  $\mu = \frac{0.25}{1}$   
29: end if
```

---

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

Sampling from the prior

In this section, we are aiming at validating 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.

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

### **Author's Response:**

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

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

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

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

(2) *Non-Bayesian methods such as TreeTime (Sagulenko et al. 2018), r8s (Sanderson 2003), and LSD (To et al. 2015) use an unrooted phylogeny as data to estimate the time tree.*

**Author's Response:**

Thank you for providing us these important literatures.

We have added some discussions about these referred works in the revised manuscript.

The details are as follows:

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.

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)
29. Cooper, A., Lalueza-Fox, C., Anderson, S., Rambaut, A., Austin, J., Ward, R.: Complete mitochondrial genomesequences of two extinct moas clarify ratite evolution. Nature 409(6821), 704 (2001)
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)

## Typos and Other Minor Comments

14. (1) While the operators as discussed in this paper are, to my knowledge, novel, others have used operators similar to the proposal on internal node heights (e.g. <https://github.com/r-revbayes/revbayes/blob/master/src/core/moves/compound/RateAgeBetaShift.cpp>)  
(2) I wonder if there may be efficiency gains by employing proposals other than a uniform, such as a bactrian proposal (Yang and Rodriguez 2013)

### Author's Response:

Thank you for providing us a new idea.

- (1) After reviewing the code in the referred link and the descriptions in the referred paper, we found that the authors define a certain distribution for the internal node height, such as a Beta distribution and a Bactrian distribution, which indicates the probability of the new node height.  
(2) However, the operators introduced in our manuscript use a Uniform distribution to move the node height uniformly on the branch. To make comparisons, we ran the simulations by using the three different proposals. The results are shown in the *Appendix3.4* in the revised manuscript.

The details are as follows.

**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 than the Category configuration. Therefore, it still needs further investigation to demonstrate the effectiveness of different proposals.

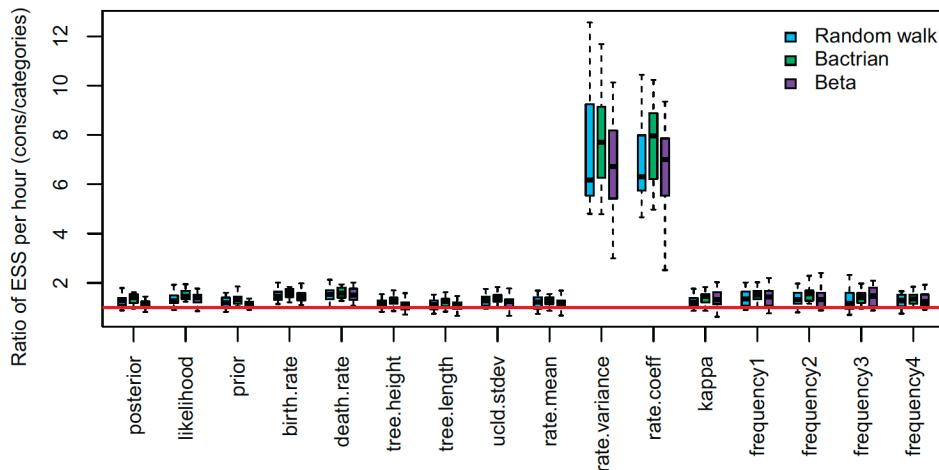
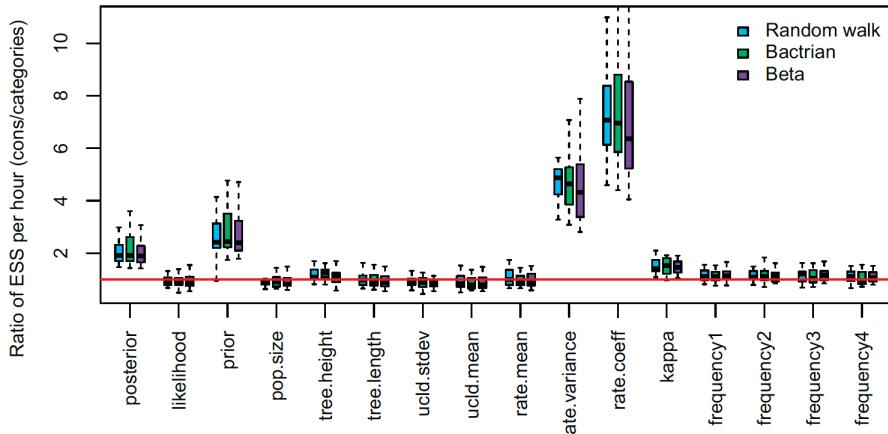
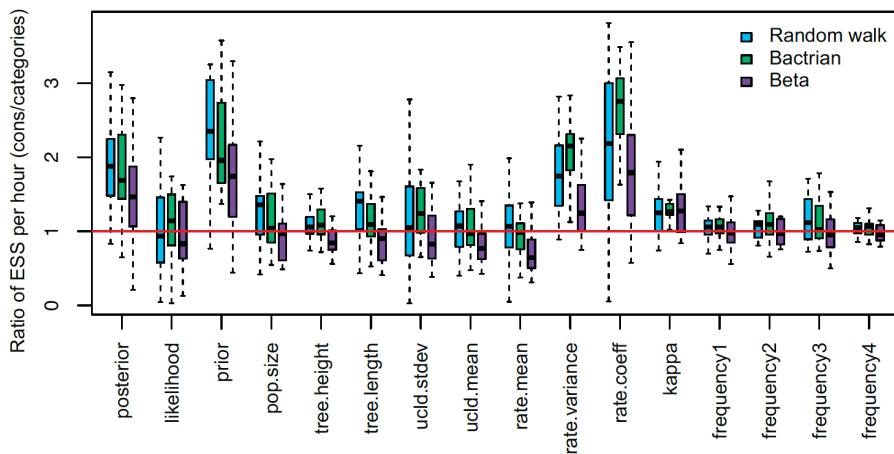


Figure 22 Efficiency comparison of proposals using Anolis data.



**Figure 23 Efficiency comparison of proposals using RSV2 data.**



**Figure 24 Efficiency comparison of proposals using Shankarappa data.**

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](https://doi.org/10.1073/pnas.1311790110)  
 46. RateAgeBetaShift.  
<https://github.com/revbayes/revbayes/blob/master/src/core/moves/compound/RateAgeBetaShift.cpp>

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

#### **Author's Response:**

Thank you for your comment.

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

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].

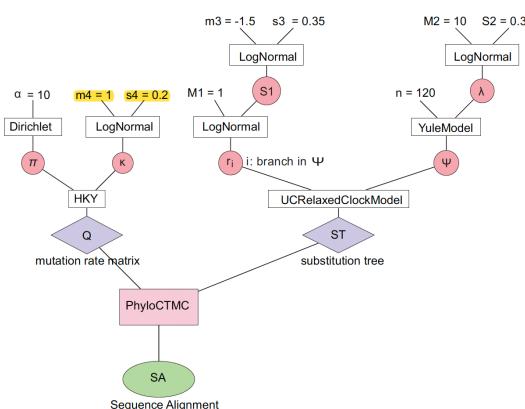
6. Thorne, J.L., Kishino, H., Painter, I.S.: Estimating the rate of evolution of the rate of molecular evolution. MolBiol Evol 15(12), 1647{57 (1998). doi:[10.1093/oxfordjournals.molbev.a025892](https://doi.org/10.1093/oxfordjournals.molbev.a025892)

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

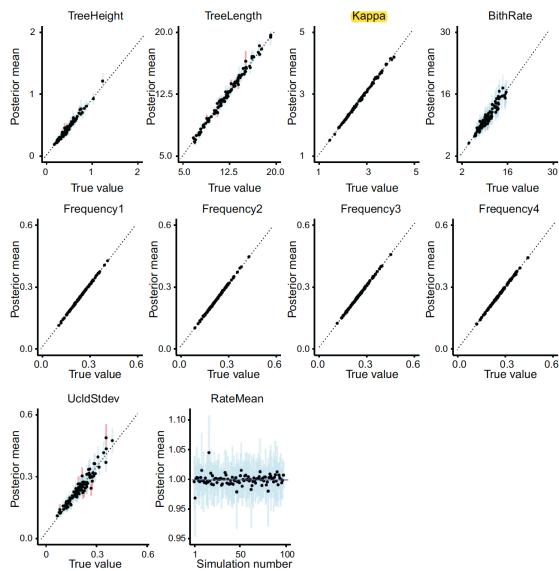
#### **Author's Response:**

Thank you for your suggestion.

In the revised manuscript, we have updated the results of calibrated-simulation study after rerunning the simulations by choosing a proper prior of kappa, i.e. LogNormal( $m_4=1.0$ ,  $s_4=0.2$ ), the mean of which is around 2.77.



**Figure 8** The models and prior distributions to simulate the sequence data. The sequence alignment (SA) is simulated through a phylogenetic continuous-time Markov Chain (PhyloCTMC) that consists of a substitution model (HKY) and an uncorrelated relaxed clock model (UCRelaxedClockModel). The random variables in HKY model construct the mutation rate matrix ( $Q$ ), including base frequencies ( $\pi$ ) and kappa ( $\kappa$ ). The time trees ( $\psi$ ) and branch rates ( $r_i$  for each branch  $i$  in  $\psi$ ) 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 UclStdDev). And the time trees have a Yule model prior with birth rate ( $\lambda$ ) having a LogNormal prior. The other prior distributions include a Dirichlet distributions of  $\pi$ , a LogNormal of  $\kappa$ , 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.

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

#### **Author's Response:**

Thank you for comment.

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

Parameters	Coverage	Parameters	Coverage
TreeHeight	89	UclStdDev	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

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

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

**Author's Response:**

Thanks for your suggestion.

We have modified our expressions in the revised manuscript to avoid unclear and misleading statements.

[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.

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

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

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

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

**Author's Response:**

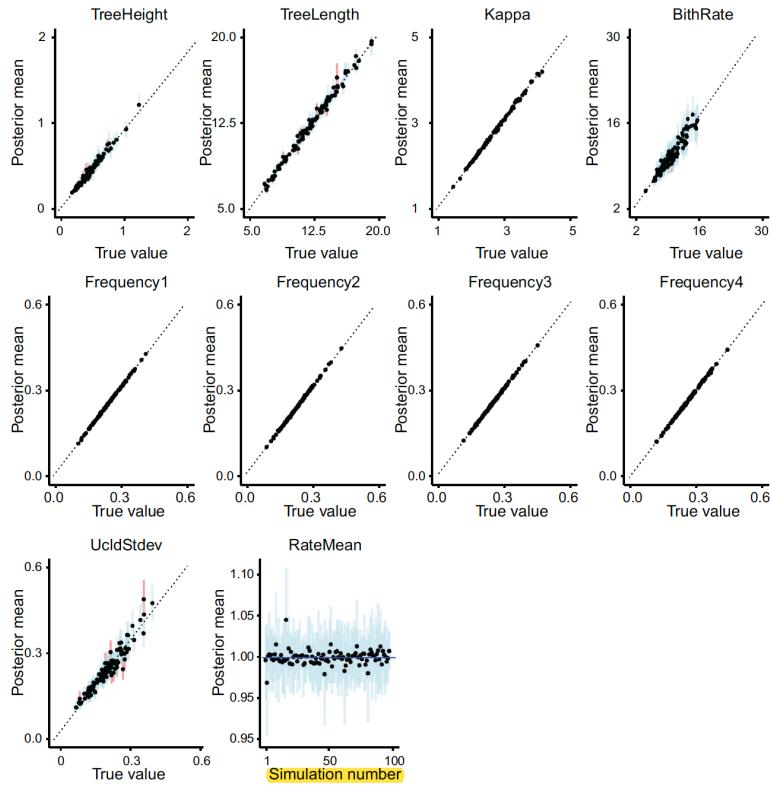
Thanks for your correction.

In the revised manuscript, we have corrected the mistakes that has been found. The details are follows:

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

*Big Pulley*

Big Pulley resamples the rates and times in a fixed unrooted tree. The genetic discredibility tree from the sampled trees in MCMC chain. This means the trees that have different shared common ancestors of each taxon in the reference tree are filtered out.



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

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

**Author's Response:**

Thanks for your comment.

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

We used a data set from Copper et al.'s work [28]. 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 conducted a pairwise comparison between each rate and branch length in order to see how they are correlated.

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)

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

**Author's Response:**

Thank you for your suggestion.

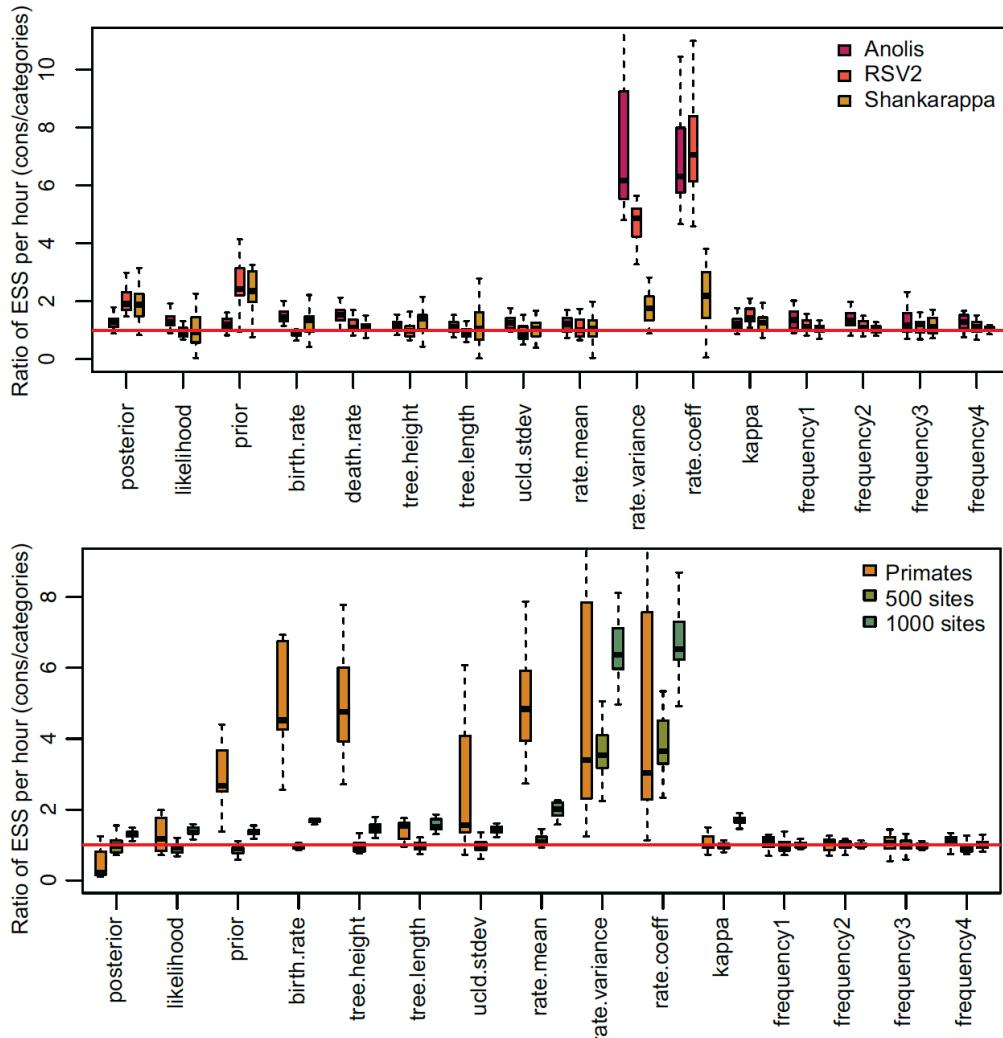
We have paid attention to modifying the colour schemes in the revised manuscript.

- (1) Figure 14 and Figure 15 in the original manuscript that shows the efficiency have become Figure 10 in the revised manuscript. Moreover, we compared the efficiency using 6 data sets and all the sampled parameters. Therefore, 6 different colours are used to represent different data sets.
- (2) In Figure 17 – Figure 21, we used another colour scheme to represent the two configurations.
- (3) In Figure 22 – Figure 24, we used the third colour scheme to represent the three different proposal methods (See Response 14(2)).

The details are as follows.

We performed the analysis on **two sets of simulated sequence alignment** (See Appendix 3.2 for more details). The simulated data sets both have 20 taxa but different sequence lengths, i.e. one data set containing 500 sites, the other containing 1,000 sites. Moreover, **we used four real data sets to further evaluate the performance of Constant Distance operator, including a primate data set 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 several parameters sampled by Cons configuration having smaller ESS per hour in some data sets, it should be noticed that the ratio is calculated by choosing random simulations in the two configurations (See Appendix 3.3 for more details). Additionally, it is worth noting that the efficiency is improved more obviously in simulated data set having 1000 sites, compared with the data 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 compared with the corresponding ESS in Cons configuration. After calculating the ESS per hour, we conclude that Cons configuration improved the efficiency of the worst estimated parameter in Category configuration by 2.26 to 15.8.



**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.

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

**Table 2** Summary of ESS and running time

**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.]

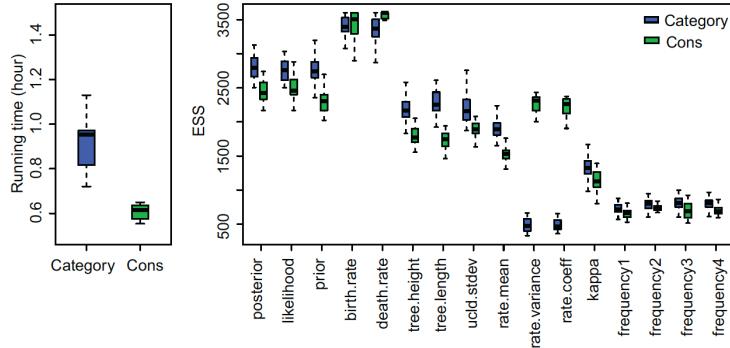


Figure 17 Running time and ESS using anolis data.

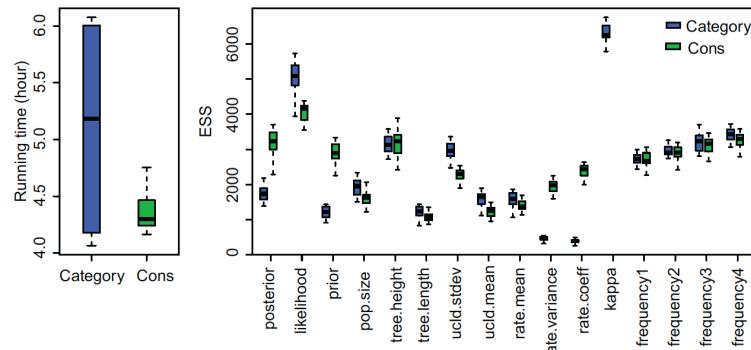


Figure 18 Running time and ESS using RSV2 data.

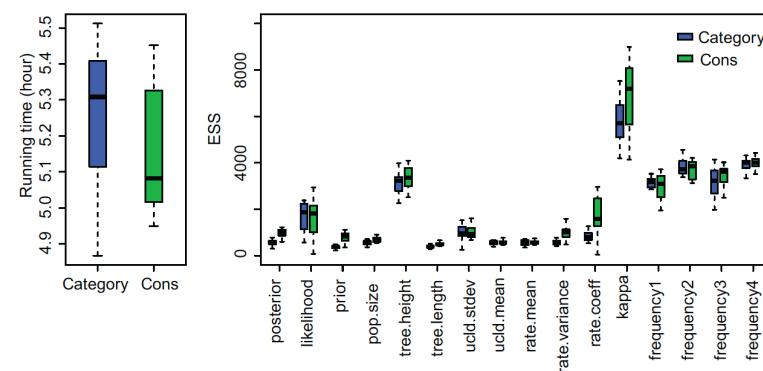


Figure 19 Running time and ESS using shankarappa data.

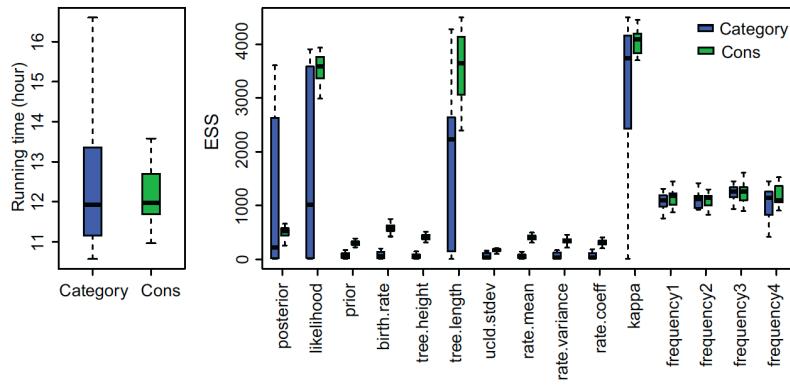


Figure 20 Running time and ESS using primates data.

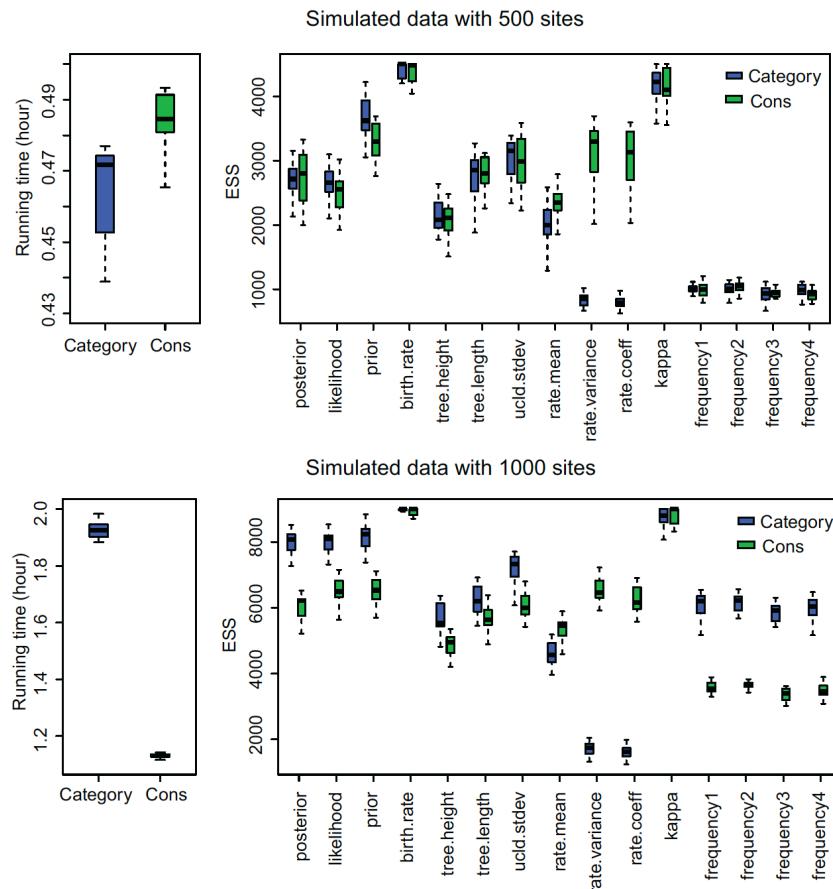


Figure 21 Running time and ESS using simulated data.

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