Detection Of Genetic Diversity of SARS-CoV-2 in Wastewater to Monitor COVID-19 Transmission

By

Subyeta Binte Sarwar (ID - 22276014)

A thesis submitted to the Department of Mathematics and Natural Sciences in partial fulfillment

of the requirements for the degree of

M.Sc. in Biotechnology

Department of Mathematics and Natural Sciences

Brac University

August 2023

Declaration

It is hereby declared that

1. The thesis submitted is my own original work while completing a degree at Brac University.

2. The thesis does not contain material previously published or written by a third party, except where this is appropriately cited through full and accurate referencing.

3. The thesis does not contain material which has been accepted, or submitted, for any other degree or diploma at a university or other institution.

4. I have acknowledged all main sources of help.

Students' Full Names & Signature:

Subyeta Binte Sarwar

Approval

The thesis/project titled "**Detection Of Genetic Diversity of SARS-CoV-2 in Wastewater to Monitor COVID-19 Transmission**" submitted by **Subyeta Binte Sarwar (ID - 22276014)** of Fall 2022 has been accepted as satisfactory in partial fulfillment of the requirement for the degree of Master of Science in Biotechnology on the 16th of August, 2023.

Examining Committee:

Supervisor:

(Member)

Iftekhar Bin Naser, PhD

Associate Professor, Department of

Mathematics and Natural Sciences

Brac University

Supervisor:

(External)

Mohammad Enayet Hossain, PhD

Associate Scientist, One Health Laboratory,

Infectious Disease Division,

ICDDR, B

Program Coordinator:

(Member) _____

Munima Haque, PhD

Associate Professor, Department of

Mathematics and Natural Sciences

Brac University

Departmental Head:

(Chairman) _____

Md. Firoz H. Haque, PhD

Associate Professor and Chairperson,

Department of Mathematics and Natural Sciences

Brac University

Acknowledgement

I would first like to express my undying gratitude to the Almighty for keeping me in good health, despite the circumstances, and allowing me to conduct my research without any major obstacles.

My profound respect and gratitude for my thesis advisor, **Dr. Iftekhar Bin Naser, Associate Professor** of the Biotechnology, School of Data and Sciences at BRAC University. The door to Prof. Naser's office was always open whenever I ran into trouble or had a question about my research or writing. He helped me shape the paper by steering me in the right direction whenever I asked.

My sincerest thanks to our laboratory head at One Health Laboratory, ICDDR, B **Dr. Mohammad Ziaur Rahman, Senior Scientist,** for giving me the opportunity to work under his guidance. His expertise and experience in the field of Zoonotic Disease and Public Health research have inspired me to move forward in this field.

More importantly, I would like to thank my supervisor at the One Health Laboratory, ICDDR, B, **Dr. Mohammad Enayet Hossain, Associate Scientist**. I could not have completed this daunting research work without his cooperation. My supervisor was not only a mentor but was also a constant support encouraging me to go beyond my capabilities. Thank you for trusting me even when I had doubts on myself.

I would like to thank **Rashedul Hasan** for being there every step of the way, from the beginning till the end. I believe you have taught me to not only be an efficient and skilled worker but also taught me what it means to be a genuinely kind person, for that I will always be indebted to you.

Additionally, I would like to express my gratitude towards the rest of the members of One Health Lab: **Mojnu Miah**, **Jenifar Quayum Ami** and **Mahmudul Hasan Joy** for their constant support.

Thank you for teaching me, correcting me when I made mistakes and looking out for me during my time at ICDDR, B. This journey truly was wonderful because of you.

I would like to especially mention **Sabik Khair** for his tremendous help while creating the visual aid for my research. Without him, I would have probably stuck with basic bar charts and tables to visualize data. Thank you for helping me regardless of it sometimes being 4am in the morning.

Lastly, I would like to say that I would not be here without the constant support of my parents and my fiancé. Thank you for pushing me, making me recognize my abilities and not letting me give up when everything seemed hopeless. I could not have done it without you.

Abstract

Coronavirus Disease 2019 (COVID-19), an infectious respiratory disease caused by the viral strain, severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), has become a distinct cause of the global burden and a rising concern to global public health. It has single-handedly destroyed economies of every country it has affected. Despite the policy makers trying to tackle this virus, SARS-CoV-2 having a rapid mutation rate makes it difficult to monitor and create effective therapeutics to completely avoid fatalities. Since the only way to detect presence of COVID-19 and circulating strain was by testing every individual, it was a tedious and very inefficient method as it required a significant amount of time.

This study aims to provide an alternate method of detecting the genetic diversity of SARS-CoV-2 and the strain currently circulating in an area by wastewater monitoring. Collecting samples from community wastewater as well as hospital discharge is faster and does not require testing hundreds of individual. This method provides an understanding of the strains currently present in a selected area as the virus is present in fecal discharge of affected individuals.

We carried out RNA extraction of over 300 samples and selected the samples with positive RT-PCR values and performed whole genome sequencing using Oxford Nanopore Technology. The sequences were then analyzed and the data were compared. It was seen that the most abundant variant of Omicron was of clades XBB and XBB.1 with majority of the mutations taking place in the Non-Structural Proteins and Spike Proteins of the virus, these are responsible for the cell attachments and play a key role in pathogenesis of the virus. This study further helps in understanding how despite mass vaccination against COVID-19, this virus seems to be present in significant concentrations in the community.

Contents

List of Figures:

List of Tables:

1. Introduction

SARS-CoV-2, which originated in Wuhan, China, is the viral strain responsible for the infectious respiratory disease, now infamously termed Coronavirus Disease 2019 (COVID-19) and since its emergence, the virus has mutated and evolved, giving rise to its multiple variants (Cui et al., 2019; Wit et al., 2016). With previously having the world at its clasp and causing the worldwide pandemic in 2020, it has caused 68 million confirmed cases of COVID-19 including 6.95 million deaths (WHO, 2023) up until July 2023, and is the third zoonotic virus after MERS-CoV and SARS-CoV in 2012 and 2002, respectively. The scenario has led to global emergencies owing to its transmission pattern, and a rising global health concern, while scientists provided their unparalleled efforts in the development of effective vaccines and discovering methods to predict possible future spillover of zoonotic diseases (Silva et al., 2021; Yin et al., 2018).

SARS-CoV-2 is reportedly known to be a single-stranded, enveloped, positive-sense RNA virus of about 30 kb in size (Lu et al., 2019) and a predominantly inherited spherical structure, with solar-crown imitating proteins protruding from its surface (Ashour et al., 2020). Similar to conventional betacoronaviruses, they also constitute a 5′ methylated cap and a 3′ polyadenylated tails (Zhou et al.,2020); the 5′ methylated cap includes a non-structural protein coding region (having non- structural proteins, nsp1-16) comprising significant genes, which are termed crucial for viral replication and the 3′-terminal region encodes the 4 structural proteins - spike protein, envelope protein, membrane protein, and nucleocapsid protein, which are absolutely necessary for the life cycle of a virus (Wu et al., 2020).

While the first case of SARS-CoV-2 was detected in China early in December 2019, it was not until March 8, 2020 when the first case of coronavirus was detected in Bangladesh and by August

10, 2020. There were 260,507 COVID-19 cases confirmed by rt-PCR while the number of deaths stood at 3438, according to IEDCR. On March 16, 2020, one week after the first case of COVID-19 was found in Bangladesh, IEDCR reported that the outbreak had spread locally, indicating that children had been infected by the virus as well (Dey et al., 2020). While initially the symptoms for COVID-19 manifested visibly through fever, cough, chest pains, fatigue, myalgia, headache, decrease in lung capacity, etc (Huang et al., 2020) current symptoms are much more asymptomatic. Despite potentially being one of the deadliest variants with high transmissibility, the Omicron variant tends to show flu like representations, causing possible positive cases to go unnoticed and untested (Kumar et al., 2022).

However, in order to take decisions, the continuous monitoring of SARS-CoV-2 mutations and its abundance in the community needs to be carried out to avoid a possible outbreak/wave. A quick and effective way to do this would be wastewater surveillance. SARS-CoV-2 and other enteric pathogen trends can be observed by wastewater surveillance, which tracks wastewater released from individual household septic tanks to public drains, surface water, and sewage treatment facilities (Meng et al., 2021). These samples not only give an overall outlook on the presence of pathogens but it is also cheaper and requires less manpower. Additionally, wastewater sample collection does not involve direct communication with people and so the difficulty in gaining samples is significantly lower (Meng et al., 2021). It had also been reported that SARS-CoV-2 can be detected in feces even before symptoms appear and is able to persist for longer periods. Other studies have shown that regardless of symptom representation, SARS-CoV-2 can be detected from feces even for patients whose rt-PCR and respiratory serum samples tested negative for the virus (Wu et al., 2020; Cheung et al., 2020; Zheng et al., 2020). Since the viral load of SARS-Cov-2 as well as other pathogens present in these samples can easily be calculated, it offers insight in to the

presence of pathogens which makes wastewater surveillance an excellent prospect for determining and quite possibly avoiding a disease outbreak (Vallejo et al., 2022).

2. Background

2.1 History Of SARS-CoV

The first coronavirus strains, HCoV-229E and HCoV-OC43, were discovered in the 1960s and were responsible for a number of minor upper respiratory tract illnesses. Following the 2003 appearance of SARS-CoV, which was in charge of one of the biggest epidemics at the time, HCoV-NL63 and HKU1 were isolated and identified in the Netherlands and Hong Kong, respectively. (Burrell et al., 2017Over time, additional strains had appeared, causing disastrous outbreaks. These widespread and potentially fatal outbreaks were caused by strains like SARS-CoV in 2003, MERS-CoV in 2012, and the continuing SARS-CoV-2 in 2019. (Stadler et al., 2003; Memish et al., 2020). The 2003 SARS-CoV emerged from the Guangdong province of southern China in November 2002 and its rapid transmission led to 29 countries being infected and a total of 8096 confirmed cases with 774 deaths by August 2003 (Costa et al., 2020; Xu et al., 2004). After May 2004, however, human SARS-CoV cases were not documented, until 2012 when, MERS-CoV had spread from Saudi Arabia and was rapidly killing people as a result of its outbreak. While MERS is still around, albeit in a more subdued form, the biggest and most recent SARS-CoV-2 outbreak first appeared in the middle of the year 2019. SARS-CoV-2, the pathogen responsible for the current worldwide pandemic, spread considerably more quickly around the world. The virus, known for its transmission through close contact, via aerosols and respiratory droplets exhaled during talking, breathing, coughing or sneezing (WHO,2021), include pathways such as direct, indirect or close contact, droplet, airborne, fomite, fecal-oral, bloodborne, mother-to-child, and animal-to-human. (Liu et al., 2020). Subsequently, symptoms may range within the non-specific range being asymptomatic or lead to severe pneumonia and death. COVID-19 affects individuals from all present age ranges however, it is far more prevalent and life-threatening among people of

ages 70 and higher, and in those with present underlying comorbidities (Prem et al.,2020; Landi et al.,2020). While this December 2019 emerged disease in Wuhan, China had been pronounced a public health emergency of international concern (PHEIC) by the World Health Organization (WHO) on 30 January 2020 and a pandemic on 11 March 2020 (WHO; 2022), countries all over the world struggled to prepare and outweigh its drawbacks. Simultaneously, countries towards their initial steps for preparation, adapted to symptomatic treatment approaches while waiting for the production and arrival of vaccines. With extended and progressive research, scientists had sequenced the first SARS-CoV-2 genome and published it on 10th January 2020 (Krammer; 2020).

2.2 Genomic Constitution and Pathomechanism Of SARS-CoV-2

As mentioned earlier, the positive sense RNA virus, SARS-CoV-2 is conjured of a spike protein, also known as the S glycoprotein, is a transmembrane protein located on the virus surface and appears to include homotrimers on the surface as well. These S proteins facilitate the binding of viral particles to the host through the angiotensin-converting enzyme 2 (ACE2) receptors, which are present in abundance in the lower respiratory tract region (Hofmann et al; 2020). Attachment leads to the cleaving of the S glycoprotein by the furin-like protease of the host cell, into an N-terminal S1 subunit consisting of a ACE2-specific receptor-binding domain (RBD), and a viral fusion-facilitating (in transmitting host cells) membrane-bound C-terminal S2 region (Astutie et al; 2020, Coutard et al;2020, Gao J. et al; 2020). The nucleocapsid protein or N protein is situated at the endoplasmic reticulum-Golgi region, attached to the viral nucleic acid material. Consequently, the N protein is known to be associated with processes implicating the genome of the virus, its cycle of replicating and the host cell response towards the viral infection (Tai W. et al;2019). Moreover, while the Membrane protein or M protein influences the shape of the viral envelope and the binding to other structural proteins, the envelope or E protein governs the viral

multiplication as well as maturation (Shoemann et al;2019). The 16 non-structural proteins also contribute to the regulation of the viral infection mechanism. Owing to these very proteins and their contributory roles, pathogenesis appears to be greatly aided. Other than the 16 NSPs on ORF1ab, there are other accessory proteins and structural proteins present on the other ORFs.

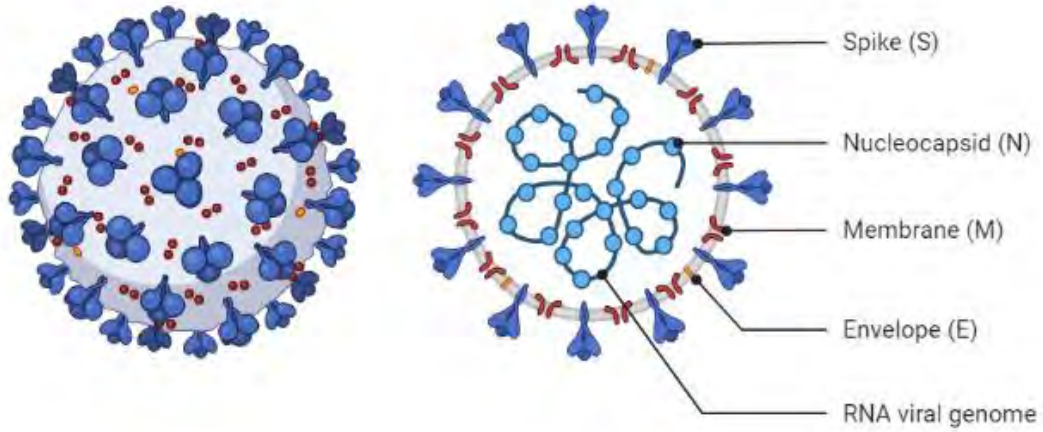

Figure 1: Structural Composition of SARS-CoV-2. Showing the genomic structures: spike proteins, nucleocapsid protein, membrane protein, envelope protein and the RNA viral genome. Retrieved from (El Demerdash et al; 2021)

Upon transmission, the virus travels to the host airways, and then attaches and enters the host cell with the help of the aforementioned spike glycoprotein. In order to progress through the viral infection, the initial step includes the binding of the virus' S1 subunit's RBD to the host cell's ACE2 peptidase domain. The virus attaches to the ACE2 receptor in sync with transmembrane serine protease 2 of the host, which are present in the airway epithelial cells and vascular endothelial cells (Cevik et al; 2020). Consequently, this leads to membrane fusion, transferring the viral genome into the cytoplasm of the host cell followed by replication, assembly, maturation and release of the virus. However, the starting mechanism implicating genome entry included the

translation process of the virus and after the viral polymerase protein translation, RNA replication followed by sub genomic transcription takes place. The structural proteins are also translated and formed, meaning that they will undergo combination with nucleocapsid. Conclusively, through an exocytosis approach, the viral particles mature and are released to infect more and more host cells (Iqbal H. et al; 2020).

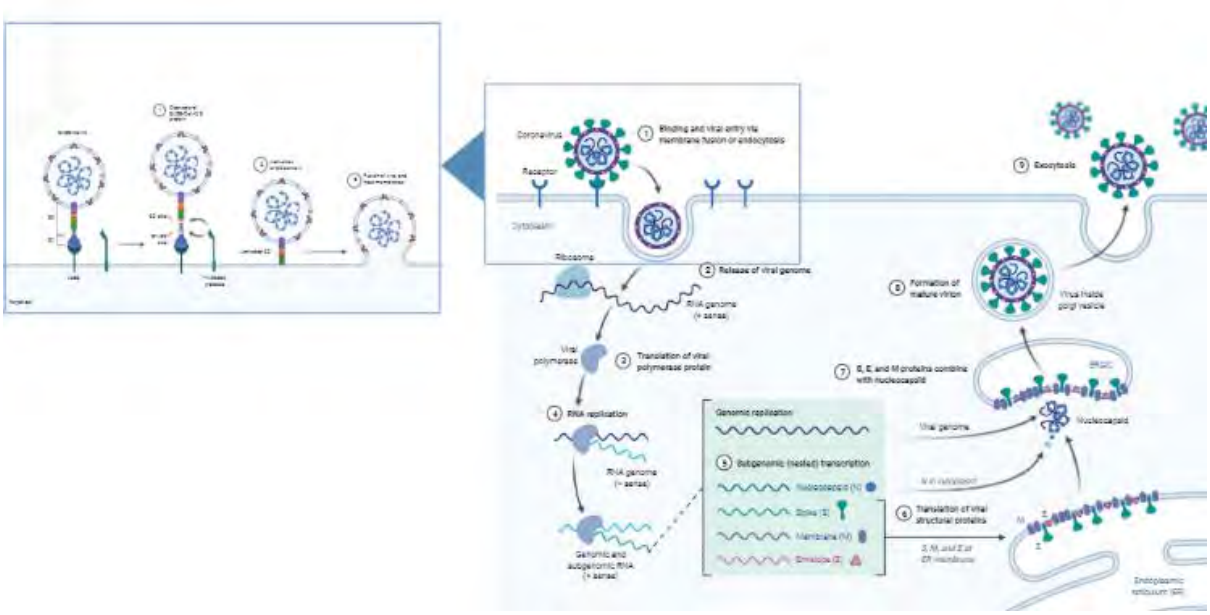

Figure 2: Viral Entry and Pathomechanism of SARS-CoV-2 retrieved from (Hajimonfarednejad et al;2023)

2.3 SARS-CoV-2 Variants

After emerging in Wuhan, China, and aCtuiring a number of mutations over time, SARS-CoV-2 was able to evolve into a variety of different forms due to high transmission and mutation rates. Due its rapid evolutionary abilities, all of the variants are categorized in to three different classifications (CDC, 2023):

1. Variant of Interest (VOI), this strain may have limited prevalence but may require public health actions, thus epidemiologic studies may be carried out to assess the risk a VOI might have.

2. Variant of Concern (VOC): strains that have a high transmissibility and causes sever diseases that might end in fatalities. Most of the times, the strain reduces antibody neutralization and thus reduces effectiveness of treatments and vaccines. Omicron has been categorized as a VOC.
3. Variants Being Monitored (VBM): Consists of lineages or strains that previously caused more severe disease but do not pose a significant or imminent threat. This includes all strains except Omicron.

It has been hypothesized that an adaptation that alters the interaction between Spike and ACE2, Furin, and TMPRSS2 is caused by a particular single nucleotide polymorphism (SNP) that results in an amino acid substitution in the Spike protein (D614G) (Eaaswarkhanth et al., 2020). This SNP has shown a greater increase in frequency than other SNPs. Furthermore, the mutation D614G is solely responsible for the evolutionary shift from the previous SARS-CoV variants as it became a permanent point mutation that is present in all of the later strains of SARS-CoV-2. Furthermore, more common mutations in Omicron were located in the Receptor-Domain Binding region (RBD) and the Spike glycoprotein. These mutations in the RBD and Spike protein denote to the increased pathogenicity and brings about changes between the viral-host attachment system.

2.4 Objectives of this Study

Omicron being the latest installment to the deadly virus, SARS-CoV-2, is significantly different from the original hCoV WUHAN strain, genetically. It is known to be deadlier due to its asymptomatic manifestations and increased virulence caused by the multiple mutations in its Spike Protein, responsible for invading the human system. As it is well known for being more powerful than its predecessors, it is possible that it might cause another mass health hazard or since SARS-

CoV-2 is able to rapidly adapt and change its genomic features, it is possible for a completely new strain to develop and wreak yet another havoc. Thus, the present study aims:

1. To detect the current strain circulating in Bangladesh.
2. To identify common mutational sites and possible new/unidentified mutation sites that might bring about a shift in its genomic characteristics.
3. To compare the mutations amongst the strains/variants we isolate
4. To assess the efficiency of wastewater as a method of virus monitoring.

2. Materials and Methods:

2.1 Sample Processing

Onsite wastewater samples were collected from sewage/water disposal sites based in local communities and hospitals. Approximately 500ml of samples was collected and sealed in zipper bags from 12 sites in Dhaka city, Cox's Bazar municipality, and Rohingya Camps. To ensure the preservation of the samples, the field assistant immediately placed them in cool boxes with temperature maintained at below 10°C. Upon arrival, we received the samples and placed them in a refrigerator set at 4°C to maintain their integrity until testing.

Throughout the week, 23 samples were collected from multiple locations in Dhaka city and Cox's Bazar. Before processing, collected samples were allowed to reach room temperature. The zipper bags were then shaken or inverted at least 5 times for the contents to homogenize and were allowed to sit for 45 seconds at room temperature. Afterwards, wastewater was transferred to a 50ml falcon for storage purposes and a 15 mL falcon containing 10 ml sample intended for use.

For processing, we used the aliquoted 10mL samples. These samples were first processed using Nanotrap® magnetic viral particle (Ceres Nanoscience), following the manufacturer's instruction, this step essentially was used to concentrate the samples to produce a supernatant of 400 µl. The kit included an enhancement reagent (ER1) and viral nanotrap particles where ER1 ensures proper binding of pathogens to the magnetic beads. Simultaneously, the nanotrap particles trap the pathogens from the samples and bind on to them as unwanted debris are filtered out and discarded. In order to further validate the quality of our work, MS2 Phage was used as an internal control. By multiplexing, detection of MS2 phage alongside SARS-CoV-2 verified our results and proved the accuracy of our study. It was also used to monitor the efficiency of RNA extraction and inhibition of the RT-PCR reaction.

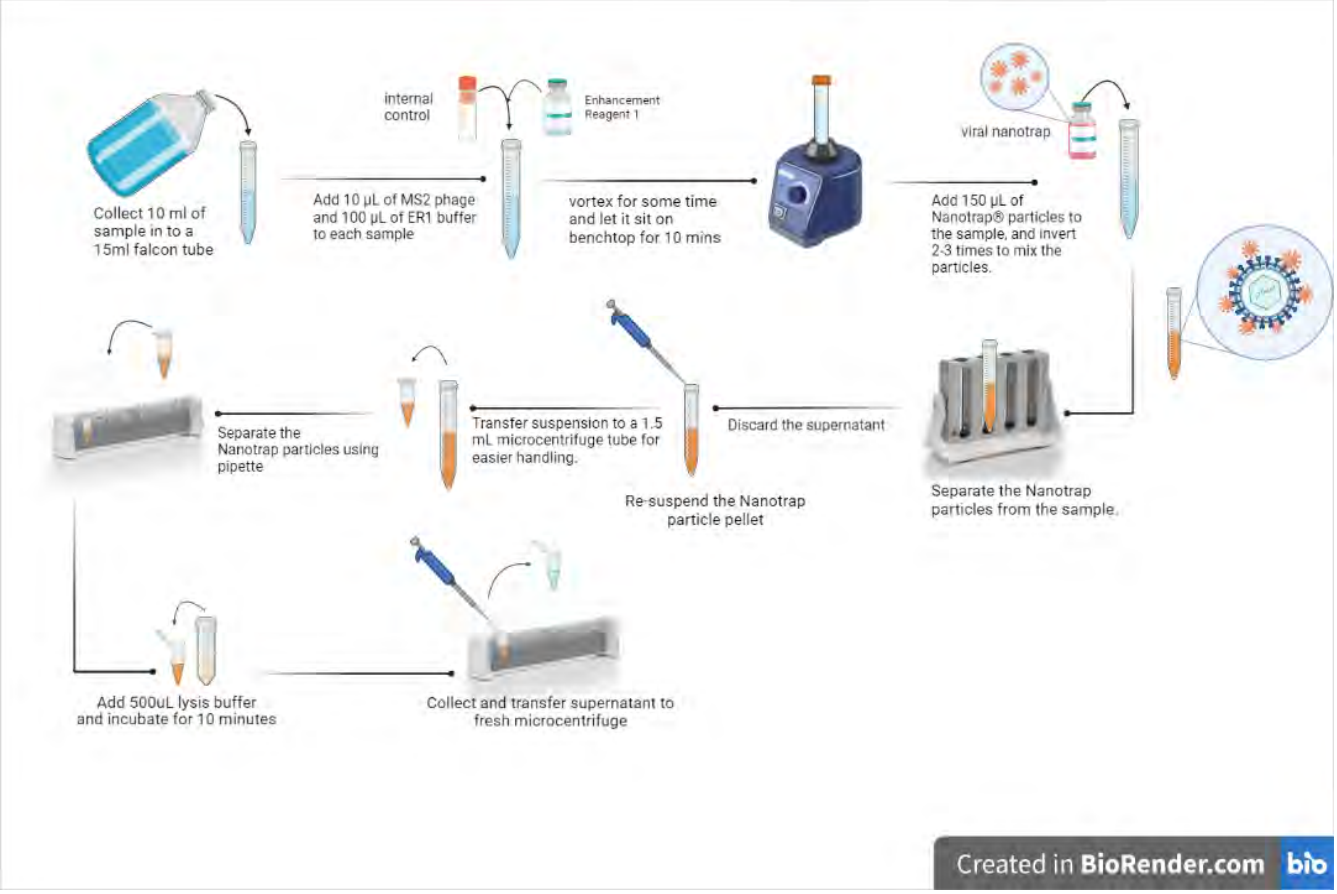

Figure 3:Illustration of Sample Processing using Nanotrap by Ceres Nanoscience to concentrate the samples. The image was created in BioRender.com

2.2 Nucleic Acid Extraction

Using the supernatant from our previous process, total nucleic acid extraction took place using Thermo-Fisher's MagMax Nucleic Acid Manual Extraction kit. Reagents such as proteinase K, Binding solution and magnetic beads were included in the kit to ensure complete separation of nucleic acid from virus particles and other contaminants. At first, a lysis buffer was used to break down cells and release nucleic acids. Binding solution and Magnetic beads were used to ensure proper binding of DNA to the beads so that they were not washed out by subsequent washing steps. Next, bead mixed samples were incubated at 65°C for 10 minutes and then placed in a magnetic

rack which aided the effective separation of these beads from the buffers, making sure the magnetic beads bound to DNA were not discarded in the process. Beads trapping the viral particle were washed using wash buffer and two 80% ethanol wash and allowed to dry to remove residual ethanol. After resuspending in elution buffer and a second incubation, viral RNA was eluted using the magnetic rack once again. Following the guidelines provided by the company, nucleic acid was eluted in approximately 50µlof elution buffer.

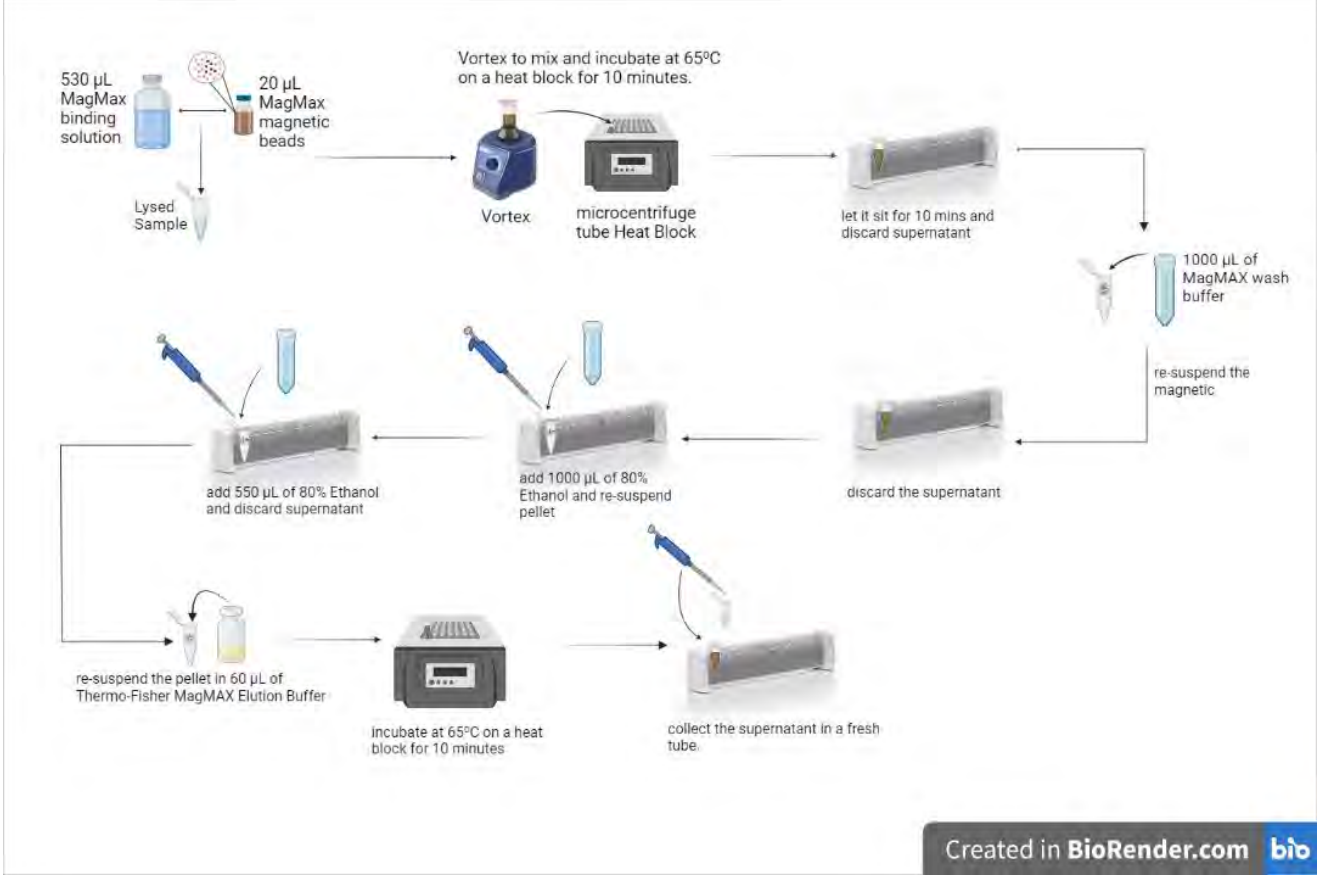

Figure 4: Illustration of manual Nucleic Acid Extraction using the Thermo-Fisher's MagMax Nucleic Acid Manual Extraction created in Biorender

2.3 Detection of SARS-CoV-2 using multiplex RT-qPCR

Duplex RT-qPCR was carried out in the Bio-Rad CFX96 Touch real-time PCR system using the TaqPath™ 1-Step Multiplex Master Mix

2.3.1 Primer and probe

Specific oligonucleotide primers were used targeting the amplification of N1 gene for the detection of SARS-CoV-2 from the samples. For MS2, primers provided by the Centre for Disease Control, USA (CDC, USA) was used. To ensure proper duplex PCR, the primer for N1 gene was conjugated to 6-carboxyfluorescein (6-FAM) fluorophore for the detection and Hexachlorofluorescein (HEX) for MS2 primer.

Table 1:Sequences of primers and probes used for Real time RT-PCR assay

Target	**Component**	**Sequence**
N1	Forward Primer	5'-GACCCCAAAATCAGCGAAAT-3'
	Reverse Primer	5'-TCTGGTTACTGCCAGTTGAATCTG-3'
	Probe (FAM)	ACCCCGCATTACGTTTGGTGGACC (5' FAM/ZEN/3' IBFQ)
MS2 Phage	MSF.F	TGGCACTACCCCTCTCCGTATTCACG
	MSF.R	GTACGGGCGACCCCACGATGAC
	MSF.P	HEX-CACATCGATAGATCAAGGTGCCTACAAGC-BHQ

2.3.2 Preparation of N1 Standard mix:

In order to create a standard curve using which we can u calculate number of viral RNA present in the sample, we prepared 4 dilutions of the SARS-CoV-2 N1 plasmid control. The concentration

of stock plasmid control was 2.1×10^{10} gene copies/µL. The quantity has been displayed in the following table:

Table 2: SARS_CoV-2 N1 plasmid control dilution

Dilution	Name	N1 control (µL)	Molecular Water (µL)	Final Volume (µL)	gc/µL*	gc/5µL	gc/ µL reaction
1	D4	100	-	100	60,000	300000	15000
2	D5	10	90	100	6,000	30000	1500
3	D6	10	90	100	600	3000	150
4	D7	10	90	100	60	300	15

2.3.3 Preparation of Duplex PCR Mastermix

Master Mix, Probe, Reverse Transcriptase enzyme mix, and molecular grade water were mixed with the primers to make up a volume of 15µl mastermix for each reaction tube as below:

Table 3: Reaction setup for Duplex RT-qPCR

Composition	Volume (uL/well)	Final concentration/well
4X TaqPath Master Mix	5.0	1×
SARS-CoV-2 N1-Forward (10µM)	0.8	400nM
SARS-CoV-2 N1-Reverse (10µM)	0.8	400nM
SARS-CoV-2 N1-Probe (10µM)	0.4	200nM
MS2-Forward (10µM)	0.8	400nM
MS2-Reverse (10µM)	0.8	400nM
MS2-Probe (10µM)	0.4	200nM
Nuclease free water	6.0	-
Total master mix	**15**	-

Template of 5µL was added per well to give a total of 20uL volume. Once every well has been properly filled, sealers were used to effectively seal the PCR plate to not let any vapors escape as the temperature was raised to 95°C.

2.3.4 RT-qPCR thermal cycling program:

Following the protocol and the thermal cycle listed below, qPCR takes place using the Bio-Rad CFX96 Touch real-time PCR system.

Table 4:Thermal cycling program

Step	Time	Temperature	Cycles
UNG Incubation	2 min	25°C	1
Reverse Transcription	10 min	53°C	1
Polymerase Activation	2 min	95°C	1
Denaturation	15 sec	95°C	45
Annealing/Extension	1 min	58°C	

2.4 Quantification of Data:

COVID-19 detection

Threshold cycle (Ct) values less than or equal to 37 were considered positive. Out of these positive samples, Ct values of less than or equal to 30 were selected for whole genome sequencing using the Next Generation Sequencing technique known as Oxford Nanopore.

2.4.1 Standard Curve Generation

The first step to quantifying data is to generate a standard curve using the values and results from our qPCR. This involves generating a series of (Positive) control dilutions where the input per gc/ul is known. As shown in the table below:

Table 5: SARS-CoV-2 serial dilution chart

Dilution	Input gc/µL	Total Input gc	Input gc/µL of reaction	Log copy	Average Ct
D7	60	300	15	1.176091259	30.785
D6	600	3000	150	2.176091259	27.635
D5	6000	30000	1500	3.176091259	24.03
D4	60000	300000	15000	4.176091259	21.06

These positive controls were added to the PCR plate alongside the extracted nucleic acids and the CT values were taken to create a standard curve. In this standard curve, the x axis denotes the Log Copy number and y axis is the CT values. The straight line was generated using MS Excel graph tool, from here, we can use the obtained CT values of our sample and match it with curve and get the corresponding Log copy number.

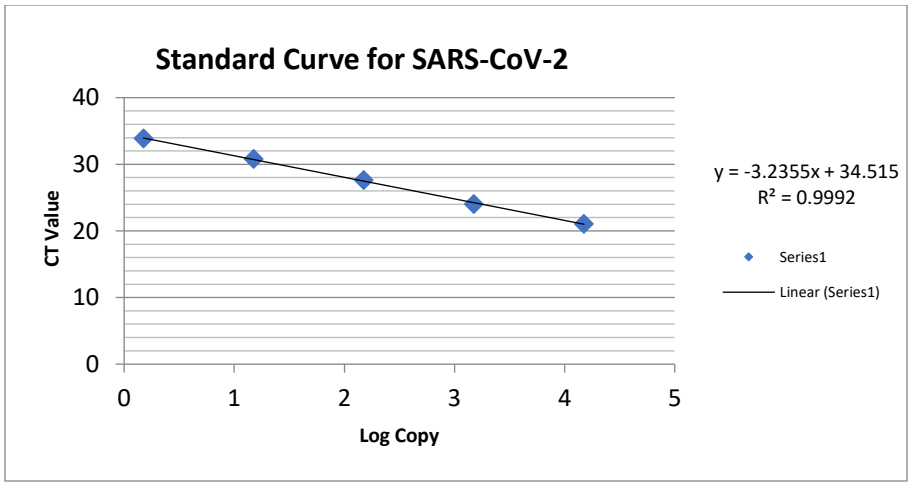

Figure 5: Standard Curve for SARS-CoV-2

The data we receive from the curve are:

- The formula: y = -3.2355x + 34.515
- The Gradient (m) = -3.2355
- The Intercept of the curve = 34.515
- Regression Value (R^2) = 0.992

Thus, another way of calculating log copy number is by placing the CT value in place of y and calculating the value of x (copy number) and using the log inverse to obtain the output copy number gc/µl.

2.4.2 PCR efficiency calculation

The percentage of target molecules replicated in a single PCR cycle is known as the efficiency (E) of PCR and is calculated from the standard curve using the following method:

%E =(10^(-1/slope)-1)*100

%E= (10^(-1/-3.2355)-1)*100 *(we use the gradient/m/slope of curve we previously got from the standard curve to calculate the percentage of PCR Efficiency*

%E= 103.76%

2.4.3 Back calculation

In order to calculate the copy number of the virus present in 1L of the collected samples, a quick back calculation is carried out. As stated in the previous sections, from 1L sample, 10 ml was processed, from

here 400ul was collected and through extraction, 60 µ l elute was achieved. 5 µ l of this elution was added as template for PCR.

As the name, back calculation, suggests, we start from the end. Therefore, the calculation starts from calculating the copy number:

Sample 1 with CT value : 23.27

In this example,
n = Ct value (from RT-qPCR)
c = Intercept =34.515 (from standard curve)
m = slope = -3.2355 from standard curve

Copy Nn = 10^{(n-c)/m}

= 10^{(23.27-34.515)/ -3.2355}
= 2988.86366/µL input template

Now for back calculation of sample 1 by following the strategy used in the Figure 4

Copies =2988.86366 /µL input template

= 2988.86366 *60 *(Elute from which PCR template was taken)*

= 179331.819 copies *(total copy numbers in 60µL elute*)

= 179331.819 copies *(equal to 400µL lysed sample, taken for extraction and eluted in 60µL elute*)

= 179331.819 for every 400µL lysed sample

= (179331.819 *500)/400 *(total copy number in 500µL lysed sample)*

= 224164.7742/500µL lysed sample

= 224164.7742/10ml wastewater *(500µL lysed sample = 10ml wastewater sample)*

= 22416.47742/ml wastewater samples

= 22416477.42/L wastewater samples.

So, for sample 1 with **Ct value 23.27**, gene **copy numbers** are **22416.47742/ml** wastewater

2.4.4 LOD Calculation

Limit of Detection (LOD) is the lowest concentration of analyte that can be detected. In this case, the positive control (SARS-CoV-2) dilutions were increased to obtain the lowest concentration of nucleic acid and then run using a real time qPCR. The lowest concentration determines the sensitivity level of the PCR

being used. In order to validate results, the PCR was repeated using the same dilutions, under the same conditions. As illustrated in the chart below, additional dilutions (D8, D9) were made with input gc/ul values of 1.5 and 0.15, respectively. From here, it can be seen that the lowest dilution, D9, was not detected in the repeated runs.

Table 6: Limit of Detection Chart of Serial Dilutions

Dilution	Input gc/µL	Total Input gc	Input gc/µL of reaction	Log copy	Average Ct
D9	.6	3	0.15	-0.823908741	N/A
D8	6	30	1.5	0.176091259	33.86
D7	60	300	15	1.176091259	30.785
D6	600	3000	150	2.176091259	27.635
D5	6000	30000	1500	3.176091259	24.03
D4	60000	300000	15000	4.176091259	21.06

2.5 Sequencing Technique

2.5.1 cDNA preparation

At first, extracted total RNA was reverse-transcripted to complementary (cDNA) using the LunaScript® RT SuperMix Kit from New England Biolabs, NEB. 2ul of Mix with 8ul of template was then mixed and placed in a thermocycler for qPCR to take place. 13 minutes later the plate was removed and immediately placed on ice for 60 seconds.

2.5.2 Multiplex PCR

Similar to PCR reactions with multiple targets, in this step we prepared 2 separate mastermix. One was for pool 1 and the other for pool 2. This was done to ensure maximum coverage of targeted sequence; in this case our target was the N1 gene on the SARS-CoV-2 virus. The prepared cDNA was used as the template for SARS-CoV-2-specific amplification with diverse multiplex PCR primers and NEB Next High-Fidelity 2X PCR Master Mix. The thermal cycling program for the multiplex PCR primer panels followed the ARTIC protocol: 30 s at 98°C; 35 cycles of 15 s at 98°C and 5 min at 63°C; then held at 4°C. Primers were then pooled and cleaned.

2.5.3 Library preparation for MinION nanopore sequencing

For nanopore sequencing, cDNA was treated with NEBNext Ultra II End repair (New England Biolabs), here dA tailing is carried out, and ligated with barcodes from a native barcoding kit (Oxford Nanopore Technologies) with NEB Blunt/TA Ligase Master Mix (New England Biolabs). The adapter from the ligation sequencing kit (Oxford Nanopore Technologies) was then ligated using the NEBNext Quick Ligation Module (New England Biolabs).

2.5.4 Running the program

After placing the FLO-MIN106D flow cell in to the Oxford Nanopore MinION, the priming port was prepared by removing bubbles and then a pipette was used to remove any remainder buffer from pervious sequence runs. Once cleaned, the Flush Buffer and Flush Tether were added. The final library was then prepared and added to the SpotOn sample port. In order to start the sequencing, the in-built software, MinKnow is opened and parameters are set and run for 24 hours. For this study, parameters such as Barcode selection was set to Native Barcoding Expansion 1-96, as 24 samples were multiplexed, Fast basecalling was selected and run time was set to 24 hours.

2.6. Data Analysis Using Bioinformatics Tools

MinKNOW v21.06 along witrh the complementary software, Guppy v.4.3.4 (Oxford Nanopore Technologies), was used for real-time base-calling as well as barcode demultiplexing. Successful FastQ files were trimmed with the aid of Porechop v.0.2.3. and Minimap2 was used for pairwise alignment with '-ax map-ont' setting. After using qualimap v.2.2.2 to assess the quality and map coverage, consensus fastq was generated using SAMtools. The quality of these fasta sequences were checked using Nextclade v2.14.1. Sequences with a sequence coverage of 50% and abover were selected and edited using Bioedit Sequence Alignment Editor v7.2 by comparing sequence with Integrated Genomics Viewer (IGV) v2.16.0. Sequences were aligned using the ClustalW tool in Bioedit and then a Neighour-Likelihood tree, otherwise known as a Phylogenetic Tree was created using both Bioedit and MEGA11.

For data visualization of Mutation frequency and locus at which mutations are taking place with respect to type of samples we have found, we used RStudio to build a Lolliplot and Heatmap for enhanced visual aid.

3. Results

3.1 Demographic Data

The study we conducted went over five months where every week we received 23 samples from multiple locations in Dhaka and Cox's Bazar. Thus, the total sample size our study was 283. Among the tested samples, only 37% (170) were found positive for SARS-CoV-2. For real-time PCR detection, we targeted the N1 gene on the SARS-CoV-2 genome as it had been proven to be the most effective target for virus detection, due to its high conservation properties (Zhang et al; 2020).

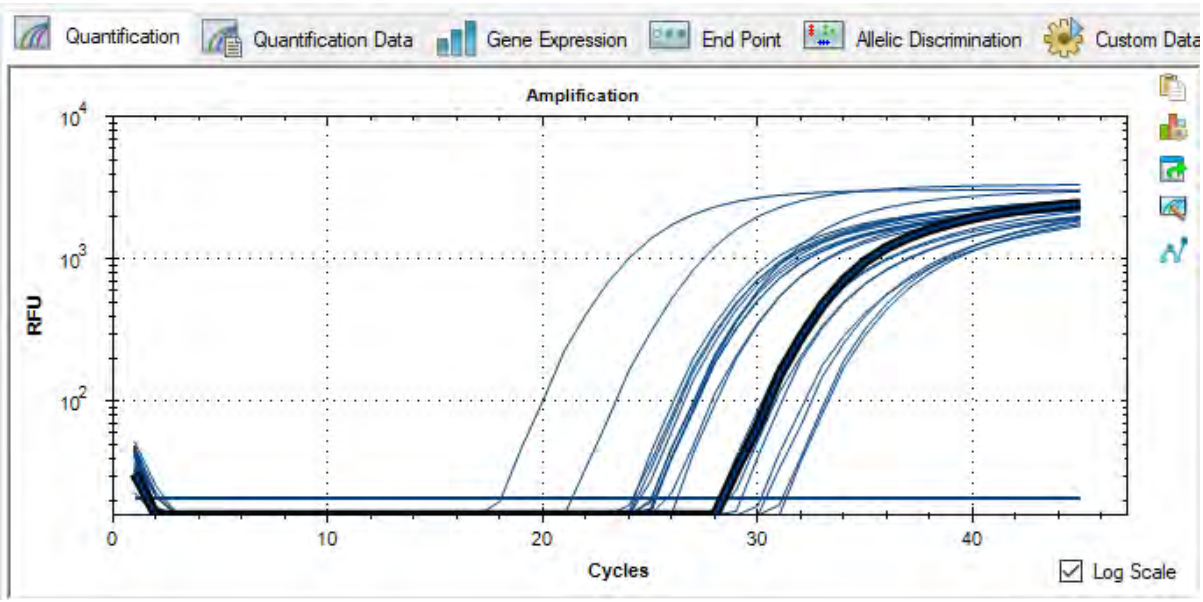

Figure 6: CFX Opus 96 Real time PCR Quantification Data of SARS-CoV-2

Therefore while we consider a Ct(Cycle Threshold) value of 37 positive (giving us a total of 170 positive samples), we only selected samples of Ct values less than and equal to 32. This is because a lower Ct value denotes the faster detection of the pathogen which means that concentration of

the pathogen in the sample is higher (Gallego). Thus, from a sample size of 170, we selected 48 samples that were less than or equal to 32.

3.2 Genomic Variation Identification

Whole Genome Sequencing of the selected samples were carried out using Oxford Nanopore's MinION. The generated genomic data was then assembled using SAMtools which gave us the final FASTq sequences. These were then analysed using Nextclade v2.14.1 (https://clades.nextstrain.org). From these 48 sequences, only 20 had passed the QC test, had a coverage of 50% or above.

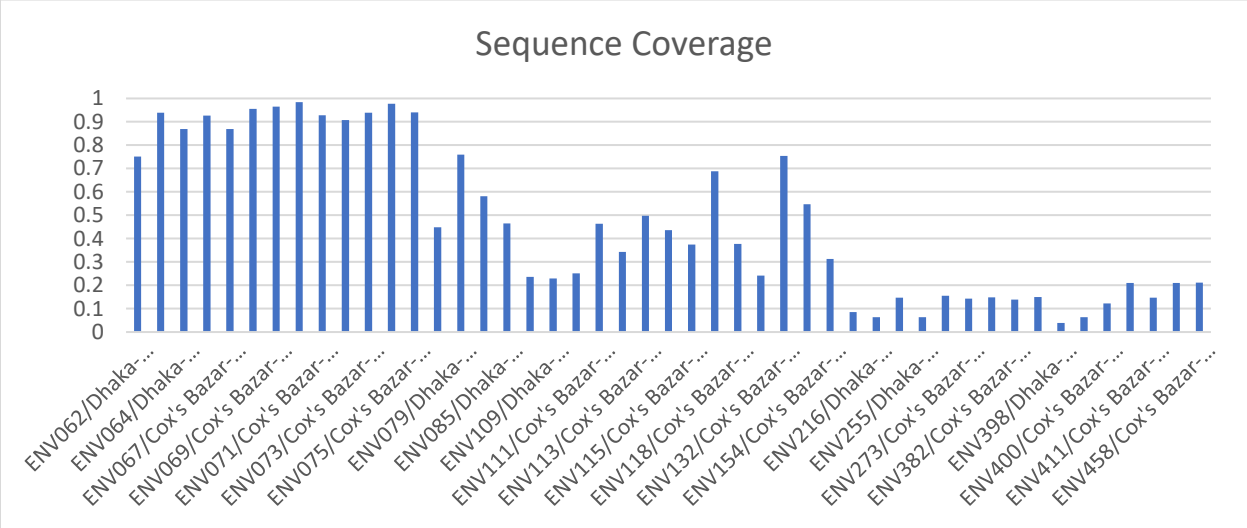

Figure 7: A bar chart illustrating the sequence coverage of the positive samples after they were blasted using Nextclade web server.

Further comparison of the data from Nextclade revealed the presence multiple PANGO lineages of SARS-CoV-2. While majority of the sequences were of Omicron clade, few were recombinants while the rest were 'unassigned' by WHO. Strains comprised of 19A, 21M, 21L, 22B, 22F and recombinant, all of which, including the frequency at which it was present, has been displayed in the following table:

Table 7: Summary of SARS-CoV-2 variants identified through Nextclade.com illustrating the WHO clade and the NextCLade pangolineage

Strain	Nextclade_Pango Lineage	Clade_WHO	Sample Count
19A	B	unassigned	20 (41.6%)
21L	BA.2	Omicron	1(2.08%)
21L	BA.2.10.4	Omicron	1(2.08%)
21M	B.1.1.529	Omicron	1(2.08%)
22B	BA.5	Omicron	1(2.08%)
22F	XBB	Omicron	10(20.8%)
22F	XBB.1	Omicron	9(18.75%)
22F	XBB.1.4	Omicron	1(2.08%)
Recombinant	XAC	Recombinant	1(2.08%)
Recombinant	XN	Recombinant	1(2.08%)
Recombinant	XW	Recombinant	1(2.08%)
Recombinant	XBN	Recombinant	1(2.08%)

Since our main objective of the study was to detect genetic diversity and possible identification of new strains circulating in Bangladesh, we decided to only select the most recent strains, which were 22F. The two main parameters we set for selecting the sequences were Omicron strain of 22F and a sequence coverage of 50% and above, thus we finally concentrated the sample size to 18 SARS-CoV-2 whole genome sequences. As we can see from the below, Omicron variants of Pangolineages XBB. XBB.1, XBB.1.14 and finally BA.2 have been further analyzed to compare their genomic changes.

Table 8: Final Sample List of WGS of SARS-CoV-2 with their Clade and Lineage names

index	**seqName**	**clade**	**Nextclade_pango**	**clade_who**

1	ENV062/Dhaka-BD/2022/October	22F	XBB	Omicron
2	ENV064/Dhaka-BD/2022/October	22F	XBB	Omicron
3	ENV072/Cox's Bazar-BD/2022/October	22F	XBB	Omicron
4	ENV079/Dhaka-BD/2022/October	22F	XBB	Omicron
5	ENV0116/Cox's Bazar-BD/2022/November	22F	XBB.1	Omicron
6	ENV0133/Cox's Bazar-BD/2022/November	22F	XBB.1	Omicron
7	ENV063/Dhaka-BD/2022/October	22F	XBB.1	Omicron
8	ENV066/Dhaka-BD/2022/October	22F	XBB.1	Omicron
9	ENV067/Cox's Bazar-BD/2022/October	22F	XBB.1	Omicron
10	ENV068/Cox's Bazar-BD/2022/October	22F	XBB.1	Omicron
11	ENV069/Cox's Bazar-BD/2022/October	22F	XBB.1	Omicron
12	ENV070/Cox's Bazar-BD/2022/October	22F	XBB.1	Omicron
13	ENV071/Cox's Bazar-BD/2022/October	22F	XBB.1	Omicron
14	ENV073/Cox's Bazar-BD/2022/October	22F	XBB.1	Omicron
15	ENV074/Cox's Bazar-BD/2022/October	22F	XBB.1	Omicron
16	ENV075/Cox's Bazar-BD/2022/October	22F	XBB.1	Omicron
17	ENV0132/Cox's Bazar-BD/2022/November	22F	XBB.1.4	Omicron
18	ENV080/Dhaka-BD/October/2022	21L	BA.2	Omicron

3.3 Mutation Frequency

3.3.1 Overall Mutation Frequency and Type

From the entire genome, analysis of our data revealed the proteins where mutations took place and the percentage frequency by which they were present. The %Frequency describes the proportion of mutations that occurred for a specific protein. So, a 25% frequency translates to the protein having 25% of the total mutations occurring in the entire genome. The proteins in question are: Envelope Protein (E), Membrane Protein (M), Nucleocapsid Protein (N), Open Reading Frames: 1a, 1b, 3a, 6, 7a, 9b (ORF1a, ORF1b, ORF3a, ORF6, ORF7a and ORF9b), Spike Protein (S). Of all the mentioned, besides the open reading frames, the rest are structural proteins that dictate the structural conformation of SARS-CoV-2 (Yadav et al; 2021). From the illustration we see that majority of the mutations occurred in the Spike Protein (38%) which was significantly higher than ORF1a which was 21.3%. ORF1b and N protein are close competitors with mutation frequency being 12.2% and 11.5%, respectively.

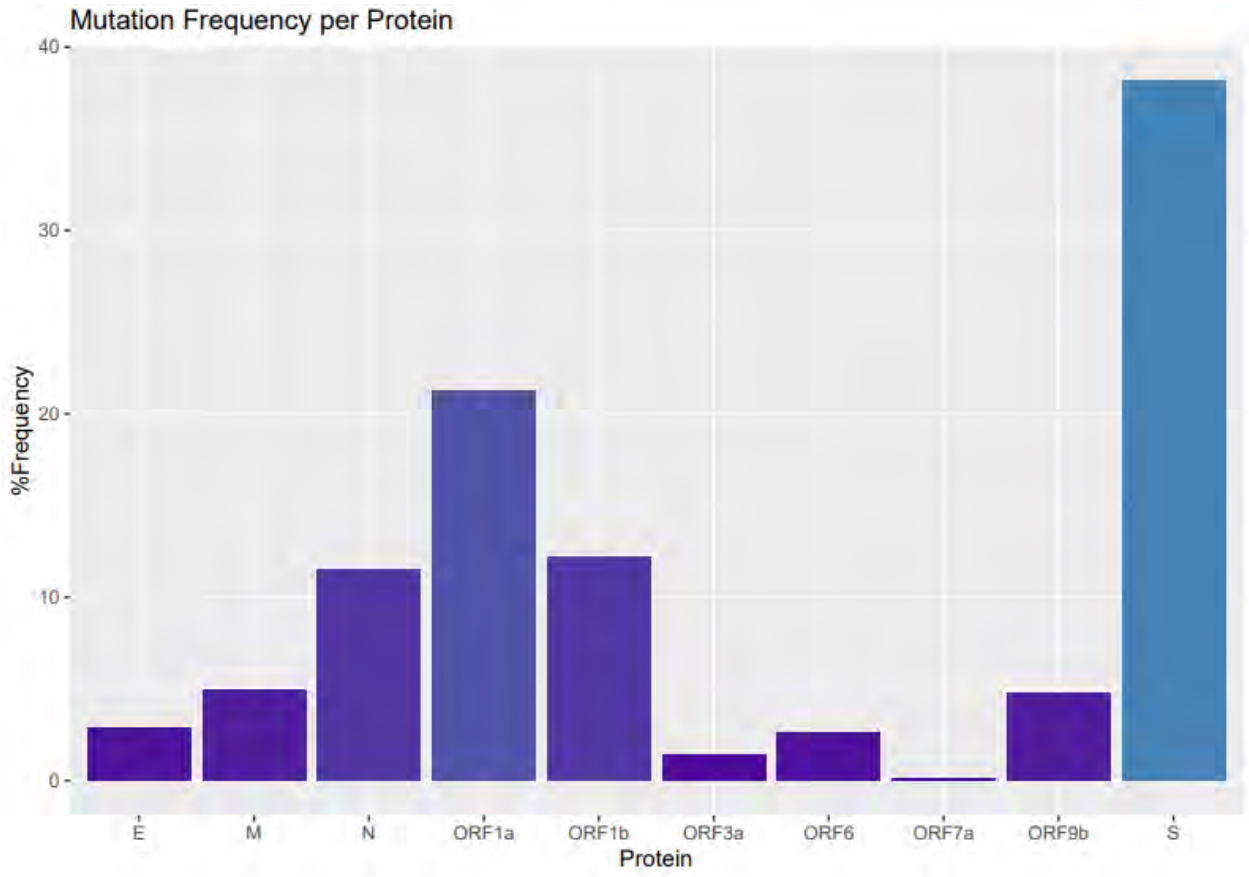

Figure 8 Frequency of Mutations Occurring Per Protein. The y axis denotes frequency of mutations occurring and the x axis portrays the different proteins in which the mutations are occurring. S protein having the highest mutation frequency with ORF7a having the least. Darker colour chart refers lower frequency while lighter shades denote higher frequency.

Additionally, a heatmap-like chart has been created to visualize the frequency of mutations with respect to proteins and their specific locus/site at which it occurs. Unlike the previous bar chart, here the frequency is the number of samples in which the mutation has occurred. 1 being the highest frequency ratio which means all 18 samples had that particular mutation occur in that specific site of the protein. For example, the spike protein mutation, G142D, was present in all 18 mutations, thus the color representing the ratio, blue, is seen. In this way, the color gradient scale represents the ratio of the mutation occurring, where, the lowest frequency is denoted by a lighter shade and deepens as the frequency increases. The presence of a tile in the column means the presence of a mutation in the protein (x-axis) concerning the site on the same position in the y-

axis. As previously seen in the bar chart, the Spike protein and ORF1a had the maximum number of mutations occurring, unlike the other proteins. This particular heatmap is an easier way to analyze mutations as we are not only seeing the site at which the mutation frequently occurs but we also know the number of mutations occurring at different sites of a protein. This is clear when observing Open Reading Frame 6, according to the bar chart, the percentage frequency was around 4% but from the heatmap we can also identify that only one site/locus on ORF6 has undergone substitution mutation which was seen in almost all the sequences.

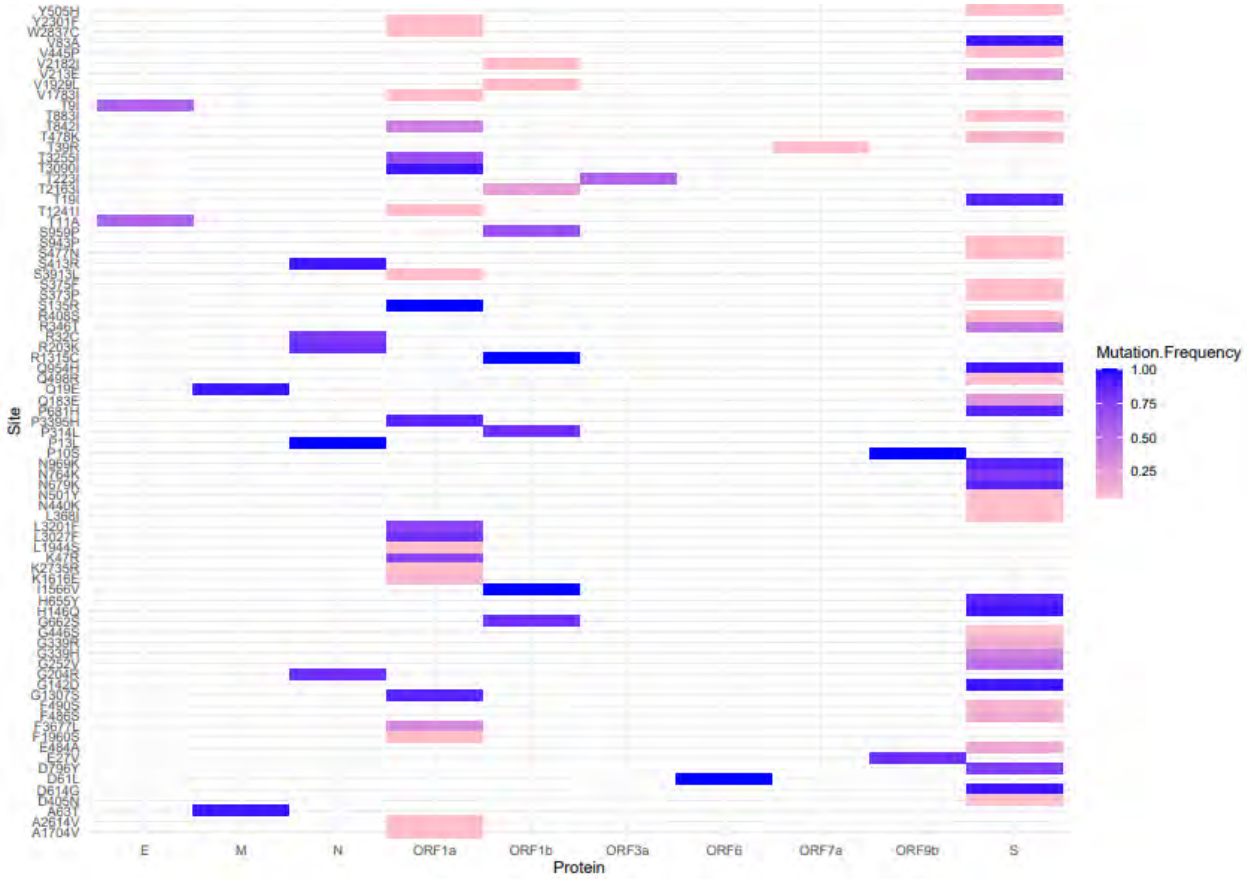

Figure 9: Mutation Frequency Heatmap

3.3.2 Envelope Protein Mutation

The Envelope protein is comprised of an N-terminal, C-terminal and a Transmembrane Protein. From the plot illustrated below (Figure 5), we can see that the Envelope Protein only had 2

mutations at sites: T9I and T11A. In addition, from the heatmap (Figure 6), we can see that only 8 samples had the T9I mutation in the N-terminal and only 7 samples had the T11A mutation in the Transmembrane. The heatmap only illustrates the presence or the absence of the mutation in a particular sample. Here the dark blue denotes the absence of the mutation whereas cyan denotes the presence of the mutation. The mutations occurring are both from polar uncharged Threonine to Alanine (T11A) and Isoleucine (T9I), both of which are aliphatic and non-polar in nature.

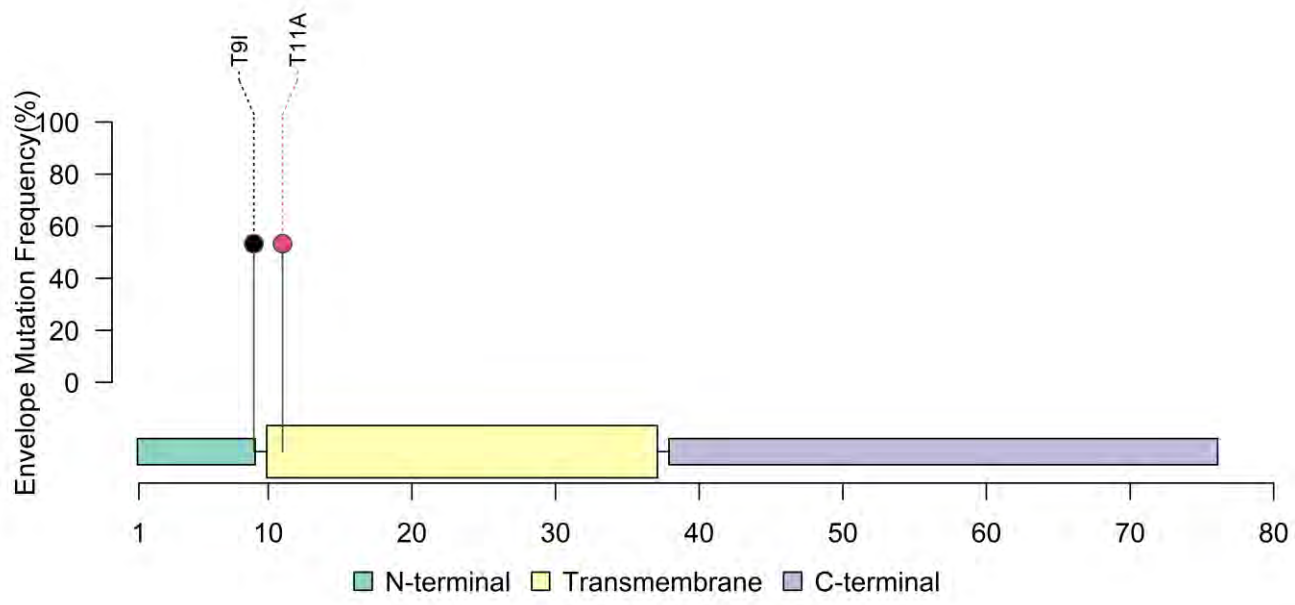

Figure 10: SARS-CoV-2 Envelope Protein Mutation site and frequency at which it occurs

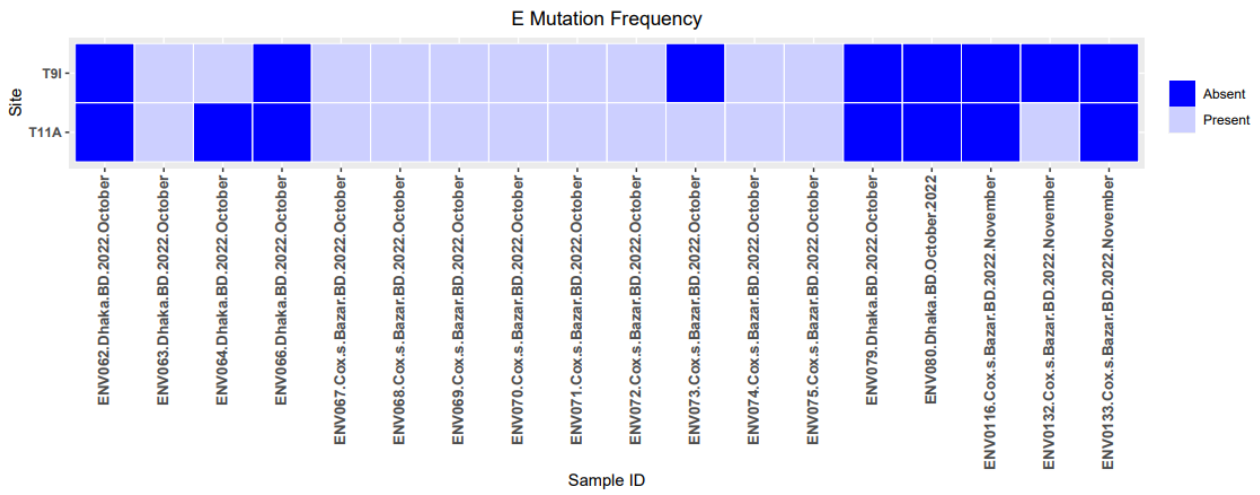

Figure 11: SARS-CoV-2 Mutation Profiling for E protein

3.3.3 Membrane Protein Mutation

The membrane protein consists of two mutations: Q19E taking place in the N-Terminal Domain and is observed in all the samples except ENV062, A63T taking place in the transmembrane protein which is also observed in all the samples besides ENV062.

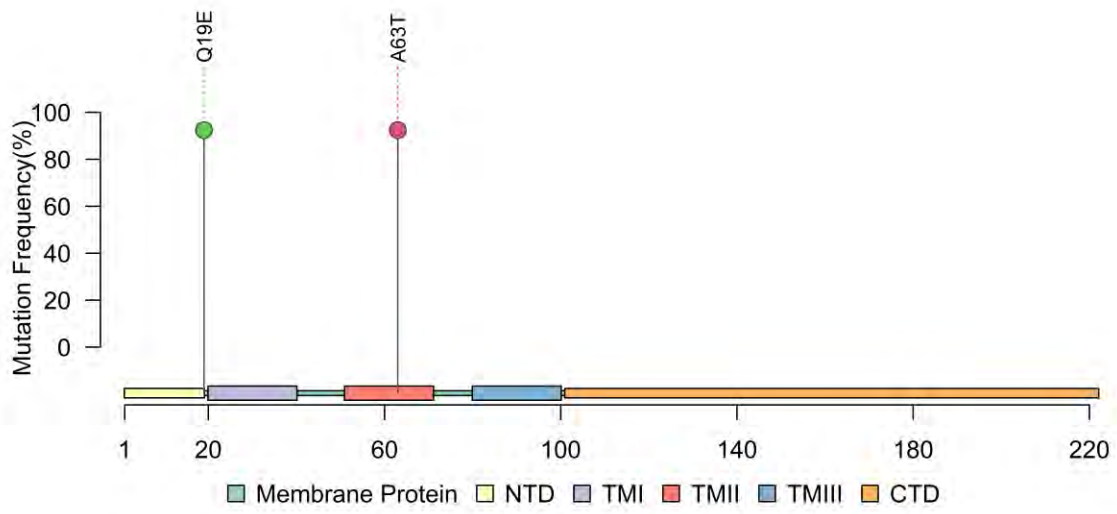

Figure 12: Membrane Protein Mutation and Frequency

In the heatmap here, the green denotes the absence of the mutation whereas cyan denotes the presence of the mutation. Despite being from the same clade (XBB), ENV062 does not have either of the mutations but both can be seen in rest of the members of the same XBB clade.

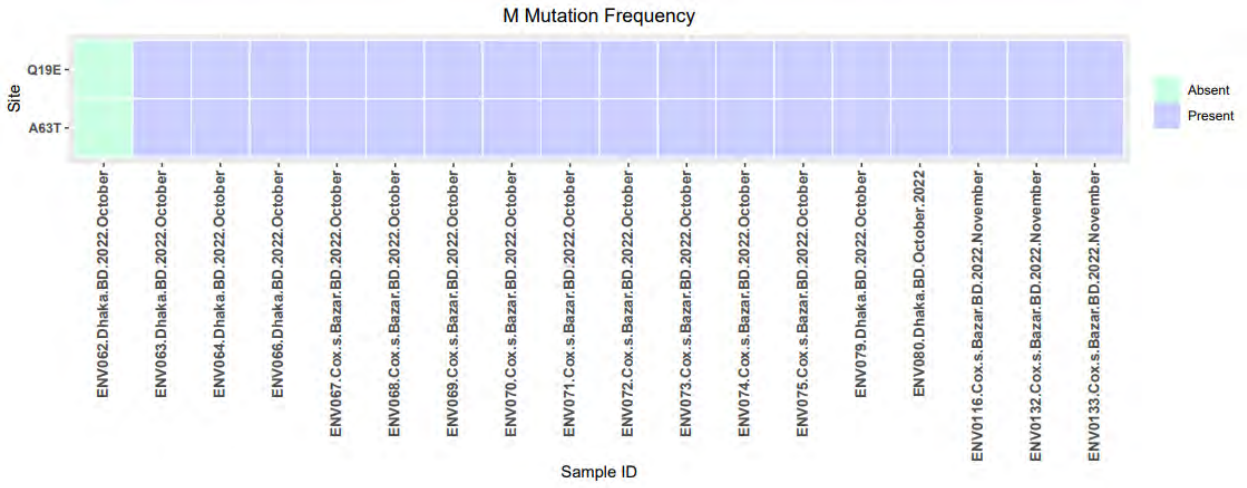

Figure 13: Membrane Protein Mutation Profiling

3.3.4 Nucleocapsid Protein Mutation

For the N protein in SARS-CoV-2, a total of 5 mutations were observed. P13L and R32C were seen in the N-terminal arm, R203K and G204R took place in the Linker Region and S413R, which was in the C-Terminal Domain. Almost all of the samples had these three substitution mutations present in their N gene but it was also seen that R32C was absent in 6 of the samples.

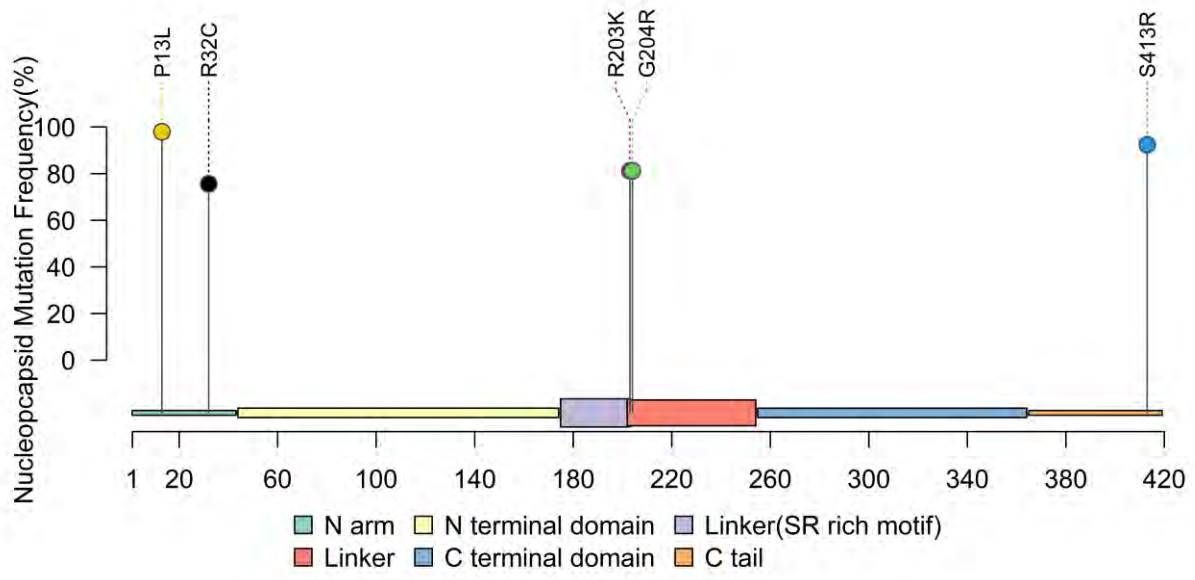

Figure 14: SARS-CoV-2 Nucleocapsid Mutation Annotation

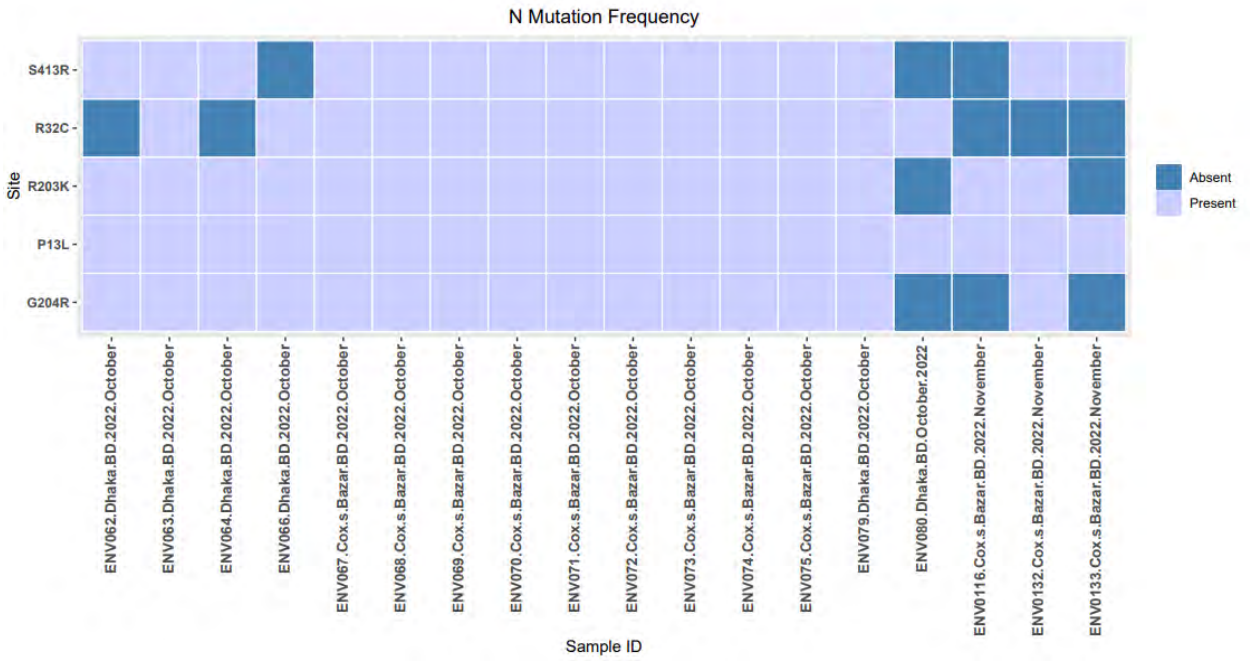

Figure 15: N mutation profiling

3.3.5.1 ORF1a Mutations

The Plot below illustrates the entire ORF1a and b and the mutations occurring in it. The Open Reading Frames 1a and 1 b are comprised of Non-Structural proteins 1 through 16 (NSP1- NSP16). As it can be clearly seen, majority of the mutations took place from NSP1 to NSP6. The mutations have been split according to the different ORF fragments and have been portrayed in two separate heatmaps (Figure 12 for ORF1a and Figure 13 for ORF1b), indicating the presence and absence of the mutations.

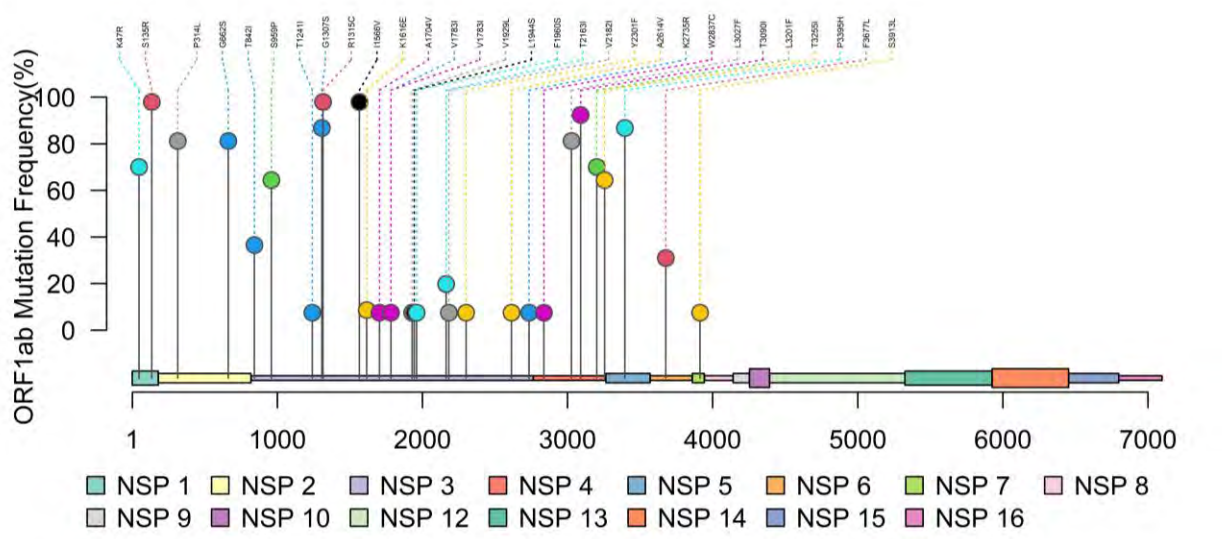

Figure 16: ORF1ab Mutation Frequency

The Mutations found in the ORF1a were: K47R, S135R, G1307S, Y2301F, A2614V, L3027F, T3090I, L3201F, P3395H, T3255I, F3677L, F1960S, K1616E, T842I, K2735R, V1783I, T1241I, A1704V, L1944S, W2837C, S3913L, V1783I. ENV064 was the only specimen which had a mutation at A2614V, the same can be said for ENV063 being the only one with a mutation at the site K2735R and ENV116 at S3913L. The most frequent mutations, on the other hand, were seen mainly for sites: G1307S, K47R, P3395H and S135R. All four of these mutations were noted for in at least one of each clade represented here: XBB, XBB.1, XBB.1.4 and BA.2.

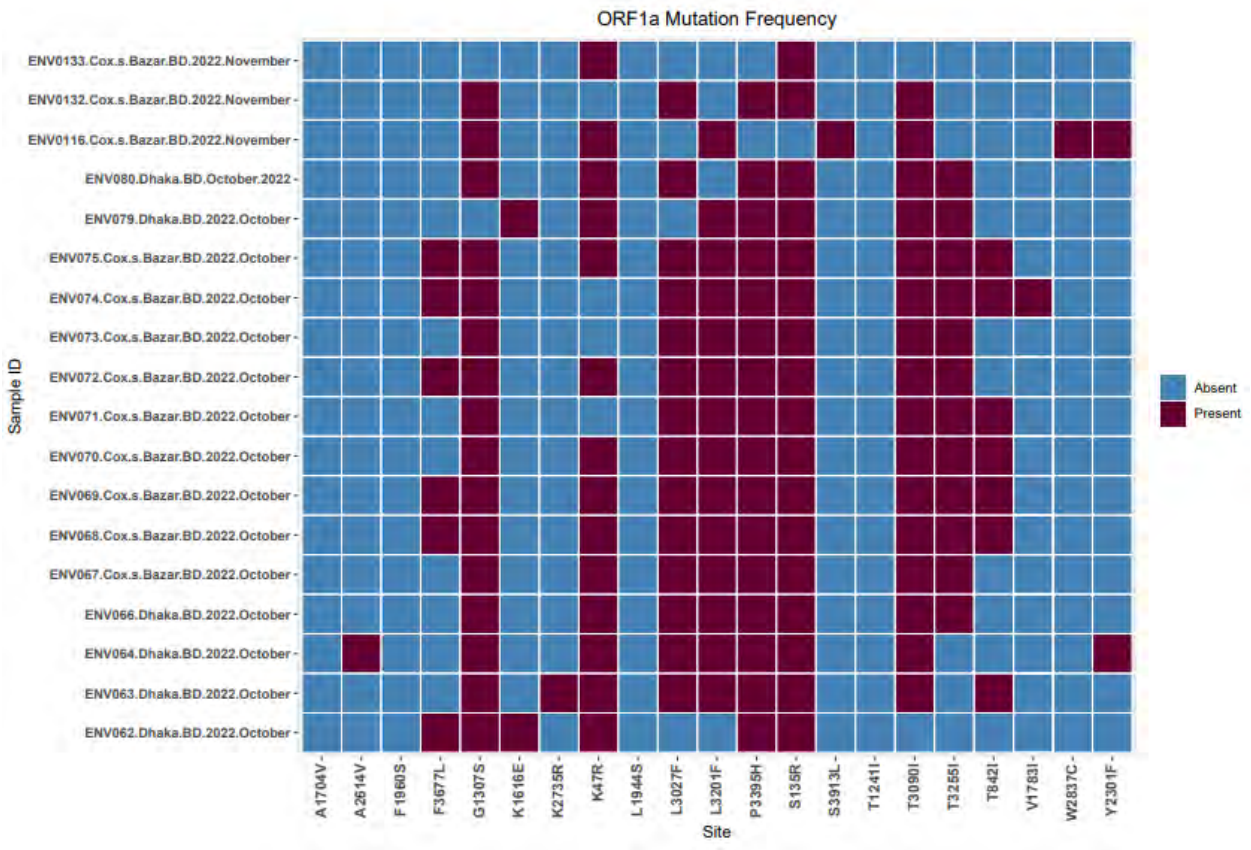

Figure 17: Mutation Profiling for ORF1a

3.3.5.2 ORF1b Mutation Profiling

The mutations observed across ORF1b were: P314L, G662S, S959P, R1315C, I1566V, V2182I, T2163I, V1929L. Here, mutation for V2182I was only seen in ENV069 and V1929L for ENV132. ENV132 being the only member of the clade XBB.1.4. Majority of the mutations took place in R1315C and I1566V, these mutations were present in all 18 samples and can be said to be a common mutation in the ORF1b.

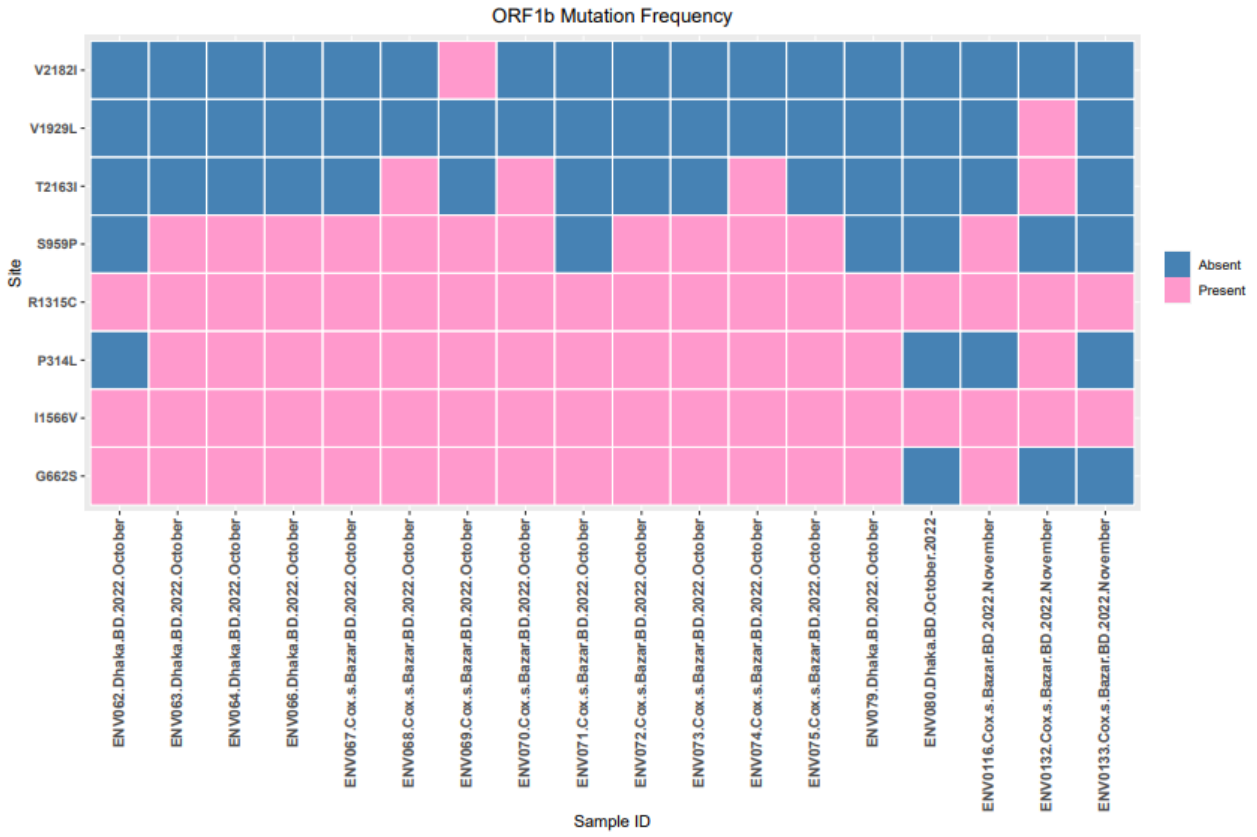

Figure 18: ORF1b Mutation Profiling

3.3.6 Spike Protein Mutations

Similar to the trend observed in Figure 3, the Spike Protein accounted for majority of the mutations occurring in the SARS-CoV-2 genome. With a total of 37 sites at which substitution mutations are occurring, there were 10 common mutations present in almost 80% of the sequences which were: D164G, G142D, N969K, P681H, V83A.

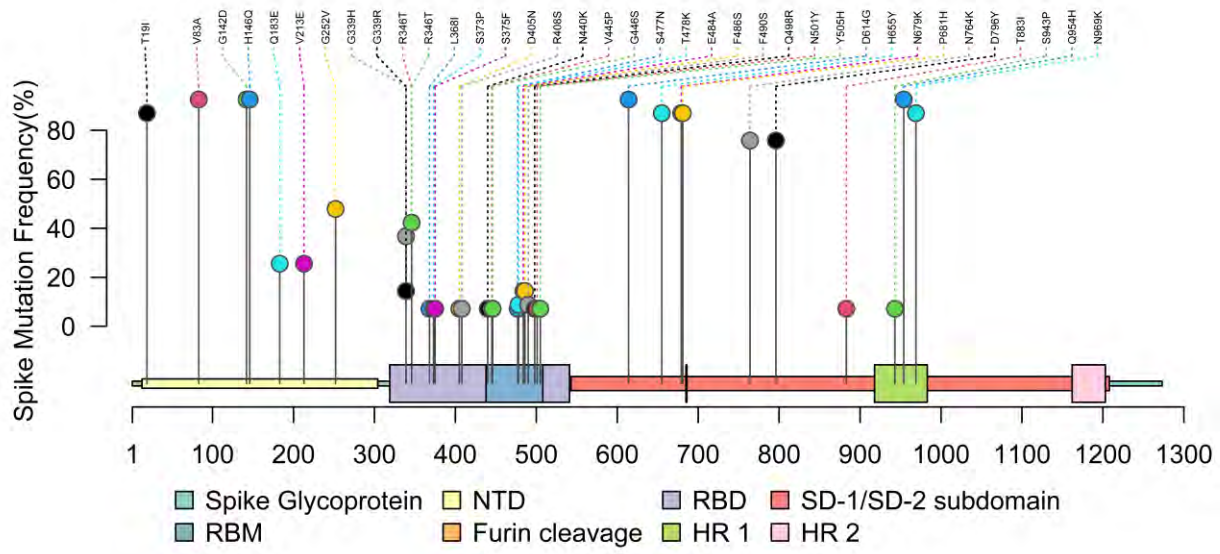

Figure 19: SARS-CoV-2 Spike Mutation Frequency

From the heatmap (Figure 15), the sequence ENV070 had multiple mutations that were absent in the 17 other sequences. This included mutation at sites: D405N, G446S, L368I, N440K, N501Y, Q498R, R408S, S373P, S375F, S477N, V445P and Y505H. Despite being from the same strain XBB.1, only ENV070 had these particular mutations.

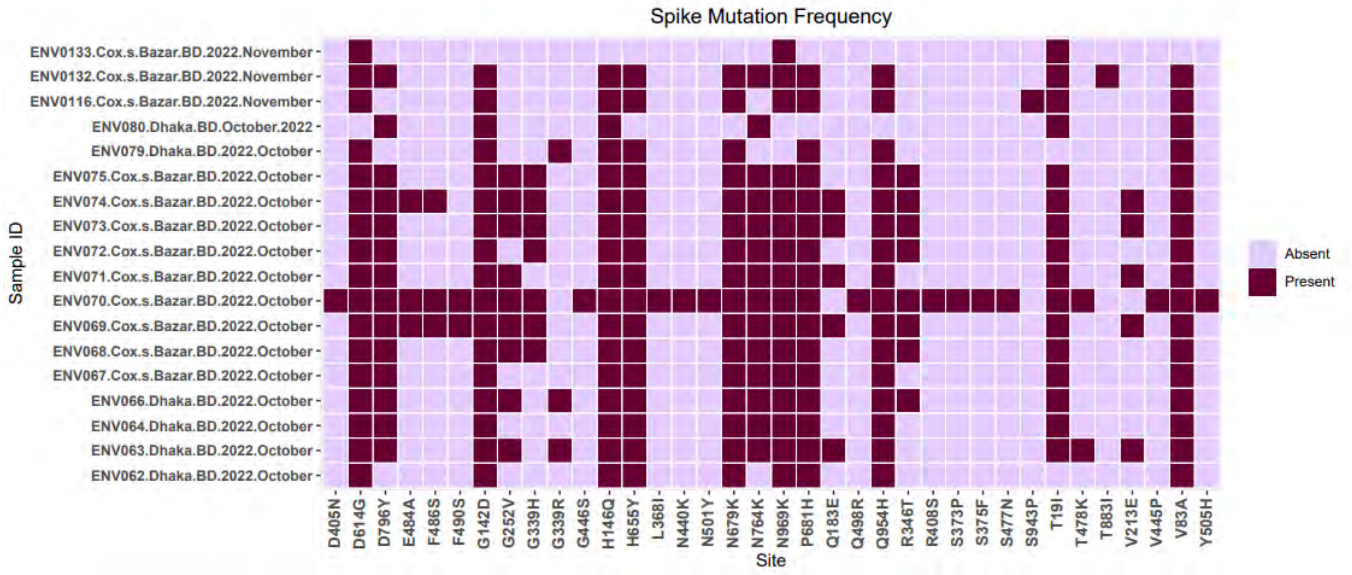

Figure 20: Mutation Profiling for SARS-CoV-2 Spike Protein

ENV080 of strain BA.2 was the only sequence that did not have its Aspartate replaced by Glycine at the 614 site. The mutation D614G was first identified in February 2020 which was believed to be the point mutation that had taken a dominant form due to its evolutionary advantage (Koening et al; 2021). Furthermore, ENV080 only had mutations at 5 sites out of the 37 occurring, this can be justified as the strain is a mutated version of the Delta strain BA.1 (WHO, 2022), clade 21L and not its descendant 22F, which had undergone several mutations to diverge in to a different strain.

3.4 Phylogenetic Tree Analysis

The phylogenetic tree was constructed taking various sequences for reference which includes NC_045512.2 or the original SARS-CoV-2 Wuhan Whole Genome isolate. EPI_ISL_16528799 was selected as it was a close match for Omicron XBB.1 and was also sequenced in Bangladesh. Other sequences selected were the Italy isolate for XBB strain: EPI_ISL_8544439, XBB 1.4 Japan isolate: EPI_ISL_17973285, the BA.2 Isolate from USA: EPI_ISL_17853102, and finally an Indian isolate closest to ENV116 of strain XBB.1. Sequences of different clades were copied and

pasted in the GIS-AID Epicov-BLAST webserver to search for similar sequences. The sequence with maximum match were selected and run to create a phylogenetic tree as shown below:

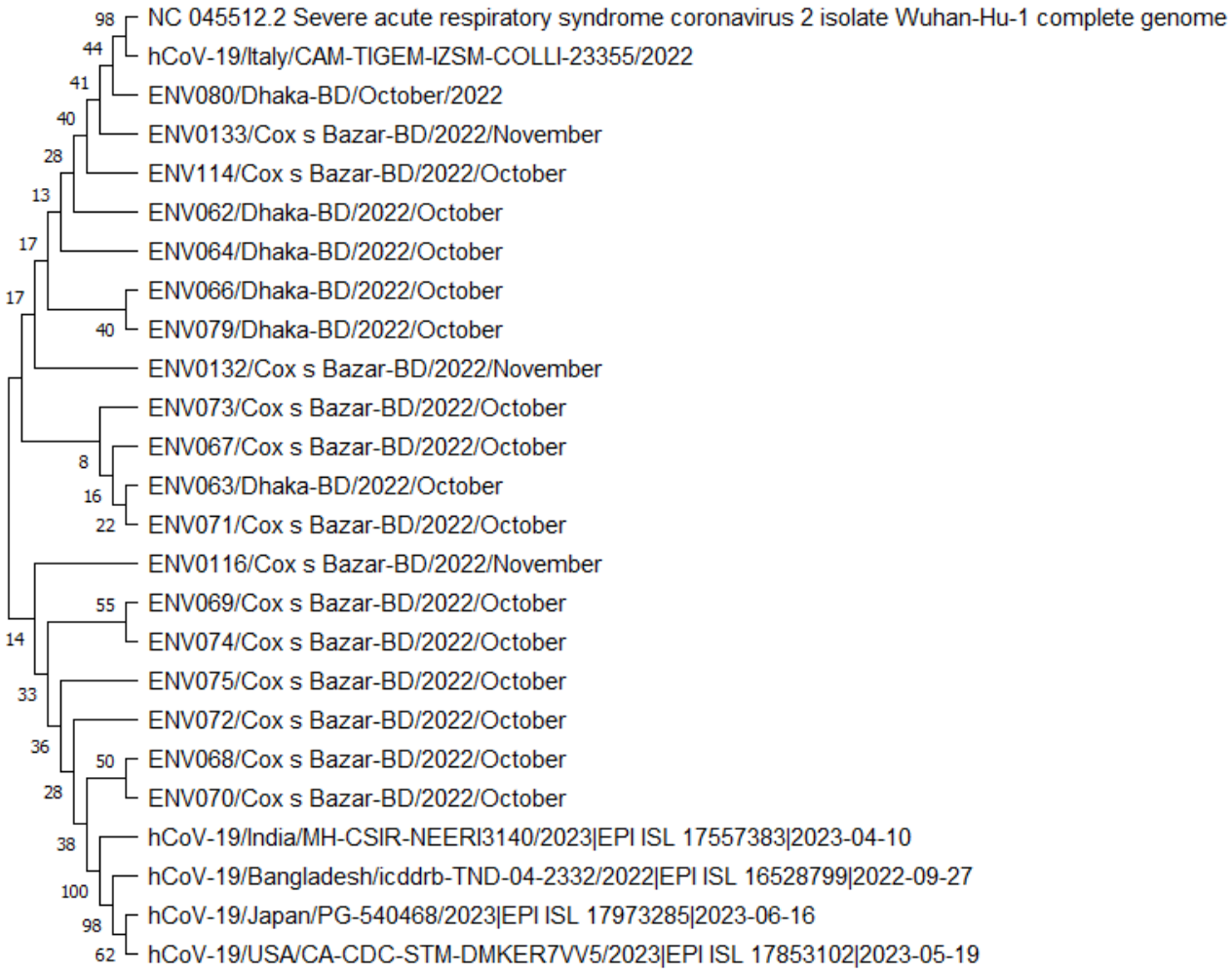

Figure 21: Maximum Likelihood Phylogenetic Tree Using Bootstrap Consensus

The percentage of replicate phylogenies that retrieved a certain clade from the original phylogeny constructed using the original alignment is known as the bootstrap value. From the song below. The bootstrap values of the nodes are significantly low, less than 70%, this means that the proportion of the replicates were very low.

As seen in the phylogenetic tree above, there are two main nodes at which the sequences separate and then further branch out. The upper branch is composed mainly of XBB, XBB.1.4, BA.2 and a few XBB.1 strains

which are more closely related to the original Wuhan HCov strain. However, the latter strains mainly composed of XBB.1 strains which are closely related to the Indian and Bangladesh isolates.

4. Discussion and Conclusion

Since it was deemed a pandemic in March 2020, Bangladesh has witnessed four waves of COVID-19 (COVID-19 Dynamic Dashboard for Bangladesh, 2022). A newly discovered SARS-CoV-2 variant with improved transmission, infection rate, and immune evasion mechanisms was the root cause of all the waves. The fourth wave, known as the Omicron surge, peaked in Bangladesh, in January 2022, and continued there through October 2022 (COVID-19 Dashboard). The current study documents the development of the Omicron variants that were circulating in Bangladesh at the time that the vast majority of the population had received vaccinations.

From our data, we can clearly see that the presence of Omicron strains was more prevalent as 20% and 67% of the samples were of 22F, XBB and XBB.1, respectively. Besides the ones mentioned, we even came across strains of lineage B, BA.2, BA.2.10.4, B.1.1.529, BA.5, XBB.1.4. Additionally, 4 strains (XAC, XN, XW, XBN) were identified as recombinant which only occurs when a patient hosts a co-infection of two different strains of the SARS-CoV-2 virus. By mixing their genetic components during replication, this enables the variants to interact and create novel combinations within the human body. Such occurrences are more likely to take place when viral cases are on the rise and spreading quickly, as is the situation with COVID19 cases, which are once again on the rise after a downward trend (Mohapatra et al; 2022).

Since the first isolated hCoV sequence, multiple mutations have taken place in the genome of SARS-CoV-2 virus which has caused the genetic evolution of this disease-causing pathogen. Few of these mutations hold a significant role in increasing virulence and host-virus interactions, usually these occur in the Spike protein, specifically in the Receptor-Binding Domain. The virus's capacity to bind to the human ACE2 receptor and penetrate human cells can likewise be impacted by RBD mutations. When compared to the

original Wuhan strain, these changes typically produce reduced binding affinity with ACE2 (Candido; 2022). As seen figure 03, majority of the mutations took place in the Spike Protein (about 38%), followed by mutations in the non-structural proteins (NSPs) located on the Open Reading Frames 1a and 1b, the mutation site that causes a change in the life cycle and pathogenesis of SARS-CoV-2 (Raj, R.; 2021). Multiple studies exploring the genome of SARS-CoV-2 have revealed several common mutations present in the omicron variant: G339D, S373P, S375F, K417N, N440K, S477N, T478K, E484A, Q493R, Q498R, N501Y, and Y505H. S371L, G446S, G496S, S371F, T376A, D405N, R408S, H655Y, N969K etc. Out of these, N440K, S447N, N460K, F490S, Q498R, and N501Y were found to have enhanced the RBD-ACE2 binding that improves viral entry and infection of SARS-CoV-2 (Qu et al., 2022). Upon examining our sequences, it was surprising to see that majority of the omicron sequences did not have the above mentioned 'common' mutations in high frequencies, on the contrary, only 2-3 sequences seemed to have had mutations at sites: E484A, N440K, N501Y, Q498R, S477N, T478K, Y505H. While the absence of N440K, N501Y, Q498R, S477N mutations suggest that despite being Omicron variants of XBB and XBB.1 lineage, these isolates do not exhibit increased SARS-CoV-2 infection. Furthermore, the same mutations were reported to decrease efficacy of neutralizing antibodies and as they are absent in over 90% of the sequences, we can assume the isolates were able to produce neutralizing antibodies. The presence of mutation T19I in 75% of the sequences suggests reduced neutralizing capabilities, but at a lower rate. sss

The second site at which most of the mutations were observed was the ORF1ab, consisting of 16 NSPs. The linker region of NSP-1 included the mutation S135R, which affects RNA stability and interacts with the 40S subunit.

The SARS-CoV-2 envelope protein (E-2) is a homopentameric channel composed of a 75-residue viroporin. It is responsible for creating a cation-selective channel across the ERGIC membrane to help in viral assembly and budding, this protein alone is strong enough to induce cell apoptosis and cause damage mimicking acute respiratory distress syndrome (ARDS). Studies have shown that in some members of coronaviridae, E deletion results in attenuated viruses, whereas E alterations that eliminate channel activity

result in viruses with lower toxicity (Wang et al; 2023). Two common mutations observed in E-2 were T9I and T11A.

T9I is a single-point mutation that alters the envelope protein's channel characteristics and reduces viral production and pathogenicity (Xia et al; 2022). The mutation T9I was found in 55% of our samples, this suggests that the virulence of these samples may be weaker than the remaining 45% of sequences missing this mutation. This particular mutation where Threonine is replaced by an Isoleucine causes attenuation of viral load when exogenous T9I expression is carried in vitro (Xia et al; 2022).

Additionally, the new dominant-negative mutation T11A resulted in weaker lung injury in mice and lowered cell lethality, cytokine induction, and virus production capacities *in vitro*. Furthermore, the combination of these two mutations, T9I and T11A was seen to have reduced cell lethality and significantly increased SARS-CoV-2 attenuation (Wang et al; 2023). 61% of the sequences carried the T11A mutation, however only 9 sequences (50%) carried a combination of both mutations. From this, we can deduce that the samples with both mutations had not affected patients with severity as strong as other strains but more study is required to conclude.

Samples from XBB.1.4 and BA.2 had only T11A and neither of the mutations, respectively. Although studies suggest that the XBB strain aCtuired T11A and T9I existed since the Alpha strains, some members of the XBB (ENV062, 079, 080) and XBB.1(ENV063, 116, 132) did not carry these mutations. This raises the question of whether these particular sequences mutated from other regions/locations than those evolving from homogenous strains. After carrying out a phylogenetic analysis, these sequences (except ENV116) were more closely related to the Italian isolate: EPI_ISL_8544439. With ENV116 having comparatively the lowest bootstrap consensus of 17, it can be said that the isolate is very genomically divergent from the rest and this happens as a majority of the characters in the sequence do not support the node from which it branches out.

M protein mutations: Q19E and A63T, have been found to play a role in viral assembly. Given that it is situated adjacent to a significant residue, Valine66, the A63T could have an impact on the stability of the M protein dimer (Schoeman et al; 2019). With 17 out of 18 sequences (94%) carrying both mutations, it can be perceived that the viral assembly might have been affected as well as the stability of the M protein dimer.

The mutation on the 13th codon of the N protein, P13L, was seen to have a relatively low global frequency, this means it was not a mutation that was present in isolates all around the world (Oulas et al; 2021). However, P13L was seen to be present in all 18 of our isolates. The P13L present on the surface of the N protein is suspected to bind with the genome via forming helical ribonucleocapsid and its interaction with the M protein suggests its key function in virion assembly. Investigation of the N-M interaction also revealed a decrease in binding affinity for the ALT-P13L-M protein complex. P13L might be responsible for fewer deaths, but the presence of this mutation in the virus facilitates increased transmissibility of the virus. (Oulas et al; 2021).

Alternatively, the mutation R32C causes changes in nucleocapsid phosphorylation, fitness, and replication may increase the virulence and transmissibility of SARS-CoV-2. A frustration analysis on the SARS-CoV-2 virus demonstrates that changes in the N protein's structure and functions give viruses flexibility in their antigenic makeup and sequencing (Haque et al; 2023). 13 out of a total of 18 samples carried this mutation.

Consecutive mutations on the N gene, R203K, and G204R were found in high frequencies (15 and 16 sequences out of the total 18, respectively). This combination of mutations seemed to have a significant effect on increasing the viral load of the SARS-CoV-2 virus by adding a new transcriptional regulatory site in the middle, enabling the development of a shortened version of the N protein (Carabelli et al; 2023). This in turn has been seen to have exacerbated the host’s immune response when experimented on VLP (virus like particle) incubated cells (Shuaib et al; 2023).

A mutational analysis was carried out for the selected strains, namely, XBB, XBB.1, XBB.1.4 and BA.2. While we were expecting the tree to branch out very less as majority of the sequences had similar mutational profiling, besides one or two, to our surprise this was not the case. According to the phylogenetic tree we created on MEGA11 adopting the Maximum Likelihood function, we obtained a tree as illustrated in Figure:21. Here, at first, we notice that the bootstrap consensus/values were significantly low for our sequences compared to the reference sequences. Not only that, but the reference sequences obtained from GIS AID seemed to be more closely related to each other rather than the sequences they were blasted against. Furthermore, besides 8 sequences originating from the same branched-out internode, the rest were not as closely related and had to branch out further to be joined. Of these, the highest values were for ENV068 and ENV070 and yet since it was less than 70%, which is considered a good replicated value, it is considered as a sequence that cannot be replicated for enough of the other reference sequences.

There are concerns that the large mutations found in the COVID-19 spike protein may compromise the efficacy of the current COVID-19 immunization and antibody therapy. Even a third booster dose could not be enough to protect against Omicron infection, despite the fact that the protection against a disease still makes booster vaccines advised. More than 139 million doses of the COVID-19 vaccine have been delivered in Bangladesh, protecting more than 75% of the population. As seen from the Omicron surge, despite the virus being able to infect vaccinated individuals, the severity of the COVID-19 sickness is much reduced as a result of the effect of vaccination that has led to herd immunity. The omicron variations' accumulation of mutations allows them to avoid immune detection and lower the host's production of neutralizing antibody titers. It can therefore spread to a person.

In order to get a quick insight in to the strain, virulence and frequency of the virus currently circulating the country, wastewater monitoring is an efficient method that can be adopted. This study proves that, due to presence of the virus in fecal matter, it can be easily detected as we were not only able to detect multiple strains present in the discarded waterbodies but we were able to detect them in high enough concentrations that would be alarming, if necessary, steps are not taken. This way, even without the going through the

difficulties of sampling specimen from human beings (i.e. such as ethical clearance and willingness to participate etc) we are able to get a detailed overview of SARS-CoV-2 virulence in the country.

However, our study has some limitations that must be taken into account in order to evaluate the analysis clearly. Due to the study's very limited sample size, samples from Dhaka city and other regions of the nation were unable to be collected. Therefore, no link between the spatial distribution and the lineage could be deduced. Additionally, as were unable to carry out other genomic tests, we cannot be a hundred percent sure that the genomic profiling we have done is sufficient enough to say that strains such as XBB and XBB.1 do not possess the common mutations that are otherwise present in majority of the GISAID uploaded sequences, neither can we deduce the actual neutralizing properties of our isolates. Further and advanced research is necessary to pass a verdict on its alleged changed genomic properties.

References:

1. Ashour, H.M., Elkhatib, W.F., Rahman, M.M. and Elshabrawy, H.A., 2020. Insights into the recent 2019 novel coronavirus (SARS-CoV-2) in light of past human coronavirus outbreaks. *Pathogens*, *9*(3), p.186.
2. Astuti I, Ysrafil (2020) Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2): an overview of viral structure and host response. Diabetes Metab Syndr Clin Res Rev 14(4):407–412. https://doi.org/10.1016/j.dsx.2020.04.020
3. Burrell, C. J., C. R. Howard, and F. A. Murphy. "Fenner and White's Medical Virology; Chapter 31—Coronaviruses." (2017): 437-446.
4. Candido, K.L., Eich, C.R., de Fariña, L.O., Kadowaki, M.K., da Conceição Silva, J.L., Maller, A. and Simão, R.D.C.G., 2022. Spike protein of SARS-CoV-2 variants: a brief review and practical implications. *Brazilian Journal of Microbiology*, *53*(3), pp.1133-1157.
5. Cevik, M., Kuppalli, K., Kindrachuk, J. and Peiris, M., 2020. Virology, transmission, and pathogenesis of SARS-CoV-2. *bmj*, *371*.
6. Cheung, K.S., Hung, I.F., Chan, P.P., Lung, K.C., Tso, E., Liu, R., Ng, Y.Y., Chu, M.Y., Chung, T.W., Tam, A.R. and Yip, C.C., 2020. Gastrointestinal manifestations of SARS-CoV-2 infection and virus load in fecal samples from a Hong Kong cohort: systematic review and meta-analysis. *Gastroenterology*, *159*(1), pp.81-95.
7. *Coronavirus disease (covid-19): How is it transmitted?* (no date) *World Health Organization*. Available at: https://www.who.int/news-room/questions-and-answers/item/coronavirus-disease-covid-19-how-is-it-transmitted (Accessed: 30 July 2023).
8. Coutard B, Valle C, de Lamballerie X, Canard B, Seidah N, Decroly E(2020) The spike glycoprotein of the new coronavirus 2019-nCoVcontains a furin-like cleavage site absent in CoV of the same clade. Antivir Res 176:104742. https://doi.org/10.1016/j.antiviral.2020.104742).
9. Cui, J., Li, F. and Shi, Z.L., 2019. Origin and evolution of pathogenic coronaviruses. *Nature reviews microbiology*, *17*(3), pp.181-192.
10. da Costa, V.G., Moreli, M.L. and Saivish, M.V., 2020. The emergence of SARS, MERS and novel SARS-2 coronaviruses in the 21st century. *Archives of virology*, *165*(7), pp.1517-1526.
11. Da Silva, P.G., Mesquita, J.R., Nascimento, M.D.S.J. and Ferreira, V.A.M., 2021. Viral, host and environmental factors that favor anthropozoonotic spillover of coronaviruses: An opinionated

review, focusing on SARS-CoV, MERS-CoV and SARS-CoV-2. *Science of the Total Environment*, *750*, p.141483.

12. De Wit, E., Van Doremalen, N., Falzarano, D. and Munster, V.J., 2016. SARS and MERS: recent insights into emerging coronaviruses. *Nature reviews microbiology*, *14*(8), pp.523-534.
13. Dey, S.K., Rahman, M.M., Siddiqi, U.R. and Howlader, A., 2020. Exploring epidemiological behavior of novel coronavirus (COVID-19) outbreak in Bangladesh. *SN comprehensive clinical medicine*, *2*, pp.1724-1732.
14. Eaaswarkhanth, M., Al Madhoun, A. and Al-Mulla, F., 2020. Could the D614G substitution in the SARS-CoV-2 spike (S) protein be associated with higher COVID-19 mortality?. *International Journal of Infectious Diseases*, *96*, pp.459-460.
15. El-Demerdash, A., Metwaly, A.M., Hassan, A., Abd El-Aziz, T.M., Elkaeed, E.B., Eissa, I.H., Arafa, R.K. and Stockand, J.D., 2021. Comprehensive virtual screening of the antiviral potentialities of marine polycyclic guanidine alkaloids against SARS-CoV-2 (COVID-19). *Biomolecules*, *11*(3), p.460.
16. Gallego, A. (no date) *How to interpret RT-qpcr results*, *GoldBio*. Available at: https://goldbio.com/articles/article/How-To-Interpret-RT-qPCR-Results#_1fob9te (Accessed: 07 August 2023).
17. Gao J, Tian Z, Yang X (2020) Breakthrough: chloroquine phosphate has shown apparent efficacy in treatment of COVID-19 associated pneumonia in clinical studies. BioSci Trends 14(1):72–73. https://doi.org/10.5582/bst.2020.01047
18. Hajimonfarednejad, M., Ostovar, M., Hasheminasab, F.S., Shariati, M.A., Thiruvengadam, M., Raee, M.J. and Hashempur, M.H., 2023. Medicinal Plants for Viral Respiratory Diseases: A Systematic Review on Persian Medicine. *Evidence-Based Complementary and Alternative Medicine*, *2023*.
19. Hoffmann M, Kleine-Weber H, Schroeder S, Krüger N, Herrler T, Erichsen S, Schiergens T, Herrler G, Wu N, Nitsche A, Müller M,Drosten C, Pöhlmann S (2020) SARS-CoV-2 cell entry depends on ACE2 and TMPRSS2 and is blocked by a clinically proven protease inhibitor. Cell 181(2):271–280.e8. https://doi.org/10.1016/j.cell.2020.02.052

20. Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y, Zhang L, Fan G, Xu J, Gu X, Cheng Z. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. The lancet. 2020 Feb 15;395(10223):497-506.
21. Iqbal H, Romero-Castillo K, Bilal M, Parra-Saldivar R. The Emergence of Novel-Coronavirus and its Replication Cycle - An Overview. Journal of Pure and Applied Microbiology. 2020;14(1):16. https://doi.org/10.22207/JPAM.14.1.03
22. IRIC. (2017). Understanding qPCR results. Institute of Research in Immunology and Cancer (University of Montreal), 1–3.
23. Koenig, P.A. and Schmidt, F.I., 2021. Spike D614G—a candidate vaccine antigen against Covid-19. *New England Journal of Medicine*, *384*(24), pp.2349-2351.
24. Krammer, F., 2020. SARS-CoV-2 vaccines in development. *Nature*, *586*(7830), pp.516-527.
25. Kumar, S., Thambiraja, T.S., Karuppanan, K. and Subramaniam, G., 2022. Omicron and Delta variant of SARS-CoV-2: a comparative computational study of spike protein. Journal of medical virology, 94(4), pp.1641-1649.
26. Landi F, Barillaro C, Bellieni A, Brandi V, Carfì A, D'Angelo M, Fusco D, Landi G, Monaco RL, Martone AM, Marzetti E. The new challenge of geriatrics: saving frail older people from the SARS-CoV-2 pandemic infection. The Journal of Nutrition, Health & Aging. 2020 Apr 6:1.
27. Landi F, Barillaro C, Bellieni A, Brandi V, Carfì A, D'Angelo M, Fusco D, Landi G, Monaco RL, Martone AM, Marzetti E. The new challenge of geriatrics: saving frail older people from the SARS-CoV-2 pandemic infection. The Journal of Nutrition, Health & Aging. 2020 Apr 6:1.
28. Landi, F., Barillaro, C., Bellieni, A., Brandi, V., Carfi, A., Cipriani, M.C., D'Angelo, E., Falsiroli, C., Fusco, D., Landi, G. and Liperoti, R., 2020. The geriatrician: The frontline specialist in the treatment of COVID-19 patients. *Journal of the American Medical Directors Association*, *21*(7), pp.937-938.
29. Liu, J., Liao, X., Qian, S., Yuan, J., Wang, F., Liu, Y., Wang, Z., Wang, F.S., Liu, L. and Zhang, Z., 2020. Community transmission of severe acute respiratory syndrome coronavirus 2, Shenzhen, China, 2020. *Emerging infectious diseases*, *26*(6), p.1320.
30. Lu, R., Zhao, X., Li, J., Niu, P., Yang, B., Wu, H., Wang, W., Song, H., Huang, B., Zhu, N. and Bi, Y., 2020. Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. *The lancet*, *395*(10224), pp.565-574.
31. Memish, Z.A., Perlman, S., Van Kerkhove, M.D. and Zumla, A., 2020. Middle East respiratory syndrome. *The Lancet*, *395*(10229), pp.1063-1077.

32. Memish, Z.A., Perlman, S., Van Kerkhove, M.D. and Zumla, A., 2020. Middle East respiratory syndrome. *The Lancet*, *395*(10229), pp.1063-1077.

33. Meng, X., Wang, X., Meng, S., Wang, Y., Liu, H., Liang, D., Fan, W., Min, H., Huang, W., Chen, A. and Zhu, H., 2021. A global overview of SARS-CoV-2 in wastewater: detection, treatment, and prevention. *Acs Es&T Water*, *1*(10), pp.2174-2185.

34. Mohapatra, R.K., Kandi, V., Tuli, H.S., Chakraborty, C. and Dhama, K., 2022. The recombinant variants of SARS-CoV-2: Concerns continues amid COVID-19 pandemic. *Journal of Medical Virology*, *94*(8), p.3506.

35. Prem K, Liu Y, Russell TW, Kucharski AJ, Eggo RM, Davies N, Flasche S, Clifford S, Pearson CA, Munday JD, Abbott S. The effect of control strategies to reduce social mixing on outcomes of the COVID-19 epidemic in Wuhan, China: a modelling study. The Lancet Public Health. 2020 Mar 25.

36. Prem, K., Liu, Y., Russell, T.W., Kucharski, A.J., Eggo, R.M., Davies, N., Flasche, S., Clifford, S., Pearson, C.A., Munday, J.D. and Abbott, S., 2020. The effect of control strategies to reduce social mixing on outcomes of the COVID-19 epidemic in Wuhan, China: a modelling study. *The lancet public health*, *5*(5), pp.e261-e270..

37. Qu, P., Evans, J.P., Zheng, Y.M., Carlin, C., Saif, L.J., Oltz, E.M., Xu, K., Gumina, R.J. and Liu, S.L., 2022. Evasion of neutralizing antibody responses by the SARS-CoV-2 BA. 2.75 variant. *Cell host & microbe*, *30*(11), pp.1518-1526.

38. Raj, R., 2021. Analysis of non-structural proteins, NSPs of SARS-CoV-2 as targets for computational drug designing. *Biochemistry and biophysics reports*, *25*, p.100847.

39. Schoeman D, Fielding B (2019) Coronavirus envelope protein: current knowledge. Virol J 16(1):69. https://doi.org/10.1186/s12985-019-1182-0

40. Stadler, K., Masignani, V., Eickmann, M., Becker, S., Abrignani, S., Klenk, H.D. and Rappuoli, R., 2003. SARS—beginning to understand a new virus. *Nature Reviews Microbiology*, *1*(3), pp.209-218.

41. Statement on the second meeting of the International Health Regulations (2005) Emergency Committee regarding the outbreak of novel coronavirus (2019-nCoV). Who.int. 2020. Available from: https://www.who.int/news-room/detail/30-01-2020-statement-on-the-second-meeting-of-the-international-health-regulations-(2005)-emergency-committee-regarding-the-outbreak-of-novel-coronavirus-(2019-ncov)

42. Statement on the second meeting of the International Health Regulations (2005) Emergency Committee regarding the outbreak of novel coronavirus (2019-nCoV). Who.int. 2020. Available from: https://www.who.int/news-room/detail/30-01-2020-statement-on-the-second-meeting-of-the-international-health-regulations-(2005)-emergency-committee-regarding-the-outbreak-of-novel-coronavirus-(2019-ncov)
43. Tai W, He L, Zhang X, Pu J, Voronin D, Jiang S, Zhou Y, Du L (2020) Characterization of the receptor-binding domain (RBD) of 2019 novel coronavirus: implication for development of RBD protein as a viral attachment inhibitor and vaccine. Cell Mol Immunol 17(6):613–620. https://doi.org/10.1038/s41423-020-0400-4
44. Vallejo, J.A., Trigo-Tasende, N., Rumbo-Feal, S., Conde-Pérez, K., López-Oriona, Á., Barbeito, I., Vaamonde, M., Tarrío-Saavedra, J., Reif, R., Ladra, S. and Rodiño-Janeiro, B.K., 2022. Modeling the number of people infected with SARS-COV-2 from wastewater viral load in Northwest Spain. *Science of The Total Environment*, *811*, p.152334.
45. *Who coronavirus (COVID-19) World Health Organization*. Available at: https://covid19.who.int/?mapFilter=deaths (Accessed: 27 July 2023).
46. Wu, A., Peng, Y., Huang, B., Ding, X., Wang, X., Niu, P., Meng, J., Zhu, Z., Zhang, Z., Wang, J. and Sheng, J., 2020. Genome composition and divergence of the novel coronavirus (2019-nCoV) originating in China. *Cell host & microbe*, *27*(3), pp.325-328.
47. Wu, Y., Guo, C., Tang, L., Hong, Z., Zhou, J., Dong, X., Yin, H., Xiao, Q., Tang, Y., Qu, X. and Kuang, L., 2020. Prolonged presence of SARS-CoV-2 viral RNA in faecal samples. *The lancet Gastroenterology & hepatology*, *5*(5), pp.434-435.
48. Xu, R.H., He, J.F., Evans, M.R., Peng, G.W., Field, H.E., Yu, D.W., Lee, C.K., Luo, H.M., Lin, W.S., Lin, P. and Li, L.H., 2004. Epidemiologic clues to SARS origin in China. *Emerging infectious diseases*, *10*(6), p.1030.
49. Yadav, R., Chaudhary, J.K., Jain, N., Chaudhary, P.K., Khanra, S., Dhamija, P., Sharma, A., Kumar, A. and Handu, S., 2021. Role of structural and non-structural proteins and therapeutic targets of SARS-CoV-2 for COVID-19. *Cells*, *10*(4), p.821.
50. Yin, Y. and Wunderink, R.G., 2018. MERS, SARS and other coronaviruses as causes of pneumonia. *Respirology*, *23*(2), pp.130-137.
51. Zhang, Xuzhao, Li, Min, Zhang, Bin, Chen, Tao, Lv, Dong, Xia, Pengfei, Sun, Zhuanyi et al. "The N gene of SARS-CoV-2 was the main positive component in repositive samples from a cohort of

COVID-19 patients in Wuhan, China." *Clinica Chimica Acta 511*, (2020): 291-297. Accessed August 7, 2023. https://doi.org/10.1016/j.cca.2020.10.019.

52. Zheng, S., Fan, J., Yu, F., Feng, B., Lou, B., Zou, Q., Xie, G., Lin, S., Wang, R., Yang, X. and Chen, W., 2020. Viral load dynamics and disease severity in patients infected with SARS-CoV-2 in Zhejiang province, China, January-March 2020: retrospective cohort study. *bmj*, *369*.
53. Zhou, P., Yang, X.L., Wang, X.G., Hu, B., Zhang, L., Zhang, W., Si, H.R., Zhu, Y., Li, B., Huang, C.L. and Chen, H.D., 2020. A pneumonia outbreak associated with a new coronavirus of probable bat origin. *nature*, *579*(7798), pp.270-273.
54. Xia, B., Wang, Y., Pan, X., Cheng, X., Ji, H., Zuo, X., Jiang, H., Li, J. and Gao, Z., 2022. Why is the SARS-CoV-2 Omicron variant milder?. *The Innovation*, *3*(4).
55. Wang, Y., Pan, X., Ji, H., Zuo, X., Xiao, G.F., Li, J., Zhang, L.K., Xia, B. and Gao, Z., 2023. Impact of SARS-CoV-2 envelope protein mutations on the pathogenicity of Omicron XBB. *Cell Discovery*, *9*(1), p.80.
56. Schoeman, D. and Fielding, B.C., 2019. Coronavirus envelope protein: current knowledge. *Virology journal*, *16*(1), pp.1-22.
57. Oulas, A., Zanti, M., Tomazou, M., Zachariou, M., Minadakis, G., Bourdakou, M.M., Pavlidis, P. and Spyrou, G.M., 2021. Generalized linear models provide a measure of virulence for specific mutations in SARS-CoV-2 strains. *PloS one*, *16*(1), p.e0238665.
58. Haque, S., Khatoon, F., Ashgar, S.S., Faidah, H., Bantun, F., Jalal, N.A., Qashqari, F.S. and Kumar, V., 2023. Energetic and frustration analysis of SARS-CoV-2 nucleocapsid protein mutations. *Biotechnology and Genetic Engineering Reviews*, pp.1-21.
59. Carabelli, A.M., Peacock, T.P., Thorne, L.G., Harvey, W.T., Hughes, J., COVID-19 Genomics UK Consortium de Silva Thushan I. 6, Peacock, S.J., Barclay, W.S., de Silva, T.I., Towers, G.J. and Robertson, D.L., 2023. SARS-CoV-2 variant biology: immune escape, transmission and fitness. *Nature Reviews Microbiology*, *21*(3), pp.162-177.
60. Shuaib, M., Adroub, S., Mourier, T., Mfarrej, S., Zhang, H., Esau, L., Alsomali, A., Alofi, F.S., Ahmad, A.N., Shamsan, A. and Khogeer, A., 2023. Impact of the SARS-CoV-2 nucleocapsid 203K/204R mutations on the inflammatory immune response in COVID-19 severity. *Genome Medicine*, *15*(1), pp.1-16.
61.