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

Akane Hara[1]

1 University of Toyama

Potential competing interests: No competing interests exist concerning this article. (The affiliation of Akane Hara, Laboratory of Pharmaceutical Quality Assurance and Assessment (University of Toyama) is an endowed laboratory supported by an unrestricted grant from Nichi-Iko Pharmaceutical Co., Ltd. (Toyama, Japan).)

Review comments on the version available on September 15th, 2023

The authors proposed the three hypotheses on COVID-19 epidemiological dynamics: the invasion of the new virus variants is controlled by 1) advantageous mutation, 2) random mutation or drift, or 3) context-dependent selection considering the both conditions of the host immunization and the viral epitopes. The three hypotheses were examined in the simulations and the public data analysis.

The three hypotheses are interesting in terms of the viral evolutional mechanisms. The classification of the three corresponding situations observed in the public datasets, as shown in Fig.4, would be one of the meaningful directions of epidemiological research. However, it is confusing to me that the aim of the SIR model simulation is not clearly explained in the manuscript.

To clarify the role of the simulation, I would like to suggest to the authors confirming the following points.

1. What kind of qualitative features of the infection did the authors aim to examine by the simulation?

At first I thought that the authors performed the simulation to test whether the situations illustrated in Fig. 2 and Fig. 4 would appear or not in the prediction by the modified SIR model. However, I noticed that the authors may have not intended it, because the situation of Fig.2b seemed not to be described by the current model. I guess the peak reoccurrence of the prior variant (a solid line in Fig.2b) cannot be reproduced, since the supply of the new susceptible people to the prior variant was not considered in the model as far as I understand ($dS/dt = -K_1\ S(t)I(t)$).

Thus, I recommend the authors to add clear explanations on the features that are intended to reproduce in the simulation.

2. The simulation method needs to be described in detail. How was the number of people susceptible to the newly appeared variant calculated? Was "1-I(t)" used as the initial value of $S_1(t)$, and then $S_1(t)$ was numerically calculated following "$dS_1/dt = -K_1\ S_1(t)\ I_1(t)$" ? Otherwise, "1-I(t)" was calculated as $S_I(t)$ at every time step? The latter way of calculation may enable the increase in susceptible population to new variant. This question may be related to the comments on question 1 about the detail of the calculation of the susceptible compartment.

3. In the simulation, it was assumed that "The probability of generating a new variant*V1* (and subsequently Vn) at a given

time is directly proportional to I ", which may directly lead to the result: "most new variants invade near the peaks of prior variants" (Fig.2). What kind of the concept were implicated in the assumption? Genetic mutation due to replication error usually occur independently from the number of the infected people, as far as I concerned about the basic mechanism. The meaning of the word "The probability" need to be clearly explained.

The minor comments are as follows:

- Does "o test the predictions..." mean "To test the predictions..." ? (at the beginning of the paragraph after the Fig.3)
- Two URLs in Data source did not work and need to be modified.