Review of: "Evolution of new variants of SARS-CoV-2 during the pandemic: mutation-limited or selection-limited?"

Kenichi Okamoto[1]

1 University of Saint Thomas (MN)

Potential competing interests: No potential competing interests to declare.

The following review applies to the first version of the manuscript.

This paper by Vajpai and Watve aims to evaluate whether mutation- or selection-limited diversification in a pathogen causes there to be qualitatively distinct dynamics. The authors use a combination of stochastic modeling and some data analysis involving SARS-CoV-2 epidemiological data to try and answer this question.

I confess to being slightly mystified by the author's chosen approach to interrogate their model with data. Under Fisher's theorems, the question the authors ask essentially boils down to whether genetic variance or the selection gradient better explains patterns of lineage diversity. Although the question is potentially of interest, I am wondering if lineage diversity is a useful indicator of the issue. Put differently, to fairly have a chance at attributing strain diversity to immune-driven selection, a quantification of phenotypic variation across strains seems warranted. Even with their model, I am not sure if mere strain diversity, per se, can answer their question. The authors allude to it, but neutral processes like genetic drift and geographic isolation can, I think, potentially explain the distribution of strains, and without more detailed records on (re-)infections, it seems on first glance there is an identifiability problem lurking here. There are also more systematic approaches to fitting time series to mechanistic models (e.g., https://doi.org/10.1038/s41586-019-1857-0) than the one the authors adopt here.

Setting the issue of data aside, the theoretical question about how qualitative dynamics might differ across the two hypotheses is of interest. Thus, I think the manuscript will benefit if the authors can be more transparent about the mechanics and implementation of their simulation, such as how the stochastic model works. At present, this is somewhat cryptic and it is difficult to infer the reasonableness of the simulation results. That said, provided the results withstand closer scrutiny, the fact that one can potentially tell whether lineage diversification is mutation or selection limited from epidemiological case data is a promising result.