Review of: "A Comparison of Performance for Different SARS-Cov-2 Sequencing Protocols"

Xuhua Xia[1]

1 University of Ottawa

Potential competing interests: No potential competing interests to declare.

This is an interesting topic. My rating of the manuscript reflects my belief that research along this direction is important.

In terms of long-reads versus short reads, it has been pointed out before (https://doi.org/10.1534/g3.117.300271) that there is no fundamental technological revolution leading to long-reads from short-reads. The longer reads generated from Illumina sequencing machines, the poorer the sequence quality of the reads. When I download NGS transcriptomic data with a read length of 250 nt, I typically would use only the first 150 nt or even the first 100 nt for characterizing genetic variation.

Similarly, for paired reads, the forward reads are of higher quality than the reverse reads, so I often ignore the reverse reads. I tend to prefer single reads if they are available.

The author highlighted the point that reference-based assembly tends to overestimate the variant represented by the reference. This point is well-made. Being aware of this point would alleviate the problem in the future, i.e., more effort would be spent in checking the possibility of horizontal gene transfer and the like that leads to the incorporation of a less related piece of sequence into a viral genome.

I am not saying that the manuscript has addressed all the questions it has raised. It is perhaps fair to label the manuscript as a rough draft. However, while the call for more studies along this line has been made before, it is always good to broadcast the call to a wider audience.