

Algorithms for Bioinformatics / Bioinformatics

Virus

Pedro G. Ferreira
pgferreira@fc.up.pt

- A virus is a tiny, infectious particle that can reproduce only by infecting a host cell.
- It is estimated that at any given moment there are 10^{31} viruses on earth (1). Most live in oceans and infect bacteria.
- They cannot reproduce by themselves and they use the cell machinery of the host to make more viruses.
- Viruses can mutate, generate genetic variation and evolve.

(1) Weitz, J. S., and Wilhelm, S. W. (2013, July 1). An ocean of viruses. In The scientist. Retrieved from <http://www.the-scientist.com/?articles.view/articleNo/36120/title/An-Ocean-of-Viruses/>.

- The virus does not have cells and are small packages of a DNA or RNA genome inside a protein shell called a capsid. Some viruses have an external membrane envelope.
- The genetic code of viruses is similar to the one in living organisms.

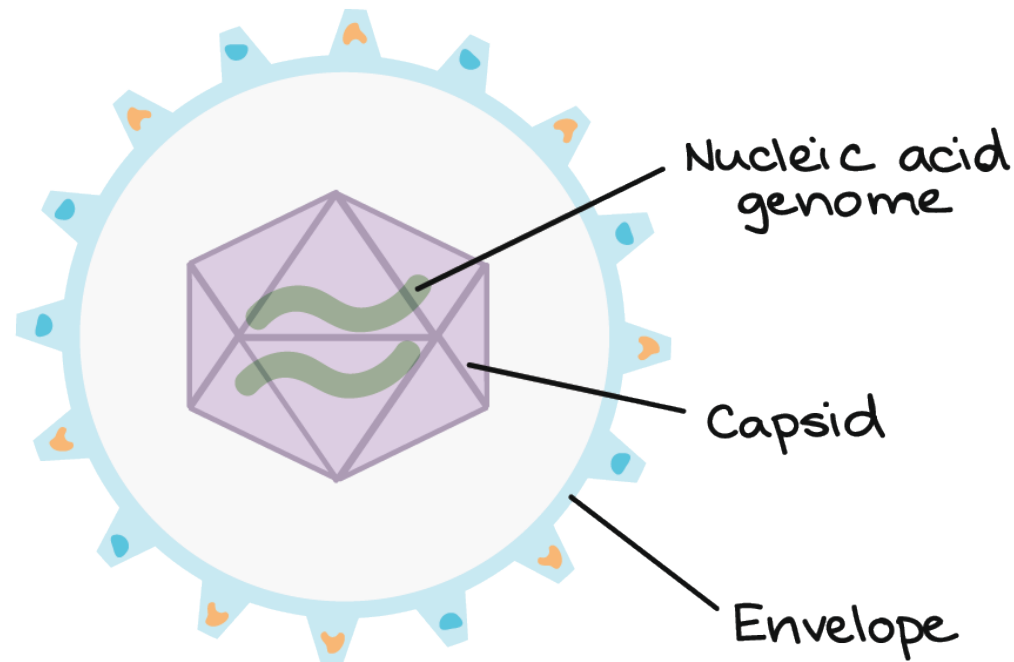


Image modified from "[Scheme of a CMV virus](#)." by Emmanuel Boutet, [CC BY-SA 2.5](#). The modified image is licensed under a [CC BY-SA 2.5](#) license.

- Capsids are protein shells made of proteins (capsomers) encoded by the virus genome. They have many forms and shapes:

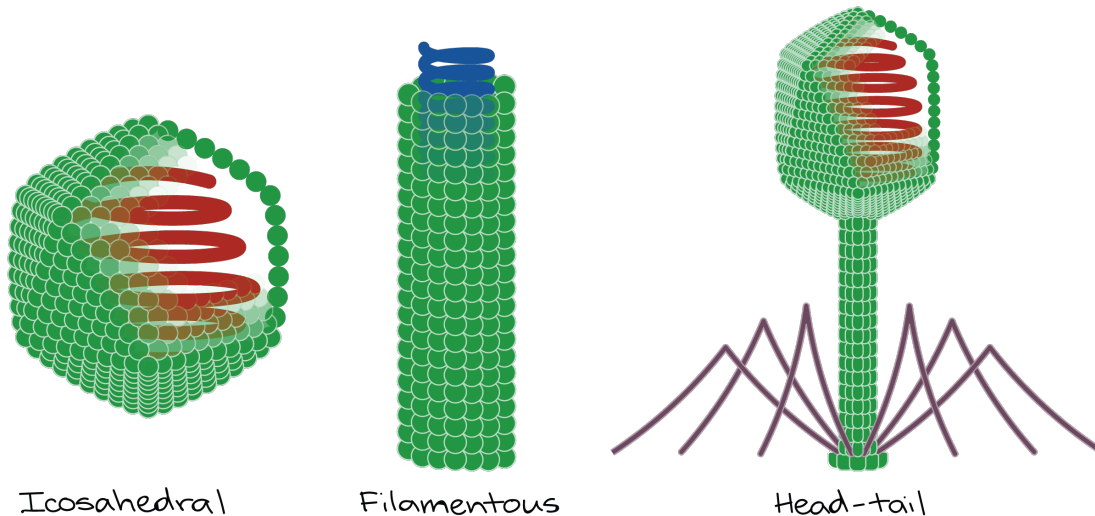
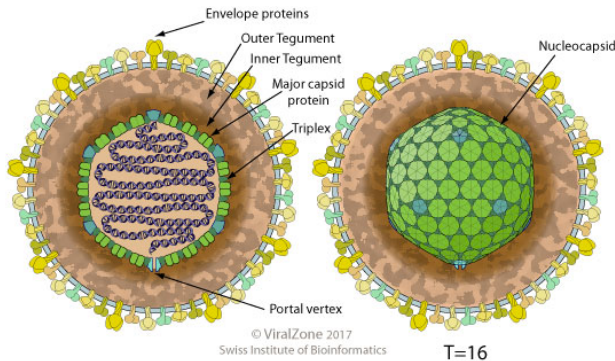


Image modified from "[Non-enveloped icosahedral virus](#)," "[Non-enveloped helical virus](#)," and "[Head-tail phage](#)," by Anderson Brito, [CC BY-SA 3.0](#). The modified image is licensed under a [CC BY-SA 3.0](#) license.

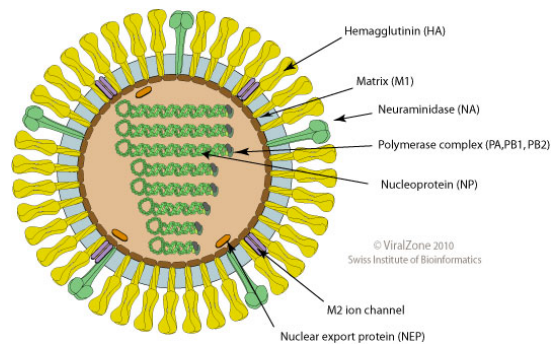
Virus Capsids

- Viruses have many sizes, shapes, and life cycles.

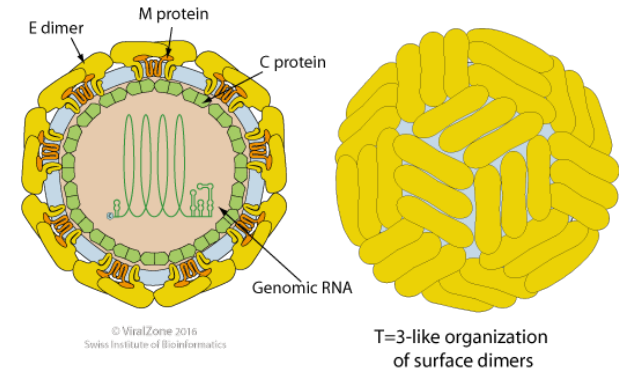
<https://viralzone.expasy.org>



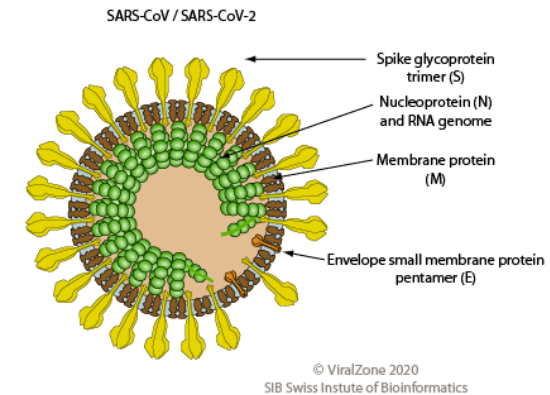
- Epstein-Barr virus has a linear, dsDNA genome of about 180 kb.



- Alphainfluenzavirus is a segmented ssRNA(-) linear genome, encapsidated by nucleoprotein (NP). Segments size range from 890 to 2,341nt. Genome total size is 13.5Kb



- The Zika virus is a monopartite, linear, ssRNA(+) genome of 10-11 kb.

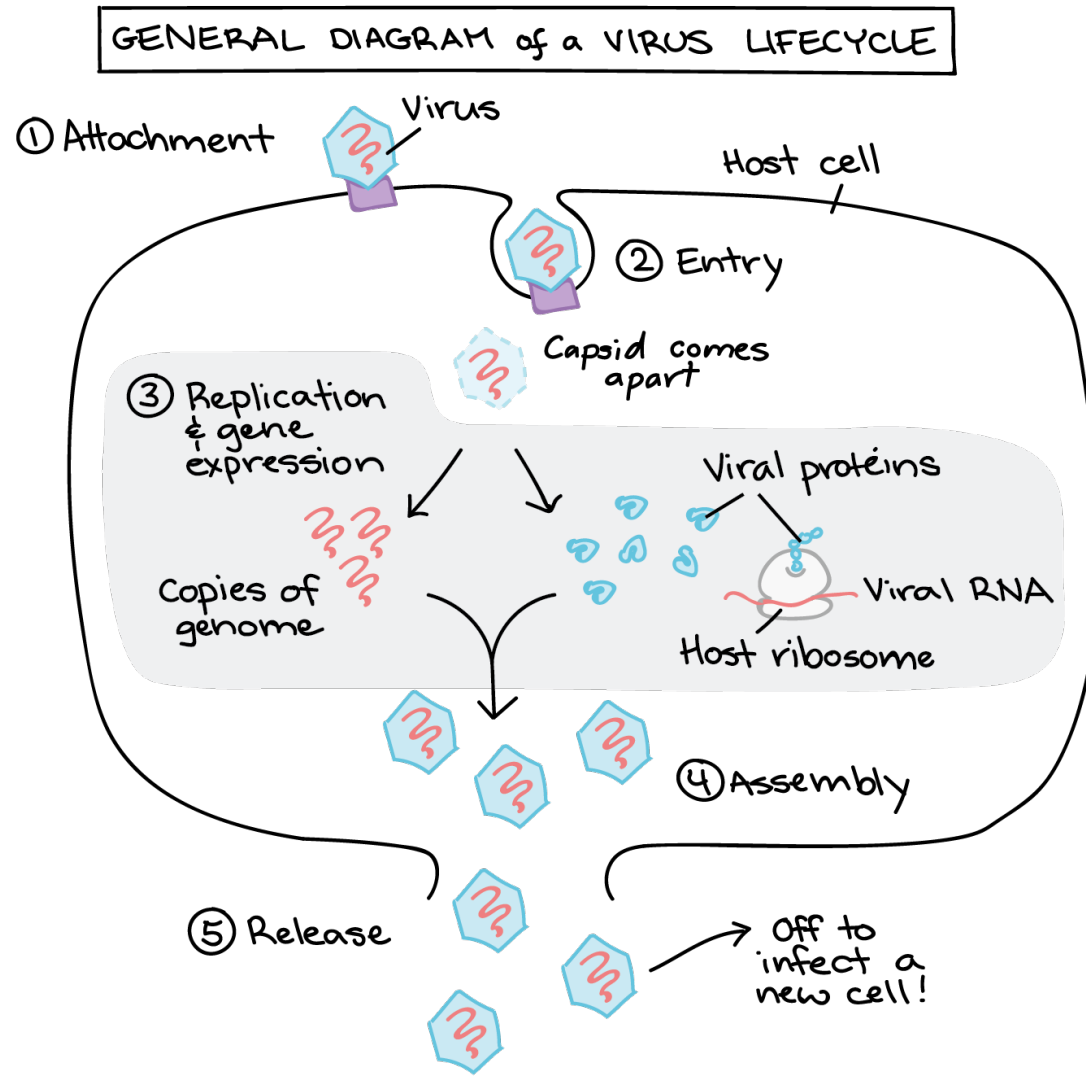


- Betacoronavirus is a monopartite, linear ssRNA(+) genome of 27-32kb in size (the largest of all RNA virus genomes).

- Virus use nucleic acids as their genetic material, so they may use either RNA or DNA.
- Virus genomes \sim 2k bps to hundred k bps in length (1).
- They present several combination of strandeness and nucleic acids: double-stranded DNA, double-stranded RNA, single-stranded DNA, or single-stranded RNA (2).

(1) Weitz, J. S., and Wilhelm, S. W. (2013, July 1). An ocean of viruses. In The scientist. Retrieved from <http://www.the-scientist.com/?articles.view/articleNo/36120/title/An-Ocean-of-Viruses/>.

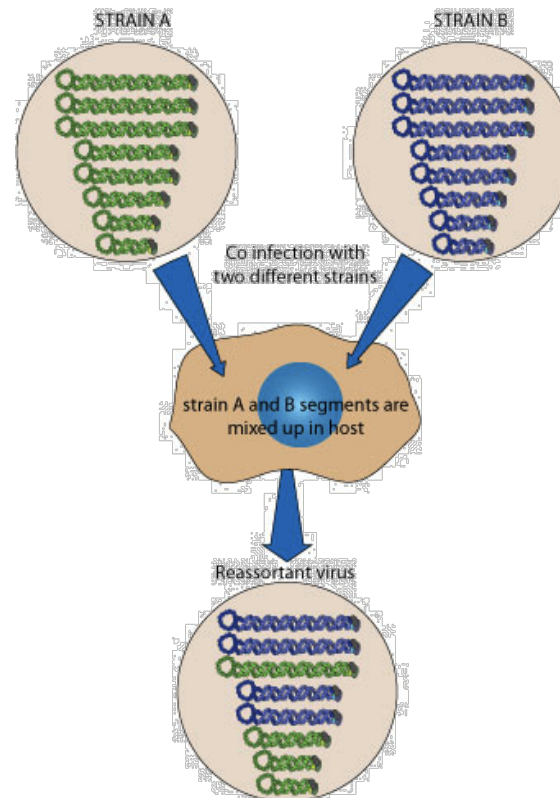
(2) <https://www.khanacademy.org/science/biology/biology-of-viruses/virus-biology/a/intro-to-viruses>



- Viruses undergo evolution and natural selection. The composition and structure of viral genes in a viral population can change over the time.
- This may confer advantages to viruses to replicate like higher infectivity (e.g. influenza) or resistance to drug (e.g. HIV).
- These benefic changes tend to become more common in the viral population over time.
- The rate of evolution in virus can be much faster than in the host.

Adapted from <https://www.khanacademy.org/science/biology/biology-of-viruses/virus-biology/a/intro-to-viruses>

- **Recombination:** viruses swap chunks of genetic material (DNA or RNA).
- **Mutations:** random changes that occur in the DNA (lower rate) or RNA (higher rate) sequence of a virus.



- Mutations can have different effects.
- Variants are viruses that show one or more mutations.
- Strain refers to a variant with a distinct physical properties of the parental strain leading to a different behavior than the parent virus.

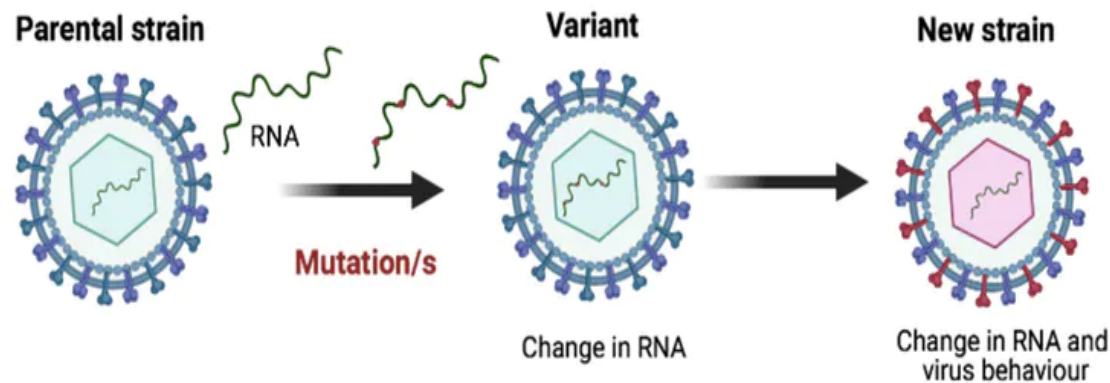
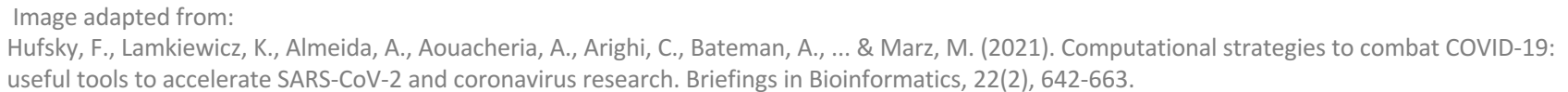


Image credit: Lara Herrero, created using BioRender adapted from <https://theconversation.com/whats-the-difference-between-mutations-variants-and-strains-a-guide-to-covid-terminology-154825>

- Virus evolve faster than humans.
- A higher mutation rate allows them to evolve quickly by providing more variation as a starting material.
- Population size and virus lifecycle are the other elements for rapid evolution:
 - The bigger the population the higher the probability that mutations arise.
 - Virus reproduce quickly ~52 hours than their hosts.

- The group of *Coronaviridae* includes viruses with very long RNA genomes, from 27kbp to 33kbp.
- Members from coronavirus family are enveloped, with a long positive-sense, single-stranded RNA genomes.
- Encode 5 major ORFs.
 - 5' frameshifted polyprotein ORF1a/ORF1ab
 - 4 canonical 3' structural proteins
 - Additional subgroups accessory genes are found interspersed among and even overlapping structural genes.
 - Accessory genes
 - are in variable number in coronavirus.
 - Not essential for virus replication
 - Involved in pathogenicity

- SARS-CoV2:
 - belongs to subgenus of *Sarbecovirus*.
 - Genome ~30Kbps.
 - 4 Main structural proteins:
 - Spike (S)
 - Envelope (E)
 - Membrane (M)
 - Nucleocapsid (N)
 - Several nonstructural proteins are encoded in pp1a and pp1ab polyproteins.



Choose Columns													Page 1 of 1			
#	Name	Accession	Start	Stop	Strand	GeneID	Locus	Locus tag	Protein product	Length	Protein Name					
1	viral segment	NC_045512.2	266	21555	+	43740578	ORF1ab	GU280_gp01	YP_009724389.1	7096	ORF1ab polyprotein					
2	viral segment	NC_045512.2	266	13483	+	43740578	ORF1ab	GU280_gp01	YP_009725295.1	4405	ORF1a polyprotein					
3	viral segment	NC_045512.2	21563	25384	+	43740568	S	GU280_gp02	YP_009724390.1	1273	surface glycoprotein					
4	viral segment	NC_045512.2	25393	26220	+	43740569	ORF3a	GU280_gp03	YP_009724391.1	275	ORF3a protein					
5	viral segment	NC_045512.2	26245	26472	+	43740570	E	GU280_gp04	YP_009724392.1	75	envelope protein					
6	viral segment	NC_045512.2	26523	27191	+	43740571	M	GU280_gp05	YP_009724393.1	222	membrane glycoprotein					
7	viral segment	NC_045512.2	27202	27387	+	43740572	ORF6	GU280_gp06	YP_009724394.1	61	ORF6 protein					
8	viral segment	NC_045512.2	27394	27759	+	43740573	ORF7a	GU280_gp07	YP_009724395.1	121	ORF7a protein					
9	viral segment	NC_045512.2	27756	27887	+	43740574	ORF7b	GU280_gp08	YP_009725318.1	43	ORF7b					
10	viral segment	NC_045512.2	27894	28259	+	43740577	ORF8	GU280_gp09	YP_009724396.1	121	ORF8 protein					
11	viral segment	NC_045512.2	28274	29533	+	43740575	N	GU280_gp10	YP_009724397.2	419	nucleocapsid phosphoprotein					
12	viral segment	NC_045512.2	29558	29674	+	43740576	ORF10	GU280_gp11	YP_009725255.1	38	ORF10 protein					
Choose Columns													Page 1 of 1			

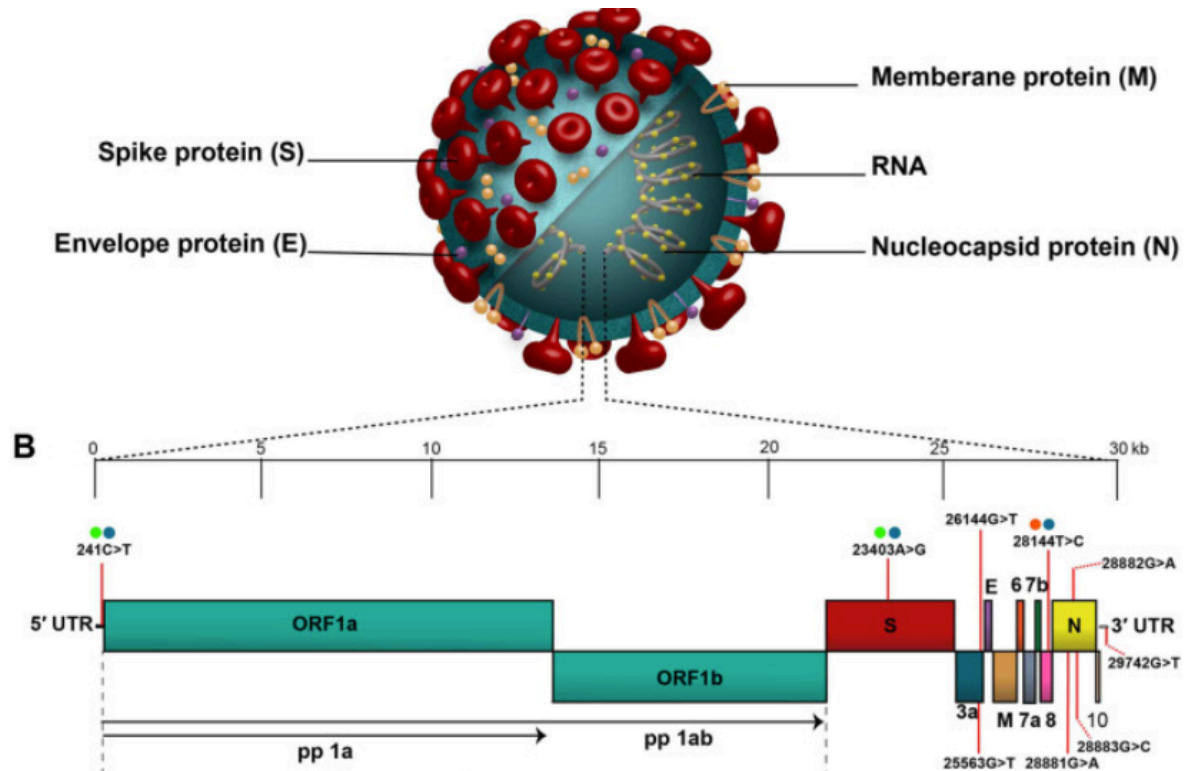


Image adapted from:

Rahimi, A., Mirzazadeh, A., & Tavakolpour, S. (2021). Genetics and genomics of SARS-CoV-2: A review of the literature with the special focus on genetic diversity and SARS-CoV-2 genome detection. *Genomics*, 113(1), 1221-1232.

- SARS-CoV2 uses the human receptor ACE2 as point of entry in the cell.
- The binding domains for ACE2 are located Spike Proteins.
- The viruses has a mutation rate lower than other RNA viruses.
- Mutations accumulate and lead to genomic diversity between infected patients.
- Genetic heterogeneity enables viral adaptation to different hosts and different environments within hosts.
- Mutations are associated disease progression, drug resistance and treatment outcome.

○ Further Readings:

- Hufsky, F., Lamkiewicz, K., Almeida, A., Aouacheria, A., Arighi, C., Bateman, A., ... & Marz, M. (2021). Computational strategies to combat COVID-19: useful tools to accelerate SARS-CoV-2 and coronavirus research. *Briefings in Bioinformatics*, 22(2), 642-663.
- Rahimi, A., Mirzazadeh, A., & Tavakolpour, S. (2021). Genetics and genomics of SARS-CoV-2: A review of the literature with the special focus on genetic diversity and SARS-CoV-2 genome detection. *Genomics*, 113(1), 1221-1232.
- Michel, C. J., Mayer, C., Poch, O., & Thompson, J. D. (2020). Characterization of accessory genes in coronavirus genomes. *Virology Journal*, 17(1), 1-13.