

## **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.

---

### **General comments**

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

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

Thanks for your comment.

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

the relaxed clock model has been widely applied, such as the study of Nothofagus [8] and flowering plants [9]. More specifically, as branch rates vary throughout a phylogenetic tree, it is able to identify regional gene bias in cooperation with fossil calibrations, according to Susanne's research in plant dispersal events [10]. Other studies such as accuracy evaluations [11] and performance comparisons [12] also suggested that relaxed clock models play an important role in estimating molecular rates and divergence times.

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

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

Thank you for pointing out this special case.

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

denoted by node **O** and its node time is  $t_O$ . Therefore, there is no deterministic relationship between  $t_O$  and  $t_Y$ .

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

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

*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.

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

#### Author's Response:

We are sorry for making these mistakes.

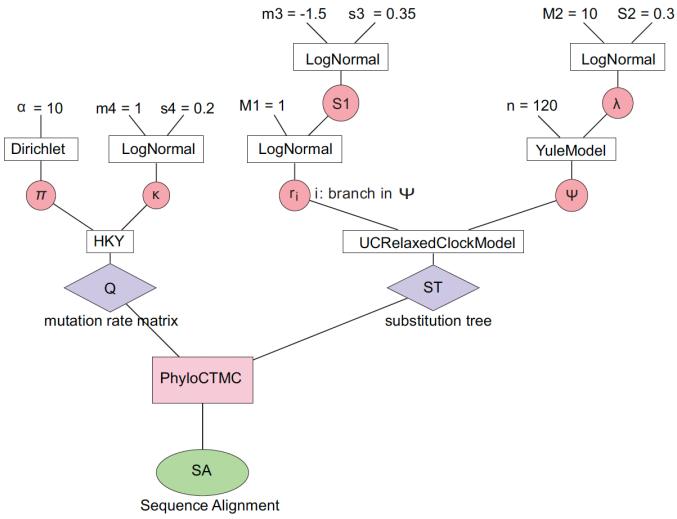
In Figure 8 (Figure 11 in original manuscript), we used S1 to denote the standard deviation of the clock, and used S3 to denote the standard deviation of the LogNormal prior of S1, so that the framework of well-calibrated simulation study can be simple enough for readers to understand. In the revised manuscript, we have added more descriptions to explain the notations below the figure.

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

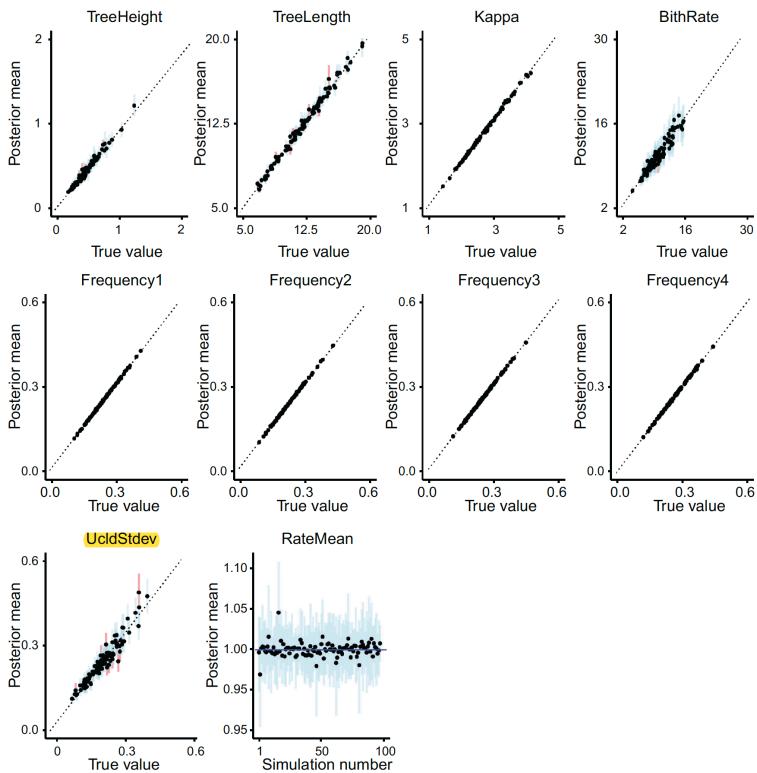
To make it clear which parameter is actually measured when comparing the efficiency, we also used the detailed term, such as "ucl.mean" (the mean of clock) and "tree.length" (total branch lengths in substitutions), in Figure 10 in the revised manuscript.

To distinguish parameters in different sections, we used uppercase letter in well-calibrated simulation study and used lower case letter in efficiency comparisons.

The details are 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.**

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

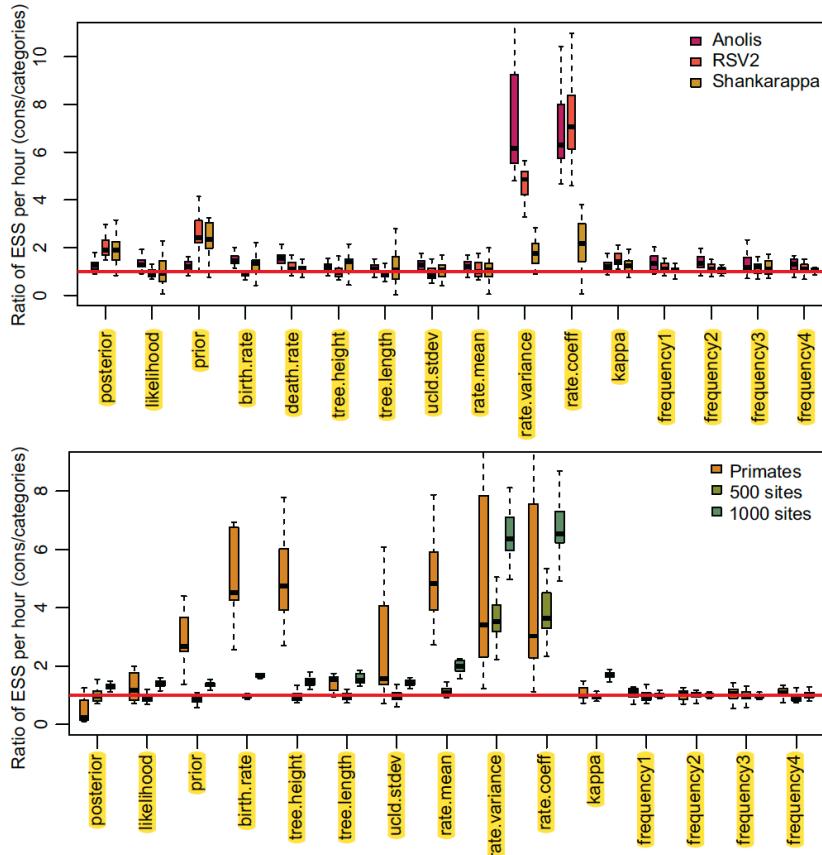


Figure 10 Comparison of ESS and running time. There are 6 data sets analysed, including 4 real data sets and 2 simulated data sets with different number of sites, as is shown in the legend. The red line represent the position where the ratio of ESS per hour is equal to 1. The horizontal axis represents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

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

#### Author's Response:

Thank you for your professional suggestion.

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

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

The details are shown below:

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

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

**Author's Response:**

Thank you for your comment.

In this section, we ran a MCMC chain with length 200000000 using BEAST2 and got 2001 sampled trees from the whole chain. Then, we used the maximum clade credibility tree (Figure 12 (a)) to filter the 2001 sampled trees. Namely, we obtained the trees that have the same topology and clades as the maximum clade credibility tree. In other words, the trees that do not match the maximum clade credibility tree are discarded. Hence, the correlation analysis was conducted by using the branch rates and node times in the tree that match the maximum clade credibility tree.

To make it clear, we have rewritten the statement of the process to avoid multiple lines describing similar content. The details are as follows:

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.

30. TreeStat2. <https://github.com/alexeid/TreeStat2>

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

**Author's Response:**

Thank you for your comment and question.

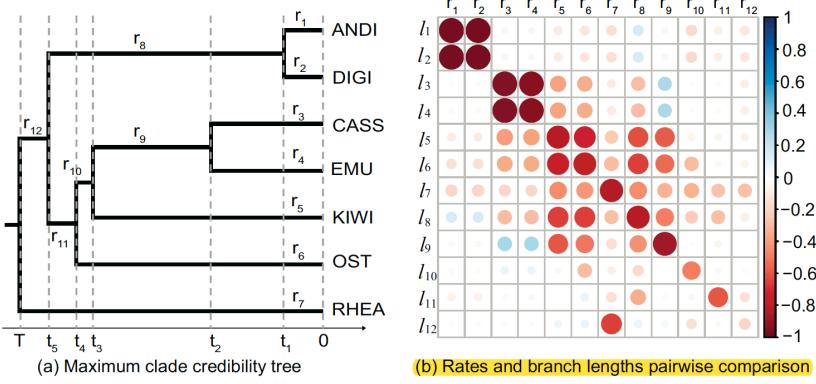
Since the new operators in our paper propose a branch rate  $r$  and a node time  $t$ , on condition that the distance  $d$  is constant, which is  $d = r \times l$ , we have a linear relationship between rate  $r$  and branch length in time  $l$ , i.e.  $r = d / l$ . If the operators work properly, we are supposed to have a negative relationship between  $r$  and  $l$  for every branch.

Moreover, for adjacent branches, let's take  $t_1, r_1, r_2$  and  $r_8$  as an example. When  $t_1$  goes up,  $l_1$  and  $l_2$  increase as well, which causes  $r_1$  and  $r_2$  to become smaller according to their negative relationships. But  $r_8$  will become larger because  $t_1$  going up means  $l_8$  goes down. It is therefore  $l_1$  and  $l_2$  have a positive relation with  $r_8$ .

By analyzing the correlation in this section, we will demonstrate whether the proposed operators are able to sample the rates and node times in consistent with the theoretical correlations. Based on the results shown in Figure 11, most rates and branch lengths tend to have expected correlations. We ascribe inconsistent correlations to the average pairwise comparison in two dimensions and the impact of other operators.

The details are as follows:

Afterwards, Figure 11(b) shows the pairwise comparison of the 12 branch rates and 12 branch lengths (in time) on these filtered trees. As can be seen from the diagonal, to a large degree, the rates are negatively correlated with branch lengths, which indicates that a larger node time will lead to a smaller rate. This is because the operators propose a branch rate  $r$  and a node time  $t$ , on condition that the distance  $d$  is constant, i.e.  $d = r \times l$ . The consequence is that we have a linear relationship between rate  $r$  and branch length  $l$  i.e.  $r = d/l$ . For example, if a larger  $t_1$  is proposed,  $l_1$  increases as well, but  $r_1$  goes down. At the same time,  $l_8$  decreases, which causes  $r_8$  to become larger, so that  $l_1$  and  $r_8$  have a positive relationship. To sum up, this dynamic change of rates and branch lengths is consistent with the mechanism of the proposed operator. Although there are some inconsistent correlations, it should be noticed that this is an average pairwise comparison in two dimensions and there are other operators sampling the rates and node times. For comparisons in higher dimensions, the results would be closer to the mechanism of the proposed operator.



**Figure 11 Correlation analysis in the ratites tree.**  $l$  represents the length of a branch, that is the time difference between a parent node and a child node, where  $l_1 = l_2 = t_1 - 0$ ,  $l_3 = l_4 = t_2 - 0$ ,  $l_5 = t_3 - 0$ ,  $l_6 = t_4 - 0$ ,  $l_7 = T - 0$ ,  $l_8 = t_5 - t_1$ ,  $l_9 = t_3 - t_2$ ,  $l_{10} = t_4 - t_3$ ,  $l_{11} = t_5 - t_4$  and  $l_{12} = T - t_5$ . The Pearson coefficients are calculated by converting the rates and branch lengths into log space, so that the coefficients range from -1 to 1, as is shown in the bar. Blue indicates positive correlations and red indicates negative correlations. The darker the colour is, the stronger the correlation tends to be.

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

#### Author's Response:

Thank you for your question.

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

Since the operators maintain the genetic distances among taxa and the underlying unrooted tree, they should be able to provide the consistent rates and divergence times with those obtained by sampling from the sequence alignment, if the genetic distances and unrooted tree are accurate. More specifically, we used an online programme PhyML 3.0 to estimate the maximum-likelihood tree based on the ratites data set. Figure 12(a) (Figure 17(a) in original manuscript) shows the unrooted estimated maximum-likelihood tree with genetic distances on the branches. Then, as an initial state of the MCMC chain, we assigned the root, divergence times and rates on the estimated maximum-likelihood tree to make it a valid rooted time tree, and make sure that the genetic distances of the rooted time tree match those of the estimated maximum-likelihood tree. Hence, these initially assigned divergence times and rates are not fixed and will be sampled by the proposed operators.

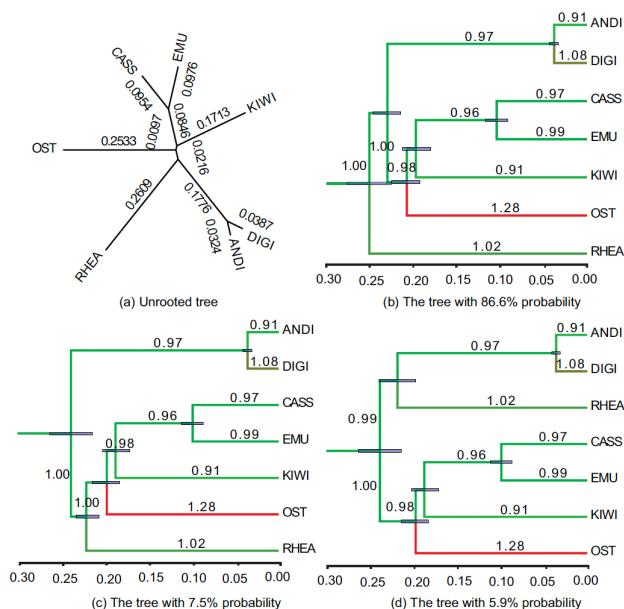
Afterwards, we used the proposed operators only to sample the initialised rooted time tree, so that the rates and divergence times are sampled but the genetic distances are always the same as the estimated maximum-likelihood tree.

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

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

The details are shown 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.



**Figure 12** Illustration of sampling a fixed unrooted tree. In subfigure (a), the unrooted tree is obtained from the ratites data set [29] by a maximum likelihood method [36] and the labeled numbers represent genetic distances. The three unique tree topologies in subfigures (b), (c) and (d) are obtained from the sampled trees by using programme TreeTraceAnalysis [41]. The branch rates and node times are summarised by using programme TreeAnnotator [42]. The labeled numbers represent the posterior mean of rates on the corresponding branches. The colour of branches from green to red indicates the rates increasing from small to large, and the blue bars represent the 95% HPD of the corresponding node times.

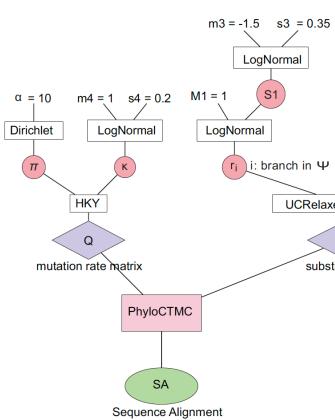
14. Reis, M.d., Yang, Z.: Approximate likelihood calculation on a phylogeny for bayesian estimation of divergence times. *Molecular Biology and Evolution* 28(7), 2161{2172 (2011)
29. Cooper, A., Lalueza-Fox, C., Anderson, S., Rambaut, A., Austin, J., Ward, R.: Complete mitochondrial genomesequences of 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)
36. PhyML3.0: New Algorithms, Methods and Utilities. <http://www.atgc-montpellier.fr/phym/>
41. TreeTraceAnalysis.  
<https://github.com/CompEvol/beast2/blob/master/src/beast/evolution/tree/TreeTraceAnalysis.java>
42. TreeAnnotator. <https://beast2.blogs.auckland.ac.nz/treannotator/>

*8. Figures 12 and 13 are identical.*

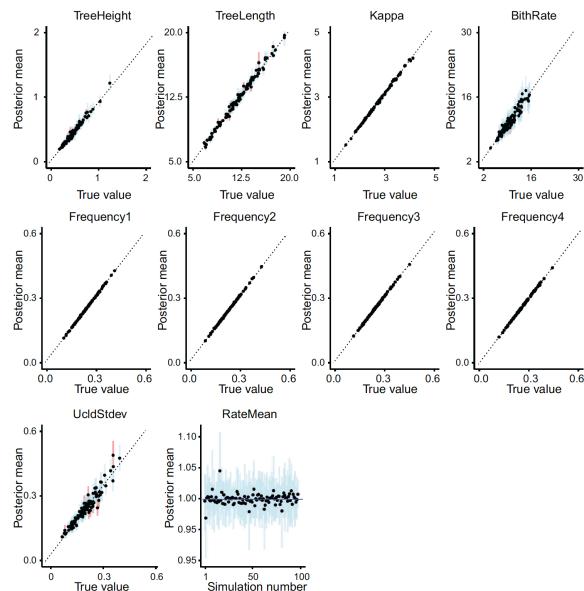
### **Author's Response:**

In the latest manuscript, the well-calibrated simulation results of 20-taxon data set (Figure 12 in the original manuscript) has been removed and the results of 120-taxon data set has been updated (Figure 13 in the original manuscript).

The details are 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 ( $n = 120$ ).



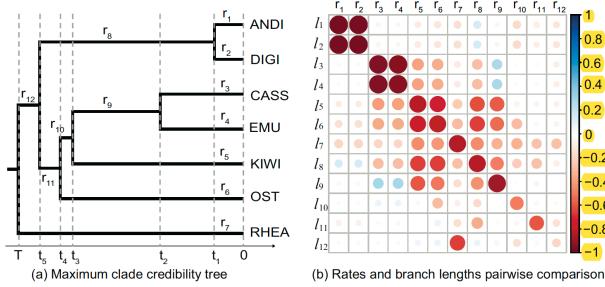
**Figure 9** Comparing the sampled parameters in simulation study with 120 taxa.

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

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

Thank you for your valuable advice.

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



**Figure 11 Correlation analysis in the ratites tree.**  $l$  represents the length of a branch, that is the time difference between a parent node and a child node, where  $l_1 = l_2 = t_1 - 0$ ,  $l_3 = l_4 = t_2 - 0$ ,  $l_5 = t_3 - 0$ ,  $l_6 = t_4 - 0$ ,  $l_7 = T - 0$ ,  $l_8 = t_5 - t_1$ ,  $l_9 = t_3 - t_2$ ,  $l_{10} = t_4 - t_3$ ,  $l_{11} = t_5 - t_4$  and  $l_{12} = T - t_5$ . The Pearson coefficients are calculated by converting the rates and branch lengths into log space, so that the coefficients range from -1 to 1, as is shown in the bar. Blue indicates positive correlations and red indicates negative correlations. The darker the colour is, the stronger the correlation tends to be.

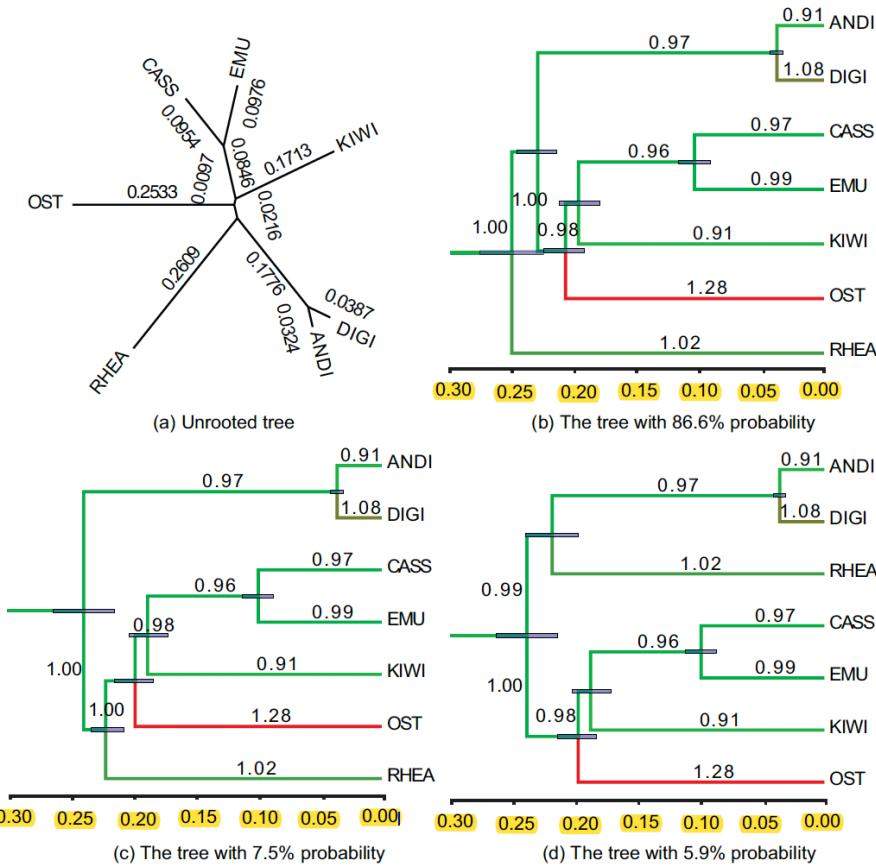
10. I would put a time axis on figure 17, as the credible intervals shown are not very meaningful without it.

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

Thank you for your comment.

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

The details are as follows:



**Figure 12 Illustration of sampling a fixed unrooted tree.** In subfigure (a), the unrooted tree is obtained from the ratites data set [29] by a maximum likelihood method [36] and the labeled numbers represent genetic distances. The three unique tree topologies in subfigures (b) (c) and (d) are obtained from the sampled trees by using programme TreeTraceAnalysis [41]. The branch rates and node times are summarised by using programme TreeAnnotator [42]. The labeled numbers represent the posterior mean of rates on the corresponding branches. The colour of branches from green to red indicates the rates increasing from small to large, and the blue bars represent the 95% HPD of the corresponding node times.

## Minor comments

### Author's Response:

After carefully reviewing the original manuscript, the following mistakes that you pointed out have been corrected in the revised manuscript. The details are listed as follows:

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

strategies discussed, named Simple Distance, Small Pulley and Big Pulley. Note that Big Pulley is able to change the tree topology, which enables the operator to sample all the possible rooted trees consistent with the implied unrooted tree. To

#### 12. p2, l39 Prelimiaries -> Preliminaries

### Preliminaries

Bayesian MCMC

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

### 13. p4, l37 $\max(tj, tk) \rightarrow \max(tL, tR)$

*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  $t_{X'} > \max\{t_L, t_R\}$  holds. Otherwise, we reject the proposal.

### 14. p4, l53 the total genetic distance $dL$ and $dR \rightarrow$ the total genetic distance $dL + dR$

based on 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. The detailed process includes the following 4 steps.

### 15. p6, l54 what happens if $t_O' < \min(t_{G1}, t_{G2})$ ?

- 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  $\mathbf{O}$  and  $\mathbf{Y}$ . (For the asymmetric tree in Figure 5, we apply Exchange ( $G1, Y$ ) and propose tree ⑦).
- if  $t_O' < \min\{t_{G1}, t_{G2}\}$ , then we reject the proposal.

### 16. p9, l26 MCMC simulation -> MCMC inference

distribution. This was achieved by examining i) the total time taken by BEAST2 to complete the MCMC inference, and ii) the effective sample size (ESS) of the

### 17. p10, l48 becomes -> become

A limiting case for the relaxed molecular clock model (and one exploited in some of our validation tests) occurs for long sequences, when the branch lengths of the unrooted tree, in units of expected substitutions per site, become known without error.

### 18. Figure 1 remove the box which only contains =

Figure 1 H3,R and H4,R should be H3,L and H4,L

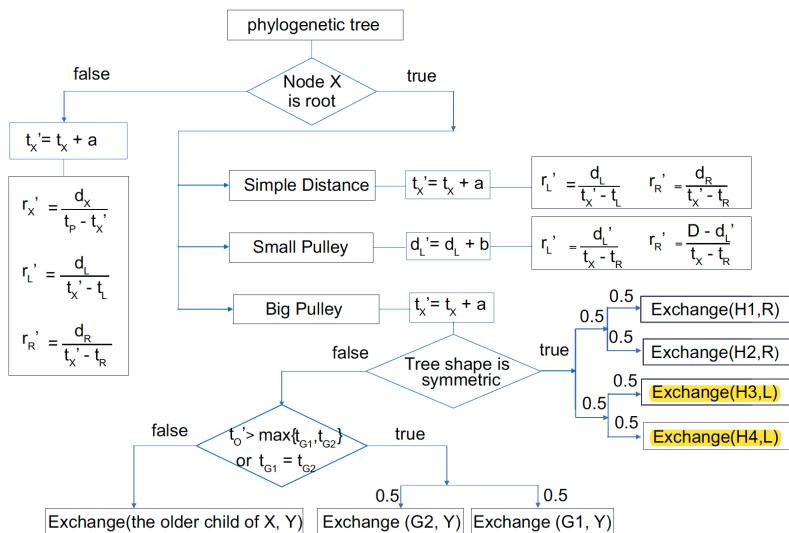
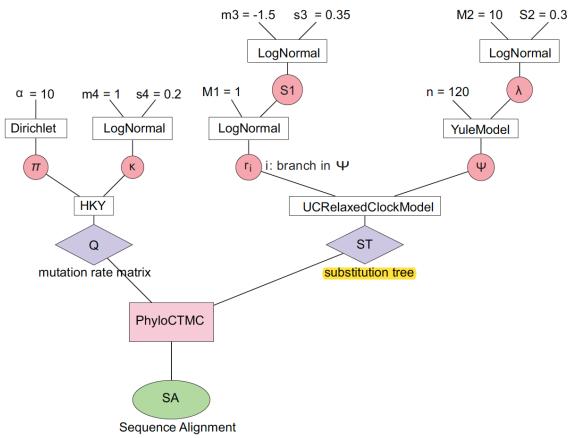


Figure 1 The flow chart of the Constant Distance operator.

## 19. Figure 11 substitution tree -> substitution tree

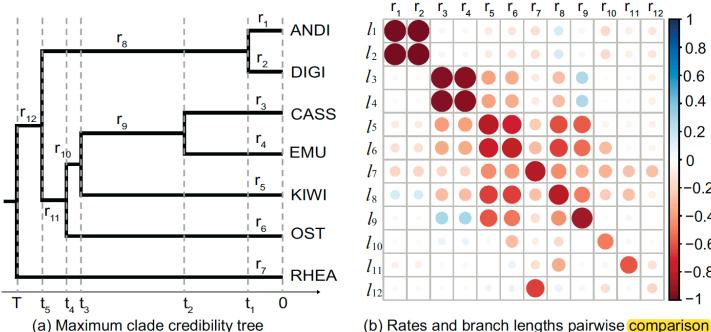
(Figure 11 in the origin manuscript becomes Figure 8 in the revised manuscript.)



**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 distribution 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 ( $n = 120$ ).

## 20. Figure 16 compare -> comparison

(Figure 16 in the origin manuscript becomes Figure 11 in the revised manuscript.)



**Figure 11** Correlation analysis in the ratites tree.  $l$  represents the length of a branch, that is the time difference between a parent node and a child node, where  $l_1 = l_2 = t_1 - 0$ ,  $l_3 = l_4 = t_2 - 0$ ,  $l_5 = t_3 - 0$ ,  $l_6 = t_4 - 0$ ,  $l_7 = T - 0$ ,  $l_8 = t_5 - t_1$ ,  $l_9 = t_3 - t_2$ ,  $l_{10} = t_4 - t_3$ ,  $l_{11} = t_5 - t_4$  and  $l_{12} = T - t_5$ . The Pearson coefficients are calculated by converting the rates and branch lengths into log space, so that the coefficients range from -1 to 1, as is shown in the bar. Blue indicates positive correlations and red indicates negative correlations. The darker the colour is, the stronger the correlation tends to be.

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.