

Rapid Reviews COVID-19 •

Review 1: "Generation and transmission of inter-lineage recombinants in the SARS-CoV-2 pandemic"

Graziano Pesole¹, Matteo Chiara

¹University of Bari: Università degli Studi di Bari Aldo Moro Department of Biosciences, Biotechnology and Biopharmaceutics ITALY"

Published on: Nov 18, 2021

License: [Creative Commons Attribution 4.0 International License \(CC-BY 4.0\)](https://creativecommons.org/licenses/by/4.0/)

RR:C19 Evidence Scale rating by reviewer:

- **Reliable.** The main study claims are generally justified by its methods and data. The results and conclusions are likely to be similar to the hypothetical ideal study. There are some minor caveats or limitations, but they would/do not change the major claims of the study. The study provides sufficient strength of evidence on its own that its main claims should be considered actionable, with some room for future revision.

Review:

In their manuscript, Jackson et al describe an accurate and stringent workflow for the identification of possible hints of recombination in SARS-CoV-2. Applied to a large collection of more than 279k genomes from the COG-UK consortium the method identifies a total of 16 candidate recombinant sequences. Different lines of evidence are provided by the authors to support the "recombinant" nature of these sequences, including the co-circulation of the lineages that form the candidate recombinant genomes in the same geographic area and interval of time, the lack of supporting evidence for co-infection in the samples from which "recombinant" genome assemblies were reconstructed, and more importantly the fact some of the recombinant sequences detected by their approach seem to be associated with community transmission.

The detection of recombination from NGS sequencing data is a very challenging task, and especially for SARS-CoV-2 since its relatively slow evolutionary rate. Currently, the extent to which recombination is ongoing for SARS-CoV-2 is not yet resolved, and different studies have reported different and sometimes contrasting conclusions see De Maio et al., 2020; Van Dorp et al., 2020; Nie et al., 2020; Tang et al., 2020; Wang et al., 2020 compared to Varabyou et al. (2020) and VanInsberghe et al. (2020), for example.

For this reason, the development of an accurate and reproducible method for the detection of recombination in SARS-CoV-2 would be required to understand and study the extent (if any) to which recombination is playing an important role in the evolution of this novel pathogen.

The method proposed by Jackson et al is promising, however, I have some potentially relevant concerns which might need to be addressed.

i) the authors report a relatively reduced number of recombinant sequences. This observation is in line with previous reports. However, the "detection power" and sensitivity of the approach proposed by Jackson et al (and of any other method that was applied previously) is not known at present. Common sense, however, suggests that a relevant proportion of recombinant sequences might be missed by this approach (due to a low level of variability). Although it is not ideal or optimal, I would suggest that the authors should perform an "in-silico" simulation to provide a lower bound estimate of the sensitivity of their method. Recombinant sequences could be generated easily in silico by admixing genomes assigned to different lineages. Additionally, previous knowledge on the most highly recombinant segments of the genome in CoVs could be used to make the simulations more realistic (see Boni et al 2020, partially from the same authors)

ii) A clear limitation of the method is that it can only detect inter-lineage recombination, this should be addressed and discussed. Since the prevalence of the B.1.1.7 increased dramatically during the interval of time considered by the authors, this also means that by definition their ability to detect recombination decreases with time. This should also be discussed.

iii) When ruling out the possibility that the candidate recombinant sequences could be the result of a mixed assembly from a sample associated with the co-infection of 2 different lineages authors state: "Firstly, the sequencing protocol used in the UK (Tyson et al. 2020) generates 98 short (~350bp) amplicons, such that long tracts that match just one lineage would be unlikely." IMHO this is not very scientific: what is it meant by unlikely? Can this be quantified? If not, it is pure speculation.

iv) I know that it is very unlikely, but on such high numbers can the possibility of barcode bleeding be completely ruled out? if so please explain how and why.

v) The variability in the genome of SARS-CoV-2 is relatively low. In the light of this consideration, it might well be that some of the recombinant segments identified by the authors in their candidate recombinant genomes could be supported only by a limited number of genetic variants. As illustrated in Figure 2, this is not the case for group A, where the two segments that compose the genome can be clearly discriminated. If possible, I would suggest the authors provide equivalent information also for the other groups identified by their analyses.

vi) By reading the methods section it seems that the proposed approach for the identification of candidate recombinant sequences is not completely automated, and

requires some manual intervention or curation for the definition of the breakpoints. This might represent a potential limitation of the method in the application of the methods to large cohorts of samples. The authors should clarify this point and discuss its implications. Additionally, authors should illustrate which criteria were used to identify "defining" variants and long contiguous tracts of B.1.1.7 and non-B.1.1.7 genomes in a more clear manner.