Response to reviews of "Modeling site-specific amino-acid preferences deepens phylogenetic estimates of viral divergence" for *Virus Evolution*

Sarah K. Hilton and Jesse D. Bloom

August 16, 2018

Below, the reviewer comments are in blue, and our responses are in black. [Change title?] [Here is what we have to do for submission/revision: We revise the manuscript by clicking "Create a Revision" under "Action" in on the author page of the submission site. We cannot make revisions to the original submitted version and instead must upload a new file. We should highlight changes using colored or boldface text. We can respond to author comments in a submission box. Try to be specific in the response to the reviewers. Our review is due on October 9th.] [I don't know if we can upload this latex document but we can at least write out the responses here.]

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 favorable evaluation of the clarity of the work, which was furthered improved by the suggestions 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"

(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 reflect the difference between sequence identity and sequence divergence.

(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 true. We have clarified the text as suggested.

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 an interesting and pertinent point. As the reviewer noted, we do not currently have a way of directly testing the effect of ExpCMs on tree topology inference. However, the modular structure of the HA tree, along with sequence annotation, gives us confidence that the nucleotide model in `RAxML` produces a reasonable tree. But our confidence in the topology inferred by the nucleotide model is certainly not universal. We have updated the text both to underscore why we believe this specific scenario gives us the correct tree. We have also commented on work others have done to show that site-specific amino-acid preferences may infer more accurate tree topologies, specifically 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. But testing this hypothesis is outside the scope of this paper.

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 analyses which quantify clock-like behavior are appropriate for our datasets. The HA sequences we used our from several different hosts. It has been shown that these hosts have their own, specific evolutionary rate (Worobey et al., 2014) but these differences are *not* accounted for in our model. Therefore, while the HA tree is a useful test case to examine branch length estimation, we do not believe that our model will improve the clock-like behavior of the sequences.

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 are agree this area of phylogenetics deserves more attention. Up until recently, the majority of model adequacy tests, including posterior-predictive methods, used the multinomial test (Goldman, 1993; Brown and Thomson, 2018). We do not believe that this test is the most appropriate test for our model because it does not test site-specificity directly. Therefore, while we believe it is possible to assess the adequacy of these models, development of the most appropriate test statistic is outside the scope of this paper. While we agree that evaluating the adequacy of the model would be important when estimating evolutionary timescales, it is not necessary to test the specific question of this paper — do models with site-specific stationary states estimate longer branches than models with uniform stationary states?

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.