

Phylogenetic analysis of the SARS-CoV-2 spike variants

Reviewed by Mudith Ekanayake

Introduction to the project is well explained and the flow of the paper is well maintained. I have some comments for the materials and methods section. You have done a good job in this section. It is better if you can cite the appropriate papers for the tools you have used. For example, paper for MAFFT is Katoh *et al.*, 2002 (Katoh, et al. "MAFFT: a Novel Method for Rapid Multiple Sequence Alignment Based on Fast Fourier Transform." *OUP Academic*, Oxford University Press, 15 July 2002, academic.oup.com/nar/article/30/14/3059/2904316). You can find these from the tutorials or by just searching in the web.

In addition to the steps you have carried out in the project, you can perform model selection using tools such as ProtTest, JmodelTest or SMS. Furthermore if you want to improve your alignments by removing poorly aligned regions you can trim your alignments by carrying out GBlocks or TrimAl. These are just additional steps for generating better output results. In the RAxML-NG phylogenetic reconstruction step, you need to mention the model you have used for constructing the phylogeny.

You have done a good job explaining the phylogenetic tree. In this section you can include the AIC/BIC scores as well. You can find them in the bottom of the RAxML-NG log file. You have nicely concluded it by addressing your question of interest. In conclusion, I found this paper well formatted and well explained. Good luck with the project and the final presentations.