Dear Melanie Hopkins, Dear PCI Paleontology Managing Board

Thank you for reaching out to another reviewer, allowing us to improve the manuscript further. Below we provide our responses to the comments made by the reviewer. We first address general comments, followed by the minor comments below.

Major comments:

One key point made was that our unexpected results were because we included OU in the set of modes of evolution. As the other modes of evolution are nested within OU, tests return disproportionately high support for OU when only AICc weights and not model parameters are considered. The reviewer suggested to exclude OU from the tested modes of evolution, and rerun the analysis.
Change made: The Ornstein-Uhlenbeck process is no longer included in the modes of evolution we test for. The results and discussion section was adjusted accordingly. The new analyses show good support for the correct mode of evolution if time series are long enough.

Another point made by reviewers in both review rounds is that the threshold of AICc weight of 0.9 is too strict and should be removed.
Change made: We removed the 0.9 AICc weight threshold from the analysis, plots, and discussion. We expanded the article with a discussion on evidence ratios, as we found that when the distribution of AICc weights overlap, there is a chance that individual lineages will be wrongly identified or no decision on the mode of evolution can be made. We think this discussion will be useful for the community, as our simulation study gives us the sample size to examine such effects that are not observable when using exclusively empirical data without knowing the underlying mode.

Minor comments:

Comment: *"line 34: here, and elsewhere in the manuscript, pulsed change is referred to as punctuated equilibrium. I don't think this is quite accurate: the punc eq model has pulsed change but it occurs at lineage splitting. A pulsed change within an unbranched lineage is more evidence against than for punc eq because it involves large changes without speciation. (Gould would sometimes try to cloud this issue.) I'd recommend using terms like pulsed or punctuated change, and not punctuated equilibrium, for unbranched lineages."*
Response: A clarification has been added.

Comment: *"Line 88: before the Fossilized Birth Death model and related approaches, there was a phase in which fossil data was used a lot (sometimes naively) to get constraints for node dating approaches"*
Response: It had been mentioned very briefly in lines 62-63 of the reviewed manuscript, under calibration of molecular clocks, but it has been made more explicit now.

Comment: *"Line 110ff: The presentation of completeness that I am familiar with (e.g., Shanan Peters' work) emphasizes that completeness will depend on the temporal scale of resolution. A section may be mostly complete when considered in 1 Myr bins, but will be much less so if the bins are 10 Kyr."*

Response: The scale-dependence of completeness is mentioned in the methods section, clarifying that we examine incompleteness on the resolution of 1 kyr (the time increment of the forward simulation).

Comment: *"Line 169: here and at a few other places, it seems to imply that previous approaches in paleo have required samples to be equally spaced in time. The model fitting approaches used here and in Hunt (2006) cited here have always allowed for arbitrary spacing of samples."*
Response: This section serves to motivate why people use such simplified age-depth models. The method developed in Hunt (2006) is analytically derived for discrete steps and parametrized using step means and step variances. The method was later numerically tested for performance under unequal sampling (Fig. 5 in Hunt (2006)). While it performs well under unequal sampling, it was originally designed for equidistant sampled time series, which is why we refer to it here.

Comment: *"Line 305: I don't think it needs to be done in this paper, but, as an FYI, it is not difficult to generate realizations of the OU model with unequal sampling. The sim.OU function in paleoTS does it one way, and there is another approach in which a whole time-series is a single draw from a multivariate normal distribution using the vector of means and covariance matrix from Hansen & Martins (1996)."*
Response: Thanks for pointing this out. With the removal of OU from the tested modes, the whole section on OU was removed.

Comment: *"Line 322: not quite right as written, as the standard deviation would be sigma *sqrt(t), not sigma. The simulation code is correct, though."*
Response: This was corrected.

Comment: *"Line 335: I would not say scenario 3 is "weakly directional". Both it and scenario 4 are strongly directional, really more so than just about any empirical sequence. This can be seen from the figures – both look almost like straight lines -- and the results are basically the same throughout for both. Calculations from Hunt (2012, Table 1) indicate that directionality accounts for 98% and 99.5% of the evolutionary change in scenario 3 and 4, respectively. I'd recommend just keeping scenario 3 as representing trends and dropping the unrealistic scenario 4."*
Response: Scenario 4 (strong Brownian drift) was removed, the weak Brownian drift was renamed into Brownian drift.

Comment: *"Line 348ff. I see the need for the distinction, but it seems odd to call them both time-series. Perhaps instead they can be stratophenetic series and time-series? The former phrase has been used occasionally in this literature."*
Response: We changed "fossil time series" to "stratophenetic" series.

Comment: *"Line 516, about stratigraphic completeness not being the driver of outcomes. This is an interesting and important point."*
Response: We agree, this is one of the main results of the study.

Comment: *"Line 640ff: this section should be reconsidered based on my comments above. The discussion about Levy flights is interesting, as I agree that kind of dynamic would probably be favored when there are unrecognized hiatuses in a section. That model isn't implemented in paleoTS, but (within-lineage) punctuations are."*
Response: tests for OU were excluded from the analysis and the results and discussion sections were adjusted accordingly.

Comment: *"Line 826: this references Hannisdal (2006). The only other example of a similar study I know of appears in one of the chapters of the Patzkowsky and Holland book. I think that should be cited somewhere in here as well. (It is possible that chapter refers to another paper that I am not remembering at the moment, too.)"*
Response: The book (fig. 7.2) uses an example from Holland (2000). However, there is no explicit identification of the mode of evolution in either. There are visual illustrations. They do not fit well in the text in the fragment to which this suggestion is made, but we referenced this example elsewhere (hypothesis 2 in the Introduction and Limitations of the study in the Discussion).

Comment: *"Line 829: you would probably see more artefactual support for stasis if the generating parameters didn't have such high rates. With lower rates, sampling noise would be a larger component, and since sampling noise looks like stasis, you should get more spurious cases of stasis. Analyzing more shorter sequences might have a similar effect."*
Response: That is true. We are aware that our sampling strategy is very optimistic. We chose this strategy to study the effects of incompleteness and irregular age models in isolation.

Comment: *"I love the forward modeling approach to investigate what happens under known conditions. I am curious if you think these sedimentation models might ever be used for the inverse problem, so as to generate more realistic age models for empirical fossil timeseries?"*
Response: We think combining empirical and synthetic data is a useful way to refine age models, and has great potential for geohistorical disciplines. Two references on this were added to the conclusions.