
Mutational Analysis of SARS-CoV-2 genomes of Nepali isolates

By

Gazi Salah Uddin Nuhash

ID: 1821599047



Department of Biochemistry & Microbiology

North South University

A thesis report submitted to the Department of Biochemistry & Microbiology, North South University for partial fulfilment of the Degree of Bachelor of Science in Biochemistry & Biotechnology.

October, 2022

Approval Letter

North South University

Dhaka, Bangladesh

We the members of Research Project Committee certify that we have carefully read and recommended to the Department of Biochemistry & Microbiology, North South University, Dhaka for approval of this project report entitled:

‘Mutational Analysis of SARS-CoV-2 genomes of Nepali isolates’

Submitted by Gazi Salah Uddin Nuhash (ID#1821599047) for partial fulfilment of the requirements for the degree of Bachelor of Science in Biochemistry and Biotechnology.

**Dr. Muhammad
Maqsud Hossain**
Director (NGRI) &
Associate Professor
Department of
Biochemistry &
Microbiology
School of Health &
Life Sciences
North South
University

**Dr. Md. Mainul
Hossain**
Associate Professor
& Chairman
Department of
Biochemistry &
Microbiology
School of Health &
Life Sciences
North South
University

**Dr. S M Bakhtiar
UL Islam**
Assistant Professor
Department of
Biochemistry &
Microbiology
School of Health &
Life Sciences
North South
University

**Kazi Fahmida
Rahman**
Lecturer
Department of
Biochemistry &
Microbiology
School of Health &
Life Sciences
North South
University

Acknowledgement

I would like to extend my gratitude to the Department of Biochemistry and Microbiology of North South University for providing me all the materials to carry out this project and for providing me the wonderful opportunity to study in the respected department.

I would also like to express my gratitude towards NSU Genome Research Institute (NGRI) for their support towards the project and my heartfelt gratitude goes to my teacher and supervisor, **Dr. Muhammad Maqsud Hossain** for giving me the flexibility, support and opportunity to do this wonderful project, all the while guiding me thoroughly and overlooking its completion. I came to know new knowledge, skills and techniques which has enabled me to understand what I want to strive for next and I am thankful for it.

I would like to extend my appreciation to all the esteemed faculties of the Department of Biochemistry and Microbiology who have given me advice and has taught me all that I know today. My deepest heartfelt respect and appreciation goes for Dr. Md. Mainul Hossain, Dr. Sohikul Islam, Dr. Zaied Ahmed Bhuiyan, Dr. Chaman Ara Keya and Dr. Muktadir Shaheed Hossain for their inspiring lectures, guidance and support throughout my time in North South University.

Lastly, a wholehearted and sincere thanks and appreciation to my mother and my close friends, for their time and effort in helping me and pushing me, so that I could successfully see this project till the end.

Gazi Salah Uddin Nuhash

September, 2022

List of Figures

Figures	Page Number
Figure 1: SARS-CoV-2	11
Figure 2: Replication cycle of SARS-CoV-2	12
Figure 3: SARS-CoV-2 has four structural proteins	14
Figure 4: Clade distribution in Nepali isolates	19
Figure 5: Pie chart showing the total number of each mutation	22
Figure 6: Bar chart of the prevalence of each mutation in genes.	23
Figure 7: Chart showing ranges of mutations in each gene	25
Figure 8: Phylogenetic tree of selected SARS-CoV-2 sequences from Nepal (Bootstrap value = 100)	26
Figure 9: Phylogenetic tree showing five major clusters based on multiple of mutations of each isolate.	27

List of Table

Table	Page Number
Table 1: Clade distribution in Nepali isolates	19
Table 2: Distribution of mutation in genes of SARS- CoV-2 isolates from Nepal	21

Table of Contents

<i>Approval Letter</i>	2
<i>Acknowledgement</i>	3
<i>List of Figures</i>	4
<i>List of Table</i>	5
<i>Abstract</i>	7
<i>Chapter 1: Introduction</i>	8
<i>Aims and objectives:</i>	9
<i>Chapter 2: Literature review</i>	10
<i>Structure:</i>	10
<i>Viral pathway into host cell:</i>	11
<i>Genomic Organization:</i>	13
<i>Characteristics and Significance of SARS-CoV-2 proteins</i>	14
<i>SARS-CoV-2 and Nepal:</i>	16
<i>Chapter 3: Materials & Methods</i>	17
<i>Sample Collection:</i>	17
<i>Data Processing:</i>	Error! Bookmark not defined.
<i>Phylogenetic Analysis:</i>	18
<i>Statistical Analysis:</i>	18
<i>Chapter 4: Results and Discussion</i>	19
<i>Clade distribution:</i>	19
<i>Classification of mutations in each gene:</i>	21
<i>Phylogenetic Tree for selected samples:</i>	26
<i>Limitations:</i>	Error! Bookmark not defined.
<i>Chapter 5: Conclusion & Future directions</i>	28
<i>Chapter 6: References</i>	29

Abstract

Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2), a highly infectious and contagious coronavirus that results in the Coronavirus Disease 2019 (COVID-19) overcame the world in 2019. Originating from Wuhan, China, it has spread all around the globe at an unimaginable pace and have overcome millions of people. One of the reasons why it was able to do so was because by the time we had found a solution, it had mutated and changed. Various mutations were uncovered in structural and non-structural proteins. These mutations change the virus's ability to infect, spread and accumulate. This project aims to identify the prevalence and significance of these mutations of the virus's proteins.

A total of 113 sample sequences of SARS-CoV-2 of Nepali strains were obtained from publicly accessible GISAID database (GISAID, 2020). The samples were mapped using Burrows-Wheeler Aligner; variant calling was carried out using freebayes and a custom manual script was used for variant annotation against the reference genome, SARS-CoV-2 genome from Wuhan (Accession ID MN908947). Visual representation was done using Excel and also displayed using a phylogenetic tree. A total of 847 mutations in various positions in the isolates was identified. Where, there was a total of 467 missense and 375 synonymous mutations. While synonymous mutations do not have much effect on the virus since it results in the similar amino acid, missense mutations tend to change the structure and hence its ability to bind, enter into host, replicate and spread. This paper helps give an overall understanding on the current mutations in Nepali isolates which may help in further studies of Nepali SARS-CoV-2 genomes and during vaccine development worldwide.

Chapter 1: Introduction

The world as we know it had changed, life had changed, every individual had to go through this experience, regardless of location, gender and even age, all because of a thing we cannot even see or touch. Moreover, just when we think it is over, it reemerges stronger and smarter, a virus. More specifically the coronavirus. Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) is a highly transmissible and pathogenic coronavirus (Hu, 2022) and is a part of the large family of coronaviruses. Possible modes of transmission for SARS-CoV-2, include contact, droplet, airborne, fomite, fecal-oral, bloodborne, mother-to-child, and animal-to-human transmission (Transmission of SARS-CoV-2: Implications for Infection Prevention Precautions, 2020) which was similar to patients with SARS and MERS; these patients showed symptoms of viral pneumonia, including fever, cough and chest discomfort, and in severe cases dyspnea and bilateral lung infiltration (Hu,2022) and hence, has had a catastrophic effect on the world's demographics resulting in more than 6 million deaths worldwide as of March 2022, emerging as the most consequential global health crisis since the era of the influenza pandemic of 1918 (Cascella et al., 2022).

A mutation (also referred to as viral mutation or genetic mutation) of the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) virus is a change in the genetic sequence of the SARS-CoV-2 virus when compared with a reference sequence such as Wuhan-Hu1 (the first genetic sequence identified) or USA-WA1/2020 (the first identified in the United States) (Center for Devices and Radiological Health, 2022b). SARS-CoV-2 has been mutating since it was first sequenced in early January 2020 (Wang, 2021). According to a new research conducted by Wang (2021), the SARS-CoV-2 has a genetic proofreading mechanism achieved by non-structure protein (NSP) 14 in synergy with NSP10 and NSP12 which means it has a

higher fidelity in its transcription and replication process than that of other single-stranded RNA viruses. Of all the mutations in different regions of the SARS-CoV-2 genome, the spike region is of major focus, as it facilitates the virus entry to the host via interaction with Angiotensin Converting Enzyme 2 (ACE2) receptors (Maurya, 2022). This genetic variation may impact the virus's properties such as transmission (for example, it may spread more easily) or the severity of symptoms on infected individuals (for example, it may cause more severe disease) (Center for Devices and Radiological Health, 2022).

The genetic variation poses quite a challenge but medicine is as evolving as diseases. These mutations, once detected, sequenced and identified can be used to further enhance our understanding of how this virus works and how we can find a way to survive and hopefully destroy it, for the betterment of human kind.

Aims and objectives:

The aim of this study is to find and analyse protein mutations in genomic sequences of 113 samples of SARS-CoV-2 in Nepali isolates and to understand the patterns of these mutations.

Chapter 2: Literature review

Structure:

Coronaviruses are a part of the Coronaviridae family. They have been classified into four genera that include α -, β -, γ -, and δ -coronaviruses (Wu et al., 2020). They are enveloped viruses having a positive, single-stranded RNA genome packed inside a capsid. The SARS-CoV-2 has four structural proteins: the E and M protein (responsible for forming the viral envelop), N protein (binds to the virus's RNA genome) and S protein (binds to host's receptors). The capsid consists of the N protein (which binds to the virus's RNA genome) enclosed by a membrane containing three other proteins: the spike glycoprotein (S), membrane (M) protein and envelope (E) protein where the E and M proteins are involved in the virus budding process and the S protein is the main part in binding to the host receptor, moderating membrane fusion and eventually entry into the host's cells. The RNA genome of SARS-CoV-2 contains 29,811 nucleotides encoding 29 proteins (Katsnelson, 2020), in addition to the 4 structural protein, the 3'-end also encodes nine accessory proteins (Orf3a, Orf3b, Orf6, Orf7a, Orf7b, Orf8, Orf9b, Orf9c, Orf10) (Mariano et al., 2022). SARS-CoV-2 proteins comprise two large polyproteins: ORF1a and ORF1ab that proteolytically cleaves by a virus-encoded protease into individual replicase complex non-structural proteins to form 16 non-structural proteins (nsp1-16) (Redondo et al., 2021).

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2)

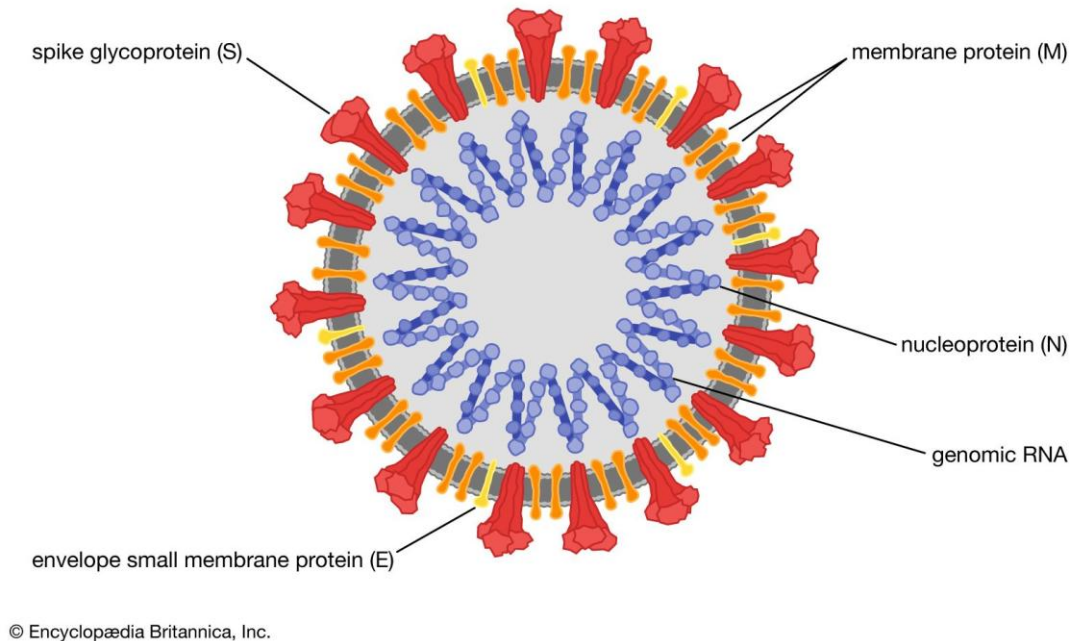


Figure 1: SARS-CoV-2

Viral pathway into host cell:

Entry into host cells is a crucial part of viral infectivity and pathogenesis. The surface-anchored spike proteins of the virus mediate entry. The SARS-CoV spike (S) protein is composed of two subunits; the S1 subunit containing a receptor-binding domain that engages with the host cell receptor angiotensin-converting enzyme 2 and the S2 subunit mediates fusion between the viral and host cell membranes (Du et al., 2009). SARS-CoV S1 subunit contains a receptor-binding domain (RBD) that specifically recognizes angiotensin-converting enzyme 2 (ACE2) as its receptor and the RBD constantly switches between a standing-up position for receptor binding and a lying-down position for immune evasion (Shang et al., 2020). In order for membrane fusion to take place, the S1 subunit dissociates and S2 subunits undergo structural changes after being proteolytically cleaved by proteases (such as TMPRSS2 and lysosomal proteases cathepsins) at the S1/S2 boundary. Therefore, to enter host cells,

coronaviruses first bind to a cell surface receptor for viral attachment, subsequently enter endosomes, and eventually fuse viral and lysosomal membranes (Shang et al., 2020).

The virion is assisted in entering by TMPRSS2 after the spike protein on the virion attaches to ACE2. The virion releases some of its RNA, some of which is translated into proteins by the machinery of the cell. A replication complex is created by some of these proteins to produce more RNA. The Golgi assembly of these proteins and RNA results in the release of a new virion.

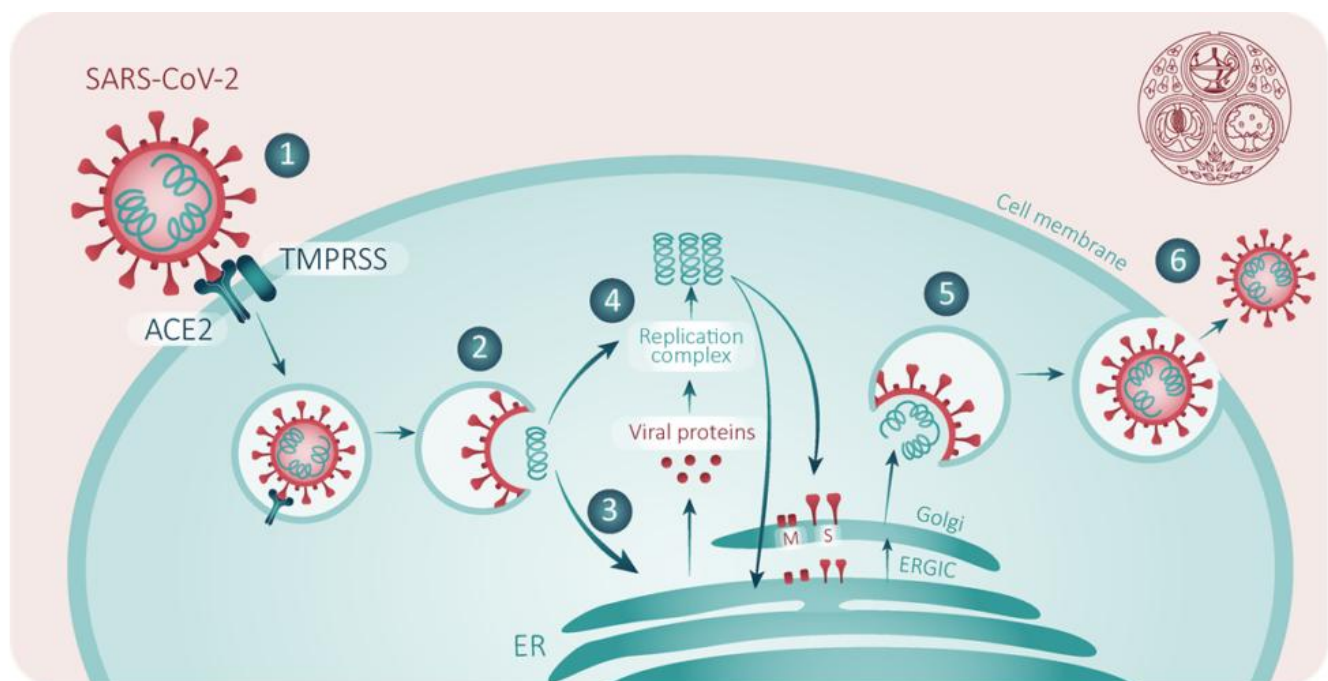


Figure 2: Replication cycle of SARS-CoV-2

Genomic Organization:

A single-stranded positive-sense RNA makes up the SARS-CoV-2 genome. The SARS-CoV-2 genome consists of 13–15 (12 functional) open reading frames (ORFs) that total about 30,000 nucleotides and 9680 amino acid polyproteins. The genome has 11 protein-coding genes and 12 expressed proteins, making up 38% of its total GC content. The ORFs are arranged as replicase and protease (1a–1b) and major S, E, M, and N proteins, which follows a typical 5'-3' order (Naqvi et al., 2020). The biggest gene, ORF1ab, has two open reading frames that overlap and encode the polyproteins PP1ab and PP1a. 16 non-structural proteins, NSP1–16, are produced after the polyproteins are subdivided and makes up roughly 67% of the genome. The remaining ORFs code for accessory and structural proteins. The hemagglutinin-esterase gene is not present in the SARS-CoV-2 genome. It does, however, have two flanking untranslated sections (UTRs) at the 3' and 5' ends of the 265 nucleotides.

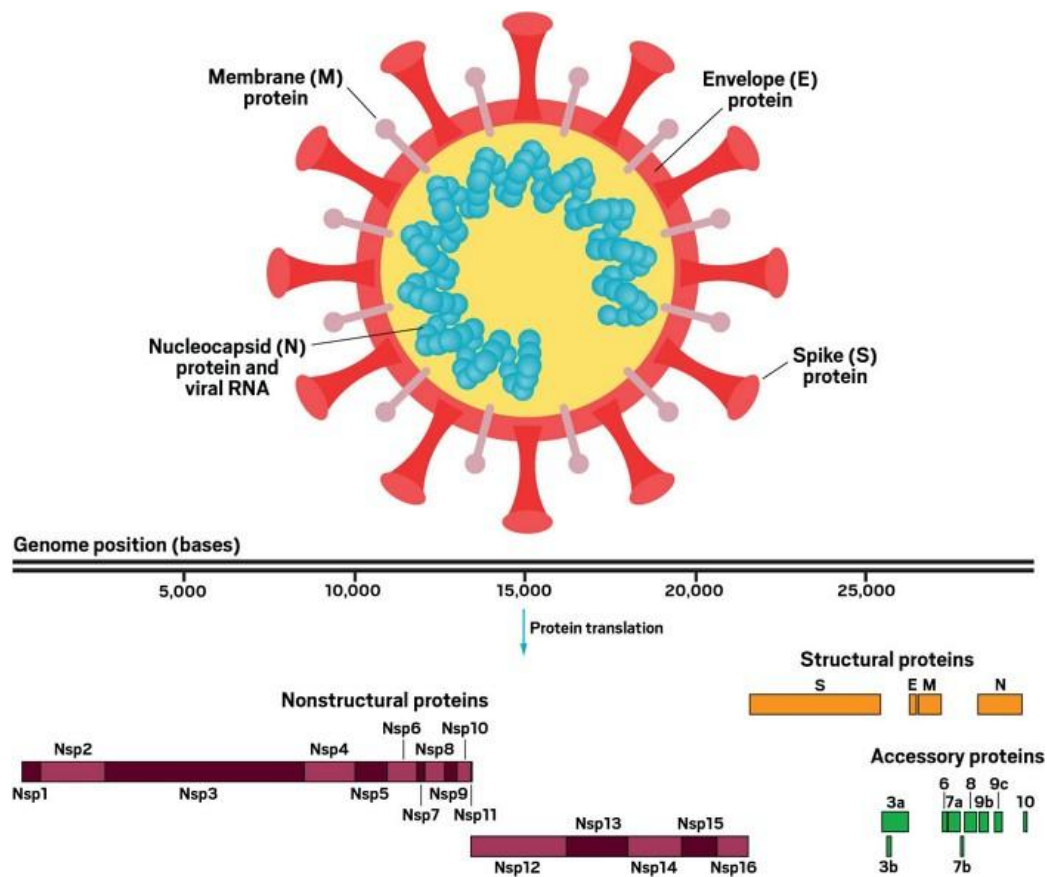


Figure 3: SARS-CoV-2 has four structural proteins

Characteristics and Significance of SARS-CoV-2 proteins

S protein

A transmembrane (TM) domain anchored in the viral membrane, an extracellular N-terminus, and a brief intracellular C-terminal portion make up the 180–200 kDa S protein. The S protein has two distinct subunits, S1 and S2, and while it typically exists in a metastable, prefusion conformation, once the virus interacts with the host cell, the S protein undergoes substantial structural rearrangement, which enables the virus to fuse with the membrane of the host cell. RBD in the S1 subunit interacts to ACE2 in the aminopeptidase N region of the cell receptor.

Mutations of key residues play an important role in enhancing the interaction with ACE2. F486 in SARS-CoV-2, instead of I472 in SARS RBD, forms strong aromatic–aromatic interactions with ACE2 Y83, and E484 in SARS-CoV-2-CTD, instead of P470 in SARS RBD, forms ionic interactions with K31, which leads to higher affinity for receptor binding than RBD of SARS-CoV (Huang et al., 2020). Hence, mutations in the spike protein can increase change viral infectivity, receptor binding capacity and immunogenicity.

N protein

The N (nucleocapsid) protein is only complexed in the nucleocapsid's structural organization. It possesses an RNA-binding domain, also known as a linker region, a C-terminal domain, and three other highly conserved domains. N protein contributes in RNA packaging, helps to assist virion assembly and improves virus transcription efficiency in addition to its role in the organization of viral genome.

Mutations in N proteins cause an increased transmission and virulence of SARS-CoV-2 variants. In addition to mutations in the spike protein, mutations in the nucleocapsid protein are important for viral spreading during the pandemic (Wu et al., 2021).

E protein

The envelope (E) protein is considered the smallest of the main structural proteins of CoVs. It is a short protein that consists of 76–109 amino acids that occur with a short hydrophilic N-terminal (7–12 amino acids), and a bigger hydrophobic transmembrane domain (25 amino acids), ending in a long hydrophilic carboxyl C-terminal that cover most of the protein (Aldaais et al., 2021).

This protein's significance stems from its crucial function in SARS. It facilitates the virus's release and entry into the host cell. The protein also modifies a few cellular functions,

supporting its function in regulating the pathogenicity of the virus. As a result, it ought to be regarded as a crucial virulence component of the new SARS-CoV-2.

M protein

A crucial part of the virus's function is played by the membrane protein (M), particularly during the budding and assembly phases. It mainly consists of a big carboxyl-terminal region at the interior of the virion, three hydrophobic transmembrane domains, and a short amino acid terminal on the virion exterior (Aldaaes et al., 2021).

Understanding the system of the new coronavirus is essential because the M protein is a dominant structural protein that can combine with other structural proteins including the spike (S) and envelope (E) proteins. Any M protein mutation is anticipated to have a significant influence on interactions with the infected cell since the M protein supports the S protein, which is essential in cell attachment and entrance to the host cell.

SARS-CoV-2 and Nepal:

Infections with the SARS-CoV-2 virus struck Nepal twice, once in 2020 and again in 2021. 127 sequences that were deposited in the GISAID database by various Nepali health authorities were analysed. Nepal had detected at least two variations of concern in 2021, alpha and delta, and one variant of interest, kappa. The delta version, like other countries had overcome the kappa and alpha variants and, by July 2021, had established itself as the dominant variant. The delta variation is thought to have been the main cause of the second wave in Nepal. Additionally, phylogenetic tree research showed that the coronavirus can spread locally and globally.

Chapter 3: Materials & Methods

Sample Collection:

A total of 113 complete sequences of SARS-CoV-2 from Nepal were collected from a total of 181 whole genome sequences using the EpiCoV™ platform from GISAID until August, 2021. Partial sequences, sequences with low coverage and genomes with long stretches of “NNN” were omitted to increase accuracy. The isolates analysed here were obtained from all regions of Nepal.

A full list of the sequences arranged in a table format according to their Accession ID, clade and mutations in all the non-structural and structural proteins are in the link below:

https://docs.google.com/spreadsheets/d/1PfUfsHJrI23JnrfKY8IR5w4_fOX3OI8A0OstHD-FQY/edit?usp=sharing

Mutation Analysis:

Burrows-Wheeler Aligner (BWA) (<https://bio.tools/bwa>) was used for aligning and mapping our sequences against the reference genome, SARS-CoV-2 genome of Wuhan-Hu-1 (Accession ID MN908947), using default parameters. Freebayes (a haplotype-based variant detector) was used for variant calling, i.e. identifying variants from our sequences when compared to our reference sequence and a custom manual script was used for the annotation of our variants.

Phylogenetic Analysis:

In order to verify the relationship among the sequences and identified mutations, a Neighbour-Joining (NJ) phylogenetic tree was constructed using the NJ (Conserved sites) method and Jukes Cantor model (Erickson, 2010) in MAFFT software (<https://mafft.cbrc.jp/alignment/software/>) and topology was evaluated using a bootstrapping resampling of 100.

Statistical Analysis:

Statistical analysis was carried out using Excel. Box plots were created using our selected sequences and reference protein structure with default parameters.

Chapter 4: Results and Discussion

Clade distribution:

All of the samples selected for this paper, i.e. 113 sequences of SARS-CoV-2 contained the clade GK and no others. Clades assist scientists in gaining knowledge about where certain strains of the virus are accumulated. We already know that the virus is mutating and evolving into separate clades with distinct geographical and time distributions (Jaroszewski et al. 2020), where the mutation rate of SARS-CoV-2 has been estimated to be about $(1.19-1.31) \times 10^{-3}/\text{site/year}$ (Paudel et al. 2021). The prevalence of the G clade increased while other clades decreased over time but according to the samples, we have collected from the months between 2020 and 2021, all the strains contain the GK clade. Furthermore, all viruses in the fifth wave were in clade GK (delta variant) (Yavarian, 2022) which emerged by the late 2020. The delta variant had already arrived in Nepal by the second-third wave.

Table 1: Clade distribution in Nepali isolates
--

Clade	Incidence	Percentage
GK	113	100%

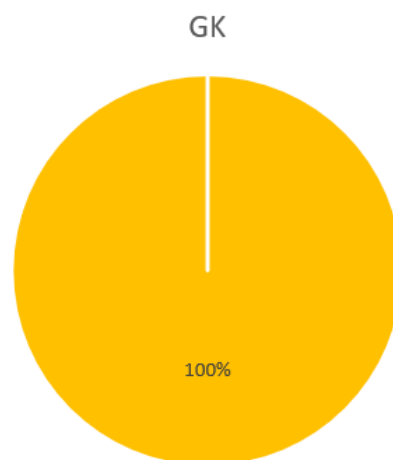


Figure 4: Clade distribution in the selected Nepali isolates
--

Classification of mutations in each gene:

A large portion of mutations observed in SARS-CoV-2 genomes were missense mutations (55.1%), followed by synonymous mutations (44.3%) the rest consist of start/stop gain and loss, as well as mutations in untranslated regions which were identified in this paper (Figure 5). The first gene (ORF1ab) expresses a polyprotein which consists of 16 non-structural proteins (NSPs). The highest number of mutations was seen to be in the polyprotein ORF1ab, consisting of a total of 321 mutations (37.7%) of which 163 were missense and the rest were synonymous. Sub-category ORF7A, 7B and 8 consist Stop-Gain mutations where ORF8 had 2 Stop-Gain mutations and the only one out of all the proteins, to contain a single inframe-del mutation. ORF1ab polyprotein gives rise to NSPs 1 to 16 and these carry out many functions. There was a total of 317 mutations (37.2%) all of the NSPs combined where, 162 were missense and 155 were synonymous with no Stop-Gain and inframe-del mutations. Furthermore, the S protein follows second in the highest number of mutations after ORF1ab, having a total of 82 mutations (9.64%), out of which 57 are missense mutations and 25 are synonymous mutations. The spike protein is a glycoprotein, which mediates the attachment of the virus to the host cell (Yoshimoto, 2020). The human angiotensin-converting enzyme 2 (ACE2) protein on the surface of the host's cell is recognized by the spike protein and so it also mediates entry into host. According to a review article by Thakur et al. (2022), mutations in the S protein will definitely affect its virulence and pathogenicity but most of the mutations lessen the virulence. The mutations have been summarised in figure 5.

The 3'UTR and 5'UTR regions, after alignment and arrangement of the data, was seen to have no mutations or may have mutations but the type was not identifiable, hence so much of it is not discussed here.

Protein	Missense Mutation	Synonymous Mutation	Stop-Gain Mutation	Inframe-Del Mutation	Total
Spike	57	25			82
Orf1ab	163	158			321
Orf3A	35	7			42
Gene E	2	0			2
Gene N	26	15			41
Gene M	1	7			8
Orf6	1	1			2
Orf7A	8	1	1		10
Orf8	4	4	2	1	11
Orf10	6	0			6
Orf7B	2	2	1		5
Nsp1	6	1			7
Nsp2	20	16			36
Nsp3	51	40			91
Nsp4	14	17			31
Nsp5	1	9			10
Nsp6	9	10			19
Nsp7	1	1			2
Nsp8	2	3			5
Nsp9	2	4			6
Nsp10	0	3			3
Nsp11	1	0			1
Nsp12	15	12			27
Nsp13	14	12			26

Nsp14	11	13			24
Nsp15	10	9			19
Nsp16	5	5			10
Total	467	375	4	1	847

Table 2: Distribution of mutation in genes of SARS- CoV-2 isolates from Nepal

The number of each type of mutation was counted after all the samples were aligned and then organized based on the specific protein. A bar chart (Fig. 6) has been generated for easy visualization of the number of mutations each protein has and the total amount of missense and synonymous mutations as they were of significance.

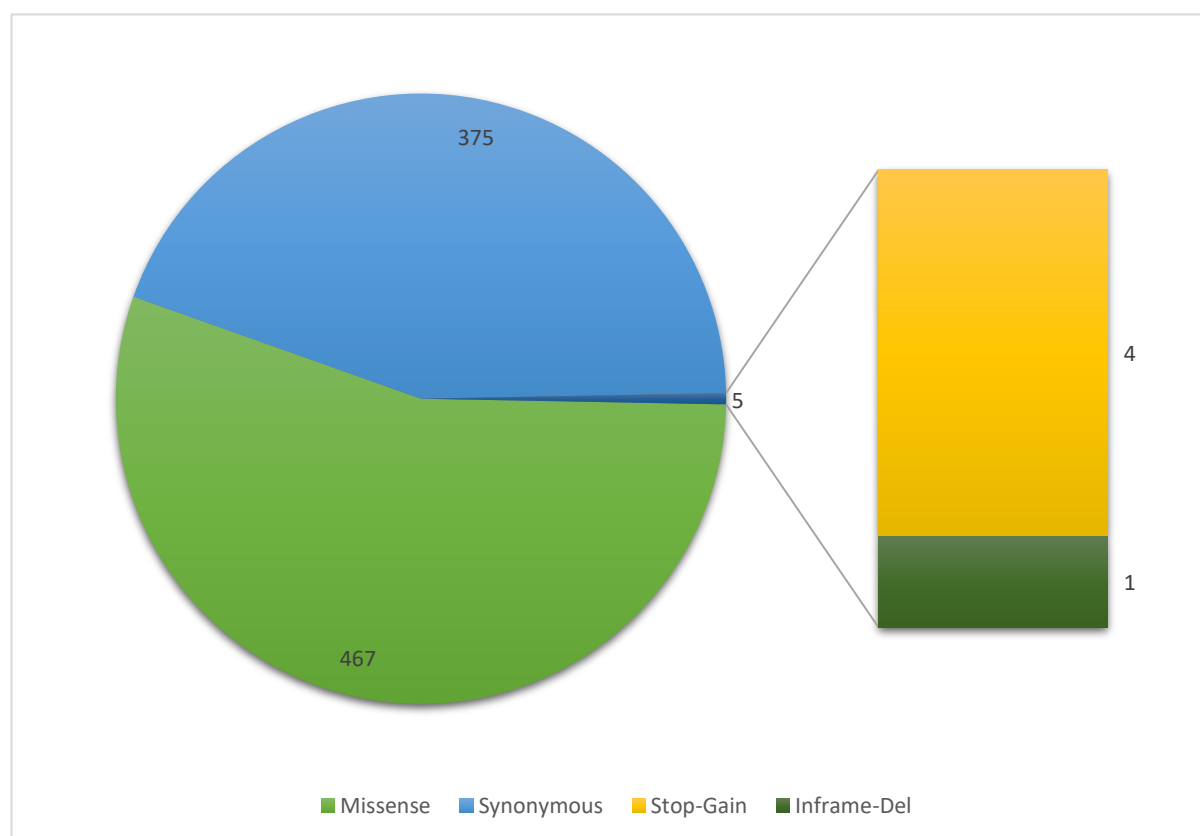


Figure 5: Pie chart showing the total number of each mutation in all samples

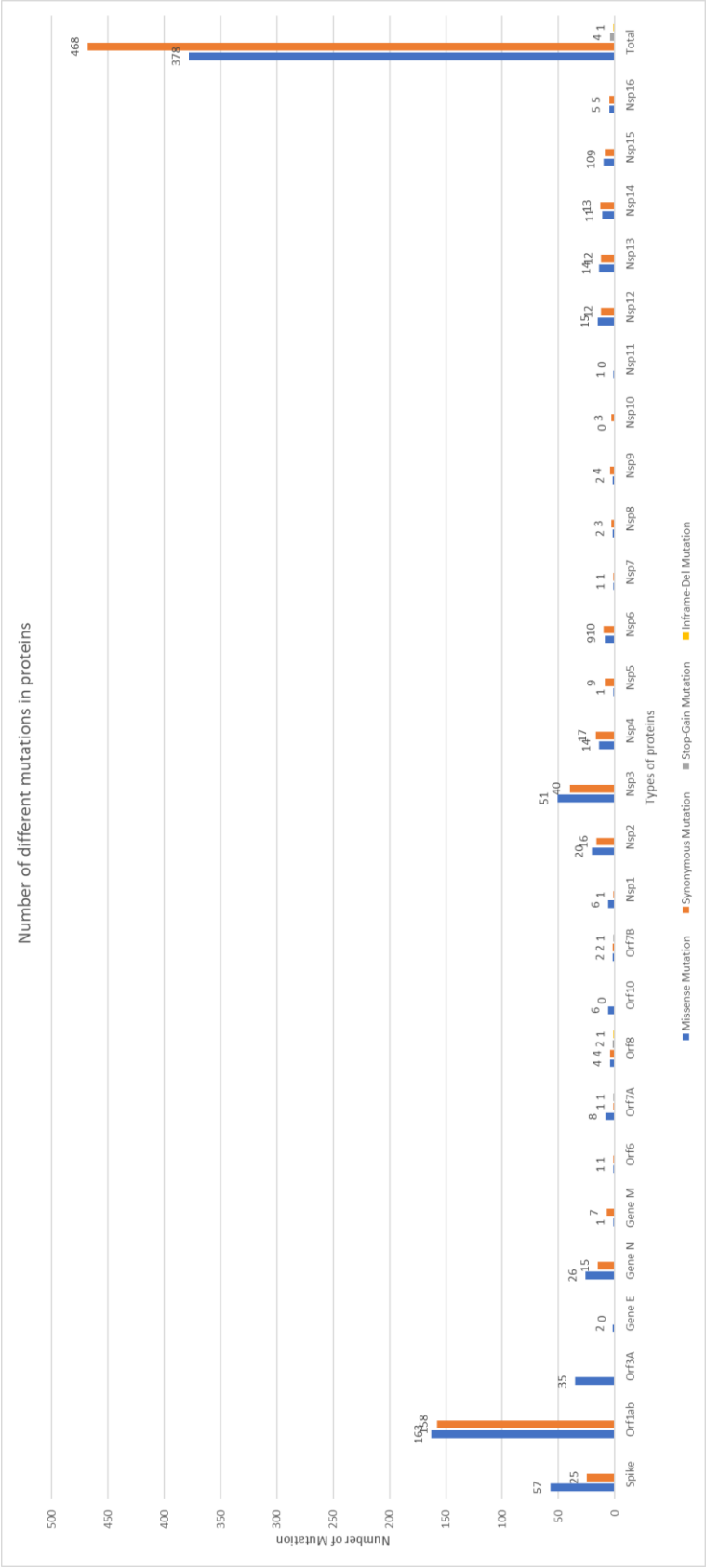


Figure 6: Bar chart of the prevalence of each mutation in genes.

Following a well-known tendency seen in many protein families from many animals, the fraction of missense mutations in structurally defined protein residues of SARSCoV-2 increases with their increasing solvent exposure. While the rate of synonymous mutations exhibits substantially lesser fluctuation, the rate of missense mutations greatly changes along the SARS-CoV-2 genome. It suggests that selection mechanisms at the protein level have a considerable impact on mutations. Figure 7 shows the ranges in which these mutations have occurred among the strains.

The clustering on the phylogenetic tree usually follows the time of isolation or prevalence of specific clades. Since all our strains are within the GK clades, for example, the clade we found in the earliest strain of our samples isolated in June, 2020 is the same as the latest isolate we have in August, 2021. Hence, we can determine that the clusters formed in our tree in figure 8 is based on the time of their isolation in Nepal.

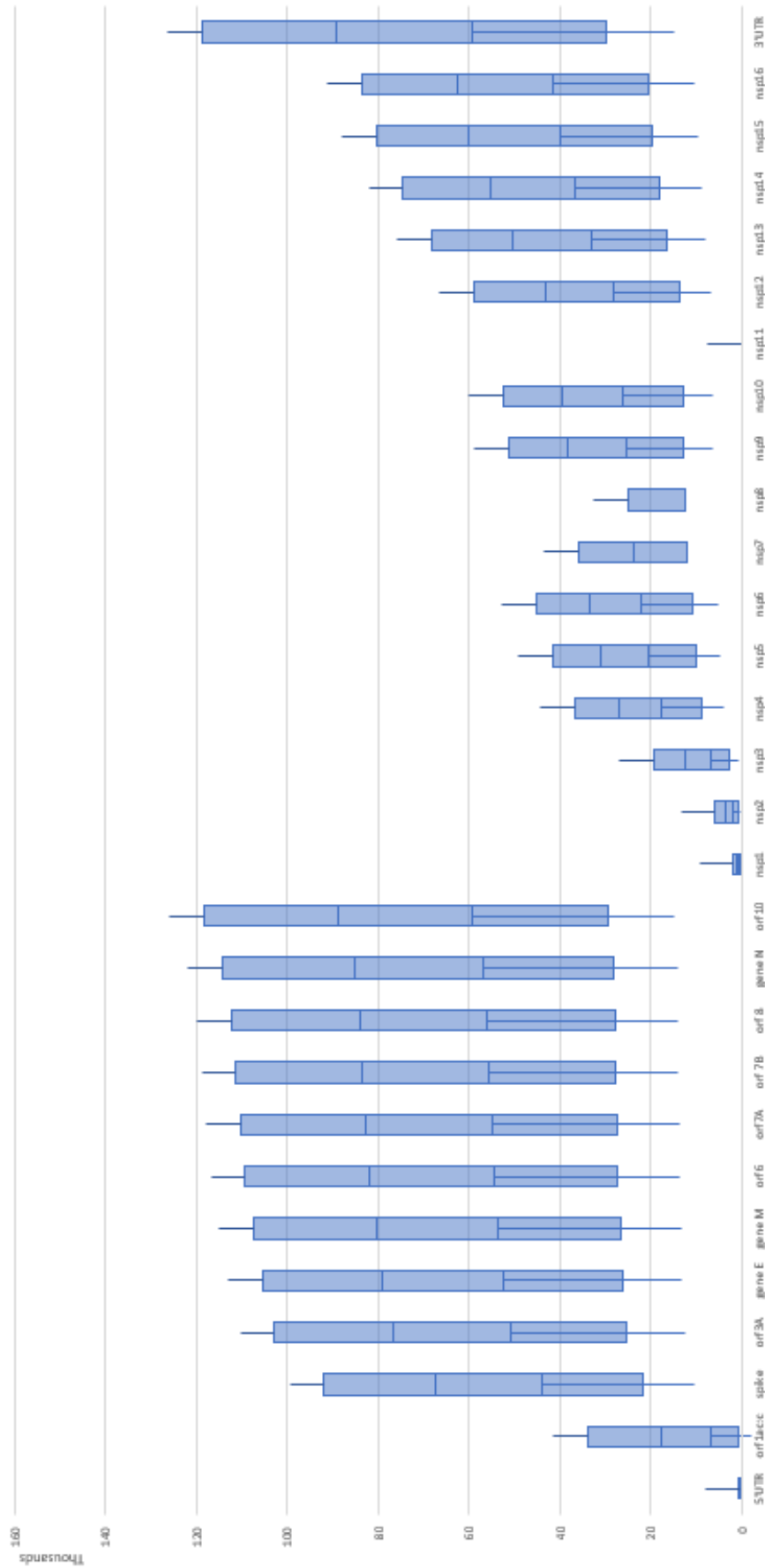


Figure 7: Box-plot showing ranges of mutations in each gene

Phylogenetic Tree for selected samples:



Figure 8: Phylogenetic tree of selected SARS-CoV-2 sequences from Nepal

(Bootstrap value = 100)

As there are a total of 113 sequences, a tree format is used to visualize due to the large number of nodes and branches.

From the tree in figure 9, we can identify five major clusters that have shown significant number of times of mutations. For example, different strains comprising of different ancestral branches and showing significant number of mutations was found in May 2021 in four out of five different clusters on the dates 9th, 28th, 28th and 31st, showing that within a short period of time, the virus had mutation many times and therefore, this indicates that although all isolates

belong to the same clade, mutations in genomes played a pivotal role in their genomic heterogeneity.

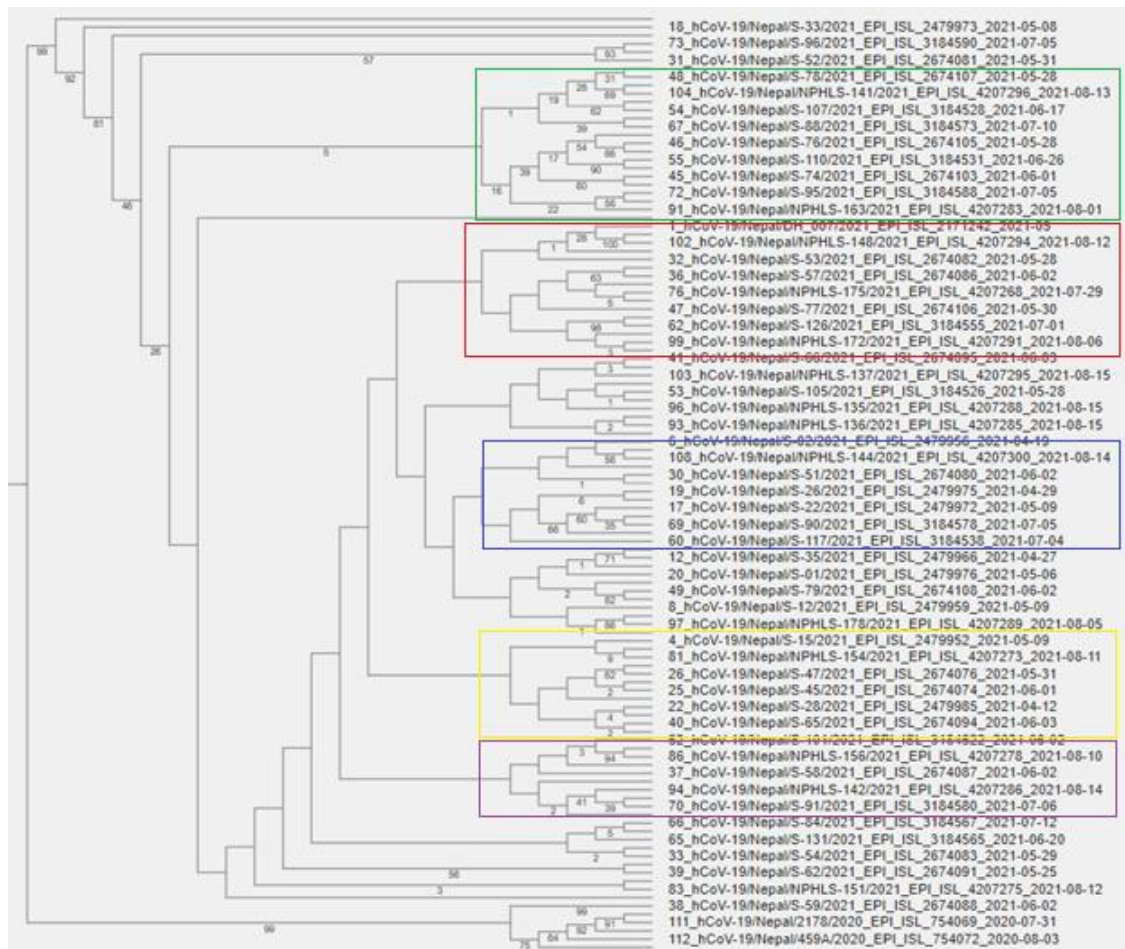


Figure 9: Phylogenetic tree showing five major clusters based on multiple of mutations of each strain.

Chapter 5: Conclusion & Future directions

Although it may possibly be of practical significance, the investigation of the SARS-CoV-2 virus's mutation pattern is intriguing from an evolutionary perspective. For example, it enables the prediction of which of the SARS-CoV-2 proteins' epitopes surfaces are more likely to experience widespread alterations in the future. Similar predictions can be helpful for regions that are the focus of PCR-based diagnostic testing for Covid-19, as there is evidence that the accumulation of numerous mutations already reduces the accuracy of some of these tests.

The SARS-CoV-2 virus developed a number of mutations, most of which affected the spike protein, which is essential for developing vaccine candidates. Numerous vaccines are undergoing clinical trials and are yielding promising results, but their effectiveness must still be verified in various SARS-CoV-2 mutants. We talked about the effectiveness of the vaccine against every SARS-CoV-2 mutation. Also considered is the question of these vaccinations' safety. Understanding coronavirus mutation is crucial for creating novel vaccines that offer long-lasting protection and can neutralize a wide range of mutant forms. When planning vaccine efficacy trials, a correct study methodology must also be taken into account, which further enhanced the study results. Equally crucial are taking preventative measures to safeguard against the virus, such as developing a vaccine.

Chapter 6: References

Aldaaais, E. A., Yegnaswamy, S., Albahrani, F., Alsowaiket, F., & Alramadan, S. (2021). Sequence and structural analysis of COVID-19 E and M proteins with MERS virus E and M proteins—A comparative study. *Biochemistry and Biophysics Reports*, 26, 101023. <https://doi.org/10.1016/j.bbrep.2021.101023>

Cascella, M., Rajnik, M., Aleem, A., Dulebohn, S. C., & Napoli, R. D. (2022). Features, Evaluation, and Treatment of Coronavirus (COVID-19) [E-book]. StatPearls Publishing. <https://www.ncbi.nlm.nih.gov/books/NBK554776/>

Center for Devices and Radiological Health. (2022, August 3). SARS-CoV-2 Viral Mutations: Impact on COVID-19 Tests. U.S. Food and Drug Administration. <https://www.fda.gov/medical-devices/coronavirus-covid-19-and-medical-devices/sars-cov-2-viral-mutations-impact-covid-19-tests>

Center for Devices and Radiological Health. (2022b, July 14). SARS-CoV-2 Viral Mutations: Impact on COVID-19 Tests. U.S. Food and Drug Administration. <https://www.fda.gov/medical-devices/coronavirus-covid-19-and-medical-devices/sars-cov-2-viral-mutations-impact-covid-19-tests>

Du, L., He, Y., Zhou, Y., Liu, S., Zheng, B. J., & Jiang, S. (2009). The spike protein of SARS-CoV--a target for vaccine and therapeutic development. *Nature reviews. Microbiology*, 7(3), 226–236. <https://doi.org/10.1038/nrmicro2090>

Hu, B. (2022, February 23). Characteristics of SARS-CoV-2 and COVID-19. Nature. https://www.nature.com/articles/s41579-020-00459-7?error=cookies_not_supported&code=d3508cb4-f4f6-4c51-83db-b41c6c237e6b

Hu, B., Guo, H., Zhou, P., & Shi, Z. L. (2020). Characteristics of SARS-CoV-2 and COVID-19. Nature Reviews Microbiology, 19(3), 141–154. <https://doi.org/10.1038/s41579-020-00459-7>

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

MAFFT - a multiple sequence alignment program. (n.d.). Retrieved September 13, 2022, from <https://mafft.cbrc.jp/alignment/software/>

Mariano, G., Farthing, R. J., Lale-Farjat, S. L. M., & Bergeron, J. R. C. (2020). Structural Characterization of SARS-CoV-2: Where We Are, and Where We Need to Be. Frontiers in Molecular Biosciences, 7. <https://doi.org/10.3389/fmolb.2020.605236>

Maurya, R., Mishra, P., Swaminathan, A., Ravi, V., Saifi, S., Kanakan, A., Mehta, P., Devi, P., Praveen, S., Budhiraja, S., Tarai, B., Sharma, S., Khyalappa, R. J., Joshi, M. G., & Pandey, R. (2022). SARS-CoV-2 Mutations and COVID-19 Clinical Outcome: Mutation Global Frequency Dynamics and Structural Modulation Hold the Key. Frontiers in Cellular and Infection Microbiology, 12. <https://doi.org/10.3389/fcimb.2022.868414>

ORF1ab ORF1a polyprotein;ORF1ab polyprotein [Severe acute respiratory syndrome coronavirus 2]. (2022, August). <https://www.ncbi.nlm.nih.gov/gene/43740578#:~:text=~%20ORF1ab%2C%20the%20largest%20gene%2C,a%20%2D1%20ribosomal%20frameshifting%20event>

Redondo, N., Zaldívar-López, S., Garrido, J. J., & Montoya, M. (2021). SARS-CoV-2 Accessory Proteins in Viral Pathogenesis: Knowns and Unknowns. *Frontiers in Immunology*, 12. <https://doi.org/10.3389/fimmu.2021.708264>

Transmission of SARS-CoV-2: implications for infection prevention precautions. (2020, July 9). World Health Organization (WHO). <https://www.who.int/news-room/commentaries/detail/transmission-of-sars-cov-2-implications-for-infection-prevention-precautions>

Wang, R., Chen, J., Gao, K., Hozumi, Y., Yin, C., & Wei, G. W. (2021). Analysis of SARS-CoV-2 mutations in the United States suggests presence of four substrains and novel variants. *Communications Biology*, 4(1). <https://doi.org/10.1038/s42003-021-01754-6>

Wu A., Peng Y., Huang B., Ding X., Wang X., Niu P., Meng J., Zhu Z., Zhang Z., Wang J. Genome composition and divergence of the novel coronavirus (2019-nCoV) originating in China. *Cell Host Microbe*. 2020;27:325–328.

Wu, H., Xing, N., Meng, K., Fu, B., Xue, W., Dong, P., Tang, W., Xiao, Y., Liu, G., Luo, H., Zhu, W., Lin, X., Meng, G., & Zhu, Z. (2021). Nucleocapsid mutations R203K/G204R increase the infectivity, fitness, and virulence of SARS-CoV-2. *Cell Host & Microbe*, 29(12), 1788–1801.e6. <https://doi.org/10.1016/j.chom.2021.11.005> [Original source: <https://studycrumb.com/alphabetizer>]

Katsnelson, A. (2020, April 1). *SARS-CoV-2 has four structural proteins* [Image]. Chemical and Engineering News. <https://cen.acs.org/biological-chemistry/infectious-disease/know-novel-coronaviruss-29-proteins/98/web/2020/04>

Ward, P. (2020, April 6). *Replication cycle of SARS-CoV-2* [Image]. Faculty of Pharmaceutical Medicine. <https://www.fpm.org.uk/blog/covid-19-sars-cov-2-pandemic/>

Encyclopædia Britannica. (2022, August 4). *SARS-CoV-2* [Image]. Encyclopædia Britannica. <https://www.britannica.com/science/coronavirus-virus-group/images-videos#/media/1/138325/245650>

Yavarian, J. (2022, May 2). Whole genome sequencing of SARS-CoV2 strains circulating in Iran during five waves of pandemic. *PLOS ONE*. Retrieved September 13, 2022, from <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0267847>