Biozentrum
Universität Basel
Klingelbergstrasse 70
4056 Basel, Switzerland

August 27th, 2020

Dear Editors,

Please find attached our manuscript entitled *Limited predictability of amino acid substitutions in seasonal influenza viruses.* We would be grateful if you considered it for publication in *Molecular Biology and Evolution.*

Surface proteins HA and NA of influenza viruses continually evolve to escape host immunity. This allows influenza to infect a significant fraction of the human population every year. A lot of effort has been directed at predicting this evolution, for example to optimize the influenza vaccine. Prediction methods are usually based on estimating fitness of circulating strains, and rely on two key assumptions: growth rate of strains or clades can be estimated from sequences, genealogy, or serology, and differences in growth rates persist long enough to be predictive of the evolution.

In this work, we challenge these assumptions by a retrospective analysis of the frequency trajectories of mutations in A/H3N2 and A/H1N1pdm populations. We quantified the extent to which frequency trajectories of mutations show sweep-like or "persistent" behavior, as would be expected in the case of mutations with positive fitness effects. We then investigated whether fixation or loss of such rising mutations can be predicted.

Surprisingly, we find that the predictability of these trajectories is very limited. A rapid rise of the frequency of a mutation in the past does not seem to contain any information about its future dynamics: on average, the frequency is as likely to keep rising as it is to decrease. Secondly, the probability that a new mutation fixes in the population differs little from its current frequency, as would be observed for neutral evolution. Furthermore, properties of sequences or of their genealogy that are usually used to estimate fitness, such as mutations at epitope sites or the *Local Branching Index (LBI)*, were found to convey little to no information on fixation probability. Finally, we find that a naïve predictor , the *consensus sequence*, performs as well as the *LBI.*

Taken together, these findings challenge our understanding of the evolution of influenza viruses.

Typical models of adapting populations tend to generate frequency trajectories that show persistent behavior, making them predictable for some time. This contrasts with our observations, suggesting that influenza evolution is poorly described by these models. We also conclude that previous methods to predict influenza evolution work, at least in part, by picking strains that represent the future population well, but not by predicting dynamics.

This manuscript is closely related to the recently published article by us entitled *Integrating genotypes and phenotypes improves long-term forecasts of seasonal influenza A/H3N2 evolution* by Huddleston *et. al.* While these two works may seem contradictory at first glance, they are in fact compatible and complementary. As opposed to the present manuscript which focuses on mutation dynamics and is model-free, Huddleston et. al. uses fitness models based on sequences and serology to predict the composition of future influenza virus population with the goal of improving vaccine selection strategies. The present manuscript thus represents a different and complementary approach for tackling the problem of influenza evolution, both in terms of methodology and of objectives.

Our manuscript sheds light on fundamental questions of viral evolution and has important implications for the predictability of influenza evolution. We thus think it is well suited for the broad readership of *Molecular Biology and Evolution.*

Sincerely,

Pierre Barrat-Charlaix (on behalf or all authors)