

OPEN REVIEW COMMENTS ON THE PAPER:

Genome Detective Coronavirus Typing Tool for rapid identification and characterization of novel coronavirus genomes. bioRxiv (2020), doi: <https://doi.org/10.1101/2020.01.31.928796>.

Authors: Cleemput S, Dumon W, Fonseca V, Abdool Karim W, Giovanetti M, Alcantara LCJ, Deforche K, de Oliveira T,

Here are the comments of the reviewers:

Reviewer: 1

Comments to the Author

General Comments: This paper describes a tool for assembling Betacoronavirus genomes from short read sequence data, and genotyping the genomes if they are within the SARS and SARS-like Betacoronaviruses, but not in the MERS family. It is of great importance to have this tool available during the current 2019-nCoV outbreak recently originating near or in Wuhan, China. Expansion of the tool to include MERS and related viruses could be done in the future, but the immediate need is to cover the 2019-nCoV outbreak.

I was able to "test drive" the tool with several SARS-related Betacoronavirus genomes and a 2019-nCoV genome and the tool was easy to use and performed very well. The classification and nomenclature for the 2019-nCoV is currently under review by the International Committee for the Taxonomy of Viruses (ICTV) Nidovirales Study Group, but it clearly falls well within the SARS and SARS-like clade of Betacoronaviruses often labeled as "Sarbecovirus". Recombination between virus lineages within this group, makes further classification of most of the bat strains problematic, but the human outbreak lineages such as 2003-SARS and 2019-nCoV are important to human public health and clearly each a monophyletic lineage.

The MERS Betacoronaviruses are in a separate clade rather than in the "Sarbecovirus" clade. But perhaps still close enough that very little change to the tool is needed to cover them. The paper is fine as it is, only the tool could be modified at some future time.

Reviewer: 2

Comments to the Author

General comments:

This paper describes a new additional function to the existing Genome Detective free, online viral analysis tool. Genome Detective provides a free, easy-to-use interface to allow classification and basic analysis of viral sequences. It has now been updated to include SARS-related coronaviruses – both the strains related to the 2003 outbreak, closely related strains often found in animals, and the current viruses (SARS-CoV-2) responsible for the ongoing epidemic. The new function of Genome Detective is timely and important, particularly for labs, clinicians, and health systems who may not be as familiar with sequencing pathogens and may benefit from easy-to-use systems which allow them to quickly perform classification and basic analysis.

This reviewer would like to particularly thank the authors for clearly and consistently listing the accession numbers (where applicable) for all the sequences mentioned. It is incredible how many papers fail to do this, and it makes reproduction or further investigation of papers and studies unnecessarily difficult!

Specific comments (all minor):

It is always tough when writing about such an emerging pathogen, but a quick update to the introduction should be performed, just to bring up-to-date the latest numbers on spread, fatality, etc. Also, the new virus name (SARS-CoV-2) and disease name (COVID-19) should be replaced throughout the document. This is a moving target and it's almost inevitable that such things will change between submission and release, so this is not at all a reflection of the paper.

Some of the sequences that have been released publicly on GISAID seem to have sequencing error leading to divergence from the majority of SARS-CoV-2 sequences of more than 40 mutations. Could the authors comment on how Genome Detective would handle such a sequence if it were submitted? Could it correctly identify that the genome seems to be more diverged than reasonable? Would it be classified into one of the sister-clusters?

There are also sequences available that are not whole-genome sequence (ex: some of just over 1,000 bases from Hong Kong). How does Genome Detective handle these, if these were submitted for some reason? The discussion does say that Genome Detective can take "whole or partial viral genomes" – but how accurate is classification and the CDS/mutation analysis for non-WGS sequences?

If it would not be too onerous, it would be very interested to see the section on the spike protein differences (page 4) expanded to comment perhaps on the similarity to the sequences isolated from Pangolins, which has currently caught some attention. Would this be shown by Genome Detective, and what does it add to this discussion?

Small typos:

Line 52 page 1 – Typo – 'a global health emergency' (not an)

Line 7 page 3 – italics runs on until end of sentence, seems unintentional.

Line 14 page 4 & line 29, page 4 – Typo – 'Envelop Small Membrane..' (instead of Envelope)

Reviewer: 3

Comments to the Author

1) The manuscript describes a new public access resource allowing to identify and analyse SARS-CoV-2 sequences, in a fast and accurate way. The tool is fast, accurate and easy to use.

2) b) Figure 1 is not clear without reading the legend. Adding some text in the figure might achieve a better understanding of it.

- The new name SARS-CoV-2 could be added in the manuscript. Page 3 speaks of possible SARS-A and B, which was a good guess, those can be replaced with SARS-1 and SARS-2.
