



Omar Ahmed Mahmoud

Third year [Medical informatics Department]

Decoding Covid-19 with the SARS-CoV-2 Genome Analysis and

Visualization: Development and Usability study

Related Works

Human CoVs Are Animal RNA Viruses Grouped in the Coronaviridae Family:

Human CoVs are zoonotic pathogens derived from animal CoVs and mainly cause respiratory diseases. All CoVs have non-segmented, large single-stranded, positive-sense RNA genomes that have a similar organisation of non-coding untranslated regions (UTRs) and coding regions or open reading frames (ORFs). They can be categorised into four separate genera according to their genome and protein sequences: alpha, beta, gamma and delta. Alpha and beta-CoVs affect humans and other mammals whereas the gamma and delta strains affect mainly birds. There are now seven known human CoVs: HCoV-229E (alpha), HCoV-NL63 (alpha), HCoV-OC43 (beta), HCoV-HKU1 (beta), MERS-CoV (beta), SARS-CoV (beta) and the most recently discovered SARS-CoV-2 (beta). All seven human CoVs have zoonotic origins linked to bats, mice or domestic animals

SARS-CoV-2 Derives from Bat CoVs and Is Closely Related to SARS-CoV

New CoVs are thought to arise from complex recombination events when two related viral genomes are found within the same cell—these events most often occur in non-human mammal species such as bats, resulting in progeny viruses that acquire the ability to infect human cells . Thus, CoV genomes often retain key features of their ancestral virus but also include new features that allow for the species jump. Not surprisingly, the genome organisation of SARS-CoV-2 is largely similar to

that of the existing human CoVs and, in particular, to SARS-CoV which is the human CoV with the most identity at a nucleotide level. The SARS-CoV-2 genome, a single-stranded positive-sense RNA molecule of approximately 29,800 nucleotides, is arranged into 14 open reading frames (ORFs) encoding 27 proteins and is shown schematically.

The majority of the genome contains the ORF1a and ORF1b that encodes 16 different non-structural proteins (nsp1-nsp16) involved in the 'replicase' complex although many have very diverse but critical functions. The final one-third of the genome houses several ORFs encoding 4 structural (spike, S; envelope, E; membrane, M; nucleocapsid, N) and 10 accessory proteins (ORF3a, ORF3b, ORF6, ORF7a, ORF7b, ORF8a, ORF8b, ORF9b, ORF9c, ORF10). Some of these ORFs are overlapping or found within a larger ORF. At either end of the genome are non-coding or untranslated regions (UTRs) known as the 5'UTR and 3'UTR. These UTRs are relatively short, being approximately 230 bases, but have important regulatory functions. The 5'UTR is highly structured and thought to contain 5 stem-loop (SL1–SL5) structures, with the SL3 structure also housing the transcriptional regulation sequence (TRS-L) important for generation of subgenomic mRNAs. Each gene in the genome also has its own upstream TRS-B sequence.

References

1.

Xu J, Zhao S, Teng T, Abdalla AE, Zhu W, Xie L, et al. Systematic comparison of two animal-to-human transmitted human coronaviruses: SARS-CoV-2 and SARS-CoV. *Viruses*. 2020;12(2).
[.https://doi.org/10.3390/v12020244](https://doi.org/10.3390/v12020244)

2.

Zhou P, Yang X-L, Wang X-G, Hu B, Zhang L, Zhang W, et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin.

Nature. 2020;579(7798):270–3. <https://doi.org/10.1038/s41586-020-2012-7>

3.

Petersen E, Koopmans M, Go U, Hamer DH, Petrosillo N, Castelli F, et al. Comparing SARS-CoV-2 with SARS-CoV and influenza pandemics. *Lancet Infect Dis.* 2020;20(9):e238–e44. [https://doi.org/10.1016/s1473-3099\(20\)30484-9](https://doi.org/10.1016/s1473-3099(20)30484-9)