Reviewer 2

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.

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.

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'.

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.

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.

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

Minor Issues

1. 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.

2. 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.

3. 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.

4. 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 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)).

(3) 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.

5. 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.

6. 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.

7. 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.

Typos and Other Minor Comments

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/revbayes/revbayes/blob/master/src/core/moves/compound/RateAgeBetaShift.cpp)>

2. 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).

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

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

5. I wonder if there may be efficiency gains by employing proposals other than a uniform, such as a bactrian proposal (Yang and Rodriguez 2013)

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

7. 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.

8. In "Simple Distance" (page 4 line 38), t\_i and t\_j should be t\_R and t\_L.

9. Page 5 line 20 should "rooted" be "unrooted"?

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

11. Page 10 line 18, taxa should be taxon

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

13. 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.