**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.

--------------------------------------------------

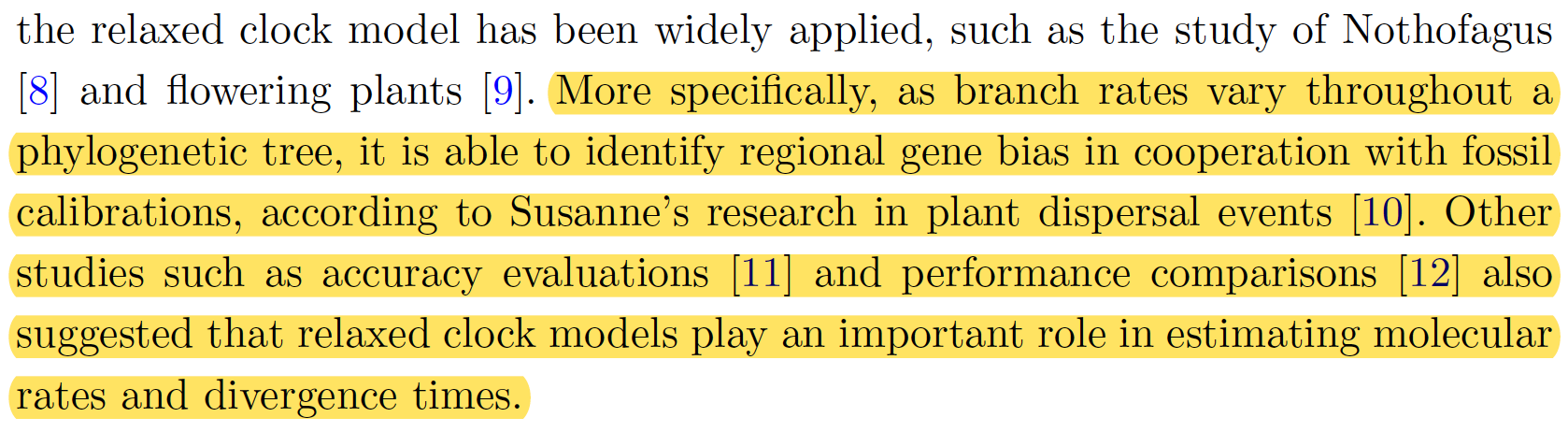
**General comments**

*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 is applied to 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.



*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 tY. Similarly, the term “extant child” refers to the node having child nodes, which is denoted by node **O** and its node time is tO. Therefore, there is no deterministic relationship between tO and tY.

However, in order to make the node times valid after the operation, it is necessary to require that tY < tO’ < tX’, because node **Y** will become the child node of node **O** in the proposed tree.

The details are as follows:

A picture containing screenshot

Description automatically generated

A picture containing indoor

Description automatically generated

A screenshot of a cell phone

Description automatically generated

*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.

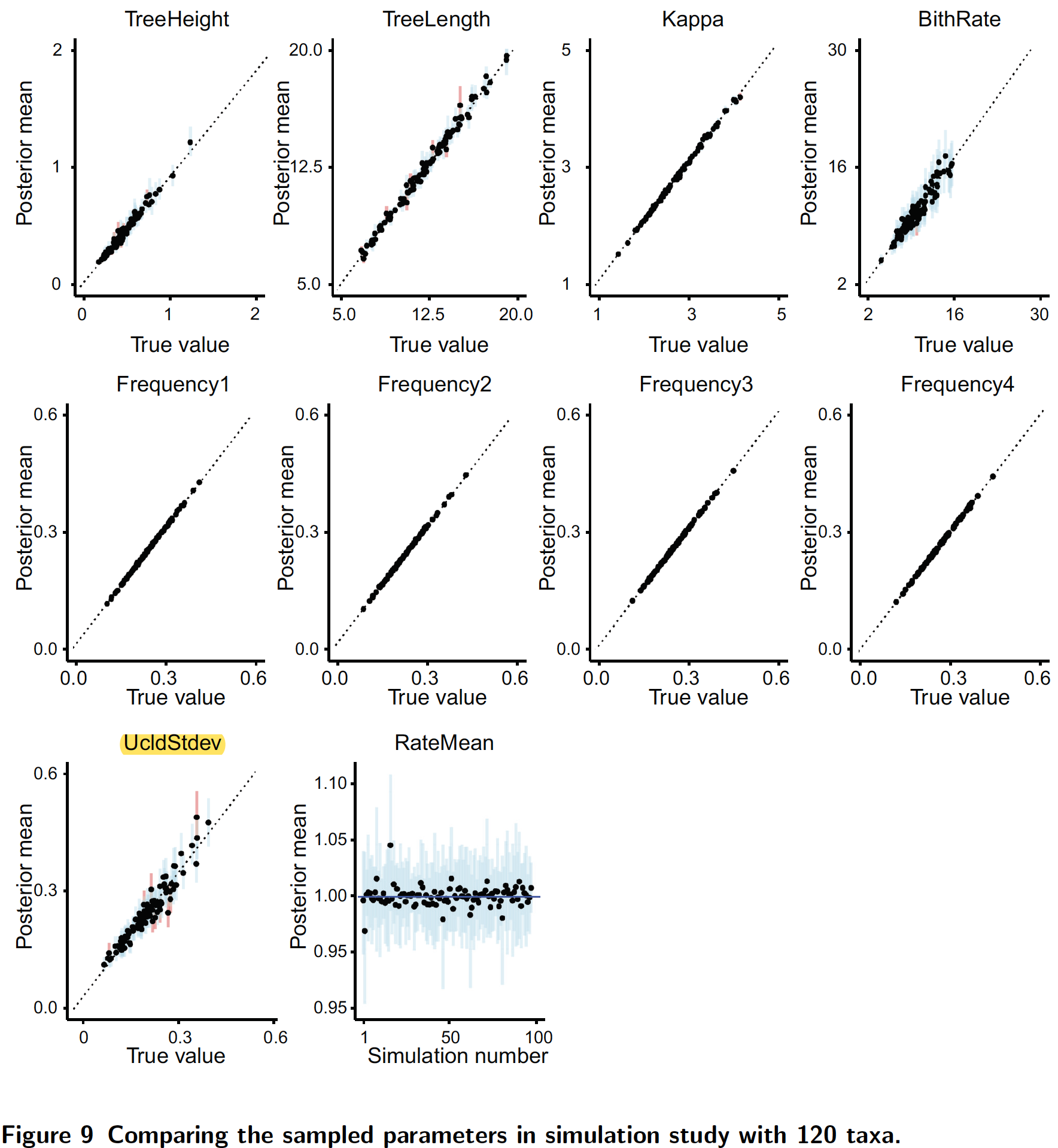
Additionally, in order to better identify what parameters are compared in Figure 9 and Table 1 (Figure 12, Figure 13 and Table 5 in original manuscript), we used the more detailed terms such as “UcldStdev”, “Kappa” and “BirthRate”.

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

The details are as follows:

A picture containing text, map

Description automatically generated



A screenshot of a cell phone

Description automatically generated

A screenshot of a cell phone

Description automatically generated

*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:

A screenshot of a cell phone

Description automatically generated

*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 11 (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, which means 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 trees 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:

A picture containing text, newspaper

Description automatically generated

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, i.e., we have a correlation equation between rate *r* and branch length in time , i.e. . If the operators work properly, we are supposed to have a negative relationship between *r* and  for every branch.

Moreover, for adjacent branches, let’s take *t1*, *r1*, *r2* and *r8* in Figure 11 as an example. When *t1* goes up,  and  increase as well, which causes *r1* and *r2* to become smaller according to their negative relationships. But *r8* will become larger because *t1* going up means  goes down. It is therefore  and  are positively correlated with *r8*.

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:

A screenshot of a cell phone

Description automatically generatedA screenshot of a cell phone

Description automatically generated

A screenshot of text

Description automatically generated

*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 by giving 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:

A screenshot of a cell phone

Description automatically generated

A close up of text on a white background

Description automatically generated

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

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/phyml/>

41. TreeTraceAnalysis. https://github.com/CompEvol/beast2/blob/master/src/beast/evolution/tree/TreeTraceAnalysis.java

42. TreeAnnotator. https://beast2.blogs.auckland.ac.nz/treeannotator/

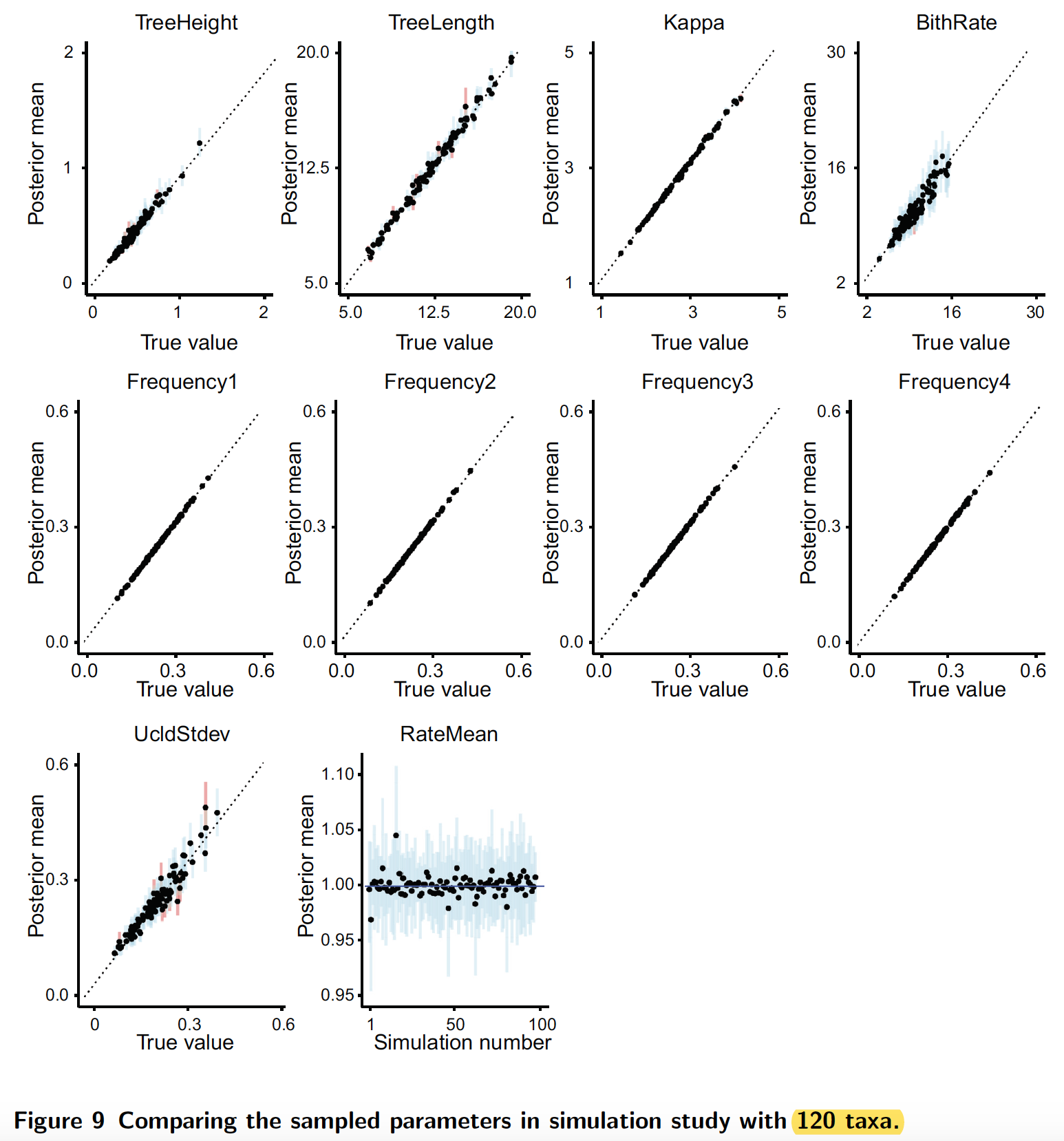
*8. Figures 12 and 13 are identical.*

**Author’s Response:**

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

The details are as follows:

A picture containing text, map

Description automatically generated

*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.

A screenshot of a cell phone

Description automatically generated

*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:

A close up of a map

Description automatically generated

**Minor comments**

**Author’s Response:**

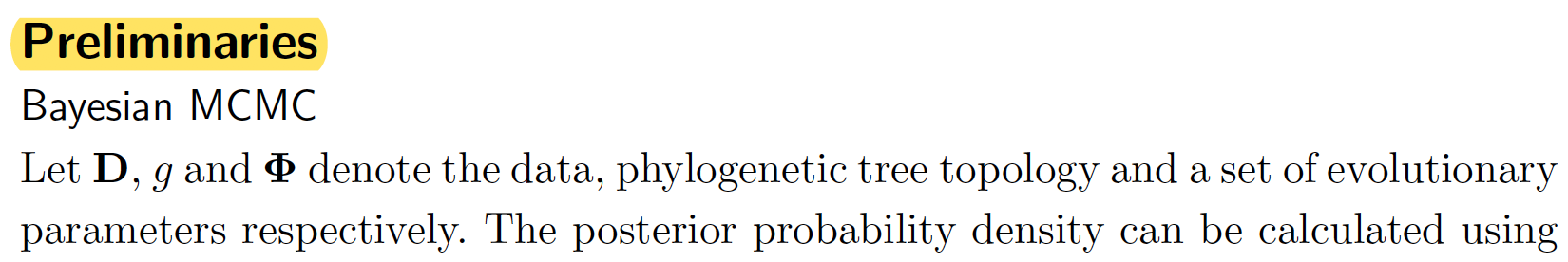
After carefully reviewing the original manuscript, the following mistakes that you pointed out have been corrected in the revised manuscript. The details are listed as follows:

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

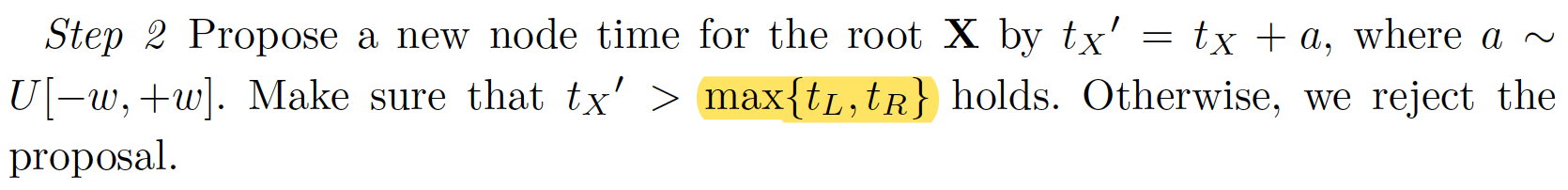
A screenshot of a cell phone

Description automatically generated

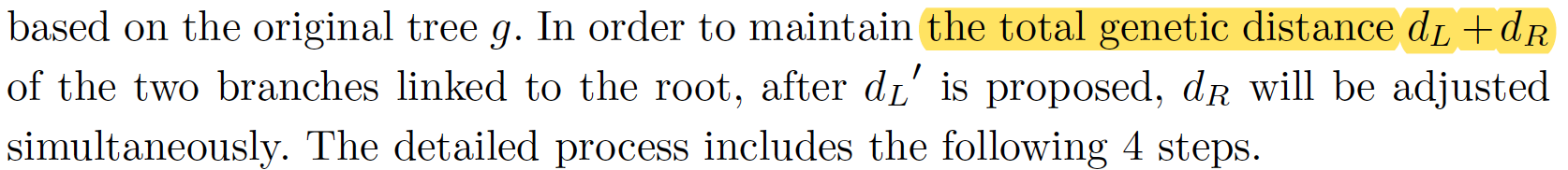
*12. p2, l39 Prelimiaries -> Preliminaries*



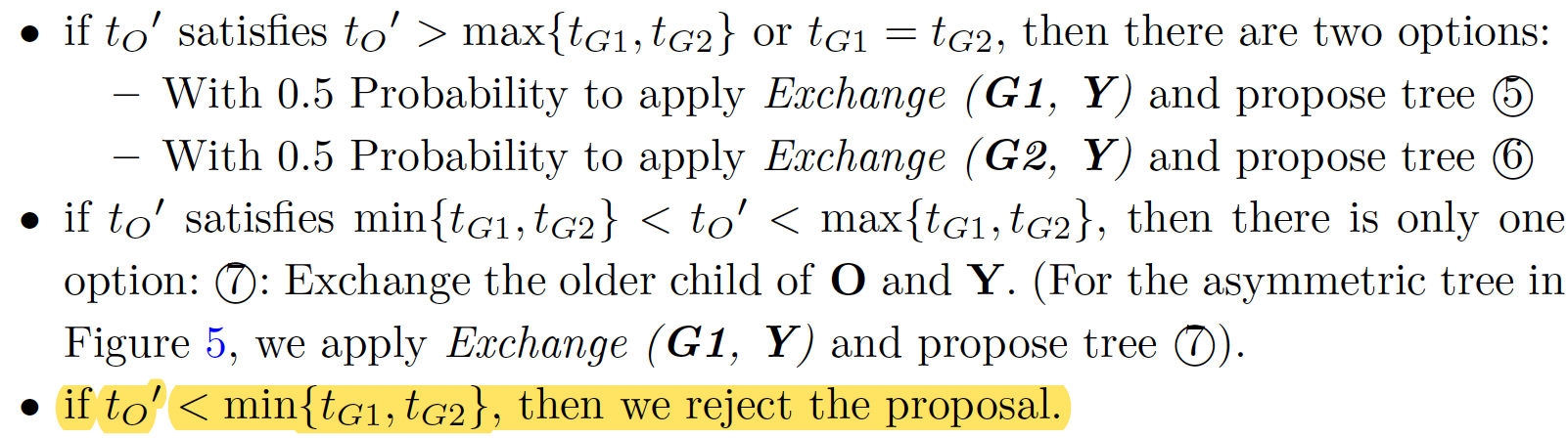
*13. p4, l37 max(tj, tk) -> max(tL, tR)*



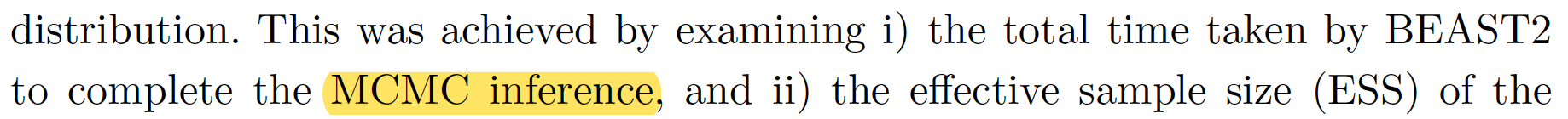
*14. p4, l53 the total genetic distance dL and dR -> the total genetic distance dL + dR*



*15. p6, l54 what happens if tO' < min(tG1, tG2) ?*



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



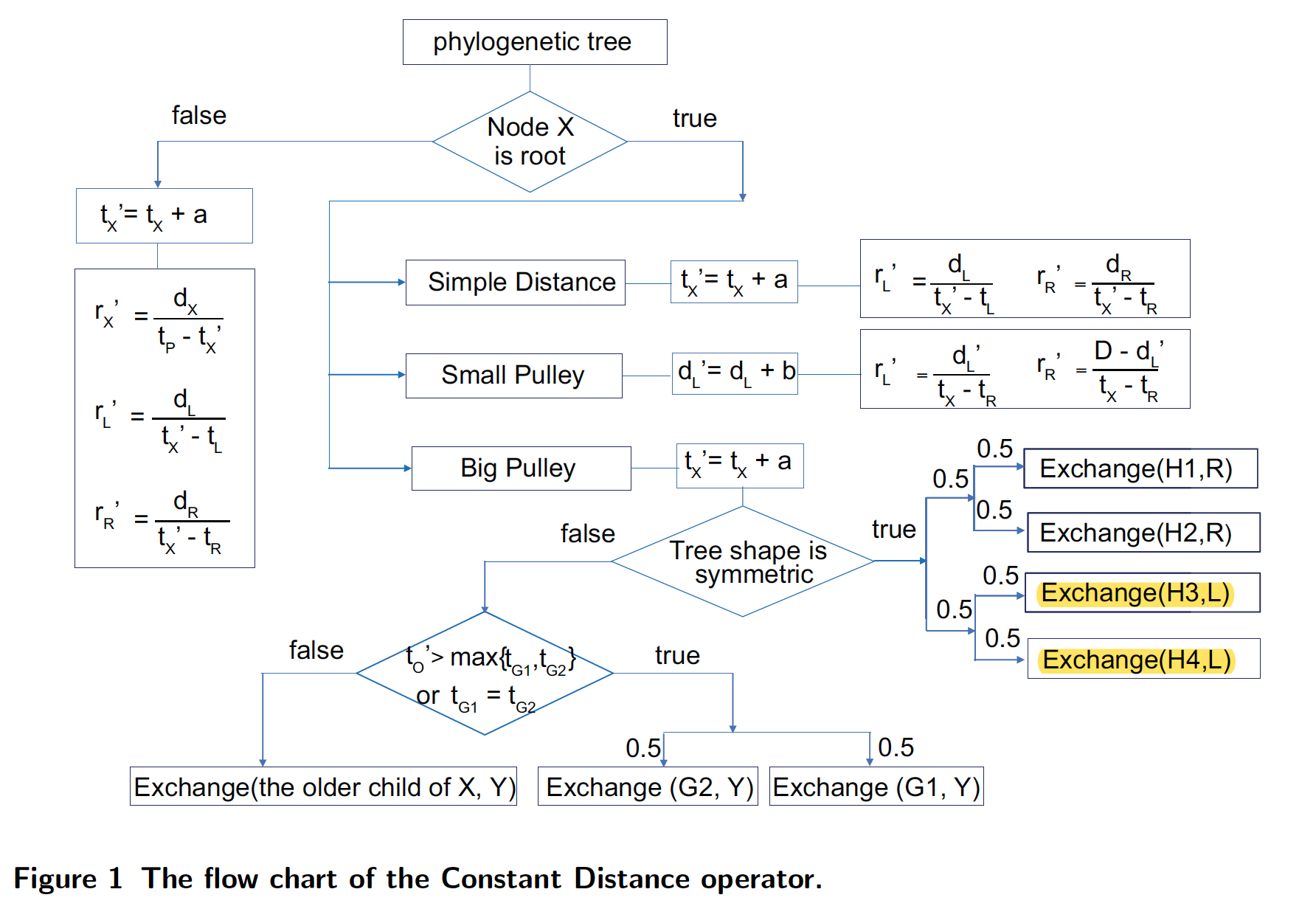
*17. p10, l48 becomes -> become*

A picture containing object

Description automatically generated

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

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



*19. Figure 11 substituion tree -> substitution tree*

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

A close up of a map

Description automatically generated

*20. Figure 16 compare -> comparison*

(Figure 16 in the origin manuscript becomes Figure 11 in the revised manuscript.)  
*A close up of a map

Description automatically generated*

**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.**