**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.*

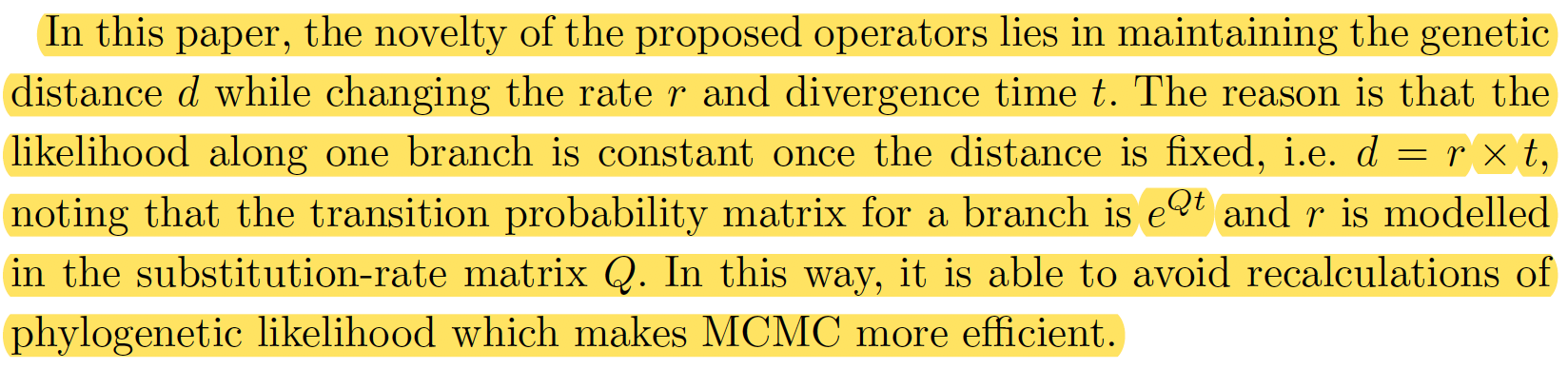
*(2) Similarly, a brief introduction to the notion of an underlying unrooted phylogenetic tree would be useful for understanding the Pulley operators.*

**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.



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

­­­­A picture containing indoor

Description automatically generated

A close up of a map

Description automatically generated

*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.

A screenshot of a cell phone

Description automatically generated

*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 tY < tO' < tX' 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.

A picture containing screenshot

Description automatically generated

A picture containing indoor

Description automatically generatedA screenshot of a cell phone

Description automatically generated

A close up of a map

Description automatically generatedA picture containing sky, text, map, indoor

Description automatically generated

*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 1. 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 (apha=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 () is the sum of interval lengths between two speciation events (). Under a Yule model,  follows an Exponential distribution with rate , where the birth rate  has a LogNormal(M2=10, S2=0.3) prior distribution. Therefore, the prior distribution of the root age can be represented by

, where  and .

A picture containing text, map

Description automatically generated

(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)

*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 in *Appendix* section in the revised manuscript showing weights on operators used in the simulations.

The details are shown below.

*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-taxa results, and the revised manuscript only shows the results of 120-taxa data set. The details are shown as follows:

A close up of a map

Description automatically generatedA close up of a map

Description automatically generated

**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.

A screenshot of a cell phone

Description automatically generated

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

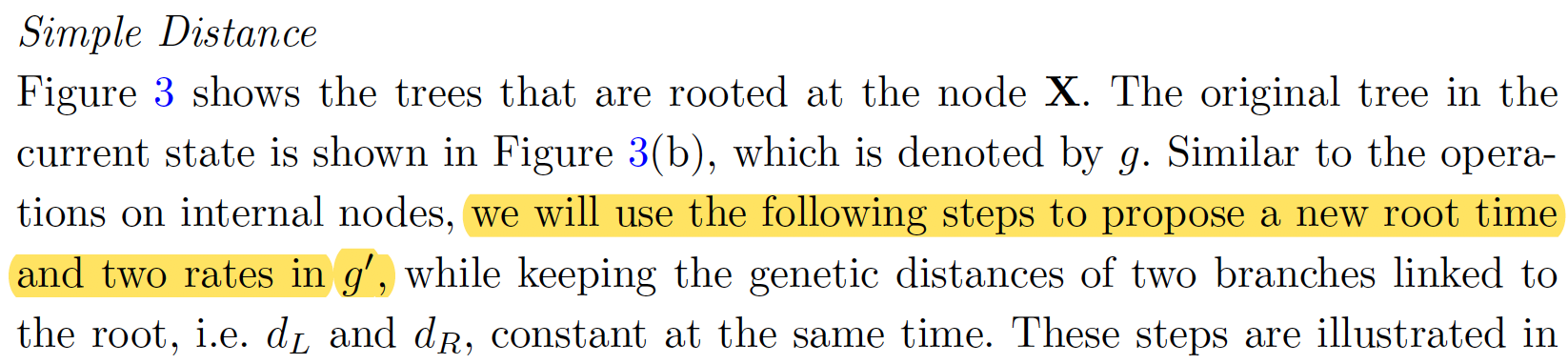
A screenshot of a cell phone

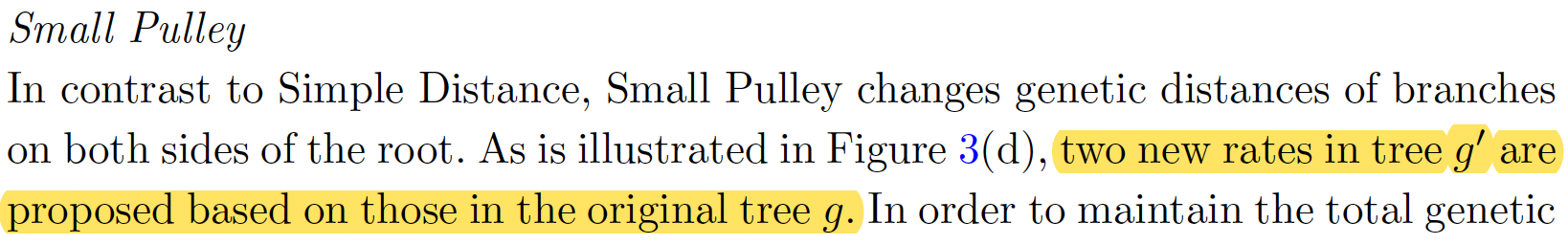
Description automatically generated

(2) To make it clear, 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.

A screenshot of a cell phone

Description automatically generated





(3) To make our manuscript more readable, we have used parameters **r**, **t**, g to represent the branch rates, divergence times and tree topology in the original state respectively. And parameters **r’**, **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 **t**.

A picture containing indoor

Description automatically generated

*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 (dL) and changes distance of the other branch (dR), so as to maintain the sum of the two distances (dL + dR). In the revised manuscript, we have modified the statement and introduced D = dL + dR.

The details are shown as follows.

A screenshot of a cell phone

Description automatically generated

*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.*

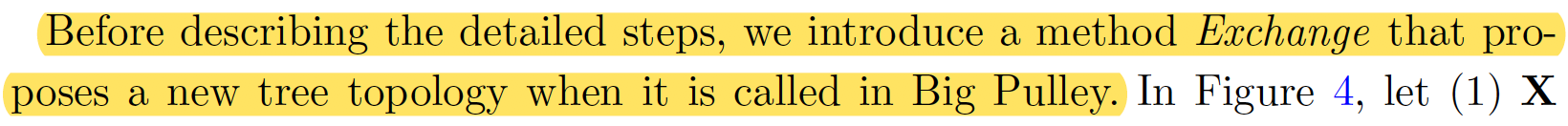
*(2) The description of symmetric tree step 3 (page 6 lines 6-7) is confusing, as 50% of the time we will apply the method to L and either child of R.*

*(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.*

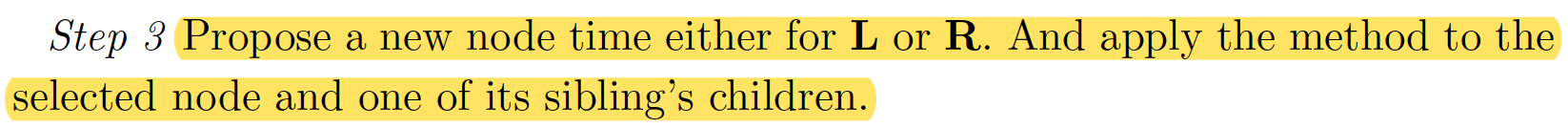
**Author’s Response:**

Thank you for your comment and suggestions.

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



(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 nephew node.



(3) In the revised manuscript, we have replaced the unclear notations of distances "d1, d2" with "dH1, dH2", so that it is explicit to understand the notations in Equation 8.

A screenshot of a cell phone

Description automatically generatedA close up of a map

Description automatically generated

*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:

A screenshot of a cell phone

Description automatically generatedA picture containing screenshot

Description automatically generated

A close up of text on a white background

Description automatically generated

(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.

A picture containing indoor

Description automatically generated

30. 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:**

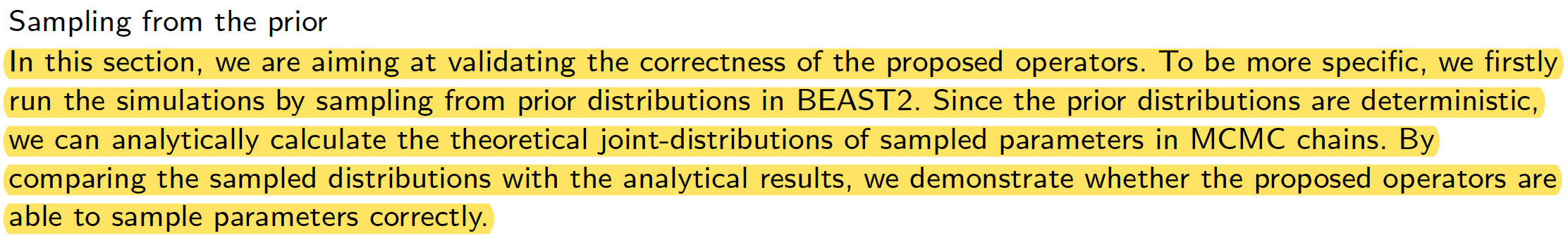
Thank you for your comment.

(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.

A screenshot of a cell phone

Description automatically generated

(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.



*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.

The details are shown below:

*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 unroted 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:

A screenshot of a cell phone

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)

31. Guindon, S.: Bayesian estimation of divergence times from large sequence alignments. Molecular Biology and

Evolution 27(8), 1768{1781 (2010)

32. To, T.-H., Jung, M., Lycett, S., Gascuel, O.: Fast dating using least-squares criteria and algorithms. Systematic

biology 65(1), 82{97 (2015)

33. Sagulenko, P., Puller, V., Neher, R.A.: Treetime: Maximum-likelihood phylodynamic analysis. Virus evolution

4(1), 042 (2018)

34. 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-evbayes/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 *Appendix* section in the revised manuscript.

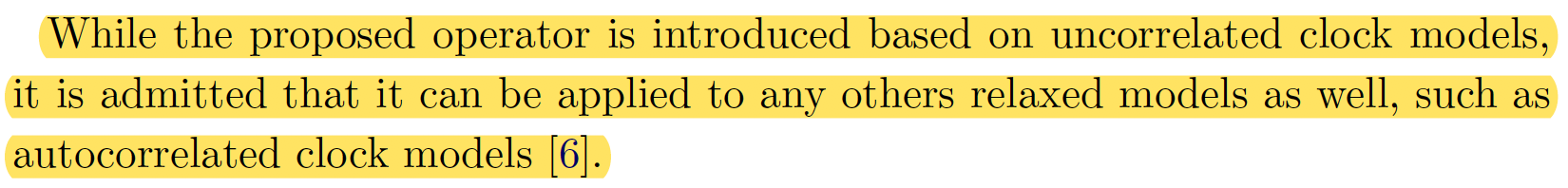
The details are as follows.

*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.



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

*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(m4=1.0, s4=0.2), the mean of which is around 2.77.

A close up of a map

Description automatically generatedA close up of a map

Description automatically generated

*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.

A screenshot of a cell phone

Description automatically generated

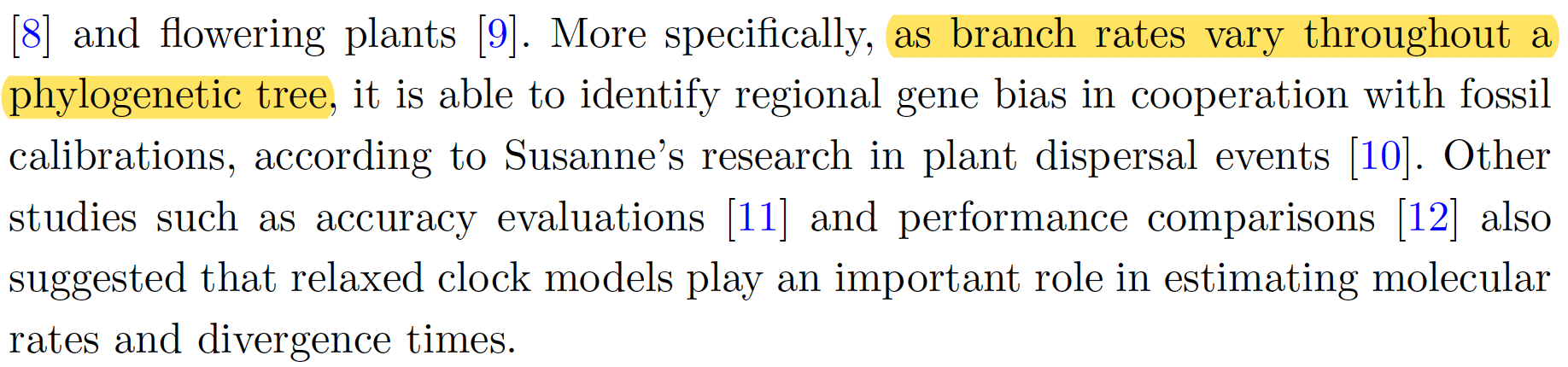
*18. (1) Page 2 lines 7-8: The sentence "By allowing rates" is somewhat unclear as currently phrased.*

*(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.



*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"?*

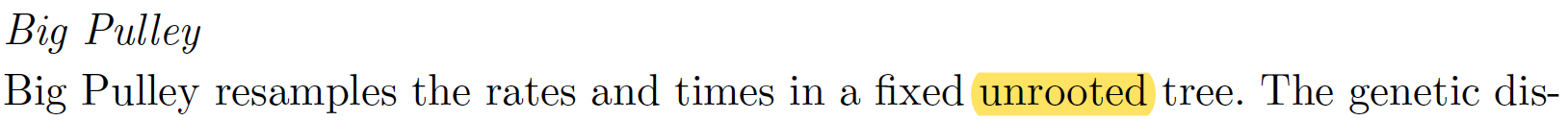
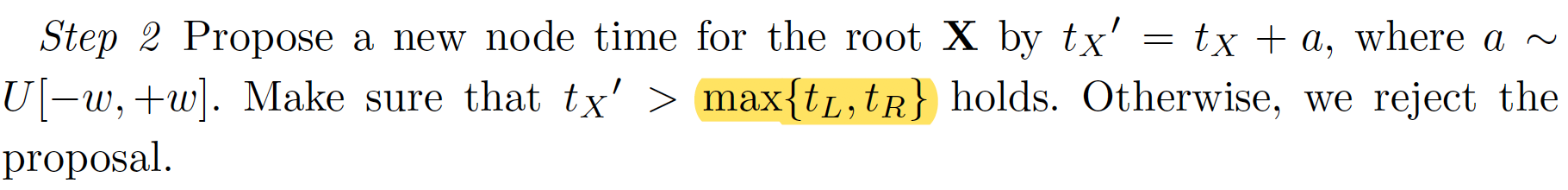
*(3) Page 10 line 18, taxa should be taxon*

*(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:

A screenshot of a cell phone

Description automatically generated

A close up of a map

Description automatically generated

*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.

A screenshot of a cell phone

Description automatically generated

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:**

**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.**