Reviewer reports:

Reviewer 1: I think this revision has addressed my previous comments adequately. I recommend it for acceptance.

Some remaining typos:

p11, l60 don't change -> doesn't change

p14, l37 cateogry -> category

Table 2 simualted -> simulated

Table 4, figure 14 senario -> scenario

Reviewer 2: I want to thank the authors for all the work they've done in updating this manuscript. This version is a substantial improvement all around. It is now much easier to understand the motivation for the operators and the experiments performed, and to appreciate the amount of work the authors have put into validating their new operator(s). This version of the paper provides a much clearer roadmap for any developers who wish to implement and validate these operators in other MCMC packages. I was pleasantly surprised to see that the authors took it upon themselves to investigate non-uniform proposals. These results show that the choice of uniform proposals is not a bad one, and set up potential comparisons in the future with various bimodal or other smart proposals.

My remaining concerns are few and minor.

Minor Issues

-The terminology around taxa with sampling times in the past is confusing.

--In algorithm 2, the use of extant/extinct does not match any common usage. The definitions on page 7 line 33 for Y and O are clearer and avoid this confusion. I suggest removing extant/extinct entirely.

--It seems that the authors are using "sampled ancestor" to mean "a tip whose sampling time is in the past," or more compactly "a heterochronous tip." This is not the standard usage of sampled ancestor, which generally means "a sampled taxon who has descendants which are also sampled taxa."

---On page 6 (l 24-25), it is stated that a sampled ancestor has no descendants (is of degree 1, having only a parent). This suggests that "sampled ancestor" is being used to mean "heterochronous tip."

---Does the "Big Pulley" algorithm apply if O is a node of degree 2 (having a parent and a single child)? It seems like it should, but I do not know if it would fall under the symmetric or asymmetric case. The authors are free to leave this case to future work, so long as the terminology surrounding what O represents is made clearer.

-Figure 10 and Table 7 appear contradictory about the models used to analyze RSV2 and HIV-1. In the table it is stated that there is an operator on the population size, but in the figure there are efficiencies listed for birth and death rates instead.

Typos and other minor comments:

-The comparison of ESS for the clock standard deviation between "cons" and "categories" doesn't seem exactly fair, since there is an entirely new operator for the standard deviation of the clock in "cons."-It may be useful to mention the "nocons" tests were run in previous analyses. This would assuage any concerns that the difference between "categories" and "cons" is about the change from discretized to continuous branch rates, rather than about the use of the constant distance operator.

-Table 7: "Substituion model"

-p26 l28: the simulations are not definitive proof, but rather a convincing demonstration