

Response to reviews for *Virus Evolution*

Sarah K. Hilton and Jesse D. Bloom

August 17, 2018

Below, the reviewer comments are in blue, and our responses are in black.

Editor Comments

Many thanks for submitting your manuscript "Modeling site-specific amino-acid preferences deepens phylogenetic estimates of viral divergence" to *Virus Evolution*. My apologies that the review process took a little longer than usual.

Your work was considered by Section Editor Aris Katzourakis, with the support of two expert peer reviewers. Their comments are included below. All conclude that your manuscript represents valuable progress on an interesting and important question for virus evolution. I agree.

We would therefore like to invite you to submit a minor revision that addresses the comments raised in the reviews.

Thanks again for submitting to *Virus Evolution* and we look forward to receiving your revision.

Thank you for the reviews. We were pleased that both reviewers liked the manuscript, and have submitted a revised version that includes the minor changes that they suggested.

Section Editor (Aris Katzourakis) Comments

Thank you for your submission to *Virus Evolution*. Your manuscript has now been evaluated by two independent reviewers both of whom found it interesting and engaging, and recommend publication. Both reviewers have some minor and easily addressable comments for clarity and presentation; please consider these and let me know how you have incorporated them into the manuscript. None of the comments require additional analysis nor should require extensive re-reviewing of the manuscript.

The minor comments by the reviewers were both helpful and straightforward to address. In this re-submission, we have made revisions that address these comments. This response details those changes, and we have also uploaded a version of the manuscript with all changes tracked.

Reviewer #1 Comments

Comments:

This is a very interesting manuscript about the consequences of not modeling site-specific amino-acid preferences when estimating sequence divergence (branch lengths in a phylogeny). What the Authors are pointing out in this article is clearly a major conceptual and methodological problem in practical phylogenetic analyses of viral sequences, and I think it is really important to raise the awareness of people interested in viral evolution about this problem. The explanations given by the Authors about the issue and about the methods that can be used to address it are cristal clear, with nice simulations and convincing analyses of empirical sequence data. For those reasons, I strongly recommend this manuscript. I would have only very minor comments and suggestions:

Thank you for the excellent summary of our work. We appreciate the kind comments about the clarity of the work, which has been furthered improved by incorporating the suggestions that you made below.

Minor points

(1) The title is slightly misleading: it seems to suggest that the article introduces a molecular dating method accounting for site-specific amino-acid preferences, which it does not do (no problem here: developing such a method would require substantial software development, way beyond the scope of this article, so this is not the question) ? but then, perhaps the title should be a bit more explicit about the exact content. What about: “Modeling site-specific amino-acid preferences deepens phylogenetic estimates of viral sequence divergence”

This is a good suggestion. We have updated the title to “Modeling site-specific amino-acid preferences deepens phylogenetic estimates of viral *sequence* divergence” as suggested.

(2) different codons for the same protein are treated as selectively equivalent: different codons for the same amino-acid?

Thank you for pointing out this typo. We have updated the text to read “Here we will consider mutation-selection models where the site-specific selection is assumed to act solely

at the protein level (different codons for the same *amino acid* are treated as selectively equivalent).”

(3) An alternative strategy of obtaining the amino-acid preference parameters via Bayesian inference (Lartillot and Philippe, 2004; Rodrigue and Lartillot, 2014) is discussed in the last section of the Results. Perhaps could be rephrased: an alternative strategy for dealing with amino-acid preferences across sites is to consider them formally, not as parameters, but as random-effects across sites, and to infer them using Bayesian inference.

This is a good suggestion. We have updated the text accordingly.

(4) when explaining the saturation problem, it could be useful to discriminate between sequence identity and sequence divergence, e.g.: a substitution model transforms sequence identity into branch length. I would say: “a substitution model transforms sequence *divergence* into branch length.” similarly: “the transformation of sequence *divergence* to branch length is trivial when sequence identity is high”.

We have updated the text to refer to sequence *divergence* rather than identity in this context, since we agree that this is a clearer way to explain the fundamental problem.

(5) The invariance of the asymptotic sequence identity under different schemes for modeling ω is a fundamental feature of the mathematics of *this type of* reversible substitution models.

This is a good point. We have revised the text as suggested to indicate that the statement applies to the types of substitution models discussed in the sentences preceding the statement in question.

Reviewer #2 Comments

Comments:

In this paper, Hilton and Bloom examine the performance of substitution models that explicitly incorporate experimental information about purifying selection. They find that they can estimate much longer branches in deep sections of phylogenetic trees, compared to traditional substitution models, and those that account for purifying selection (such as those implemented in PAML and HyPhy). I find that this is an interesting and important piece of work that is well written. Indeed, I find that the introduction is illuminating in explaining the difference between substitution models for protein coding data. I am happy to recommend this paper for publication, but I found some key aspects were neglected that can be improved with simple analyses and some clarifications in the text:

Thank you for the accurate and fair summary of our work.

Minor points:

In page 9 it is explained that a fixed tree topology was used for the subsequent analyses. I certainly agree with the authors that the model used in RAxML to estimate this tree probably underestimates branch lengths and that topological errors are negligible, but can they comment on how this could be tested using the ExpCM or codon models, given that they require substantially more computation? In other words, how suitable are these models to infer the tree topology, in addition to branch lengths? It is likely that most virus phylogenetic analyses can estimate a reasonable topology, but what about cases where there are very short branches next to very long branches (Felsenstein zone)? I do not recommend additional analyses for this part, but I think that discussing this would be very useful.

This is a highly relevant point when thinking about the larger context of our work.

As the reviewer noted, our current work only assesses how the site-specific models affect branch length, but not the topology of the tree itself. For the specific HA phylogenetic tree analyzed in our paper, the different models seem very unlikely to substantially affect tree topology since the tree is clearly divided into widely separated clades (a “modular” tree structure), meaning that even the nucleotide model in RAxML produces a reasonable tree. We have updated the text to more clearly explain why we think that the substitution model is unlikely to have much affect on topology *for the specific HA tree that we analyze*.

However, as the reviewer notes, for other trees the model may markedly affect the topology as well as the branch lengths. We now describe this point in detail, and refer to other work showing that models can affect the accuracy of the tree topology, particularly in the Felsenstein Zone (Lartillot et al., 2007). Given this work and some leeway for speculation, we would hypothesize that the ExpCMs would be less sensitive to the Felsenstein Zone than uniform stationary state codon models. However, directly testing this is beyond the scope of the current paper, since it requires the non-trivial computational implementation of software that can use the ExpCMs and estimate tree topology. We have therefore taken the reviewer’s suggestion and added careful discussion of this issue.

One key point in this study, which the authors acknowledge, is that estimating branch lengths accurately is key to infer evolutionary timescales. In this respect, is it possible to show for the simulations and empirical data how a root-to-tip regression (i.e. TempEst/pathogen) differs between the different substitution models? I would imagine that using more realistic models leads to better clocklike behaviour. However, it may be that a time-dependent rate phenomenon will be clearer, such that deeper branches will exhibit a predictably slower rate than those that are more recent.

This is an interesting suggestion. It is certainly true that a clock-like behavior is necessary in order to estimate evolutionary timescales. However, we do not believe that root-to-

tip regression is informative when applied to our datasets. The reason is that the HA sequences we use are from influenza viruses from several different hosts, such as humans, pigs, and birds. It has been shown that these hosts each have their own specific evolutionary rate (Worobey et al., 2014) (e.g., HA evolves *much* faster in humans than birds, probably due to immune selection). However, our current implementation does *not* incorporate host-specific rates, which therefore will confound clocklike behavior. We agree with the reviewer that in principle accounting for the substitution model issues discussed in our manuscript in conjunction with the host-specific rate issues discussed by others should improve clocklike behavior. Since there is not currently software that does both of these, we are unable to actually do this. However, we have added mention of this point in the Discussion.

Finally, the authors refer to the adequacy of substitution models. There has been some work in this field, also known as absolute model fit. This raises the question as to whether it is possible and valuable to assess the adequacy of these models, for example, using posterior predictive methods. Moreover, can the authors comment on the adequacy of the models described here for the flu empirical data?

Model adequacy is indeed a very interesting area of future research and we fully agree this area of phylogenetics deserves more attention. However, most model adequacy tests, including posterior-predictive methods, typically use the multinomial test (Goldman, 1993; Brown and Thomson, 2018). However, the multinomial test itself is not applicable to the models we use. The reason is that this test is explicitly designed for substitution models that are i.i.d. (independent and identically distributed) as stated by Goldman (1993) in his original introduction of these tests (second paragraph of page 184 of his paper). But the models we use are not identically distributed, since there is a different stationary state for each site. Therefore, the multinomial test cannot be used to evaluate model adequacy.

The general framework of adequacy tests should still be applicable to the models we use. However, new specific implementations other than the established multinomial test will be necessary to assess the adequacy of these models. We are indeed interested in developing computationally feasible implementations of model adequacy tests for codon models with site-specific stationary states, but have not yet succeeded in doing so. Therefore, we are not yet in a position to address this important question. So the best we can do for now is briefly mention the topic of model adequacy in the text.

References

- Brown JM, Thomson RC. 2018. Evaluating model performance in evolutionary biology. *Annual Review of Ecology, Evolution, and Systematics*. .
- Goldman N. 1993. Statistical tests of models of DNA substitution. *Journal of molecular evolution*. 36:182–198.
- Lartillot N, Brinkmann H, Philippe H. 2007. Suppression of long-branch attraction artefacts in the animal phylogeny using a site-heterogeneous model. *BMC Evolutionary Biology*. 7:S4.
- Worobey M, Han GZ, Rambaut A. 2014. A synchronized global sweep of the internal genes of modern avian influenza virus. *Nature*. 508:254.