INTERNATIONAL UNIVERSITY OF SARAJEVO

FACULTY OF ENGINEERING AND NATURAL SCIENCE

DEPARTMENT OF GENETICS AND BIO-ENGINEERING

**BIO402: Molecular Evolution**

**Logo

Description automatically generated**

**Submitted to: Asst. Prof. Dr. Mr. Mohamed Adilović**

**Submitted by: Amina Karić, Kishor Mani Karunakaran, Majed Alwis**

**and Nadia Islam**

9th January 2022 Sarajevo

***Disclaimer***

*Amina was responsible for discussing similarities and different variants, Kishor did the gaps, Majed discussed mutation on 20 strains, and the rest of the paper was organized and written by Nadia Islam.*

*However, the members do not wish to take an individual approach, rather we take full responsibility entirely as a group and give full consent upon teacher’s discretion regarding grading.*

Table of Contents

[**1.** **Introduction** 4](#_Toc92657974)

[**1.1.** **SARS-CoV-2** 4](#_Toc92657975)

[**1.2.** **Mutations of SARS-CoV-2** 5](#_Toc92657976)

[**1.2.1.** **Alpha variant B.1.1.7.** 5](#_Toc92657977)

[**1.2.2.** **Delta variant B.1.617.2** 6](#_Toc92657978)

[**1.2.3.** **Omicron variant B.1.1.529.** 7](#_Toc92657979)

[**1.3.** **Databases and Tools** 7](#_Toc92657980)

[**2.** **Method** 8](#_Toc92657981)

[**3.** **Result and Discussion** 8](#_Toc92657982)

[**3.1.** **Gaps** 10](#_Toc92657983)

[**3.2.** **Mutation** 11](#_Toc92657984)

[**3.3.** **Similarities** 13](#_Toc92657985)

[**3.4.** **TT Ratio** 13](#_Toc92657986)

[**4.** **Observation** 17](#_Toc92657987)

[**5.** **Conclusion** 18](#_Toc92657988)

[**Reference** 18](#_Toc92657989)

Strain Comparison of SARS-CoV-2

1. **Introduction**

In 2019 the world experienced something it never experienced before. The entire world was shut down for a virus called SARS-CoV-2, people were terrified and global market economies went crashing down. As of December 2021, over 5 million people died worldwide when this paper was written (*Cumulative Confirmed COVID-19 Deaths*, n.d.). It felt like an Armageddon or maybe a scene from a Hollywood Sci-Fi movie.

Chart

Description automatically generated

**Figure 4‑1: Confirm SARS-CoV-2 Death Worldwide, Source: John Hopkins University CSSE COVID-19 Data- Last updated 13 December, 08:05 (London time), (Cumulative Confirmed COVID-19 Deaths, n.d.)**

However, the virus did not just pop up in 2019. It was here on earth for a long time before the Global Pandemic started.

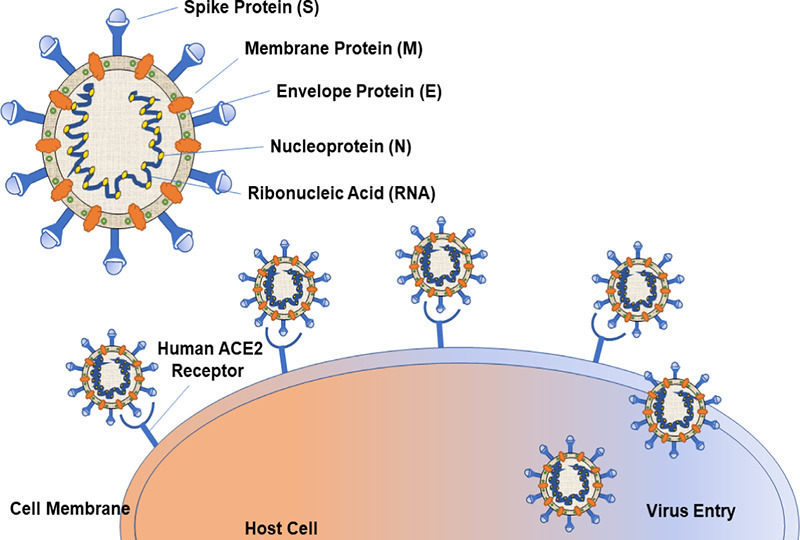
* 1. **SARS-CoV-2**

COVID-19 is a pneumonia like acute respiratory syndrome with symptoms like cough, upper airway congestion, myalgia, headache and intermittent fever (Chang et al., 2020) caused by the virus called SARS-CoV-2 which was first detected in Wuhan, China in December 2019. The virus was previously called 2019-nCoV- before the International Committee on Taxonomy of Viruses changed the name into SARS-CoV-2. Various studies have shown that the viruses are “are a group of enveloped pathogenic viruses, with a positive single-stranded RNA genome of approximately 30,000 bases with 5′-cap structure and 3′-poly-A tail, belonging to the Coronaviridae family of the order Nidovirales, and this subfamily includes four genera: alphacoronavirus (α), betacoronavirus (β), gammacoronavirus (γ), and deltacoronavirus (δ)” (Biscayart et al., 2020; Giovanetti et al., 2021; Naqvi et al., 2020).

After the Chinese government informed World Health Organization (WHO) about the pneumonia like new respiratory disease in 2019, WHO declared that this “epidemic is a public health emergency of international concern (PHEIC)” (M.-Y. Wang et al., 2020).

SARS-CoV-2 has high transmission efficiency (Biscayart et al., 2020) and symptoms similar to two bat-derived severe acute respiratory syndrome bat-SL-CoVZC45 and bat-SL-CoVZXC2 and it belongs to the “betacoronavirus 2B lineage" (Lai et al., 2020). The virus can infect “respiratory, gastrointestinal, hepatic and central nervous system tracts” among human and other various species (Biscayart et al., 2020; Zhou et al., 2020). Although the disease is thought to be transmitted from wild animals to human, recent studies have shown that it is transmitted through human-to-human contact (Li et al., 2020; D. Wang et al., 2020).

SARS-CoV-2 genome is around 26 to 32 kb long and it is 82% similar to SARS-CoV (2002) and Middle East respiratory syndrome coronavirus (MERS-CoV) (2013) viruses and it is 90% similar to them in terms of essential enzymes and structural protein (Naqvi et al., 2020). The virus contains 4 structural proteins: “spike (S), envelope (E), membrane (M), and nucleocapsid (N)” proteins (Naqvi et al., 2020).



**Figure 4‑2: Schematic representation of SARS CoV-2 and mechanism of host entry, Source: (Naqvi et al., 2020)**

* 1. **Mutations of SARS-CoV-2** 
     1. **Alpha variant B.1.1.7.**

Alpha variant had 10 mutations compared to original Sars-Cov-2 variant. Precisely, there were 7 mutations and 3 deletions in the spike protein. Spike protein allows for the attachment of host cells to enable entry into the human cells. Mutation and deletions in the spike protein can disrupt, change structure which can affect its function and stability. These mutations later will mainly lead to expression in lungs and kidneys (Ostrov, 2021).

The mutations are: D614G, T716I, S982A, N501Y, D1118H, A570D, P681H.

The deletions are: H69, V70, Y144 (Davies et al., 2021).

What happened in D614G mutation is that aspartic acid at position 614 was replaced with glycine which led to change of spike protein’s structure making it easy for the furin (enzyme) to be split. In the receptor binding domain, this is a N501Y mutation. Receptor binding domain interacts with ACE2 receptor and deletions present on the N-terminal domain of the spike protein. In P681H mutation, change from proline to histidine, happened (Khan et al., 2021). As previously mentioned, all of these mutation lead to change of spike protein structure making it hard for immune response to recognize it and fight it. If virus is not fought, it will lead to health problems and over time to serious severe diseases. Some of the health problems caused by alpha variant are cough, shortness of breath, headache, fatigue, vomiting, muscle aches, fever, chills, sore throat and some more.

* + 1. **Delta variant B.1.617.2**

Compared to the first Sars-Cov-2 alpha strain, Delta variant's genome had 23 mutations. Out of those 23 mutations, 12 were in the spike protein. Spike protein allows for the attachment of host cells to enable entry into the cells. It is also a primary target for vaccines. Immune system targets this spike protein for the removal of virus. Once spike protein is targeted, B cells will produce antibodies which will bond with spike protein for virus removal. Spike protein has 2 subunits: S1, which binds to ACE2 receptor, and S2 subunit which helps with integration of the virus to the host cell. After each mutation in the spike protein, it will be harder for immune system to identify them. It will be also harder for antibodies to attach to the spike protein and to remove a virus. Delta variant with mutated spike protein, leads to better attachment to cells and that will help with infection (Huang et al., 2020). The most significant mutations to mention are: T478K, T19R, D614G, P681R, L452R and D960N including deletions at positions 157 and 81. The L452R changes an arginine for leucine. This happens at position 452. This change will help spike protein to bind stronger to the ACE2 receptor. This receptor’s role is to allow spike protein to attach. This may cause remove possibility for antibodies to bind to the spike protein because spike protein is already tight bound with ACE2 receptor.

Another research showed that L425R mutation helps delta variant to avoid being attacked by CD8 T cells, which are the cells that fight and remove virus. P681R mutation change arginine for the proline at position 681. This mutation allows better integration of the virus to the host cell (Starr et al., 2021). These mutations are important to mention and explain because it will cause high affinity of the virus and human cell which can lead to different health problems. The delta variant infection has many health problems: sore throat, loss of smell and taste, fatigue, headache, cough, shortness of breath, fever, vomiting, diarrhoea and rhinorrhoea (*How Dangerous Is the Delta Variant (B.1.617.2)?*, n.d.). There are some reports in UK about delta variant causing auditory impairment because of bad blood clots. (UCdavis, 2021). Compared to alpha variants, health effects are similar. Differences is that people infected with delta variant become faster and more ill and have more respiratory tract problems. The delta variant spreads much faster than the alpha variant. Report says that there is around 50% increase in in transmissibility compared to the alpha variant. Death rates increased more rapidly and in shorter timeframe; 133% higher chance of death than the wild type variant. Also, there are reports saying that there is also 109% increase in hospital risk and 235% increase in ICU admission compared to alpha variant.

* + 1. **Omicron variant B.1.1.529.**

Compared to original Sars-Cov-2 virus strain, Omicron variant's genome had 53 mutations. Out of those 53 mutations, 30 were in only one part- Spike encoding S gene which is 3-5 times more mutation than in any previous variant. This part connects to human cells before it comes into the cell. This is also a primary target for vaccines. Since mutations on spike protein determine how virus attaches, infects human cells and also how antibodies attach to the spike protein to fight the virus, those mutation can have big effect on the way virus is transmitted, how body will fight it off as well as effectiveness of treatments (Torjesen, 2021). The Omicron variant infection has many health effects such as: sore throat, sneezing, headache, body aches, runny nose, and fatigue. Children were also diagnosed with pneumonia and has severe gastrointestinal symptoms as well as dehydration. Loss of smell and taste was not reported like in previous cases with other variants. This data is from UK and South Africa study (Iacobucci, 2021).

The main focus for research is receptor binding domain RBD of the Spike protein but some research currently suggest that Spike N terminal domain may help virus to enter the cell via sialic-acid receptor binding. Compared to previous variants, Omicron variant have increased sialic acid binding energy which can lead to increased transmission (Lam et al., 2021; Xia, 2021). Another study conducted that some mutations make new salt bridges and hydrogen bonds between the spike protein and the human cell receptor ACE2 which can make binding affinity stronger so virus will better attach to the cell and make this bond stronger. Those mutations are R493, R498 and S496. Mutation K417N will make this bond weaker. This data indicates that Omicron variant has better binding affinity than the original SARS-CoV-2 virus despite the fact that it has so many mutations (Barton et al., 2021).

According to South Africa data, the Omicron is connected to an increased ability to evade immunity from previous infection or vaccines compared to the Beta or Delta variants. So, the two Omicron characteristics are: strong binding with human cells and increased antibody evasion and increased immunity evasion from previous infections. Both those characteristics lead to better transmissibility of the Omicron variant (Dyer, 2021).

According to WHO, more research and data are needed to understand better clinical severity of Omicron.

Until now, some researches are conducted which tell us that severity of the disease might be lower than with previous variants. Data from Scotland suggest that severity is around 29% lower than previous wave which means that adults, who are infected with Omicron, are 29% less likely to be hospitalized when compared to infection from other variants for example Delta variant. Also, another research in China, focused on respiratory tract, says that Omicron multiplies and infects much faster than wild type SARS-CoV-2 but it is 10x less effective in the lung. This may lead us to idea that less effective replicationin lungs means lower severity, but disease severity is not determined only by virus replication; what is important is host immune response to the infection which may lead to dysregulation of the innate immune system. So, the conclusion is that threat from Omicron should be taken seriously because this very infectious variant which can cause severe disease and lead to death even though the virus itself may be less pathogenic (*HKUMed Finds Omicron SARS-CoV-2 Can Infect Faster and Better than Delta in Human Bronchus but with Less Severe Infection in Lung - News | HKUMed*, n.d.).

* 1. **Databases and Tools**

For this paper, 20 different strains of SARS-Cov-2 virus were collected in FASTA format from National Centre for Biotechnology Information (NCBI) website. The strains were then analysed through a program written on Python ®, which was then run for a step-by-step compilation of implementing applications of major analytical, theoretical, and applicable bioinformatics tool of MSA using the online program Clustal Omega ®.

FASTA format is a text-based ‘language’ format consisting of single-line description of AA or nucleotide sequences starting with a ‘>’ symbol (*BLAST TOPICS*, n.d.), which can be used to run multiple bioinformatics tools. BLAST is an investigative method which uses the FASTA format and works by searching and matching for similarities between the query and database sequences. The phylogenetic tree is a branching diagram that can be drawn converting either Multiple Sequence Alignment (MSA) or the BLAST result based upon either maximum likelihood ratio or neighbour joining or UPGMA method into prediction of the evolution of a set of taxa from the organism’s most recent common ancestor(s).

1. **Method**

Command: NCBI<genome\_SARS-CoV-2< Genome Assembly and Annotation report. The data collected is listed below:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **collection date** | **name** | **locus** | **protein coding gene** | **link** |
| 5-Jan-20 | sars\_cov\_2\_2020\_01 | MN908947 | 12 | <https://www.ncbi.nlm.nih.gov/nuccore/NC_045512.2> |
| 27-Feb-20 | sars\_cov\_2\_2020\_02 | MT121215 | 5 | <https://www.ncbi.nlm.nih.gov/nuccore/MT121215.1> |
| 28-Feb-20 | sars\_cov\_2\_2020\_03 | MT123292 | 10 | <https://www.ncbi.nlm.nih.gov/nuccore/MT123292.2> |
| 11-Feb-20 | sars\_cov\_2\_2020\_04 | MT039890 | 10 | <https://www.ncbi.nlm.nih.gov/nuccore/MT039890.1> |
| 9-Feb-20 | sars\_cov\_2\_2020\_05 | MT049951 | 12 | <https://www.ncbi.nlm.nih.gov/nuccore/MT049951.1> |
| 2-Mar-20 | sars\_cov\_2\_2020\_06 | MT135042 | 10 | <https://www.ncbi.nlm.nih.gov/nuccore/MT135042.1> |
| 2-Mar-20 | sars\_cov\_2\_2020\_07 | MT135044 | 10 | <https://www.ncbi.nlm.nih.gov/nuccore/MT135044.1> |
| 6-Mar-20 | sars\_cov\_2\_2020\_08 | MT163718 | 12 | <https://www.ncbi.nlm.nih.gov/nuccore/MT163718.1> |
| 6-Mar-20 | sars\_cov\_2\_2020\_09 | MT163716 | 12 | <https://www.ncbi.nlm.nih.gov/nuccore/MT163716.1> |
| 6-Mar-20 | sars\_cov\_2\_2020\_10 | MT163719 | 12 | <https://www.ncbi.nlm.nih.gov/nuccore/MT163719.1> |
| 2-Mar-20 | sars\_cov\_2\_2020\_11 | MT135041 | 10 | <https://www.ncbi.nlm.nih.gov/nuccore/MT135041.1> |
| 2-Mar-20 | sars\_cov\_2\_2020\_12 | MT135043 | 10 | <https://www.ncbi.nlm.nih.gov/nuccore/MT135043.1> |
| 4-Feb-20 | sars\_cov\_2\_2020\_13 | MT019531 | 10 | <https://www.ncbi.nlm.nih.gov/nuccore/MT019531.1> |
| 4-Feb-20 | sars\_cov\_2\_2020\_14 | MT019529 | 10 | <https://www.ncbi.nlm.nih.gov/nuccore/MT019529.1> |
| 6-Mar-20 | sars\_cov\_2\_2020\_15 | MT163717 | 12 | <https://www.ncbi.nlm.nih.gov/nuccore/MT163717.1> |
| 30-Jan-20 | sars\_cov\_2\_2020\_16 | MT007544 | 12 | <https://www.ncbi.nlm.nih.gov/nuccore/MT007544.1> |
| 21-Jan-20 | sars\_cov\_2\_2020\_17 | MN975262 | 10 | <https://www.ncbi.nlm.nih.gov/nuccore/MN975262.1> |
| 27-Jan-20 | sars\_cov\_2\_2020\_18 | MN996528 | 10 | <https://www.ncbi.nlm.nih.gov/nuccore/MN996528.1> |
| 27-Feb-20 | sars\_cov\_2\_2020\_19 | MT123290 | 10 | <https://www.ncbi.nlm.nih.gov/nuccore/MT123290.1> |
| 14-Mar-20 | sars\_cov\_2\_2020\_20 | MT192772 | 10 | <https://www.ncbi.nlm.nih.gov/nuccore/MT192772.1> |

The FASTA sequences were then aligned by MSA tool on Clustal Omega ®.

1. **Result and Discussion**

coding sequence length: 29132

non-Coding sequence length: 816

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Similarity** | **Mutations** | **TT\_ratio** | **Gaps** | **Insertions** | **Deletions** | **CDS\_Similarity** | **CDS\_Mutations** |
| sars\_cov\_2\_2020\_01 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| sars\_cov\_2\_2020\_02 | 0.9985308 | 44 | 0 | 42 | 0 | 42 | 0.9999313 | 2 |
| sars\_cov\_2\_2020\_03 | 0.9984974 | 45 | 0.5625 | 20 | 0 | 20 | 0.9998627 | 4 |
| sars\_cov\_2\_2020\_04 | 0.9996995 | 9 | 0.8 | 0 | 0 | 0 | 0.9996911 | 9 |
| sars\_cov\_2\_2020\_05 | 0.999833 | 5 | 0.6667 | 0 | 0 | 0 | 0.9998627 | 4 |
| sars\_cov\_2\_2020\_06 | 0.9998664 | 4 | 3 | 0 | 0 | 0 | 0.9998627 | 4 |
| sars\_cov\_2\_2020\_07 | 0.9998664 | 4 | 3 | 0 | 0 | 0 | 0.9998627 | 4 |
| sars\_cov\_2\_2020\_08 | 0.9996327 | 11 | 1.2 | 0 | 0 | 0 | 0.9998284 | 5 |
| sars\_cov\_2\_2020\_09 | 0.9995659 | 13 | 0.8571 | 0 | 0 | 0 | 0.999794 | 6 |
| sars\_cov\_2\_2020\_10 | 0.9997997 | 6 | 0 | 0 | 0 | 0 | 0.999794 | 6 |
| sars\_cov\_2\_2020\_11 | 0.9998664 | 4 | 3 | 0 | 0 | 0 | 0.9998627 | 4 |
| sars\_cov\_2\_2020\_12 | 0.999833 | 5 | 1.5 | 0 | 0 | 0 | 0.9998284 | 5 |
| sars\_cov\_2\_2020\_13 | 0.999833 | 5 | 0 | 4 | 4 | 0 | 0.9999657 | 1 |
| sars\_cov\_2\_2020\_14 | 0.9997663 | 7 | 2 | 4 | 4 | 0 | 0.999897 | 3 |
| sars\_cov\_2\_2020\_15 | 0.9995325 | 14 | 3 | 6 | 6 | 0 | 0.9998284 | 5 |
| sars\_cov\_2\_2020\_16 | 0.9995659 | 13 | 0.5 | 10 | 10 | 0 | 0.999897 | 3 |
| sars\_cov\_2\_2020\_17 | 0.9994323 | 17 | 0 | 12 | 12 | 0 | 0.9998284 | 5 |
| sars\_cov\_2\_2020\_18 | 0.9995993 | 12 | 0 | 12 | 12 | 0 | 1 | 0 |
| sars\_cov\_2\_2020\_19 | 0.9991652 | 25 | 2.5 | 18 | 15 | 3 | 0.9999313 | 2 |
| sars\_cov\_2\_2020\_20 | 0.9995659 | 13 | 0 | 12 | 12 | 0 | 0.9999657 | 1 |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **CDS\_TT\_ratio** | **CDS\_Gaps** | **CDS\_Insertions** | **CDS\_Deletions** | **nonCDS\_Similarity** | **nonCDS\_Mutations** |
| sars\_cov\_2\_2020\_01 | 0 | 0 | 0 | 0 | 1 | 0 |
| sars\_cov\_2\_2020\_02 | 0 | 0 | 0 | 0 | 0.948529412 | 42 |
| sars\_cov\_2\_2020\_03 | 0 | 0 | 0 | 0 | 0.949754902 | 41 |
| sars\_cov\_2\_2020\_04 | 0.8 | 0 | 0 | 0 | 1 | 0 |
| sars\_cov\_2\_2020\_05 | 1 | 0 | 0 | 0 | 0.99877451 | 1 |
| sars\_cov\_2\_2020\_06 | 3 | 0 | 0 | 0 | 1 | 0 |
| sars\_cov\_2\_2020\_07 | 3 | 0 | 0 | 0 | 1 | 0 |
| sars\_cov\_2\_2020\_08 | 0 | 0 | 0 | 0 | 0.992647059 | 6 |
| sars\_cov\_2\_2020\_09 | 1 | 0 | 0 | 0 | 0.991421569 | 7 |
| sars\_cov\_2\_2020\_10 | 0 | 0 | 0 | 0 | 1 | 0 |
| sars\_cov\_2\_2020\_11 | 3 | 0 | 0 | 0 | 1 | 0 |
| sars\_cov\_2\_2020\_12 | 1.5 | 0 | 0 | 0 | 1 | 0 |
| sars\_cov\_2\_2020\_13 | 0 | 0 | 0 | 0 | 0.995098039 | 4 |
| sars\_cov\_2\_2020\_14 | 2 | 0 | 0 | 0 | 0.995098039 | 4 |
| sars\_cov\_2\_2020\_15 | 0 | 0 | 0 | 0 | 0.988970588 | 9 |
| sars\_cov\_2\_2020\_16 | 0.5 | 0 | 0 | 0 | 0.987745098 | 10 |
| sars\_cov\_2\_2020\_17 | 0 | 0 | 0 | 0 | 0.985294118 | 12 |
| sars\_cov\_2\_2020\_18 | 0 | 0 | 0 | 0 | 0.985294118 | 12 |
| sars\_cov\_2\_2020\_19 | 0 | 0 | 0 | 0 | 0.971813725 | 23 |
| sars\_cov\_2\_2020\_20 | 0 | 0 | 0 | 0 | 0.985294118 | 12 |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **nonCDS\_TT\_ratio** | **nonCDS\_Gaps** | **nonCDS\_Insertions** | **nonCDS\_Deletions** |
| sars\_cov\_2\_2020\_01 | 0 | 0 | 0 | 0 |
| sars\_cov\_2\_2020\_02 | 0 | 42 | 0 | 42 |
| sars\_cov\_2\_2020\_03 | 0.3125 | 20 | 0 | 20 |
| sars\_cov\_2\_2020\_04 | 0 | 0 | 0 | 0 |
| sars\_cov\_2\_2020\_05 | 0 | 0 | 0 | 0 |
| sars\_cov\_2\_2020\_06 | 0 | 0 | 0 | 0 |
| sars\_cov\_2\_2020\_07 | 0 | 0 | 0 | 0 |
| sars\_cov\_2\_2020\_08 | 0.2 | 0 | 0 | 0 |
| sars\_cov\_2\_2020\_09 | 0.75 | 0 | 0 | 0 |
| sars\_cov\_2\_2020\_10 | 0 | 0 | 0 | 0 |
| sars\_cov\_2\_2020\_11 | 0 | 0 | 0 | 0 |
| sars\_cov\_2\_2020\_12 | 0 | 0 | 0 | 0 |
| sars\_cov\_2\_2020\_13 | 0 | 4 | 4 | 0 |
| sars\_cov\_2\_2020\_14 | 0 | 4 | 4 | 0 |
| sars\_cov\_2\_2020\_15 | 0.5 | 6 | 6 | 0 |
| sars\_cov\_2\_2020\_16 | 0 | 10 | 10 | 0 |
| sars\_cov\_2\_2020\_17 | 0 | 12 | 12 | 0 |
| sars\_cov\_2\_2020\_18 | 0 | 12 | 12 | 0 |
| sars\_cov\_2\_2020\_19 | 1.5 | 18 | 15 | 3 |
| sars\_cov\_2\_2020\_20 | 0 | 12 | 12 | 0 |

From the data table above, we can see clear relationships among the strains.

* 1. **Gaps**

An alignment gap is representing the missing bases (one, two or more) in the sequences. Other than missing bases, gaps are created by the inserting a nucleotide which is known as insertion and by deleting the nucleotide, called as deletion. These gaps are made in specific location, either insertion or deletion of single nucleotide to loss or gain large section of DNA.

Here we analysed 20 strains of SARS-CoV-2 for the alignment gaps, where the sequences are going to be compared and the differences are recorded.

Before going into the sequences, let us have a look on the coding sequences of those 20 strains. We know that it is rare to find gaps on the coding region of the genome. As per expectation, the entire 20 sequences showed the same result for the coding sequences which has zero gaps in it.

Now, sequence 2 is different from sequence 1 by 42 gaps, which did not happen by insertion, it is because of deletion in the non-coding sequences of the sequence.

Followed by the next sequence 3, it showed us the difference with 20 gaps from the sequence 1, which is also like sequence 2. The gaps are made on non-coding sequences with the deletion.

From sequence 4 to 12, we were expecting gaps like sequence 2 & 3. But in those sequences, the entire genome, coding region and non-coding region were obtained with zero gaps like sequence 1.

So far, the gaps are identified, and the reason is deletion in non-coding sequences. From the sequence 13 to 20 except sequence 19, the alignment gaps are discovered in the reason of insertion took place in the non-coding sequences.

Starting with sequence 13, there is a gradual increase in the gaps due to the insertion on the non-coding- sequences (from sequence 13 to 18, 20). Initially, the gap was 4 for the sequence 13 and 14,

Sequence 15- 6, Sequence 16- 10, Sequence 17- 12, Sequence 18- 12, Sequence 20- 12 gaps are identified due to insertion on non-coding region.

After finding out those gaps in the sequence collection we took either by insertion or deletion, there is one sequence that remains in our list which showed the gaps in non-coding regions like all other sequence with both insertion and deletion. In total, there are 18 gaps, where 15 is due insertion and 3 gaps happened because of the deletion. And this discovered on sequence 19.

Since the gaps are located on the sequence, the gene is either going through mutation or gene silencing depending on the process and condition.

* 1. **Mutation**

Before analysing other sequences, we would like to go over the Mutations in the Coding sequences in all the strains, we found that there are no gaps in all the coding sequences which that means that there is no insertion or deletion of the nucleotides, that’s lead to say that there is no frame shifting mutations happen in the coding sequences, however the mutations are caused by the substitution of the nucleotides either by transition or transversion, and that’s lead to point mutations.

Now there is a big difference in the number of mutations in the second sequence and third sequence compared to the first sequence, where the second sequence has 44 mutations where 2 of them are accumulated in coding region and 42 mutations accumulated in the noncoding region. Also, the third sequence has 45 mutations in the entire sequence where they are distributed to 4 mutations in the coding region and 41 mutations in the non-coding region.

From sequence 4 to sequence 12 the mutation occurred by substitution of the nucleotides either by transition or transversion. Sequence from 4 to 12 except sequence 5, 8 and 9, the point mutations only accumulated in the coding region of the strain, so sequence 4 has 9-point mutations, sequence 6, 7 and 11 have 4-point mutations, sequence 10 has 6 point mutations, and sequence 12 has 5 point mutations. Sequence 5 accumulate 5-point mutations where 4 mutations accumulated in the coding region and 1 mutation in the noncoding region, furthermore, sequence 8 has 11-point mutations where 5 mutations in coding region and 6 mutations in the noncoding region, sequence 9 has 13-point mutations where 6 mutations in the coding region and 7 mutations in the noncoding region.

After analysing all the mutations in the coding regions for all strains, we analysed the mutations for noncoding regions for the sequence starting from 13 to sequence 20 except sequence 18. All the mutations in those noncoding regions are accumulated due to the insertion of nucleotides except sequence 15 and 19. Sequence 13 has 5 with 4 frameshift mutations in noncoding regions, sequence 14 has 7 mutations with 4 frameshift mutations in the noncoding region, sequence 16 has 13 mutations with 10 frameshifti mutations in the noncoding region, sequence 17 has 17 mutations with 12 mutations occur due to insertion, sequence 20 has 13 mutations with also 12 mutations occur due to insertion. Sequence 15 has 14 mutations with 9 mutations in noncoding region divided to 4 mutations due to point mutations and 6 mutations duo to frameshift mutations, sequence 19 has 25 mutations with 23 mutations in the noncoding region divided to 18 mutation due to frameshift mutation and 5 mutations due to point mutations.

After analysing the mutations for all strains and sort them in the two types, we were left with one sequence which the mutation in this sequence only accumulated in the noncoding region, where there is 12 frameshift mutations and the is due to the insertion of the nucleotides in the noncoding sequences.

We can use the table to further simplify our result:

|  |  |
| --- | --- |
| **strain** | **Comment** |
| sars\_cov\_2\_2020\_01 | There are no mutations in the whole genome (no mutations in coding region and non-coding region) mutation frequency is zero |
| sars\_cov\_2\_2020\_02 | There are 44 mutations in the whole sequence (2 from the coding region and 42 from the noncoding region) |
| sars\_cov\_2\_2020\_03 | There are 45 mutations in the whole genome (4 from CDS and 41 from non-CDS) |
| sars\_cov\_2\_2020\_04 | There are 9 mutations in the whole genome (9 from CDS and 0 from non-CDS) no gaps |
| sars\_cov\_2\_2020\_05 | there are 5 mutations in the whole genome (4 from CDS and 1 from non-CDS) |
| sars\_cov\_2\_2020\_06 | There are 4 mutations in the whole genome (4 from CDS and 0 from non-CDS) |
| sars\_cov\_2\_2020\_07 | There are 4 mutations in the whole genome (4 from CDS and 0 from non-CDS) |
| sars\_cov\_2\_2020\_08 | There are 11 mutations in the whole genome (5 from CDS and 6 from non-CDS) |
| sars\_cov\_2\_2020\_09 | There are 13 mutations in the whole genome (6 from CDS and 7 from non-CDS) |
| sars\_cov\_2\_2020\_10 | There are 6 mutations in the whole genome (6 from CDS and 0 from non-CDS) |
| sars\_cov\_2\_2020\_11 | There are 4 mutations in the whole genome (4 from CDS and 0 from non-CDS) |
| sars\_cov\_2\_2020\_12 | There are 5 mutations in the whole genome (5 from CDS and 0 from non-CDS), no gaps |
| sars\_cov\_2\_2020\_13 | There are 5 mutations in the whole genome (1 from CDS and 4 from non-CDS) |
| sars\_cov\_2\_2020\_14 | There are 7 mutations in the whole genome (3 from CDS and 4 from non-CDS) |
| sars\_cov\_2\_2020\_15 | There are 14 mutations in the whole genome (5 from CDS and 9 from non-CDS) |
| sars\_cov\_2\_2020\_16 | There are 13 mutations in the whole genome (3 from CDS and 10 from non-CDS) |
| sars\_cov\_2\_2020\_17 | There are 17 mutations in the whole genome (5 from CDS and 12 from non-CDS) |
| sars\_cov\_2\_2020\_18 | There are 12 mutations in the whole genome (0 from CDS and 12 from non-CDS) |
| sars\_cov\_2\_2020\_19 | There are 25 mutations in the whole genome (2 from CDS and 23 from non-CDS) |
| sars\_cov\_2\_2020\_20 | There are 13 mutations in the whole genome (1 from CDS and 12 from non-CDS) |

The table for the mutation frequency is listed below:

|  |  |  |  |
| --- | --- | --- | --- |
| **mutation fequency** | **mutation frequency for CDS** | **mutation frequency for non CDS** | **Comment** |
| 0 | 0 | 0 |  |
| 0.001469 | 6.87E-05 | 0.051471 | smaller |
| 0.001503 | 0.000137 | 0.050245 | smaller |
| 0.000301 | 0.000309 | 0 | bigger |
| 0.000167 | 0.000137 | 0.001225 | samller |
| 0.000134 | 0.000137 | 0 | bigger |
| 0.000134 | 0.000137 | 0 | bigger |
| 0.000367 | 0.000172 | 0.007353 | smaller |
| 0.000434 | 0.000206 | 0.008578 | smaller |
| 0.0002 | 0.000206 | 0 | bigger |
| 0.000134 | 0.000137 | 0 | bigger |
| 0.000167 | 0.000172 | 0 | bigger |
| 0.000167 | 3.43E-05 | 0.004902 | smaller |
| 0.000234 | 0.000103 | 0.004902 | smaller |
| 0.000467 | 0.000172 | 0.011029 | smaller |
| 0.000434 | 0.000103 | 0.012255 | smaller |
| 0.000568 | 0.000172 | 0.014706 | smaller |
| 0.000401 | 0 | 0.014706 | smaller |
| 0.000835 | 6.87E-05 | 0.028186 | smaller |
| 0.000434 | 3.43E-05 | 0.014706 | smaller |

* 1. **Similarities**

When it comes to the similarity between sequences of the whole sequence analysis, there was not some significant change to mention. Between all of the 20 variants it differs in 0.000x%. Analysis of coding sequence region showed that SARS-CoV-2\_01 variant and SARS-CoV-2\_18 variant have completely same coding region. Other variants differ in small, but not significant number; again, it is 0.000x%.

Non-coding sequence region analysis is a little bit more different. Data shows that variants: SARS-CoV-2\_01, SARS-CoV-2\_04, SARS-CoV-2\_06, SARS-CoV-2\_07, SARS-CoV-2\_10, SARS-CoV-2\_11 and SARS-CoV-2\_12 has the same non-coding region. Other variants differ in 0.0x%. The most different non-coding regions are in these variants: SARS-CoV-2\_02 (0.9485294 %) and SARS-CoV-2\_03 (0.9497549 %).

* 1. **TT Ratio**

The elaboration of TT Ratio is Transversion & Transition Ratio in Genetics. The ratio stands for the number of probability by which nucleotide transitions are more likely to occur over transversions (Stoltzfus & Norris, 2016). In short, “Transitions are interchanges of two-ring purines (A→G), or of one-ring pyrimidines (C→T): they therefore involve bases of similar shape. Transversions are interchanges of purine for pyrimidine bases, which therefore involve exchange of one-ring & two-ring structures” (*Transitions vs Transversions*, n.d.). From various studies, we know that TT Ratio in SARS-CoV-2 is biased towards C>U indicating a rapid mutation (Matyášek & Kovařík, 2020).

One of the limitations of our data is that it was not capable enough of producing substantial information on nucleotide, so it was difficult for us to decode the relationship between the TT Ratio with the other data we gathered. However, this was sufficient to see that the strains did not emerge very far from each other. The dataset did not provide any outliers and data points were within ranges. The standard deviation of the dataset did not also show any discrepancy either.

A picture containing chain, metalware, necklet, toggle

Description automatically generated

**Figure 3‑2: TT Ratio, Source:** [**https://www.mun.ca/biology/scarr/Transitions\_vs\_Transversions.html**](https://www.mun.ca/biology/scarr/Transitions_vs_Transversions.html)

For our study, we compared the TT ratio in all 20 strains, then we compared the coding and the non-coding regions. The table below shows the Ratio in reference to strain 1 (sars\_cov\_2\_2020\_01).

|  |  |  |  |
| --- | --- | --- | --- |
|  | **TT\_ratio** | **CDS\_TT\_ratio** | **nonCDS\_TT\_ratio** |
| sars\_cov\_2\_2020\_01 | 0 | 0 | 0 |
| sars\_cov\_2\_2020\_02 | 0 | 0 | 0 |
| sars\_cov\_2\_2020\_03 | 0.5625 | 0 | 0.3125 |
| sars\_cov\_2\_2020\_04 | 0.8 | 0.8 | 0 |
| sars\_cov\_2\_2020\_05 | 0.666666667 | 1 | 0 |
| sars\_cov\_2\_2020\_06 | 3 | 3 | 0 |
| sars\_cov\_2\_2020\_07 | 3 | 3 | 0 |
| sars\_cov\_2\_2020\_08 | 1.2 | 0 | 0.2 |
| sars\_cov\_2\_2020\_09 | 0.857142857 | 1 | 0.75 |
| sars\_cov\_2\_2020\_10 | 0 | 0 | 0 |
| sars\_cov\_2\_2020\_11 | 3 | 3 | 0 |
| sars\_cov\_2\_2020\_12 | 1.5 | 1.5 | 0 |
| sars\_cov\_2\_2020\_13 | 0 | 0 | 0 |
| sars\_cov\_2\_2020\_14 | 2 | 2 | 0 |
| sars\_cov\_2\_2020\_15 | 3 | 0 | 0.5 |
| sars\_cov\_2\_2020\_16 | 0.5 | 0.5 | 0 |
| sars\_cov\_2\_2020\_17 | 0 | 0 | 0 |
| sars\_cov\_2\_2020\_18 | 0 | 0 | 0 |
| sars\_cov\_2\_2020\_19 | 2.5 | 0 | 1.5 |
| sars\_cov\_2\_2020\_20 | 0 | 0 | 0 |

We know that transversions are more conservative than transitions (with few exceptions) (Stoltzfus & Norris, 2016). Mathematically, 8 types of transversions and 4 types of transitions can happen in nature (Vogel, 1972). And not surprisingly, nature follows mathematical rules. In the first column, we can see that strain 3,4,5, 9, 16 is slightly higher than 0 (0.5625, 0.8, 0.666666667, 0.857142857, 0.5 respectively) whereas strain 6, 7, 8, 11, 12, 14, 15 and 19 are significantly higher than 1 (3, 3, 1.2, 3, 1.5, 2 and 2.5 respectively). In terms of TT ratio, strain 1, 2, 10, 13, 17, 18 and 20 are exactly same. In the coding region, however, we can see strain 1, 2, 3, 8, 10, 13, 15, 17, 18, 19 and 20 are exactly the same, whereas strain 4 is 0.8%, 5 is 1%, 6 and 7 are 3%, 9 is 1%, 11 is 3%, 12 is 1.5%, 14 is 2% and finally 16 is 0.5% higher than strain 1.

The frequency of the strains is given below:

|  |  |  |  |
| --- | --- | --- | --- |
|  | **entire seq (%)** | **Cds (%)** | **n-cds (%)** |
| 0 | 0.35 | 0.55 | 0.75 |
| 0.1-0.9 | 0.25 | 0.1 | 0.2 |
| 1.0-3.0 | 0.4 | 0.35 | 0.05 |

From the frequency table, we can see that the strains are 100% similar at 35% of the time in the entire sequence. It is 55% in the coding region. 25% times they diverge from each other in the entire sequence, and comparing the info with mutation data, we can say that these strains are highly similar and most probably emerged from the same region. The phylogenetic tree created from the Clustal Omega website also support the hypothesis. The phylogenetic tree is given below:

A picture containing diagram

Description automatically generated

**Figure 3-2: This is a Neighbor-joining tree without distance corrections. Cladogram created at** [**http://www.ebi.ac.uk/Tools/services/web/toolresult.ebi?tool=clustalo&jobId=clustalo-E20220109-172652-0550-54723398-p2m**](http://www.ebi.ac.uk/Tools/services/web/toolresult.ebi?tool=clustalo&jobId=clustalo-E20220109-172652-0550-54723398-p2m)**, only available from 9th January 2022 to 16th January 2022**

We can see the relationship among the three groups in the chart below:

And when we can compare mutation rate with TT ratio, we can get a more holistic picture.

# **Observation**

If we take a look at the guide tree, we can understand how these strains are connected to each other.

A picture containing diagram

Description automatically generated

**Figure 4-1: Cladogram created at** [**http://www.ebi.ac.uk/Tools/services/web/toolresult.ebi?tool=clustalo&jobId=clustalo-E20220109-172652-0550-54723398-p2m**](http://www.ebi.ac.uk/Tools/services/web/toolresult.ebi?tool=clustalo&jobId=clustalo-E20220109-172652-0550-54723398-p2m)**, only available from 9th January 2022 to 16th January 2022**

The Interesting points that we found in this study is that we saw the highest number of mutation (45 in total, 4 in CDS and 41 in N-CDS) in strain 3 comparing to strain 1. However, we did not see that reflected in the TT Ratio. We observed a 0.5625% divergence from strain 1, which was nor the lowest or highest % in the dataset.

Strain 1 (NC\_045512) was collected on 05-JAN-2020, which is a Betacoronavirus.

Strain 3 (MT123292) was collected on 28-FEB-2020, which is also a Betacoronavirus.

From figure 4-1, we can see that even if they were both Betacoronavirus collected from a month apart in China, they differ massively. The phylogenetic tree shows that the strain 3 diverges 0.000562094% from strain 1. That is a minute difference for sure. But this provides insight how SARS-CoV is a highly mutative virus which could do potential damage to human (and other species).

# **Conclusion**

This study provided insight on SARS-CoV-2 strains and how did they evolve. Like it or not, virus and microbiome are parts of our lives. Studying virus and their evolutionary pathway are crucial for human evolution and survival. In this paper, we discussed about 20 random strains we collected from the NCBI website and compared them with each other. We used a web based MSA tool to align the data and then used a Python based program to analyse the result. This study helped us understand how quickly viruses can evolve and how much do they differ in terms of mutations, transversion and transition ratio, similarities, and gaps from each other within a short period of time. However, the study would be much more interesting if we had a chance and time to predict their protein structure and compare newer strains like Omicron and Delta instead of just Beta variants.

# **Reference**

Barton, M. I., MacGowan, S. A., Kutuzov, M. A., Dushek, O., Barton, G. J., & van der Merwe, P. A. (2021). Effects of common mutations in the SARS-CoV-2 Spike RBD and its ligand, the human ACE2 receptor on binding affinity and kinetics. *ELife*, *10*, e70658. https://doi.org/10.7554/eLife.70658

Biscayart, C., Angeleri, P., Lloveras, S., Chaves, T. do S. S., Schlagenhauf, P., & Rodríguez-Morales, A. J. (2020). The next big threat to global health? 2019 novel coronavirus (2019-nCoV): What advice can we give to travellers? – Interim recommendations January 2020, from the Latin-American society for Travel Medicine (SLAMVI). *Travel Medicine and Infectious Disease*, *33*, 101567. https://doi.org/10.1016/j.tmaid.2020.101567

*BLAST TOPICS*. (n.d.). Retrieved December 20, 2021, from https://blast.ncbi.nlm.nih.gov/Blast.cgi?CMD=Web&PAGE\_TYPE=BlastDocs&DOC\_TYPE=BlastHelp

Chang, D., Lin, M., Wei, L., Xie, L., Zhu, G., Dela Cruz, C. S., & Sharma, L. (2020). Epidemiologic and Clinical Characteristics of Novel Coronavirus Infections Involving 13 Patients Outside Wuhan, China. *JAMA*, *323*(11), 1092–1093. https://doi.org/10.1001/jama.2020.1623

*Cumulative confirmed COVID-19 deaths*. (n.d.). Our World in Data. Retrieved December 13, 2021, from https://ourworldindata.org/grapher/cumulative-covid-deaths-region

Davies, N. G., Abbott, S., Barnard, R. C., Jarvis, C. I., Kucharski, A. J., Munday, J. D., Pearson, C. A. B., Russell, T. W., Tully, D. C., Washburne, A. D., Wenseleers, T., Gimma, A., Waites, W., Wong, K. L. M., Zandvoort, K. van, Silverman, J. D., Group1‡, C. C.-19 W., Consortium‡, C.-19 G. U. (COG-U., Diaz-Ordaz, K., … Edmunds, W. J. (2021). Estimated transmissibility and impact of SARS-CoV-2 lineage B.1.1.7 in England. *Science*. https://doi.org/10.1126/science.abg3055

Dyer, O. (2021). Covid-19: South Africa’s surge in cases deepens alarm over omicron variant. *BMJ*, *375*, n3013. https://doi.org/10.1136/bmj.n3013

Giovanetti, M., Benedetti, F., Campisi, G., Ciccozzi, A., Fabris, S., Ceccarelli, G., Tambone, V., Caruso, A., Angeletti, S., Zella, D., & Ciccozzi, M. (2021). Evolution patterns of SARS-CoV-2: Snapshot on its genome variants. *Biochemical and Biophysical Research Communications*, *538*, 88–91. https://doi.org/10.1016/j.bbrc.2020.10.102

*HKUMed finds Omicron SARS-CoV-2 can infect faster and better than Delta in human bronchus but with less severe infection in lung—News | HKUMed*. (n.d.). Retrieved January 9, 2022, from https://www.med.hku.hk/en/news/press/20211215-omicron-sars-cov-2-infection

*How Dangerous Is the Delta Variant (B.1.617.2)?* (n.d.). Retrieved January 9, 2022, from https://asm.org/Articles/2021/July/How-Dangerous-is-the-Delta-Variant-B-1-617-2

Huang, Y., Yang, C., Xu, X.-F., Xu, W., & Liu, S.-W. (2020). Structural and functional properties of SARS-CoV-2 spike protein: Potential antivirus drug development for COVID-19. *Acta Pharmacologica Sinica*, *41*(9), 1141–1149. https://doi.org/10.1038/s41401-020-0485-4

Iacobucci, G. (2021). Covid-19: Runny nose, headache, and fatigue are commonest symptoms of omicron, early data show. *BMJ (Clinical Research Ed.)*, *375*, n3103. https://doi.org/10.1136/bmj.n3103

Khan, A., Zia, T., Suleman, M., Khan, T., Ali, S. S., Abbasi, A. A., Mohammad, A., & Wei, D. (2021). Higher infectivity of the SARS‐CoV‐2 new variants is associated with K417N/T, E484K, and N501Y mutants: An insight from structural data. *Journal of Cellular Physiology*, 10.1002/jcp.30367. https://doi.org/10.1002/jcp.30367

Lai, C.-C., Shih, T.-P., Ko, W.-C., Tang, H.-J., & Hsueh, P.-R. (2020). Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) and coronavirus disease-2019 (COVID-19): The epidemic and the challenges. *International Journal of Antimicrobial Agents*, *55*(3), 105924. https://doi.org/10.1016/j.ijantimicag.2020.105924

Lam, S. D., Waman, V. P., Orengo, C., & Lees, J. (2021). *Insertions in the SARS-CoV-2 Spike N-Terminal Domain May Aid COVID-19 Transmission* (p. 2021.12.06.471394). https://doi.org/10.1101/2021.12.06.471394

Li, Q., Guan, X., Wu, P., Wang, X., Zhou, L., Tong, Y., Ren, R., Leung, K. S. M., Lau, E. H. Y., Wong, J. Y., Xing, X., Xiang, N., Wu, Y., Li, C., Chen, Q., Li, D., Liu, T., Zhao, J., Liu, M., … Feng, Z. (2020). Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus–Infected Pneumonia. *The New England Journal of Medicine*, *382*(13), 1199–1207. https://doi.org/10.1056/NEJMoa2001316

Matyášek, R., & Kovařík, A. (2020). Mutation Patterns of Human SARS-CoV-2 and Bat RaTG13 Coronavirus Genomes Are Strongly Biased Towards C&gt;U Transitions, Indicating Rapid Evolution in Their Hosts. *Genes*, *11*(7), 761. https://doi.org/10.3390/genes11070761

Naqvi, A. A. T., Fatima, K., Mohammad, T., Fatima, U., Singh, I. K., Singh, A., Atif, S. M., Hariprasad, G., Hasan, G. M., & Hassan, Md. I. (2020). Insights into SARS-CoV-2 genome, structure, evolution, pathogenesis and therapies: Structural genomics approach. *Biochimica et Biophysica Acta. Molecular Basis of Disease*, *1866*(10), 165878. https://doi.org/10.1016/j.bbadis.2020.165878

Ostrov, D. A. (2021). Structural Consequences of Variation in SARS-CoV-2 B.1.1.7. *Journal of Cellular Immunology*, *3*(2), 103–108. https://doi.org/10.33696/immunology.3.085

Starr, T. N., Greaney, A. J., Dingens, A. S., & Bloom, J. D. (2021). Complete map of SARS-CoV-2 RBD mutations that escape the monoclonal antibody LY-CoV555 and its cocktail with LY-CoV016. *BioRxiv: The Preprint Server for Biology*, 2021.02.17.431683. https://doi.org/10.1101/2021.02.17.431683

Stoltzfus, A., & Norris, R. W. (2016). On the Causes of Evolutionary Transition:Transversion Bias. *Molecular Biology and Evolution*, *33*(3), 595–602. https://doi.org/10.1093/molbev/msv274

Torjesen, I. (2021). Covid-19: Omicron may be more transmissible than other variants and partly resistant to existing vaccines, scientists fear. *BMJ*, *375*, n2943. https://doi.org/10.1136/bmj.n2943

*Transitions vs transversions*. (n.d.). Retrieved January 9, 2022, from https://www.mun.ca/biology/scarr/Transitions\_vs\_Transversions.html

Vogel, F. (1972). Non-randomness of base replacement in point mutation. *Journal of Molecular Evolution*, *1*(3), 334–367. https://doi.org/10.1007/BF01653962

Wang, D., Hu, B., Hu, C., Zhu, F., Liu, X., Zhang, J., Wang, B., Xiang, H., Cheng, Z., Xiong, Y., Zhao, Y., Li, Y., Wang, X., & Peng, Z. (2020). Clinical Characteristics of 138 Hospitalized Patients With 2019 Novel Coronavirus–Infected Pneumonia in Wuhan, China. *JAMA*, *323*(11), 1061–1069. https://doi.org/10.1001/jama.2020.1585

Wang, M.-Y., Zhao, R., Gao, L.-J., Gao, X.-F., Wang, D.-P., & Cao, J.-M. (2020). SARS-CoV-2: Structure, Biology, and Structure-Based Therapeutics Development. *Frontiers in Cellular and Infection Microbiology*, *10*, 587269. https://doi.org/10.3389/fcimb.2020.587269

Xia, X. (2021). Domains and Functions of Spike Protein in SARS-Cov-2 in the Context of Vaccine Design. *Viruses*, *13*(1), 109. https://doi.org/10.3390/v13010109

Zhou, P., Yang, X.-L., Wang, X.-G., Hu, B., Zhang, L., Zhang, W., Si, H.-R., Zhu, Y., Li, B., Huang, C.-L., Chen, H.-D., Chen, J., Luo, Y., Guo, H., Jiang, R.-D., Liu, M.-Q., Chen, Y., Shen, X.-R., Wang, X., … Shi, Z.-L. (2020). *Discovery of a novel coronavirus associated with the recent pneumonia outbreak in humans and its potential bat origin* (p. 2020.01.22.914952). https://doi.org/10.1101/2020.01.22.914952