Syrian Arab Republic **Syrian Virtual University**

Master Bioinformatics **Ministry of Higher Education**

Identification and genetic characterization of Sars-CoV-2 in Syria (2023) by using Oxford Nanopore Technology

Project submitted for Master's Degree in Bioinformatics

Student ID

Shaza_166781

Supervised by

Dr. Hayan Hasan

2024

Acknowledgments & Dedication

To my beloved husband, Salem, to the spirit that lives within me and that has never been stingy in supporting me. Increase my strength, self-esteem, and passion for trying whenever I am stranded, and that nothing is impossible with the strength of faith and patience.

To my dear father, the beacon of knowledge and light that surrounds me in the darkness of my days, and anchors me in a port of safety, to the lofty one who taught me the highest and most precious values, rising above all sins.

To my dear mother, a symbol of love and inner peace, to the one who gave of herself for us.

To my dear children, Ali and Adam, you are the soul and fragrance of life, my hope, and the secret of my happiness. In your eyes I see life, and in your presence, I find myself, the most precious thing I have and the most beautiful thing I have given,

To my brothers and sisters, Hasan, Ali, Doaa, Zainab, my eternal peace, and my companions, may God bless your hearts with all goodness.

To my dear friend Marah Meftah, words cannot return the favor in the face of the goodness of your generosity and the kindness of your soul, which elevates the purity of your beautiful soul and the unlimited giving of your heart that you have showered upon me and which I will never forget as long as I live.

To my loyal friends, who supported me throughout my studies. To Shaza Al-Freijat for her giving and scientific assistance whenever I needed it.

All thanks and appreciation To Dr. Majd Al-Jamali for his giving and passion in conveying and presenting the concept of bioinformatics, introducing this new field in our country, and opening the door wide to new horizons of science, and thanks go to Dr. Hayyan Hassan for following the course of the research and his valuable comments.

I am honored to extend my sincere thanks and appreciation to Dr. Muhannad Khalil, Director of Public Health Laboratories and Head of the National Polio Laboratory, for allowing me to have an important experience in the field of Next Generation Sequencing, which he strives to provide and introduce into the laboratory work system.

I also extend my sincere thanks and gratitude to Dr. Jamil Al-Khabbaz, Head of the Emergency Diseases Laboratory, for the hard work and dedication made to complete the work in the field of biotechnology and for providing the necessary data related to research, and for supporting me in this direction

Summary

Background: The COVID-19 virus represented a difficult stage in modern history that the world has gone through since the beginning of its spread in Wuhan province in China on December 12, 2019, and quickly turned into an epidemic in March 2020, leaving millions of victims and injured. Vaccines and monoclonal antibodies have contributed to reducing the number of deaths and alleviating the severity of the symptoms until the WHO announced the end of the public health emergency in an important plan towards the end of the Corona pandemic https://www.who.int/. The first Syrian case of COVID-19 was reported in March 2020. over time, Measures like lockdowns, social distancing, and vaccination drives have been enforced in Syria to manage virus spread and safeguard the populace despite the difficult health and living conditions. The Omicron variant (B.1.1.529) has become the dominant type by sublineages and the only VOC circulating until now.

Methods: we used Oxford Nanopore Technology to detect COVID-19, generating consensus sequences data for 24 samples that consist of the base for analyzing data and extracting results using many bioinformatics tools that varied between processing data, studying mutations, and phylogenetic trees. determine the genetic lineage for samples in Syria using three platforms Pangolin, Nextclade, and GISAID database.

Results: out of a hundred samples only 24 samples were sequenced after giving high CT values**.** Through tracking viral variants in Syria, the XBB recombinant Omicron variant circulated at time Jun, July, and August. XBB.142.1(75%) in 18 sample was designed as Syrian lineage, XBB.2.3.11(16.67 %) in 4 samples, XBB.1.9.1(4.17%) and FL.12 (4.17%). the clades were 22F, 23E, and 23D. The clade in GISAID is GRA unless one sample GR. 19 samples from this study were submitted to GISAID (Global Initiated on Sharing All Influenza Data). We apply a genome-based survey for mutations and evaluate the genetic diversity and the extent of its relationship with the changes recorded in the world through the phylogenetic tree with neighboring countries showing the relation in genetics and the prevalence.

Conclusion: This is the first study that has provided epidemiological descriptive and valuable insights into the genetic variation in the Syrian population, hoping to introduce valuable analysis used in more studies to evaluate the vaccines and medicines in our country.

Keywords: SARS-CoV-2, GISAID, Phylogenetic tree, clades

Contents

LIST OF FIGURES

LIST OF TABLES

List of Acronyms

Item	Name
NGS	Next Generation Sequencing
ONT	Oxford Nanopore Technology
SARS-CoV-2	Severe Acute Respiratory Syndrome coronavirus2
MERS-CoV	Middle East Respiratory Syndrome Coronavirus
WHO	World Health Organization
PCR	Polymerase Chain Reaction
VTM	Viral Transport Medium
CT	Cycle Threshold
FAST5	Container file using the HDF5(Hierarchical Data Format) to describe information of Data
FASTQ	Text file format for storing both usually nucleotide sequencing and corresponding Quality score
FASTA	Text-based format for representing nucleotide or peptide sequencing
GISAID	Global initiated on sharing all influenza data
NCBI	National Center for Biotechnology Information
CDC	Central of Diseases Control and Prevention
ORF	Open Reading Frame
NSP	Non-Structural Protein
PDB	Protein Data Bank
BAM	Binary Alignment Format
NRP1	Neuropilin-1

TMPRSS2	Transmembrane Protease Serine 2
MHC-I	Major Histocompatibility Complex Class I
GC	Guanine-Cytosine
Indels	Insertion and Deletion
S	Spike glycoprotein
M	Membrane
N	Nucleocapsid
E	Envelope
RdRp	RNA-directed RNA polymerase
RBD	Receptor Binding Domain
SNP	Single Nucleotide Polymorphism
bp	Base pair
MSA	Multiple Sequence Alignment
VOC	Variant Of Concern
VOI	Variant Of Interest
VUM	Variant Under Monitoring
UTR	Untranslated Region
ER	Endoplasmic Reticulum
ACE2	Angiotensin Converting Enzyme
mAbs	Monoclonal Antibodies
CSV	Comma Separated Values
RMSD	Root Mean Squared Distance

Introduction

The Coronavirus has constituted a qualitative shift in the world since the beginning of its spread on 31 December 2019 centered in Wuhan, Hubei Province, China [1]. since this date, more than 1975 residents flocked to hospitals with respiratory complaints. these cases of pneumonia were found to be related to a large seafood and animal market in Wuhan and later revealed to be a novel coronavirus, known as severe acute respiratory syndrome coronavirus 2. Now nearly 17 million sequences have been shared online in the GISAID database. analysis of this new virus showed 89.1% similarity to SARS-like coronaviruses (genus, Betacoronavirus, subgenus, Sarbecovirus) that existed before in bats in China. [2] Outbreaks the causative pathogen is severe acute respiratory syndrome coronavirus 2 known as (SARS-COV-2) and the WHO announced it Coronavirus Infectious Disease (COVID-19) which is the strain of coronavirus disease caused convicting pandemic.

The first confirmed sample of this virus was reported on 5 January 2020 and was released on the open-access virology website for researchers a month after COVID-19. on 10 January 2020 take MN908947 an accession number in GenBank [3]. The WHO announced the first coronavirus pandemic COVID-19 on 11 March 2020 [2] and its rapid spread constituted great concern and difficult challenges that faced the world, as quarantine rules were applied and complete closure of many sectors, which led to negative effects on the economy and social life and posed a difficult challenge to face the epidemic. The first Syrian case of COVID-19 was reported on 22 March 2020 for a woman infected who traveled to the country, while the first death occurred on the 29th of the same month. Initially, the diagnosis of COVID-19 depends on the PCR approach. Late in 2023, Oxford Nanopore technology was used.

Health sectors in Syria suffered under a war that ravaged it and lasted for years, during which it lost many hospitals and health centers which made the health infrastructure unready.

Great challenges faced by the Syrian government and international organizations to confront the coronavirus and limit its spread, the effect of sanctions that included many sectors in light of an economic crisis, limited resources, and the devaluation of the Syrian pound, the country suffering from a deterioration in the health care system and a lack of

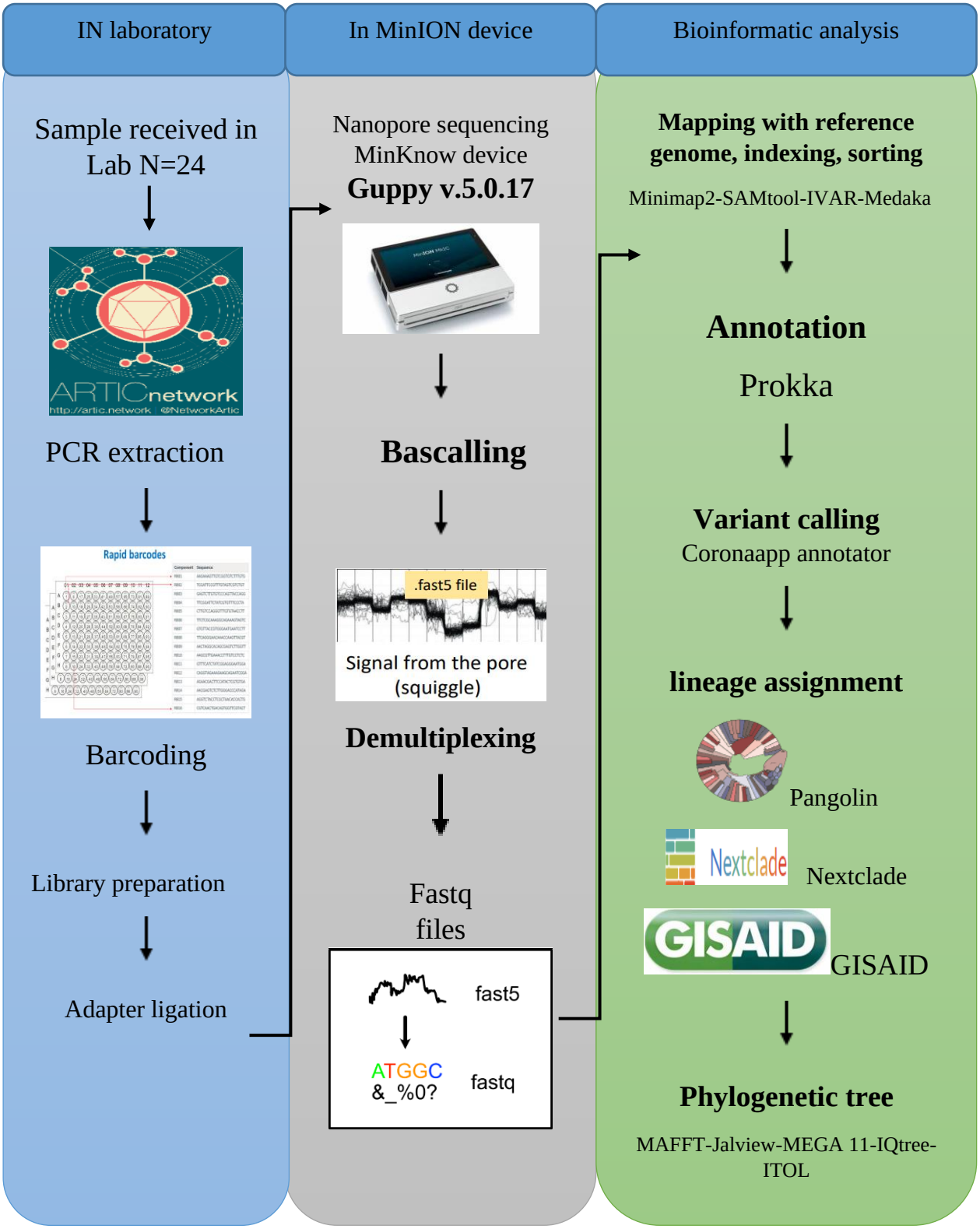
resources services and medical devices such as ventilators, medical equipment, and pharmaceuticals. According to the statistics announced by the Syrian Ministry of Health, the number of people infected with COVID-19 in Syria was 57743 cases until 26 August 2023. Most of them in Damascus governorate, Lattakia, and Homs, while the number of deaths reached 3165 persons. On 22 April 2021, Syria received the first shipment of the AstraZeneca vaccine which is a glimmer of hope for people and Health workers in the country exhausted by the decade-long war. www.unicef.org , syria.un.org

Aim of the study

In light of the lack of genetic studies since the pandemic of SARS-CoV-2 regarding the population in Syria, this study provides comprehensive genomic insight into the whole genome of SARS-CoV-2 variants circulating in Syria in June, July, and August 2023 for 24 isolates of Syrian individuals from major cities using Oxford Nanopore Technology. through which we determine and analyze the mutations and SNPs in the complete genome of SARS-CoV-2 using the appropriate bioinformatics tools, in terms of quality sequencing mapping, assembly, calling variants, lineage, and phylogenetic relationships. This study highlights the prevalence of the Omicron variant, through the population. submitting the sequences on the GISAID platform, putting Syria on the genome map of SARS-CoV-2 distribution by sharing our sequences with the world.

Research Hypothesis

Looking for lineages of VOC circulating in Syria and determining the trajectory of SARS-CoV-2 through Oxford Nanopore Technology. The results are processed and analyzed using a series of bioinformatic tools that provide A detailed epidemiological explanation of the mutations and their positions on the genome, and determine the circulating lineages and spread of the virus in Syria.



Workflow for detailed steps of SARS-CoV-2 analysis

Chapter one: Theoretical section

NGS technologies (brief history)

The sequencing genome journey began when Watson and Greck found a Structure for Deoxyribose Nucleic Acid (DNA) in 1953 [4] the initial beginning of sequencing methods was in late 1970 when Sanger developed a method for sequencing based on the selective incorporation of chain-terminating fluorescently tagged dideoxynucleotides (ddNTPs) by DNA polymerase. The first-generation sequencing was with reads less than one kilobase. in 1975 sequencing the first genome phage ϕX174, it was a 5368 bp. [4][2]

The initial sequencer of NGS relied on Pyrosequencing and this was later commercialized by 454 Life Science [4] in 2003 Solexa developed Sequencing by Synthesis (SBS) engineering DNA polymerase [4] and released Illumina sequencer in 2006 where both strands of DNA were sequenced based on bridge amplification followed by SOLID sequencer which is based on a ligate fluorescent oligonucleotides. Both sequencers generate 35 bp long only which is announced as the next generation sequencing. In 2010 Ion Torrent released the Personal Genome Machine depends on proton detection in semiconductors that generate 100 bp-long [5][6][7].

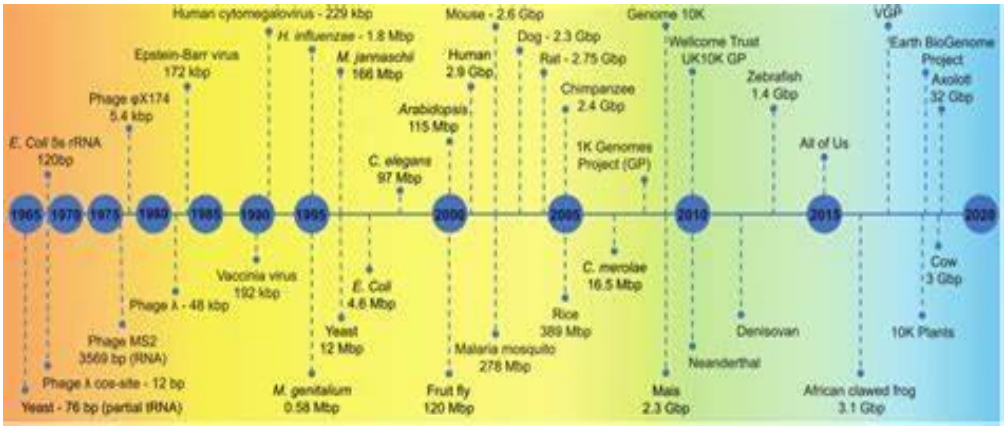


Figure 1 A brief history of NGS from the beginning of sequencing to the large-scale genome projects. [7]

2011 saw the development of sequencing methodologies as Pacific Biosciences (PacBio) released single molecule Real Time (SMRT) sequencing technology that consists of the initial of Third-Generation Sequencing (TGS), technologies capable of sequencing single DNA molecules without amplification [2][4]. And allows it to produce long reads of 30-50 kb These attempts culminated in the development of technology that was able to sequence millions of genomes per run able to generate a huge amount of data depending on nanopore technology released in 1990 which was named fourth generation sequencing combined significant properties precision in sequencing a huge size of genome with low cost and short time and can greatly control and directed investigation and follow-up operations for SARS-CoV-2. shortly has become a strategy capable of overcoming NGS sequencing drawbacks and pitfalls. [8]
Oxford Nanopore Technology (ONT) introduced nanopore sequencing in 2005 [4] enabling direct real-time analysis of individuals and all forms of living beings producing whole genome data used to describe several different massively parallel with large volumes of raw data [5].

Bioinformatics and NGS

NGS technology is closely linked to bioinformatics as generates a massive amount of raw data that requires Bioinformatics pipeline tools and algorithms to process huge data and extract results. Brings benefits and good management for issues, that vary between alignment, assembly, annotation, and visualization. Process vary in thus taking the necessary health measures and achieving good management, especially in health crises, as happened in the COVID-19 epidemic. Bioinformatics tools constituted an important factor in managing the epidemic, developing diagnostic, and monitoring tools, and deriving results and presenting them to researchers. Regarding the Coronavirus, studying the genomic sequence, knowing the origin of the virus and its family, and following up on its spread globally. This leads us to follow its evolution and changes occurring in the genome, thus mutations, and new patterns, and study the rate of their occurrence in the future. [9][2]

Output results require computer processors and digital algorithms to draw conclusions and devise solutions.

The NGS will not stop here, but also study immune responses providing a molecular basis for the development of SARS-CoV-2 vaccines, and the detection of drug-resistant genes which decrease the time for researchers and a large segment of scientists around the world and allows them to identify unknown infection routes. That makes it a powerful tool in genomic epidemiology and transcriptome analysis. Nanopore Technology is a high throughput sequencing that was untenable before in short reads which causes errors and ambiguities for read mapping. Nanopore has recently been used in many fields like bacteria, viruses, and plants. The great credit of ONT for providing valuable information regarding the transmission, development, and evaluation of the pathogenicity of the virus. And will not forget one of the applications Building a new reference genome for instance ONT used to assemble the first genome of Rhizoctonia solani. ONT and Illumina was used to assemble the first draft genomes of Maccullochella Peelii, and Amphiprion ocellaris (kinds of fish) and so many reference genomes. [10][11]

Limitations of NGS

Despite the power of Next Generation Sequencing methods such that we obtain from a small amount of DNA the entire genome sequence, including genomic variants, the technology carries many drawbacks, the first of which is its high costs including equipping laboratories and raw materials, providing a work and training environment, in addition to preparing the virtual environment in terms of computer and software resources capable of processing huge amount of data using appropriate bioinformatics tools and ability to store a large data processing.

The challenge is that NGS requires high-quality DNA or RNA samples to obtain results with high accuracy, and the shortness of reads generated limits the number of SV and difficult-to-sequence genomes with long lengths like the SARS-CoV-2 genome. [8]

On the other hand, accuracy plays a significant role in the quality of the results. many factors affect the reading and extracted results. The first obstacle is the ability to map reads to the reference genome and read

length itself. as the percentage of CG content, high or low where sequence alignment can be biased and error-prone (i.e. aligning the less complex and no-repeat reads preferentially over the tricky regions) that led to obtaining wrong or inaccuracies in interrupting results. [5][12]

Although these limitations might be acceptable for research projects for prescreening or discovery purposes, they are not suitable in clinical practice, where small errors may lead to severe consequences for patient diagnosis and/or treatments. Likewise, significant bioinformatic tools rigorously validate within the intent to ensure it is accurate and reproducible. Hence the importance of using bioinformatics tools to detect variants within the intended assay evaluate the results issued, and examine the accurate reporting of analytics and specificity. [5][12] Contamination: NGS influence with degradation, can impact the quality of genetic material, whether from the laboratory environment, workers, or the reagents used, which poses a challenge in applying standard protocols and methods in sequencing. Also, a need for trained and educated labor to analyze and interpret the released data. [12]

Flow cell of Oxford nanopore devices

Depending on the ability of DNA to pass through holes produced by protein membrane. Two proteins could detect recognizable ionic current blockades of strands of the genome, α-Hemolysin protein from Staphylococcus aureus and design nanopore protein Mycobacterium smegmatic porin A (MspA) engineered to allow the four bases on oligonucleotide molecules to be distinguished [13]. Improving the signals to slow the DNA translocation through the flow cell. Control this moving by the addition of motor protein that controls the rate of DNA passage through the pores. such as a highly processive DNA polymerase (phi29) and other accessory proteins, such as a DNA helicase, exonuclease I, or oligonucleotides to bind DNA strands. [14][13]

Using ARTIC Network protocol (V3) (https://artic.network/) which is developed as a protocol to be used on Oxford Nanopore Sequencing, can be applied to different libraries preparation kits and sequencing platforms environmental and metagenomics samples. it has been applied to pathogens with larger genomes like Ebola virus 18-19Kb in 2016, Zika

virus 11 Kb, and last widely used for whole genome sequencing of SARS-CoV-2 30 Kb [7]. The first protocol was released for rapid sequence for clinical samples such as nasopharyngeal or oropharyngeal swabs of SARS-CoV-2 RNA, and integrated in MinION MK1C [15]. MinION is the first long-read nanopore-based sequencer, it was released for users in 2014 and commercialized in 2015 by Oxford Nanopore Technology. ONT constantly evolving, improving flow cells and motor protein. Other versions as GridION, and a high throughput PromethION (48 flow cells of 3000 nanopores each). MinION platform permits real-time analysis because individual DNA strands are translocated through the nanopore allowing decisions to be made during the sequencing run. [1][14][13]

Flow cell containing 2048 nanopores arranged in 512 channels with 4 nanopores in each channel. sequencing occurs by reading the changes in ionic current when a strand of DNA or RNA passes through the protein nanopore which is a tiny hole on an array and translating the ionic current changes across the membrane into nucleotide type at ~450bp. MinION nanopore technology is portable, handy, simple to use, and rapid thus is highly applicable for real-time sequence analysis [10] able to generate 10 - 20 GB of data in about 48 hours.

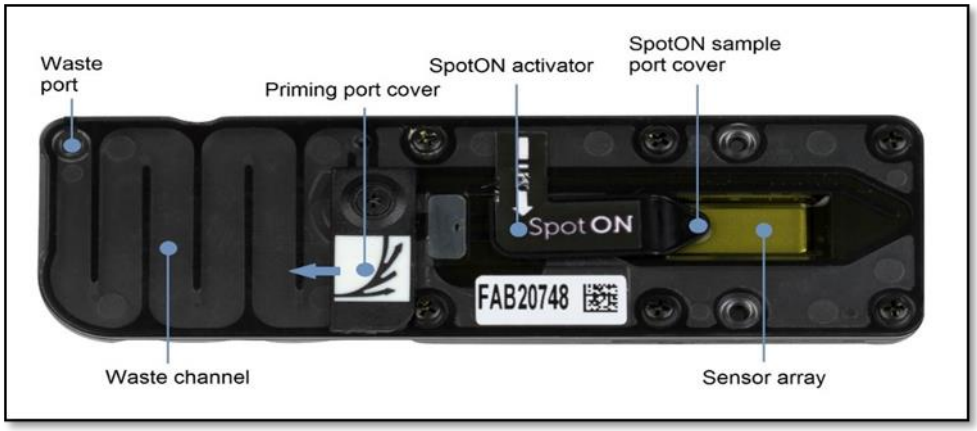


Figure 2 Flow cell of MinION MK1B

Overview and origin of SARS-CoV-2 and Molecular biology

Human coronaviruses (CoVs) pose a major threat to human health, group family contains seven viruses that are known to affect humans, and has been classified into three generations (Alpha, Beta, and Gamma), namely, HCoV-NL63 and HCoV- 229E from α alpha genus, have receptor binding affinity Aminopeptidase N (APN), and HCoV-OC4, HCoV-HKU1, SARS-CoV, MERC-CoV, SARS-CoV-2 from the βcoronavirus family Lineage B that infects humans. [16] The most interest in this family spike protein consists of the virus surface, which gives the virions a crown-like shape. B-coronaviruses use receptors: dipeptidyl-peptidase 4 (DPP4), and angiotensin-converting enzyme 2 (ACE2). [17]

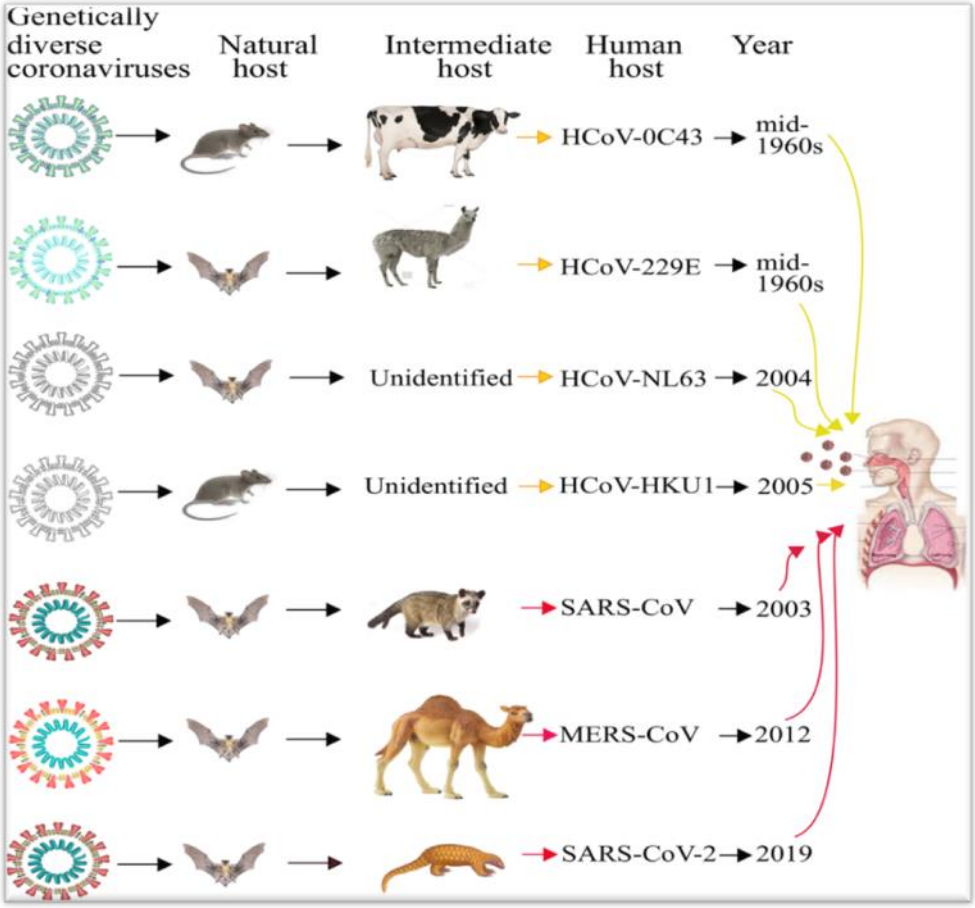


Figure 3 Overview of beta coronaviruses and transmission [20]

SARS-CoV: in November 2002 SARS began spreading in China for patients exposed to animals. And quickly spread as a global epidemic in

more than 30 countries SARS-CoV belongs to the genus beta coronavirus, lineage B. an envelope, positive-strand RNA virus for a long 29.727 nucleotides, 41% of which are guanine or cytosine. [16]

Scientists reported that the bat strain WIV16 is the natural host of genetically diverse coronaviruses, suggesting that SARS-CoV was produced by recombination within bats and other mammals (civet cat) before transmission to humans. [16]

MERC-CoV: Middle East Respiratory Syndrome virus The six virus belongs to beta coronavirus, mediated respiratory diseases, lineage c with genome size 30.119 nucleotides [16], the first reported in humans in Saudi Arabia in 2012 and has since spread to other countries in the Middle East, Qatar, Asia, Europe, and the United States, still infects humans causing a fatality rate of 35.5%. [18]
MERS-CoV strains isolated from camels were identical to those from humans (camel workers and patients not exposed directly to camel) and that supports the belief that camels' immediate source of most primary human infections. [19] And like SARS-COV, MERS-CoV is believed to have originated from bats as Sequence analysis indicates related to Tylonycteris bat coronavirus HKU4 and transmitted to camels. [20] MERS-CoV is a zoonotic virus transferred through many animals from bats and camels to Humans. MERS-CoV enters the cell via the human dipeptidyl peptidase 4 (DPP4) receptor also known as CD26 occurs in MERS-CoV infection. MERS-CoV is a zoonotic from animals to humans. [16][18]

Effective vaccines play an important role in preventing MERS-CoV infection depending on the S protein properties and DPP4 receptor via a receptor binding domain RBD. According to a research report MERC-CoV N protein-based vaccine induces $CD8^+$ T cell response in a mouse model. [21]

SARS-CoV2: At the end of 2019, the world witnessed a dangerous transformation and global concern was represented by the outbreak of the Covid-19 epidemic from 2019 to 2023 that infects humans, which was called Sever Acute Respiratory Syndrome (SARS-CoV-2). The sequence isolates from the patient were analyzed using NGS technologies, the tree of emergence and evolution showed that the new virus belonged to a

Sarbecovirus Subtype of βcoronavirus, lineage B. SARS-CoV-2 is the seventh virus that belongs to the βcoronavirus family that infects humans. [16]

Returning to metagenomic analysis shows SARS-CoV-2 is the most closely related virus to RaTG13 with an identity of 96.2% sharing a common ancestor, including the S gene. [22][16] [23]
That led the researchers to think that SARS-CoV-2 was produced by the selection and recombination of bat and pangolin coronavirus. [8]
Coronaviruses like SARS which spread in 2002-2003 and MERS in 2012 occur as infectious diseases, especially respiratory diseases, sharing similarities in RBD region on the genome. and sharing characteristics like affecting respiratory ways, they are airborne and transmitted from animal to human. [24] But SARS-CoV-2 remains less pathogenic than MERS-CoV and spreads more and appears to have higher transmissibility. [25]
While SARS-CoV-2 has a higher reproductive rate, every replication cycle takes ~10 hr, which reflects the mutation rate arriving at 10^{-3} mutations per site per year assuming neutrality and neglecting the effects of evolutionary selection, controlled by the viral exoribonuclease nsp14. [26][27] This large genome size of about 29.903 bp suggests there is a proofreading mechanism called ExoN to reduce the mutation rate (~10^{-6} sites per cycle) and stabilize the genome. That enhances its ability to transfer rapidly, escape from the immune system, and easy transmissibility. [27]
Both arboviruses and SARS-CoV2 initiate infection by binding to receptors in the human cell-surface protein angiotensin-converting enzyme 2(ACE2) via the spike protein determined in a region called the receptor binding domain (RBD), this binding in SARS-CoV-2 consists dissociation constant (KD) of 14.7 nM, and the deference in S protein is ~ 24% indicating SARS-CoV-2 is more sensitive to ACE2 than SARS-CoVs. [28][29][30]

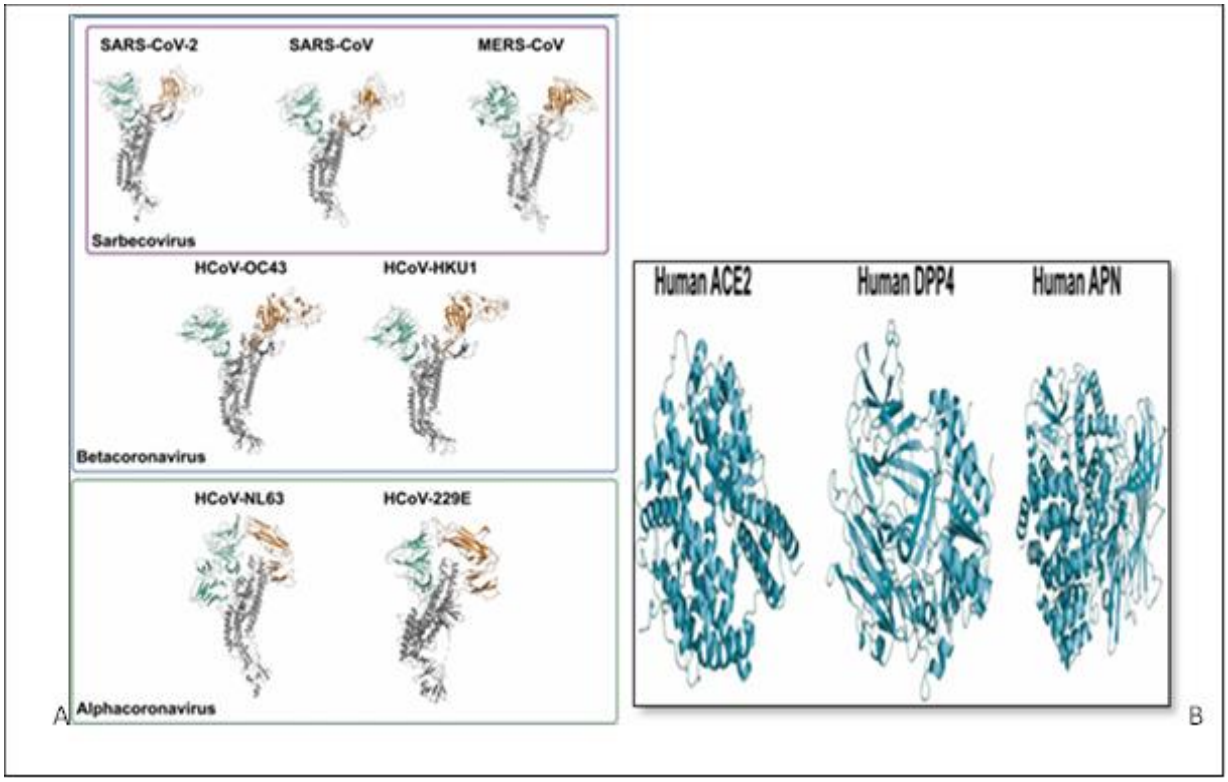


Figure 4 (A) Structure of spike protein from seven human-infecting coronaviruses. (B) Structure of monomers of human ACE2 (PDB: 1R42), human DPP4 (PDB: 2ONC) and human CAN (PDB: 4FYQ). [17]

Molecular biology of SARS-CoV-2

SARS-CoV-2 is an enveloped, spherical-single strand-positive sense RNA virus, which is 29.903 bp and encodes 9860 amino acids, SARS-CoV-2 enters the host cell through spike glycoprotein by attaching the cell which has a high affinity to human Angiotensin Converting Enzyme 2 receptor hACE2 producing infectious. it is 29.903 kb in length, with 14 ORF and 5 major structure proteins. the **5′** frameshifted poly proteins (Orf 1a and Orf 1ab) encode coterminal polyproteins pp1a and pp1ab for viral RNA synthesis which are later converted by proteolytic cleavage into 16 nonstructural proteins (NSP1-NSP16). [20][31]

(NSPs) known as the replica transcriptase gene consists two two-thirds of the genome which encodes many enzymes for RNA processing and viral replication and methylation. [32]
These nonstructural proteins names as Host translation inhibitor (nsp1), p65 homolog (nsp2), Papain-Like-Proteinase (nsp3), (nsp4), 3C-Like

Proteinase (nsp5), (nsp6), (nsp7), (nsp8), (nsp9), Growth factor-like peptide (nsp10), RNA-directed RNA polymerase (nsp12), Helicase (nsp13), Guanine-N7 methyltransferase (nsp14), Uridylate-specific endoribonuclease (nsp15), and 2′-O-methyltransferase (nsp16). and four 3′ four structural proteins named spike (S) for viral entry, envelope (E) forms an ion channel in the viral membrane and participates in viral assembly, membrane (M) is critical for incorporating essential viral components into new virions during morphogenesis, and nucleocapsid (N) associated with the viral genome and M to direct genome packaging into new viral particles. [33]

These proteins consist of one-third of the genome and are shared with all coronaviruses. Differ from coronaviruses, the SARS-CoV-2 genome has accessory proteins that vary in number, location, and size named (ORF 3a,3b, 6, 7a, 7b, 8a, 8b, 9b, and 10) arranged in a way that two are located between S and E genes (Orf3a, Orf3b), ORF3a the largest of accessory proteins act as ion channel promote virus release while ORF3b protein is a potent interferon (IFN) antagonist, suppressing the induction of type 1 interferon. five proteins are between the N and M genes, and ORF9b is arranged within the N gene. that are not required for virus replication related to pathogenicity in the natural host. [34]

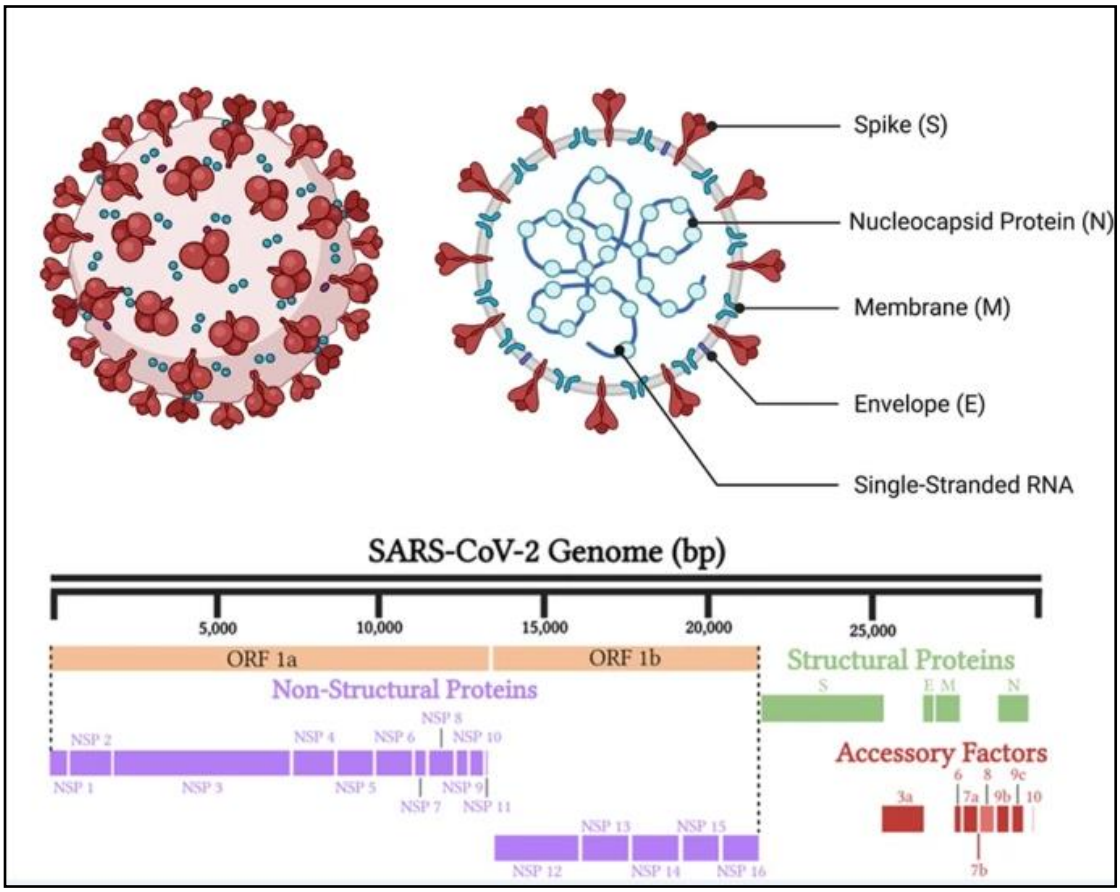


Figure 5 Structure of SARS-CoV-2 genome [29]

Characteristics of infection

Non-structural protein Nsp1 is one of the least mutated proteins during the SARS-CoV-2 pandemic, with a rate mutation lower than 0.01 for all amino acid residues. [35] NSP3 has two jobs: cutting viral proteins to free and remove tags from other proteins which set for the destruction that loses the ability of cells to fight the virus. [35] While Spike protein a trimeric class I TM glycoprotein that has a spotlight due to its exposure to mutagenesis more than other viral proteins and plays important roles in viral entry and infection. spike protein is common to all HCoVs located on the surface of the virus and consists of two subunits:

(subunit 1) S1 is responsible for binding and recognizing the human cell, where there are two receptor binding domains N-terminus domain which binds to carbohydrates, and the C-terminus domain (CTD). receptor binding domain (RBD) residues (333 to 526), found as a pocket inside the

S protein. [36] When the virus replicates, the S protein activated enables the RBD region (the most variable part on the surface of the virus) to recognize the human host cell receptor Angiotesin-converting Enzyme (ACE2) making it the primary target for the development of monoclonal antibodies because bosses the ability to escape such antibodies. [28]

RBD is composed of a receptor-binding motif (RBM) that interacts with the hACE2 to form a gently concave surface that accommodates the N-terminal α-helix of the hACE2 and a series of hydrophilic residues that form a solid network of H-bond and salt bridge interactions dominate polar contacts and a core made of interconnected loops and helixes. [23][33] [35]

ACE2 is a homolog of ACE, which converts angiotensin I to angiotensin 1-9. ACE2 is located in the lung, intestine, heart, and kidney. [29][37][38]

(subunit 2) S2 controls the fusing between the viral and cellular membranes of ACE2 which is exposed on the cell surface of Human lung tissue and makes it easy to enter the cell by three critical domains:

1-fusion peptide FP composed of GLY and Ala, allowing the interaction with the membrane.

2-heptapeptide repeated sequence 1 (HR1) forms a homotrimeric assembly that exposes three highly conserved hydrophobic grooves on the surface that bind to heptapeptide sequence 2 (HR2) which they interact during the fusion process and form a six-helix bundle (6-HB) domain core region prepares the cell and viral membrane to enter the cell. [39][5][38]

In addition to protease Furin a (membrane-bound), described as a secreted protein, contains four redundant cut sites (PRRA motif) to facilitate virus entry into the cell. [40]

3-transmembrane protease serine 2 (TMPRSS2) which is located in the host cell membrane plays a scientific role by priming the S proteins of SARS-CoV-2 and facilitating its endocytosis and the release of its viral genome in the infected host cell for subsequent replication. [14]
The interaction between the S and ACE2 shows a high variation especially in its RBD region along with three Nonstructural proteins the

Papain-like protease (nsp3), 3C Like protease (nsp5), and the RNA-directed RNA polymerase (RdRp)(nsp12) are important targets for researchers to observe these proteins apply homology modeling and molecular dynamic to design appropriated antiviral drugs. [41]

SARS-CoV-2 cycle replication begins by translating the structural proteins into the endoplasmic reticulum (ER)-Golgi, which ultimately integrates to form elaborate webs of convoluted membranes, translating the template RNA to structural proteins. and accessory proteins. After formation, the virions are transferred to the cellular surface. [34]

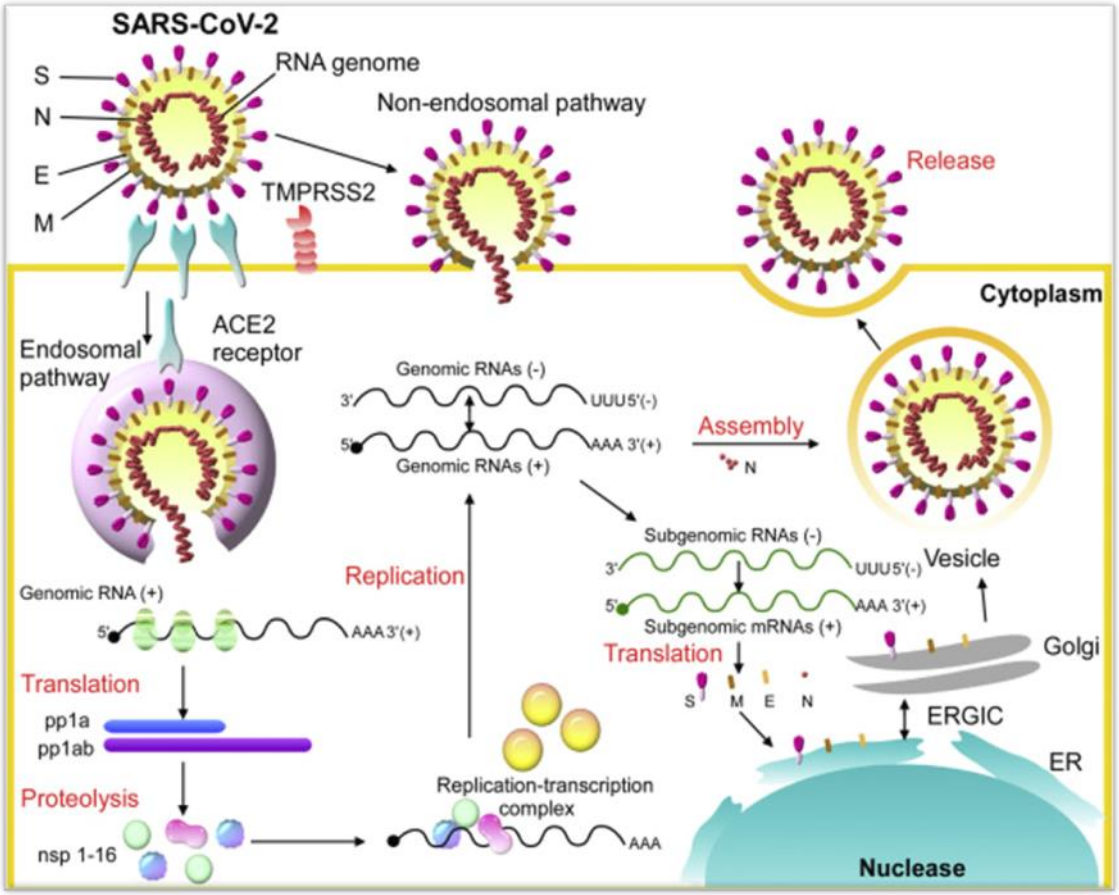


Figure 6 Replication cycle for SARS-CoV-2 [26]

Variants of SARS-CoV-2:

Because coronavirus is an envelope, spherical RNA virus it has a high rate of mutation under replication, that analysis shows developing in rate ~1.1×10^{-3} nucleotide substitutions and increases over time causing one substitution per ~11 days, while S protein considerate one of the highest

mutated protein with 4% mutations observed in the first quarter of the pandemic as reported in Koyama. [42][43][44] There generate a new genetic vibration by affecting the amino acid sequence of the protein Study virus sequence gives a chance to deal with the pandemic and attract the spread of this virus. Because of this high ability to mutate, the World Health Organization (WHO) classified the variants according to their various effects on the characteristics of the virus in terms of spread, diagnostic methods, its association with other diseases, effectiveness with the vaccines, and protocols of treatment. This variety in substitution and mutations led the WHO and Centers for Disease Control and Prevention (CDC) to classify the variants of the SARS-CoV-2 to sort the more significant of these variants and grouped them for ease into three different classes: variant of concern (VOC), variant of interest (VOI), and variant under monitoring (VUM) www.who.int. The classification of variants may differ according to each country's situation.

1) Variants Of Interest **(VOI)** SARS-CoV-2 variant with genetic changes that are predicted or known to affect how the virus behaves or impacts human health include the ability to spread or cause infection and how easily to treat global health, immune escape, and disease severity Whereas

2) Variant Under Monitoring **(VUM)** with genetic changes that are suspected to affect virus characteristics with some indication that it may pose a future risk, evidence of phenotypic or epidemiological impact is currently unclear.

3) Variants of concern **(VOC)** SARS-CoV-2 variant are detrimental changes that have clear evidence of an increase in transmissibility, increase in severe disease, increased hospitalizations or deaths, and/or impact on immune responses and drug efficacy. Significant by antibodies generated during previous infection or vaccination, making it difficult for the health system to provide health care for patients. VOC labels using Greek letters corresponding to Pango lineages and the country which it originated from as follows Alpha, Beta, Gamma, Delta, and Omicron. Alpha (B.1.1.7_United Kingdom_ September 2020) the first VOC with high transmissibility and pathogenicity, includes 17 mutations in the viral genome. Beta (B.1.351_South Africa_

December 2020) contains nine mutations in Spike protein, associated with the E484K mutation, which increases with K417N and N501Y variants the binding affinity to ACE2 receptor. [45][46] Gamma (P.1_Brazil_ January 2021) similar to previous VOCs, variants enhance the infectious and binding to hACE2 and disease severity with a sudden increase in the number of COVID-19 infected hospitals and deaths. Another VOC appeared but not for a long time Mu, Zeta, Kappa, Lota, Eta, and Epsilon, however, those variants with Alpha, Beta, and Gamma all became VBM on 21 September 2021 according to www.cdc.gov. Delta (B.1.617_India_March 2021) has already evolved into three sub-lineages, defined as B.1.617.1, B.1.617.2, and B.1.617.3. possess the ability to escape immunity, and risk of reinfection. a new infection mutation appears in a spike protein like L452R, P681R, E484Q, and G142D.[46] Omicron (B.1.1.592_ South Africa_ November 2021) is a new variant recognized as the fifth VOC. Known as higher transmissibility, and immune escaping.

On 15 March 2023, WHO updated its tracking system and definitions for variants the Greek letters will only be used for VOC while VOI will be referred to as used by Next strain and Pango. WHO classified XBB as an Omicron subvariant under monitoring www.who.int/en/activities/tracking-SARS-CoV-2-variants. Currently, Omicron is the only one circulating worldwide.

Omicron (B.1.1.592) variant

Omicron (B.1.1.592) is recognized as the fifth VOC by the WHO on 26 November 2021 in South Africa, and the only VOC from 9 October 2022. [47][48] Raising the average number of COVID-19 cases to 44%. [49] The rate of infection is faster than the other four variants, the omicron variant may be over 10 times more contagious than the SARS-CoV-2 Wuhan virus or about 2.8 times more infectious when compared to the delta variant. [50][51] The Omicron RBD part has a weaker or similar binding affinity to ACE2 [52][53] causes less severe infections, a low rate of mortality, and is lower than other strains (VOCs). [51]
However, Omicron variants are highly transmissible, have high

environmental stability, and evaded immunity even for those vaccinated or previously infected that causes concern for the community. [54] What distinguishes it is that Omicron spike protein is not accepted to the solvents and higher fractions of alpha-helix composition cause a property of hydrophobic and nonpolar amino acids. [55] Omicron has a large number of mutations in comparison with other VOCs, with more than 50 known mutations 32 of them in the Spike protein in the virus with continued increase, and 15 out of them are in the RBD region. [56] These mutations enhance the binding affinity with ACE2 more than the wild type of SARS-CoV-2. [57][58] Omicron has one insertion in the NTD region and three deletions compared to the wild type of SARS-CoV-2. Omicron variants made selective balance by increasing transmissibility and making it difficult for antibodies to bind with the S region of the virus. [48] [53]

The five sub-lineages of Omicron variants have been detected namely BA.1, BA.2, BA.3, BA.4, and BA.5. The most interesting is the ability of these variants to escape from the immune system and decrease the neutralization efficiency of the vaccine. During this wave, two discovered sub-lineages appeared BA.4 and BA.5 in South Africa followed by BA.2.12.1 sub-variant distinguished by its ability to escape from immunity. [59]

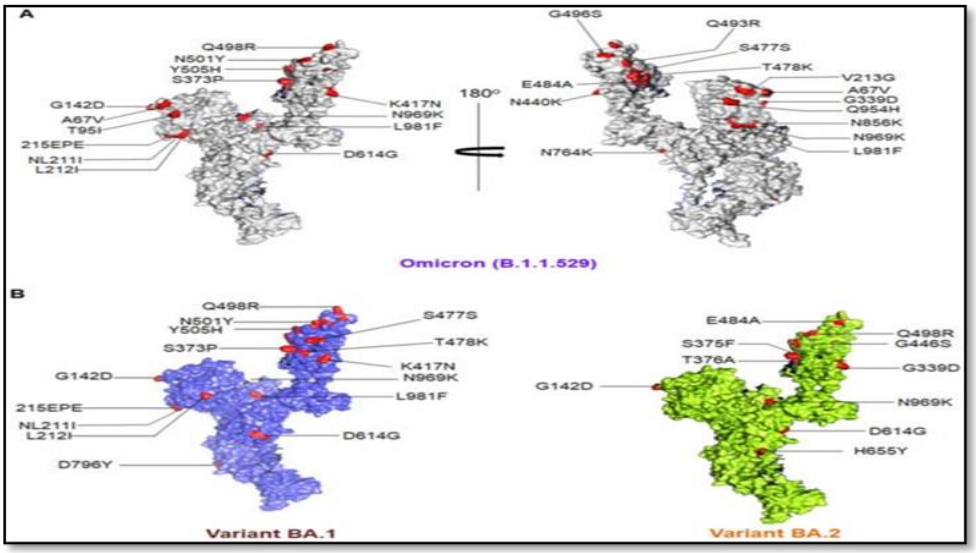


Figure 7 3D model of the S-glycoprotein of the Omicron and Omicrons sub-variants, BA.1, BA.2 [46]

New parts of sub-lineages consist of numerous recombinant sub-variants that evolve namely XBB, XBD, XBF, etc. The Omicron XBB variant likely originated through recombination of two BA.2 lineages (BJ.1 and BM.1.1.1). [60] Also, XBB.1.5 subvariant significantly evades humoral immunity induced by vaccinations and natural infection, increasing fitness through recombination. In October 2022, the WHO classified XBB as an Omicron subvariant under monitoring. [61][60]

Symptoms

A COVID-19 patient suffers from several symptoms from the beginning of the epidemic. This diversity is due to the presence of ACE2 in many tissues, including the liver, heart, pancreas, kidneys, and lungs symptoms are not limited to respiratory symptoms, as there are other symptoms such as headache, dizziness, disturbances in the sense of smell and taste, headache, nausea. However, these symptoms differed during the outbreak of the Omicron epidemic, disease severity appears to be lower than compared with previous epidemics like Delta. There was a decrease in hospital admissions and a lower death rate decrease in the severity of symptoms was also observed, It was limited to sore throat and a hoarse voice. Omicron tends to stay in the upper respiratory tract, such as in the nose, throat, and bronchi, rather than settling in the lungs. also vomiting, and diarrhoea are the most common in children while less severe symptoms are in infants. It must be noted that the number of people vaccinated during the Omicron period 78 million greatly exceeds the 1.6 million vaccinated during the Delta period. [59]

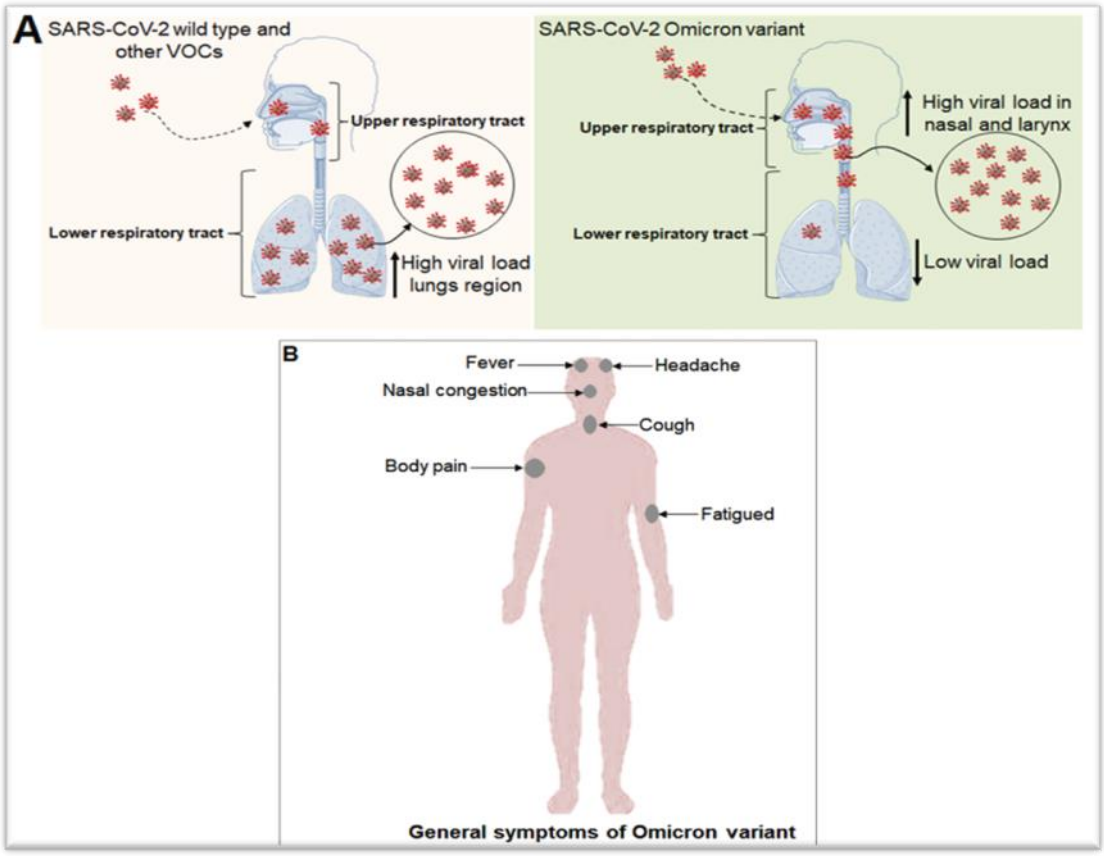


Figure 8 Viral load in the respiratory tract and the lungs during an infection with the wild strain of SARS-CoV-2 and Omicron [46]

Diagnosis

There are multiple diagnostic methods for SAR-CoV-2 at several levels including two major types**:** serological- or immunological-based assays.

1. the molecular-based methods: Molecular methods depend on detecting viral RNA using techniques such as RT-PCR which is considered the gold standard for diagnosis. the most common clinical diagnosis tool to detect the presence of SARS-CoV-2 at the beginning of the epidemic. relying on the enzymatic activity of DNA polymerase and short DNA fragments called primers. [62]

Digital PCR measures the positive fraction of samples that reflect an actual number of molecules used by digital droplet PCR. Where samples are divided into tens of thousands of nanodroplets. dPCR uses the same primers as RT-PCR. [5]

Isothermal nucleic acid Amplification works at a fixed temperature of 37-

42 C° only allowing simplicity and less cost without thermal cyclers, This technology has two advantages: fast, it requires only 15 minutes for amplification using exonuclease probes with a linked quencher (Exo-IQ). Second, it does not need high temperatures. And **Reverse Transcription Loop-Mediated RT-LAMP technology**, which is a one-step amplification in less time. Highly sensitivity to a specific region, that developed primers for ORF1ab, E, and N genes to detect SARS-CoV-2 [63][5]

CRISPR-based assay associated with CAS enzymes (Cas12, Cas13) includes RNA extraction and RNA amplification. Efforts continue to simplify detection processes as well as SHERLOCK testing in one STOP which is a detection method combining virus RNA extraction, isothermal amplification, and CRISPR-based detection. Dings group utilized a CAS12a-based detection specific for N genes after Isothermal amplification. [5][64]

2. The serological-based methods: Protein antigen tests mainly utilize blood serum or plasma-like secondary anti-human IgM, and IgG antibodies, the initiate response of the body to defence against infection**,** increases during the first week after infection, but IgG level continues for a long time**. Immune-based methods:** there is a variety of techniques to detect antibodies such as enzyme immunoassays like **Enzyme-Linked Immunosorbent Assay** (ELIZA), investigating present antibodies against the receptor binding domain (RBD) in the sample. Also, immunofluorescent assays release fluorescent signals when antibodies interact with viral protein. This technology is low-cost and simple. Another technology is under development.

Rapid antigen detection: uses specific monoclonal antibodies to detect Spike and N proteins of SARS-CoV-2 through antibody-antigen interaction. Moreover, far from these methods **Imaging** aims in the diagnosis of SARS-CoV-2, including chest computed tomography (CT) relative to the damage in the lungs. AI allows for analyzing chest radiography, and lung ultrasound and extracting scientific results quickly and accurately. [62]

Treatments of SARS-CoV-2

The great service provided by NGS technology, and the availability of the SARS-CoV-2 sequence, saved time and effort for researchers to

investigate potential treatments and medications whose main goal is to block and close the receptor ACE2 and prevent it from binding to the S protein, thus inhibiting the fusion of unit S2 with the membrane and preventing infection. Also, when we talk about the immune system, and predict how the body develops an immunity response against the viral genome.

Once the SARS-CoV-2 virus enters the cell, the innate and adaptive immune response is generated (IgM continues for 2-5 weeks, and IgG for 3-7 weeks) this incites a storm of cytokines and chemokines to be released such as tumor necrosis factor-alpha (TNF-α), interleukin-6 (IL-6), interleukin-8 (IL-8), and interleukin-10 (IL-10), then adaptive immunity against Covid-19, but soon a storm of inflammatory cytokines begins to work and causes damage to the cell and failure of the respiratory tissues. What governs these treatments is the participation of many proteins in the fusion process of the protein complex S protein and the receptor ACE2. [65]

chemical drugs: The initial treatments depend on inhibiting spik-ACE2 interaction like the use of chemotherapy drugs such as dexamethasone, remdesivir, hydroxychloroquine, lopinavir, and Ritonavir which had no effective effect on hospitalized patients' these drugs did not achieve the desired benefit in influencing the virus and preventing infection. It was necessary to search for other solutions. [65]
Aside from protein S, there is a protein ORF8 that plays an important role corresponding to the role of INL- 1. It secures the immune environment for virus replication by disrupting the work of macrophages, B cells, CD8+, and killer expression cells. This makes inhibiting ORF8 and stopping gene expression a therapeutic goal. Rapamycin from Polyketide drugs was used, which gave positive results in MERC-CoV. [66]

monoclonal antibodies (mAbs): The first use of this approach treatment was in 1986 (muromonab-CD3) approved by the US Food and Drug Administration (FDA) and continued until covered many diseases like cancer, inflammatory and autoimmune diseases. [65]

The primary target of neutralizing antibodies is the spike protein. Not only the RBD receptor, or N protein which acts as antigens by having special sites for binding antibodies, but also Host factors or pathogenesis

of COVID-19 a potential targets for monoclonal antibodies like mAb3: a monoclonal against neuropilin-1 (NRP1). [65]

Therapeutic Monoclonal antibodies have only a single target recognizing one epitope making it more effective. Takashita et al. has assessed the antibodies against Omicron which are Bamlanivimab (LY-CoV555), Imdevimab (REGN10987), Casirivimab (REGN10933), Tixagevimab (COV2-2196), Cilgavimab (COV2-2130), and Sotrovimab precursors (S309). [59] while Polyclonal antibodies are considered more resistant and effective against the variants and mutations of SARS-CoV-2 because they bind more than one epitope. Although therapeutic with Cocktail Antibodies that combine two or more monoclonal antibodies, bind multiple antigens like REGN-COV2 combine two neutralizing antibodies (Casirivimab and Imdevimab). [65][50]

Recent years have witnessed the development of monoclonal antibodies as one of the therapeutic against SARS-CoV, MERS-CoV, and SARS-CoV-2 according to the Coronavirus Antibody Database (CoV-AbDab). https://opig.stats.ox.ac.uk/ The number of therapeutic antibodies for the coronaviruses family has reached 12,916 entries until February 2024. That is offered by Oxford Protein Informatics Group. Most antibodies are IgG, while other classes exist such as IgA. [67]

Vaccines: The Corona vaccine represents a turning point in the history of the epidemic because of its amazing speed in production and marketing as soon as the pandemic's release. After developments of clinical and preclinical for 200 candidates, a few reached phase III, and began vaccinated from December 2020 into January 2021. Where WHO releases (mRNA vaccine) Pfizer COVID-19 vaccine, and (Adenovirus vector vaccine) AstraZeneca/Oxford COVID-19 vaccine, Johnson and Johnson COVID-19 vaccine. However, it created widespread controversy in scientific circles about the effectiveness of these vaccines. [68]
in the context of COVID-19, researchers found that after ~6 months post-vaccinated with two doses retained T cell responses (84%-CD4+ and 85% - CD8+) to the SARS-CoV-2 Omicron variant with decreases for memory B cell response (42%). [50]
That is compatible with the study (Martinez and Ooi 2022) showed that

delaying the second dose reduces the number of deaths and allows the production of antibodies. [69]

However, the double-dose- vaccinated people not provide protection against Omicron, still increasing the immunity of population by production of vaccines reduced the severity of SARS-CoV-2 when vaccinated people infected with Omicron. while patients previously infected with COVID-19 have 54–79% protection against the variant in confronting genomic mutations. [70]

Regarding the transmissibility and effectiveness of the vaccines against the variants, the Alpha variant is 50 % more transmissible than the reference strain the Beta variant has 50 % more transmissibility than the Alpha variant but is reduced efficiency by the antibodies, The Gamma variant is 1.7–2.4 times more transmissible than the non-VOCs are, The Delta variant is 40–60% more transmissible than the Alpha one is. Vaccines are less effective Compared with other VOCs Delta variant and Omicron sub-lineages, especially the BA.4 and BA.5 ones, that several studies have confirmed that individuals vaccinated with mRNA-based vaccines elicit a lower immune response against the Omicron strain, but it is still sufficient. [51][59][70][71]

Chapter Two: Practical section

The work in this study is divided into departments including the laboratory procedures from received samples to get ready to input to the sequencer MinION then obtain the consensus sequencing as FASTA format. The second part is to analyze the outputs and process it to obtain and extract results using bioinformatics tools. The third part is to interrupt the results in light of the variants of SARS-CoV-2 and related to neighboring countries

Material and Methods

Sample collection

The 24 Nasopharyngeal Swabs samples 12 females and 12 Males patients were infected with SARS-CoV-2 and accepted in hospitals to receive the necessary treatment. the samples were collected by specialized health teams and preserved in a viral transport media (VTM) until received in the EDL Emerging Disease laboratory which is located in Damascus within central public health laboratories where samples were kept at -70 C° until processing. date of collection for samples took from June, July, and August in 2023 these samples were analyzed using PCR ZY BIO, and Gene-expert, and giving high positivity on PCR with low CT value. Samples were collected from several Syrian governorates (Damascus, Swidaa, Homs, Hama, Lattakia, and Tartos) and the patients were aged between 5 to 75 years. EDL used the MinION MK1C instrument to detect Covid 19 and influenza Recently it has been working on Oxford Nanopore Sequencing to determine the characterization dynamic of the virus in Syria.

In laboratory

1-RNA extraction and reverse transcription (cDNA):

Swab samples are assessed for PCR testing, that was incubating in VTM to maintain the viability of the virus, avoiding contamination until received in the lab. We use the Zybio detection kit (PCR-Fluorescent Probe Method), using quantitative real-time PCR(QPCR) that developed for the detection of SARS-CoV-2 which targeted three SARS-CoV-2 genes: N, RdRp, and E. Only 24 samples of hundreds were successful in giving low CT values. CT values range for the N gene (11 to 25), RdRp gene (12 to 26), and E gene (20-36). Then we move to begin sequencing. After we obtained RNA from SARS-CoV-2 samples turned to reverse transcribed cDNA Synthesis by adding 2ml of Luna Script RT SuperMix (LS-RT) to 8ml of sample RNA, mixing and pipetting 10 times, placing in thermal cycle in (PCR Applied 7500 instrument) using the following program:

CYCLE STEP	TEMP	TIME	CYCLE
Primer annealing	25 C°	2 minute	1
cDNA Synthesis	55 C°	20 minute	
Heat Inactivation	95 C°	1 minute	
Cooling	8 C°	Forever	

Table 1 PCR cycle cDNA synthesis

2-PCR (whole genome amplification):

The whole genome of the SARS-CoV-2 virus is about ~30 kp this constitutes an obstacle to sequencing the entire genome at once, so the genome is broken into small fragments when these fragments are sequenced a small part of the genome is created. The most popular method used in sequencing the genome of SARS-CoV-2 is ARTIC Network v3 protocol release on 98 overlapping. we use PCR (Rotor-Gene Q-real-time PCR) tailing protocol of SARS-CoV-2 based on Midnight primer scheme (RT PCR) Expansion (EXP-MRT001). To get full genome coverage and avoid overlapping primers in the same PCR reaction, we apply two PCR master mix reactions and split Midnight primers into two MIXs pools A and B for each sample to be sequenced, one PCR reaction contains thirty primers that generate the odd-numbered amplicons (pool1-A) and the second PCR reaction contains twenty-eight primers to generate the even-numbered amplicons (pool2-B) with amplicon lengths. For 24 samples, each pool contains 92.5 µl of nuclease-free water, 1.25 µl of Midnight Primer Pool A or Pool B, 2.4 µl of SYBR dye, and 156.25 µl of Q5®HotStartHF2x Master Mix.

Reagent	Pool A	Pool B
Q5®HS Master Mix	156.25	156.25
Midnight Primer Pool A	1.25 µl	–
Midnight Primer Poor B	–	1.25 µl
Nuclease-free water	92.5 µl	92.5 µl
SYBR dye	2.4 µl	2.4 µl

Table 2 PCR amplicons by two master

transfer 10 µl to every mix A and mix B into two new tubes, then add 2.5 µl of cDNA sample to the corresponding well-containing mix A, the

same for Mix B, mix well by pipetting. and place samples in a thermal cycler on the following program:

STEP	TEMP	TIME	CYCLES
Initial denaturation	98 C°	30 second	1
Denaturation	98 C°	15 second	35
Primer annealing	61C°	2 minute	
Extension	65 C°	3 minute	
Melting carve	61-90 C°		1
Cooling	8 C°	30 minute	1

Table 3 PCR cycle after preparing pools and primers

3-Library preparation and sequencing by MinION MK1C:

To form the genomic library we transferred the whole volume of PCR PoolA and Pool B to new wells, mixed 10 times by pipetting, and prepared a clean PCR plate containing nuclease-free water, for Rapid Barcode addition, in a plate 96 well, we added 2.5 µl of RNAse-free water to each well and transfer 5 µl of PCR product to the corresponding well, then add Rapid Barcoding kit 96 (SQK-RBK110.96). Each barcode from the 96-well Barcode plate is added to a corresponding well. Taking into account the change tip after each addition, mix by pipetting and incubate in a thermal cycler at 30 C° for 2 minutes and then at 80 C for 2 minutes. In this way, we give identity for each sample.

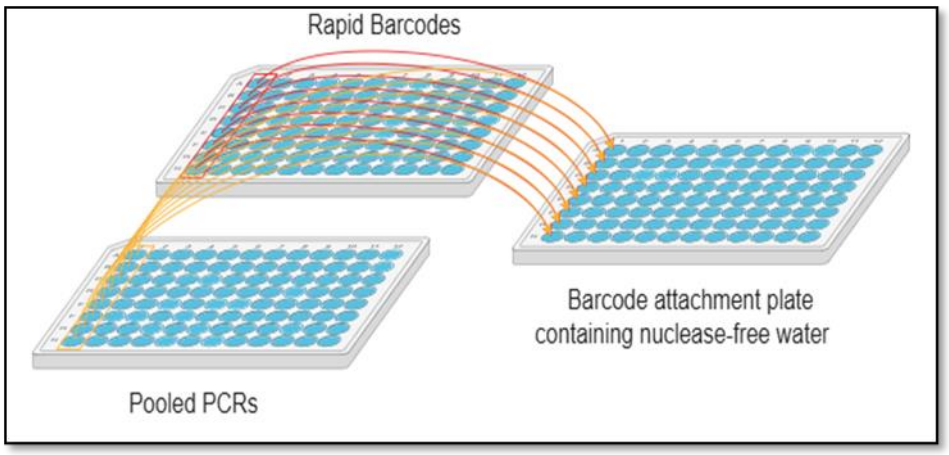


Figure 9 Barcode attachment for the whole SARS-CoV-2 genome

Then combine all resulting barcoded samples in a clean 1.5 ml Eppendorf DNA tube, and we have obtained the whole genome library.

To examine the quality of the library before running samples we use Thermo Fisher Scientific Qubit Fluorometer, ds DNA HS Assay Kit Q3285H (Invitrogen) to measure the fluorometric of DNA concentration, for standard take 190µl Buffer with 10µl reagents, then take 199 µl from it with 1µl from sample, we adjust the concentration of the DNA to be 100 ng/µl by dilution 1:10.

4-Clean-up:

After combining the whole volume of all barcoded samples in one tube, purified using AMPure XP Beads (AXP, SPRI), we add an equal volume from it to the half-pooled barcode samples and preserve the other half at -20 C°, mix by flicking or pipetting then incubate at room temperature for 10 minutes. On a magnetic rack wash the beads that attract the DNA by using 80% ethanol, after that remove the supernatant. Repeat this step once again then collect the DNA library by adding elution buffer to the beads., Add 15 µl from Elution Buffer (EB) incubate for 10 minutes then obtain 15 µl of the DNA library.

5-Rapid adapter addition:

Take 800 ng of the library and make the volume to 11 µl with elution buffer and then we add 1 µl of rapid adapter RBK004 Kit) to 11 µl of barcoded sample. The role of the adapter is like transposase to attach barcodes and Motor proteins to DNA molecules (and incubate for 5 minutes for the T4 DNA ligase enzyme to attach the first part of the adapter to the ends of the fragmented DNA molecules. Store the library in ice until loading on the flow cell (FLO-MIN106D) type (R9.4.1).

6-Flow cell priming and loading:

By using Flow cell washing Kit EXP_WSH004, add 30 µl Flush Tether (FLT) to 1170 µl of Flush Buffer (FB), first put MinION flow cell (FLO-MIN106D) at room temperature for 10 minutes. open the MinION device, and insert the flow cell into the device under the clip. We checked the procedure of active pores to run the flow cell and that was 1264 active

pores which is a sufficient number, we made sure there were no air bubbles. We load the 800 µl flush mix we prepared before to the flow cell via the priming port avoiding the introduction of air bubbles, wait for 5 minutes, and prepare the library for loading in the new tube as follows:

Component	Volume in µl
DNA library	12
Sequencing Buffer II (SBII)	37.5
Loading Beads II (LBII) mixed by vortexing immediately before use	25.5

Table 4 library preparation for loading into the flow cell

In MinKNOW v21.10.9: after the exam of the pores, we put our parameters, name output folder for our results then chose our kit Rapid barcoding Kit 96 SQK-RPK110-96 with Midnight RT PCR primers 1200 amplicon (EXP_MRT001) (Artic Network). Run options for 72 hours as default, choose high accuracy and barcoding enabled to adjust the Phred value score > 7 which indicates a very good sequencing run. Any read above 7 is passed to the Fastq folder and any read less than 7 is sent to fail Fastq folder. Switch on the option Barcode trim also performed through Guppy barcoder integrated into the MinKNOW software, and the option requires barcodes on both ends Add 200 µl of the priming mix into the flow cell priming port, then Load 75 µl of the prepared library to the flow cell via SpotON sample port while avoiding Bubbles entering the array, close SpotON then press start bottom on the screen. The run terminated after achieving sufficient data. Sequencing began when a strand of DNA through across nanopore protein changed the electricity of the surface of the array Guppy bascaller receives this signal and converts it to letters, and demultiplexing completes the process of sorting these letters by their barcodes so we obtain FAST5 raw data files which are a container file using HDF5 format contain base-calling information and modified-base detection, generated during run. Start computational pipeline analysis using FASTQ files format second step is to check the quality reads, mapping the reads to the SARS-CoV-2 reference downloaded from GISAID to obtain a consensus sequence for each barcode the device. We evaluate the sequence data generated by MinKnow, using tools for each step. MinKnow work depends on two steps:

Base-calling (multiplexed)

Basecalling is a major step in nanopore sequencing technology based on translating the raw current measurement into bases (A, T, G, C). Guppy V5.0.17 is a neural network-based bascaller that applies a filter of low-quality reads, trim adapters, and barcodes. Estimating methylation probabilities per base. The biggest challenge for ONT is accurate mapping and base modification of the raw voltage signals with to bases to generate a sequence. ONT developed base callers which were implemented to the available software packages such as Albacore and Guppy which is (the official ONT) continued now. [10]
The result of MinION is raw data in the FAST5 file. The visualized voltage signal of nanopore raw reads is commonly called (squiggle). Bascalling is process of identifying the bases (A, T, G, C) sequences based on translating the raw current signals generating DNA reads using algorithms based on Hidden Markov Model (HMM)-based methods. This operation is done during the sequencing process or offline with the sequencer Bascaller (Guppy v.5.0.17) which is a part of MinKNOW software because of it is superiority in accuracy and speed. Guppy translates these signals to decode the sequences into Fastq files and identity barcodes in the sequence, in our operation we choose a High accuracy base calling model of 1200 bp long amplicons with a quality score for reads above 7. The resulting Reads aligned to the reference genome WIV04 using Minimap2. Need GPU to enhance Guppy run.

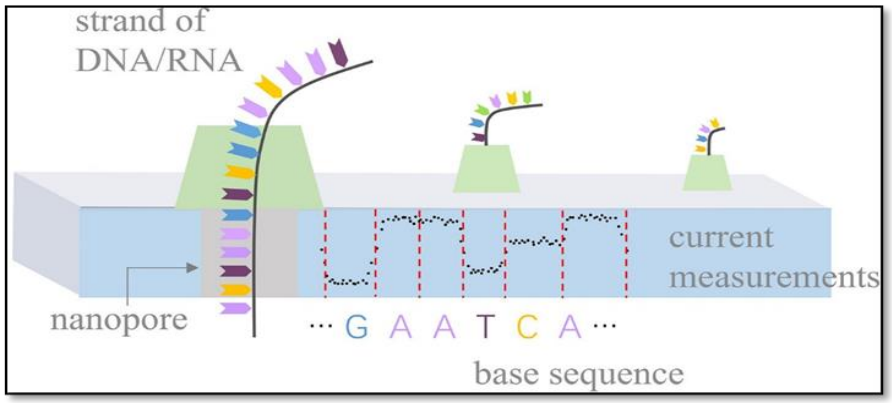


Figure 10 Principle of nanopore sequencing [11]

Demultiplexing (debarcoding)

After the Guppy stores the primary sequencing data as a FAST 5 file it is concatenated into a single file per basically sequence, want to classify sequences pass to their barcodes by demultiplexing or debarcoding the reads into individual FASTQ files (one per sample) based on the barcode sequences

The result of Bascalling (at the instrument) storing the information of sequence reads in binary hierarchical files with groups, datasets, and attributes is FAST 5 (a variant of HDF5 files) (1 read= 1 single read FAST 5) enable demultiplexing which can render up to 20 % of the reads unusable due to base calling errors [2] After demultiplexing, the sequence read is processed by ARTIK Field bioinformatics software to analyze FASTQ files

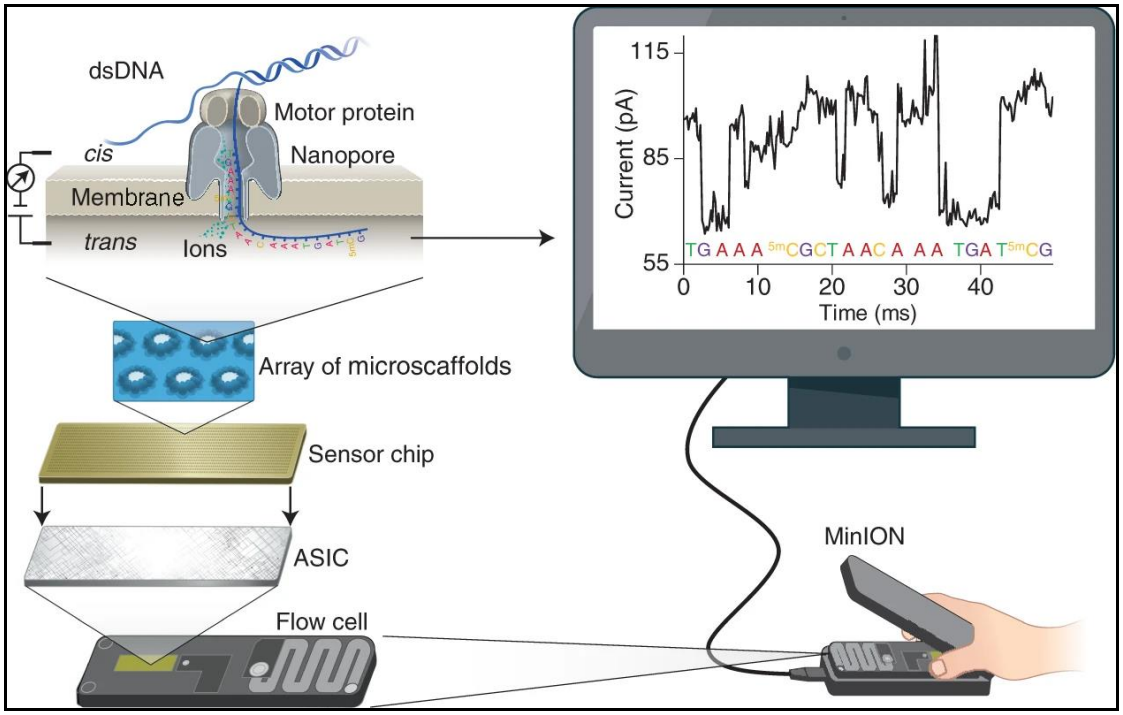


Figure 11 Principle of nanopore sequencing [10]

Chapter three: Bioinformatics analysis and software

Next-generation sequencing was conducted by using Oxford Nanopore Sequencing MinION and analysis data using many bioinformatics tools That require Miniconda3, Python3.11 23.5.2 for Linux-64 applied on VMware Workstation 17 Player. Download data from the instrument on the desk, then check quality and Mapping sequences to extract consensus sequences, for assembly and alignment, we use minimap2 with the reference genome from the GISAID accession number (hcov-19/Wuhan/WIV04/2019(EPI_ISL_402124), followed by SAMtool software for sorting sequences and MEDAKA for correction sequences. Analyze the amino acids changes using the coronaapp annotator to analyze the data, three platforms (Pangolin, Nextclade, and GISAID) were used to identify genomic lineages, variants, and clades)

1-Mapping (Assembly/Alignment) by Minimap2

There are two main methods for assembling the consensus sequence from the raw data, reference mapping and De novo assembly. Mapping reads aligns the reads against the reference genome. Past ONT used GraphMap aligner was developed in 2016 refines candidate alignment to handle high error rates and use the fast graph to align long reads in a short time to enhance the aligner push ONT to develop a minimap2 aligner which is faster and more accurate from Graph Map increase read length beyond 100kb. [10] Minimap2 is a fast sequence mapping and alignment program, that is the first RNA-seq aligner specifically designed for long sequence reads and assembly contigs. Produced by single Molecule Real-Time (SMART) sequencing technology and Oxford Nanopore Technology (ONT) that produces long reads of about 10 Kbp at an error rate of ~15%, in a few minutes to generate a minimizer index for the reference in a hash table before mapping, then for each query sequence, minimap2 takes query minimizers as seeds, find exact matches to the

reference, identifies sets of colinear anchors as chains. [72]
Finding overlap between long reads. Mapping DNA sequence against the reference for our data, we align MinION reads to the full genome using Minimap2 which is A pairwise aligner program that uses sequences in FASTQ format to map DNA against the reference hcoV-19/Wuhan/WIV04/2019(EPI_ISL_402124). the output is in BAM file format for every sample.

2-SAMtool: this is widely used in genomic workflows for post-processing alignment data from nanopore sequencing. It aids in the manipulation and analysis of data in the sequence alignment/Map (SAM) format and provides many functions including indexing sorting merging sorted files, we obtain index tabs files that reflect the Depth of bases and how many times they sequenced, improving the number and distribution of mapped reads for each barcode show the length of genome per each barcode sample comparison to the whole genome of SARS-CoV-2. This indexing offers facilities for downstream analysis and interpretation of the data. During alignment, a quality score identification for each base call depends on two trends: **Depth**: is the amount of times a base within a genome has been sequenced so the greater depth the greater confidence in the sequenced base, and **Coverage**: detects the percentage of the whole genome that has been sequenced we obtained Average sequencing Depth using Samtool implemented in the signify. Then we use iVar

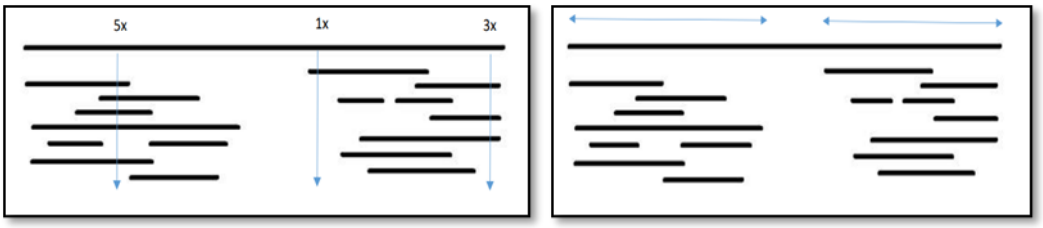


Figure 12 Depth and Coverage

3-iVar: Software is a computational package for viral amplicon-based sequencing for draft consensus contains functions from multiple tools that are required to obtain consensus sequences from data and then apply

4-MEDAKA: is a bioinformatic tool developed by ONT to create high-quality consensus sequences and variant calls from our data and used for polishing which is a process to correct assemblies by correcting

the homopolymer-rich parts of the genome and increase base calling accuracy using deep learning algorithms. Medaka run needs SAMtool and minimap2 to work, and basically in FASTA or FASTQ format to improve the assembly quality.

Downstream analysis

The first step is to check the quality of sequences by FASTQC software V.0.12.1. After obtaining our fasta files, we need to align our sequences with SARS-CoV-2 reference genome hcoV19/Wuhan/WIV04/2019 (EPI_ISL_402124) using molecular evolutionary genetic analysis (MEGA 11). Which is a program that collects a group of methods and tools of comprehensive tools for building time trees of species, pathogens, and gene families using rapid relaxed-clock methods. Also, MEGA estimate evolutionary distance-based offers maximum parsimony methods for molecular phylogenetic analysis allowing to selection of the best-fit substitution models it allows researchers to investigate relationships among different sequences, construct a phylogenetic tree, and predict ancestral sequences and mutations.

Annotations:

Annotation is a process whereby describing the variants in the genome in terms of the changes occurring and their locations in the genome. Assemblies are still just A, T, G, and C, to get descriptive information about our sequences, understand gene function, and show changes in genetic bases and amino acids identifying these variants like insertions, deletions, and SNP mutations. We performed annotation with the Prokka command software tool v1.14.6, accepted DNA sequences in Fasta format, and produced output results in many types. [73]

Variant calling:

Variants (mutation) for viruses in general as per the WHO definition, changes became significantly different to a previously detected virus or parent strain, and show these differences in samples by comparison with reference genome sequence. That generates types like insertion-deletion (indels), and large structural alterations (insertion, inversions, and translocations).
Mutations may occur naturally during viral replication leading to changes in the virus characteristics. Since 2020, SARS-CoV-2 has been changing and mutating globally, some of these changes have not affected the virus properties. Still, some changes affect these mutations making the virus more resistant to receptor-binding inhibitors and some antibodies. [15]

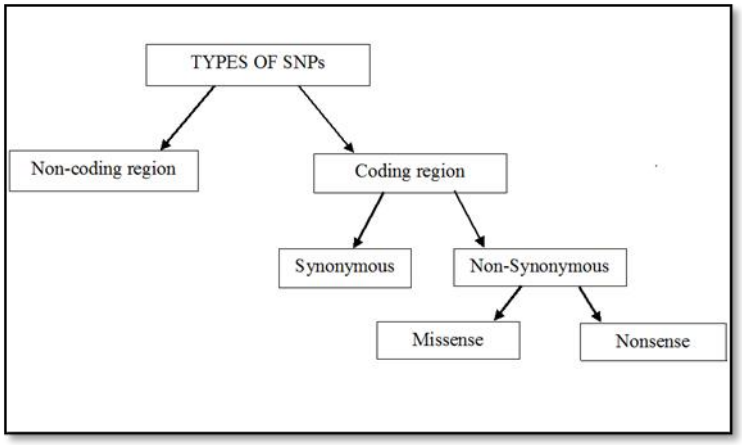


Figure 13 Types of SNP mutations

Variant Calling identified mutations appear like (SNPs), insertions, and deletions (Indels). SNP Single Nucleotide Polymorphism is a source variance in a genome, that occurs at a single position in building blocks of genetic material Adenine(A), Guanine(G), Thymine(T), and Cytosine(C). SNPs in the coding region have two different effects **non-synonymous** that change the protein by causing a change in the codon. And **synonymous** the substitution causes no amino acid change to the protein also called a silent mutation. [74]
Mutation analysis with particular crucial parts of the virus-like S or other structural proteins gives statistically more significant insights rather than considering the complete genome of the SARS-CoV-2 virus. By studying SNPs, we can determine their genetic diversity, and potentially gain insights into variations in virulence, and transmissibility, explain the changes that happened in a wide shape. Non-synonymous mutation

analysis with a particularly crucial part of the virus-like S or other structural proteins gives statistically more significant insights rather than considering the complete genome of the SARS-CoV-2 virus the classification of variants depends on s protein mutations. [23]
Also, The more mutations change permanently, the less effective the vaccines are against protein S, because the antibodies will not recognize the new mutagen site, and this reduces the effectiveness of the vaccines against the Omicron variant evaluating the performance of vaccines and the results of applying therapeutic ways in society. We use **COVID-19 genome Annotator** a web tool that provides a fast, simple tool to annotate the SARS-CoV-2 genome, allowing identity and visualize all mutations in selected proteins using the language R package Google Vis and the shiny server to monitor SARS-CoV-2 mutations. [75]
Export many results including figures, diagrams, and tables of mutations and their positions. It is useful to extract and predict the effect of these variants on genes with two purposes: real-time tracking of SARS-CoV-2 mutational status and annotation of user-provided viral genomic sequences. It requires a multi-fasta file as input and displays the results as a CSV file. Using this tool allows for annotation of SARS-CoV-2 sequences to identify mutations. The application is available at http://giorgilab.unibo.it/coronannotator/.

For a deeper understanding of the mutations present in the spike protein and to predict the influences of these mutations, we used the **PredictSNP** server a consensus classifier, which includes a set of tools to study the impact of these mutations on the function of S protein using a set of computational tools for prediction effects of amino acid substitution on the function MAPP (Physico-chemical properties and alignment score), PhD-SNP(Support vector machine), PolyPhen-1(Expert sets of empirical rules, PolyPhen-2 (Naïve Bayes classifier, SNAP (Neural network), and SIFT(Alignment score). These tools mainly employ machine learning methods to derive their decision rules based on training datasets of annotated mutations.[76] It is free and available at http://loschmidt.chemi.muni.cz/predictsnp
For visualization, we use **(IGV) Integrative Genomics Viewer v2.16.0** a high-performance tool for visual exploration of genetic data,

especially provided by Next Generation Sequencing (NGS) data, making it easy to use aligning the reads mapped by the BAM file format [77] providing a clear vision for variants specifically highlighting and shading mismatched bases in individual reads in color expression their quality and that by comparison with reference genome Wuhan/EPI_ISL_402124. It is commonly used for variant calling mutations detection enabling researchers to study the genomes and interpretation results.

Homology modeling for RBD structure:

S protein is an important target protein for treatment and drugs. The binding affinity of the receptor binding domain (RBD) region in the S protein S1 and human angiotensin-converting enzyme 2 (ACE2) is considered a major determinant of viral infectivity, and the target for most neutralization antibodies. Inhibiting this association leads to important treatments, making the RBD region significant for the development of Drugs and vaccines [78] that might increase the ability of infection and transmissibility or resistance against neutralizing antibodies. bind RBD region resulting in high affinity to monomeric of RBD reference. Homology modeling aims to build a model from sequences with enough accuracy, compatible with experimental analysis, and is considered a guide for experimental testing for discovering drugs. [78] Build a model of the RBD of our samples using the Alpha fold tool a deep learning system by DeepMind that predicts the 3D structure of a protein based on its amino acid sequence. It utilizes a neural network model to generate accurate structural models, which can provide valuable insights for research in biology and medicine.
Then visualize the model on UCSF ChimeraX the next-generation interactive visualization program. [79] From the Biocomputing, Visualization, and Informatics (RBVI), following UCSF ChimeraX includes full user documentation and is free for noncommercial use, with downloads available for Windows, Linux, and macOS.

Lineage assignment and Clades

Because SARS-CoV-2 genomes are being generated continuously at a similar space in the virus they allow continual the process of lineage generation and extension through time. According to WHO Lineage is a group of variants evolved from a common progenitor and are categorized based on the accumulation of specific genetic mutations. For lineage

analysis and knowing the strain of COVID-19 circulating in Syria we can use **PANGOLIN V4.3** (Phylogenetic Assignment of named Global Outbreak Lineages) is a software tool that uses a Python-based assignment pipeline, that accepts one or more query sequences as FASTA files provided by Pango Network https://cov-lineages.org/ which process sequences taking only the coding region (positions 265-29674) trims the remains and mask the **5′, 3′** regions by using N and generates results the minimum length is (10.000 base)and the maximum percentage of N bases in the genome sequence (50 %) criteria. Each query sequence is mapped against an anonymized lineage A genome from Wuhan using minimap2. The output CSV file classifies genetic lineages for SARS-CoV-2, identifying and tracing the different lineages of the virus based on mutations in its genetic material. Pangolin provides us to understand the evolutionary relationships and spread of different variants in the population, and access to actionable information about the pandemic's transmission lineages. [80][81]

Nextclade v3.7.4: A powerful open-source tool for clade assignment, developed by the Nextstrain team https://nextstrain.org/, used for analyzing genetic sequences of SARS-CoV-2 identifying mutations, tracing the evaluation, and understanding how these mutations may impact the virus behavior such as transmissibility or immune evasion. We take the consensus sequences from our population in fasta format and compare them with the reference SARS-CoV-2(Mature proteins) Wuhan-Hu-1/2019 (MN908947). Nextclade provides a tree explaining genetic relatedness between sequences and the important branches integrated in the clade reflect the common specifications between sequences. The first clade named 22A which was the first clade began in 2022.

In a previous study, many lineages including the main variants of concern VOC can identified by spike alone allowing the development of software tools to assign newly generated spike nucleotide sequences to Pango lineage sets. [8]

GISAID database: A platform containing collection sequences of Influenza viruses and coronaviruses, including SARS-CoV-2. The registered number reached over 17 million sequences providing sharing and analyzing genomic data. GISAID database requires its users to log in to the site and accept the terms of the agreement so that the site sends a

message to the previously entered e-mail, allowing the user to access all data available (EpiCoV, EpiFlu, EpiRSV, and EpiPox) databases. EpiCoV database: enables real-time monitoring of SARS-CoV-2 genomic data where the first virus genomes and associated data were shared via GISAID on 10 January 2020. Allow tracking of the distribution of variants and relevance around the world. GISAID provides several analyses to support the efforts in sequence alignment, phylogenetic tree, diagnostic primer and probe coordinates, 3D protein models, and drug targets. [82] Offered many tools like CoVsurver to perform sequence alignment and annotations highlighting phenotypically or epidemiologically interesting candidate amino acid changes with 3D structure mapping. Blast tool finds related sequences to the inputs to provide nucleotide alignment. Also, CoVsurver provides AudacityInstant to find closely related sequences, such as their clade, lineage, location, variant, and collection date. GISAID updates twice weekly.

Phylogenetic analysis

A phylogenetic tree is of great importance in understanding evolutionary relationships and diversity for SARS-CoV-2 between countries identifying and isolating a specific genetic region and knowing the changes and mutations occurring. To study the association of Syrian genomes with the neighboring country's genomes in mid-2023, we use GISAID to download 140 full-length SARS-CoV-2 sequences from many countries neighboring and far from Syria in the Middle East Region. The collection date was from April to September of the 2023 Year to align them with 10 sequences chosen for this study. Align 140 sequences with the reference EPI_ISL_402124 (hcov-19/Wuhan/WIV04/2019) using MAFFT online service version 7 for Multiple Sequence Alignment (MSA). [83] MAFFT An open-source project and powerful analytic tool for aligning nucleotide or amino acid sequences, providing real-time evolution of SARS-CoV-2, and improving outbreak response. It provides a fast and accurate reading using the default method (FFT-NS-2), which improve the accuracy by recalculating the tree using a new distance matrix based on MSA. [83]
MAFFT provides a URL for results to visualize this alignment in Jalview www.jalview.org a platform program for editing, visualization, and

analyzing multiple sequence alignment. use it to align, with functions to edit sequence alignments, calculate the consensus, and remove redundant sequences. The resulting alignment was input in IQ-TREE2 v2.1.2 a free and open-source package for Phylogenetic and evolutionary trees using the maximum likelihood (ML) criterion. [84]

Results and Discussion

All sample sequences used in the study were received in the EDL laboratory analyzing samples, first reverse transcript by Luna Script RT Super Mix Kit for PCR tailing, using Midnight RT PCR Expansion (EXP-MRT001). The criterion used to Evaluate the PCR results depends on the temperature of the melting curve. Samples with peaks on temperature greater than 80 C° can be processed while samples whose melting curve is low are excluded from sequencing, adding Rapid Barcoding Kit 96 (SQK-PBK110.96), adjusting genome concentration by Qubit fluorimeter and dsDNA HS Assay Kit Q3285H device then loading library on the flow cell (FLO-MIN106D) type (R9.4.1). After genome assembly by comparing with reference WIV04 using minimap2, we apply SAMtool to get index FASTQ files for its barcodes, then we use iVar tool and MEDAKA for polishing. The run took a maximum of 24 hours and generated 2.5 million reads with estimated Bases of 2.35 GB data. The proportion of GC content is (37.9%). The Fastqc tool results the quality of for sample was not very good and the number of N is 7627 pb, the length-Max is 29694 pb and the length-Min is 28436 pb. visualization by SAMtool for every sample shows the distribution of reads along genome position and Depth of reads. For example, EPI_ISL_18869841 shows the total reads as 343358.

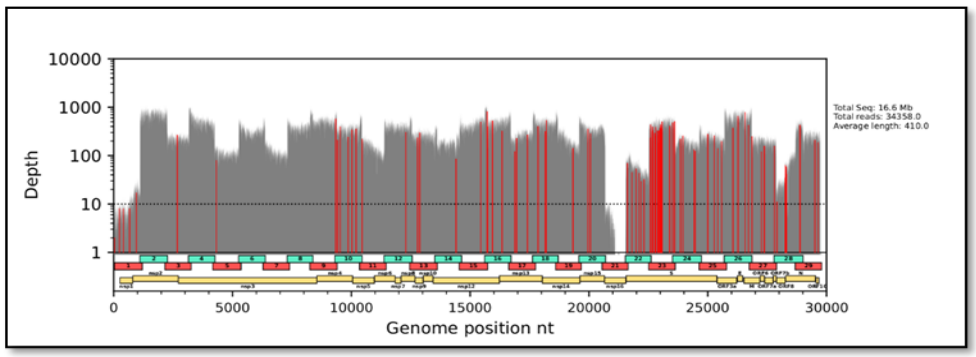


Figure 14 Depth of sample (EPI_ISL_18869841)

Prokka generates 12 outputs in different formats and generates standard output files, containing the protein sequences of the gene and their length in every sample (Replicase polyprotein, Replicase polyprotein, spike, protein 3a, Envelope, membrane, hypothetical protein, Protein 7a, hypothetical protein, Nucleoprotein), in addition to the information and summary about the gene annotated.

We do alignment for sequences in the fasta file with the reference sequence by using MEGA 11 and applying MUSCLE alignment to evaluate the sequences and determine the changes in everyone.

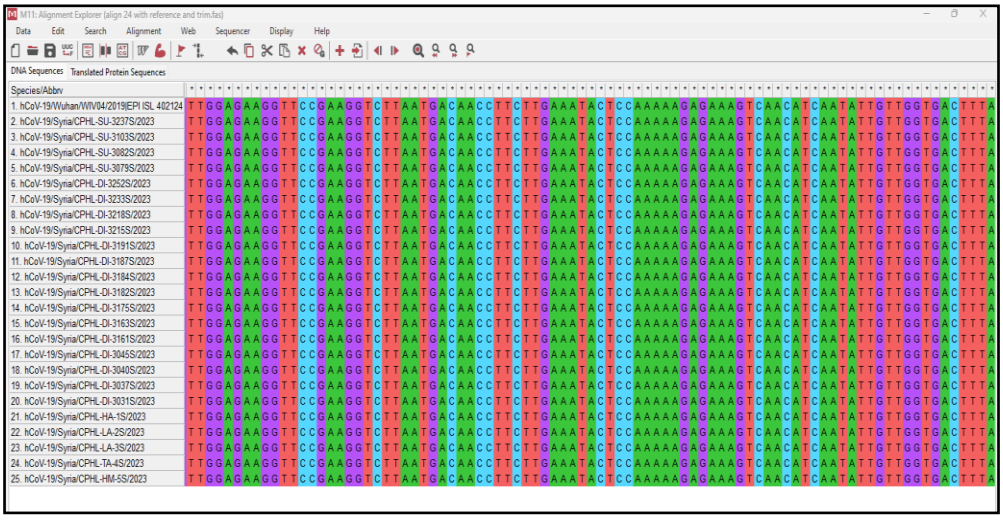


Figure 15 Alignment 24 fasta sequence with reference

Mutation profile

To identify mutations, their locations on the genome, and changes occurring in Syrian sequences by comparing them with the reference positions. We apply **Coronaapp software** to provide us with statistics analysis for samples annotate every mutation and sort its type to show changes in sequences. The results show that the most common changes and substitutions occurred at the nucleotide level which was the most frequent type C >T, A > G, G >A, and C >A.

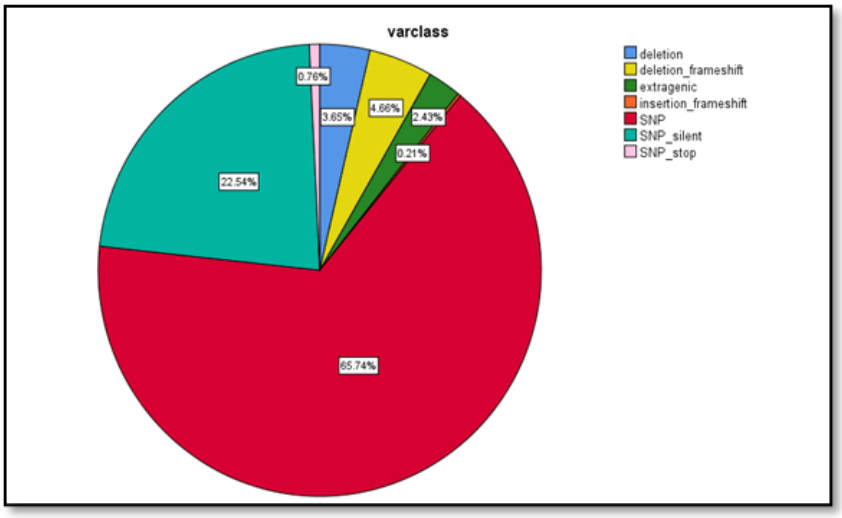


Figure 16 Percentage of mutation classification on 24 samples

The variant (SNP) class obtained the largest percentage in classification achieving 65.7% followed by variant SNP-Silent at 22.5%. afterward, deletion-frameshift 4.7%, deletion 3.7%, Extragenic 2.4%, SNP-stop 0.8%, and last an insertion frameshift mutation of only 0.2%.

Mutation data allow us to observe the evolution of the virus through the circulating in a population. Mutation analysis applied using coronaapp an online web tool to annotate COVID-19 and CoVsurver (v1.22.06) a part of GISAID provides an evolutionary vision of the epidemic through the changes in the amino acids of the virus, as it provides. The analysis determines many amino acid mutations, interrupting results using SPSS V.26 statistics. Mutations distribution take place in these areas in the

genome (5’UTR, NSP1, NSP2, NSP3, NSP4, NSP5, NSP6, NSP7, NSP8, NSP9, NSP10, NSP12b, NSP13, NSP14, NSP15, NSP16, ORF3a, E, N, ORF6, ORF7a, ORF8, 3’UTR, and N). The results of the Omicron variant have shown 232 mutations in all samples. Their frequency varies between sequences.

The majority of mutations occurred in order starting with S spike glycoprotein consisting (41.4%) in all isolates, then NSP3 (6.5%) predicted phosphoesterase papain-like proteinase, his job cutting other viral proteins to free them and removing tags from old and damaged proteins. NSP12b RNA-dependent RNA-polymerase (RdRp) (5.6 %) was interact with NSP7, and NSP8 to facilitate the RNA synthesis machinery for incorporation into viral replication complex. [35] Likewise, NSP4 (5.1%) assists in building double-membrane vesicles for favorable replication inside infected cells. [35] N nucleocapsid (5.1 %) protects the genome RNA and plays a role in the replication and regulation of viral RNA synthesis, transcription, and regulation of infected cell metabolism. [35] The lowest mutations are two mutations in NSP16 and ORF10, one mutation in NSP10, five mutations in ORF7a, and eight mutations in NSP7 and NSP8. On the other hand, deletion mutations consist of 3.7% of all mutations in amino acids found in **NSP6**: [S106-, del/TCTGGTTTT,], **N**: [E31-, del/GAGAACGCA], and **S**: [L18-, del/ TTACAACCA] [L24-, del/ TACCCCCTG], [Y145-, del/ TTA], and [Y144, del/ATTA] occurred only in Latakian sample.

The most frequent nonsynonymous mutations detected in the S protein: (T19I, L24, V83A, T95, G142D, **Y144-**, Y145-, H146Q, Q173K, Q183E, R190, V213E, **G252V**, K310, **G339H**, **R346T**, L368I**, S371F, S373P, S375F**, T376A, D405N, R408S, K417N, **N440K**, V445P, **G446S**, N460K**, S477N**, T478R, **E484A**, F486P, F490S, **Q498R**, N501Y, **Y505H**}, Q613H, D614G, H655Y, **N679K**, P681H, N764K**, D796Y**, **Q954H**, **N969K**, D1146D). These variants are special for the 22F clade. The Bold mutations are the signature of Omicron variants.
Considering other structural protein mutations, the N protein is the second of the structural proteins that has the most frequent mutations (N8N, P13L, E31del, RG203KR, D377D, K369, P67, and S413R).
The other structural proteins E and M proteins are more conserved than N. The E protein has two substitution mutations [T9I, T11A] found in all

isolates while the M protein has five frequent mutations [Q19E, A63T, F112F found in all isolates, K14 occurs in five isolates, and N117N in one isolate.

Nonsynonymous mutations are the most prevalent type across the variants, mostly founded in the S protein. The RBD structure comprised a substructure of the S protein between the 319 re and Gly526 residues of 1273 AA length of S protein and contains 22 most frequent mutations. For accessory proteins, ORF8 contains (7 mutations) and one of them is an insertion-frameshift mutation (D119) that inserts the C nucleotide in position 28250 from the genome.

ORF3a is considered the largest accessory protein.[35] contains 5 mutations in samples. The most frequent proteins (100 %) common in the omicron variant, existing in our samples are as follows: E: T11A, E: T9I, M: A63T, M: F112F, M: Q19E, N: RG203KR, N: S413R, NSP12b: D824D, NSP12b: G662S.

Here we review all the proteins that are subject to substitution mutations in Table (6) and their frequencies. Explain the most and the least protein regions exposed to variants.

Protein	Number SNP nonsenses mutation %	Number of all mutations
NSP1	2 (50%)	4
NSP2	8 (61.5%)	13
NSP3	16 (51.6%)	31
NSP4	4 (57.1%)	7
NSP5	1 (20%)	5
NSP6	2 (33.3%)	6
NSP7	1 (25%)	4
NSP8	0 (0%)	3
NSP9	1 (33.3%)	3
NSP10	0 (0%)	1
NSP12b	2 (20%)	10
NSP13	4 (50%)	8
NSP14	4 (36.3%)	11
NSP15	4 (40%)	10
NSP16	1 (50%)	2
S	51 (71.8%)	71
N	3 (37.5%)	8
E	2 (100%)	2
M	2 (40%)	5
ORF7a	1 (50%)	2
ORF7b	1 (33.3%)	3
ORF3a	1 (20%)	5

ORF6	1 (50%)	2
ORF8	2 (28.5%)	7
ORF10	2 (100%)	2

Table 5 Non-synonymous mutations in SARS-CoV-2 proteins for 24 samples

The Coronaapp annotator tool provided us with diagrams explaining the distributions of every mutation in SARS-CoV-2. Every dot indicates mutation, red dots are aa changes and blue dots are mutations silent, these dots are distributed along the length of the genome, and amount of these mutations in all 24 sequences.

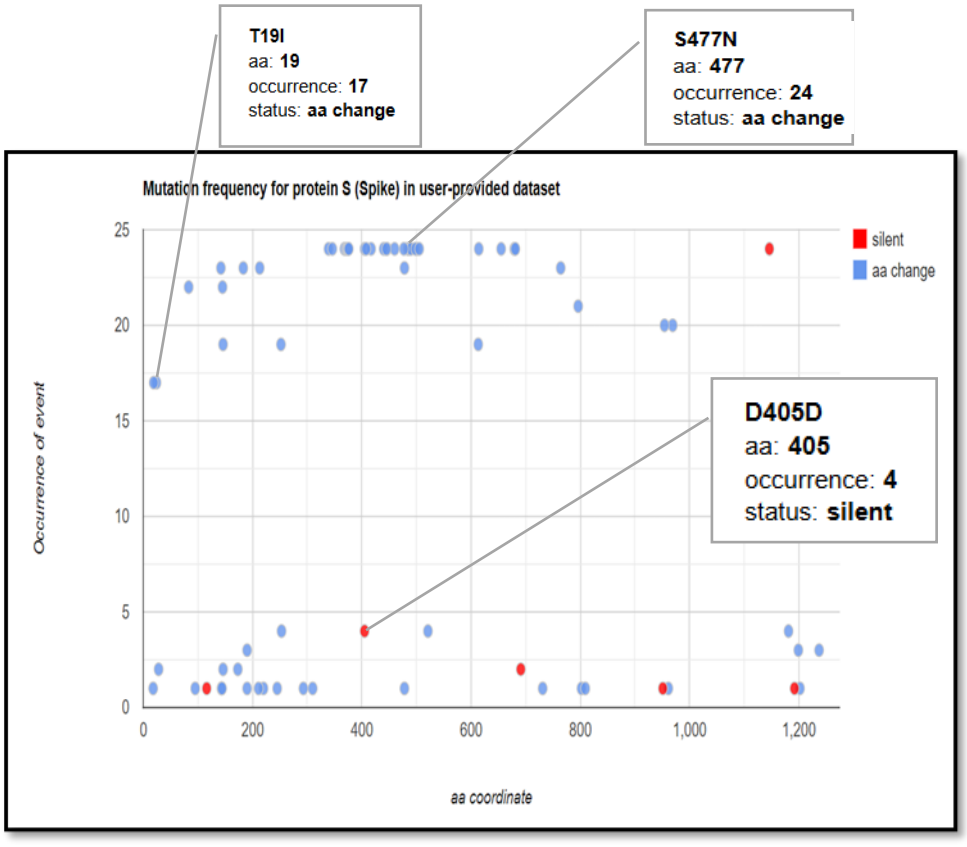
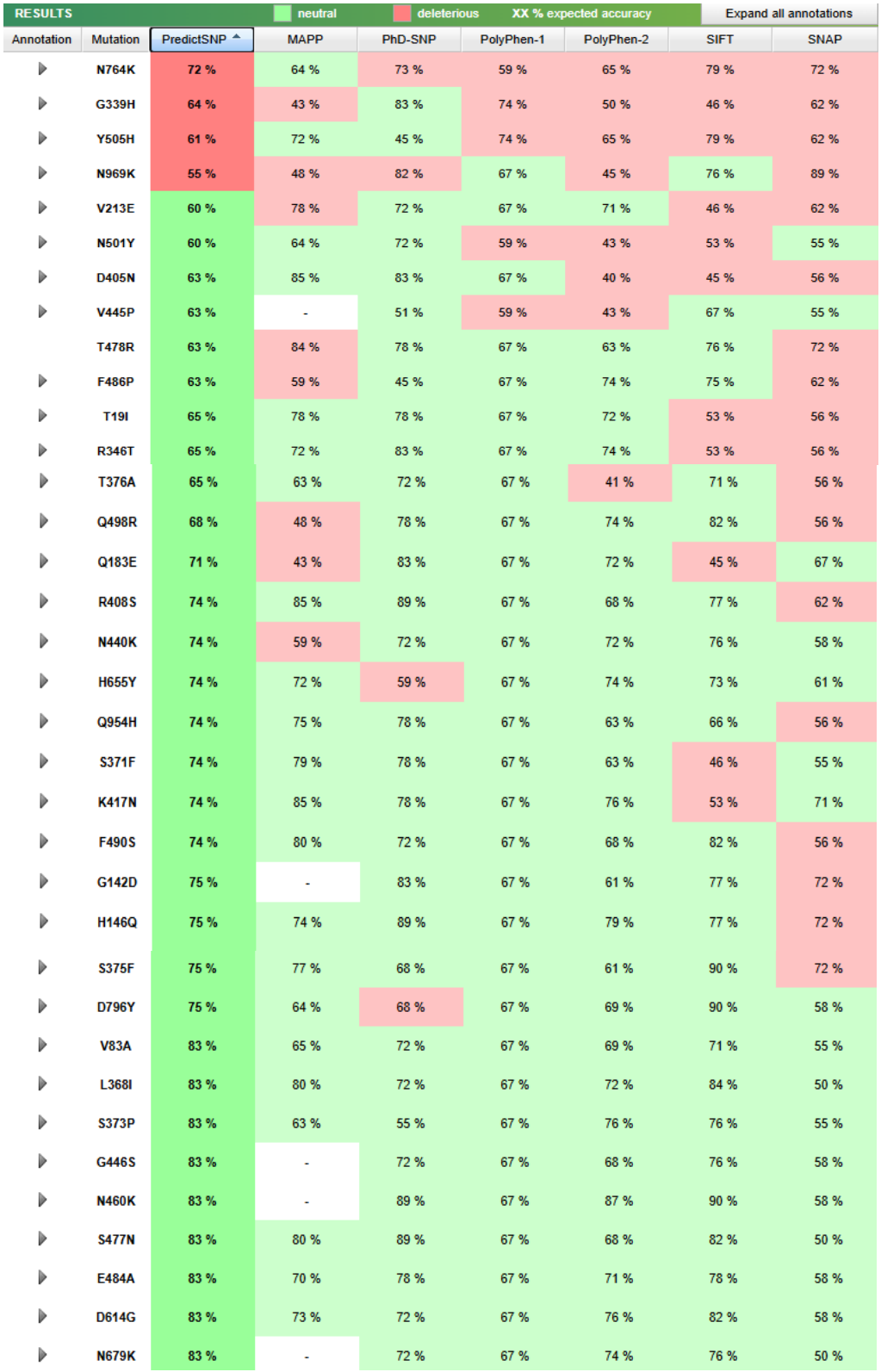


Figure 17 Mutation frequency for structural protein S

To evaluate these SNPs in the Spike protein, and explain the extent of their association with diseases we use the PredectSNP server which predicts SNP severity using six predicted SNP tools: MAPP, PHD-SNP, PolyPhen-1, PolyPhen-2, SIFT, and SNAP. represents an accurate and robust alternative to the predictions delivered by individual tools.
I downloaded the Spike protein of SARS-CoV-2 obtained from the UniProt database with accession number: P0DTC2 in fasta format, submitted the sequences to PredictSNP, and then manually determined mutations of the S protein of our samples.
The results show that among 35point mutations four mutations N764K, G339H, Y505H, and N969K were predicted as the most deleterious by five tools that affect the spike protein function. G339H and Y505H are located in the RBD region. N764K and N969K are in the S2 subunit. These mutations are related to the transmission of the virus and infection. Here it can be noted that all substitutions are Histidine His/H and Lysine Lys/K are positively charged amino acids making it hydrophilic preferring the surface of the protein to its interior. while the remaining mutations are classified as neutral.



RESULTS | neutral | deleterious | XX % expected accuracy | Expand all annotations

Annotation	Mutation	PredictSNP	MAPP	PhD-SNP	PolyPhen-1	PolyPhen-2	SIFT	SNAP
▶	N764K	72 %	64 %	73 %	59 %	65 %	79 %	72 %
▶	G339H	64 %	43 %	83 %	74 %	50 %	46 %	62 %
▶	Y505H	61 %	72 %	45 %	74 %	65 %	79 %	62 %
▶	N969K	55 %	48 %	82 %	67 %	45 %	76 %	89 %
▶	V213E	60 %	78 %	72 %	67 %	71 %	46 %	62 %
▶	N501Y	60 %	64 %	72 %	59 %	43 %	53 %	55 %
▶	D405N	63 %	85 %	83 %	67 %	40 %	45 %	56 %
▶	V445P	63 %	-	51 %	59 %	43 %	67 %	55 %
	T478R	63 %	84 %	78 %	67 %	63 %	76 %	72 %
▶	F486P	63 %	59 %	45 %	67 %	74 %	75 %	62 %
▶	T19I	65 %	78 %	78 %	67 %	72 %	53 %	56 %
▶	R346T	65 %	72 %	83 %	67 %	74 %	53 %	56 %
▶	T376A	65 %	63 %	72 %	67 %	41 %	71 %	56 %
▶	Q498R	68 %	48 %	78 %	67 %	74 %	82 %	56 %
▶	Q183E	71 %	43 %	83 %	67 %	72 %	45 %	67 %
▶	R408S	74 %	85 %	89 %	67 %	68 %	77 %	62 %
▶	N440K	74 %	59 %	72 %	67 %	72 %	76 %	58 %
▶	H655Y	74 %	72 %	59 %	67 %	74 %	73 %	61 %
▶	Q954H	74 %	75 %	78 %	67 %	63 %	66 %	56 %
▶	S371F	74 %	79 %	78 %	67 %	63 %	46 %	55 %
▶	K417N	74 %	85 %	78 %	67 %	76 %	53 %	71 %
▶	F490S	74 %	80 %	72 %	67 %	68 %	82 %	56 %
▶	G142D	75 %	-	83 %	67 %	61 %	77 %	72 %
▶	H146Q	75 %	74 %	89 %	67 %	79 %	77 %	72 %
▶	S375F	75 %	77 %	68 %	67 %	61 %	90 %	72 %
▶	D796Y	75 %	64 %	68 %	67 %	69 %	90 %	58 %
▶	V83A	83 %	65 %	72 %	67 %	69 %	71 %	55 %
▶	L368I	83 %	80 %	72 %	67 %	72 %	84 %	50 %
▶	S373P	83 %	63 %	55 %	67 %	76 %	76 %	55 %
▶	G446S	83 %	-	72 %	67 %	68 %	76 %	58 %
▶	N460K	83 %	-	89 %	67 %	87 %	90 %	58 %
▶	S477N	83 %	80 %	89 %	67 %	68 %	82 %	50 %
▶	E484A	83 %	70 %	78 %	67 %	71 %	78 %	58 %
▶	D614G	83 %	73 %	72 %	67 %	76 %	82 %	58 %
▶	N679K	83 %	-	72 %	67 %	74 %	76 %	50 %

Figure 18 Nonsynonymous amino acid substitutions that affect the function of the Spike protein using PredictSNP server.

IGV software visualization for samples uses bam and bam.bai files as input. For example, determine SNP mutations in the sample hCoV-19/Syria/CPHL-SU-3237S/2023|EPI_ISL_18869836. The coverage area paragraph indicates the number of reads aligned to the reference. The panel Shows the quality of this base and for every mutation explains the percentage of every base according to the reference. In the coverage region, every color indicates a base, compared with the reference WIV04 sequence. For example, the Q498R mutation occurs in the ref position (22055), changing the nucleotide from Adenine in red to Guanine in orange. The bold color refers to the bases are high quality with reference while the Faint color refers to lower quality or mismatches.

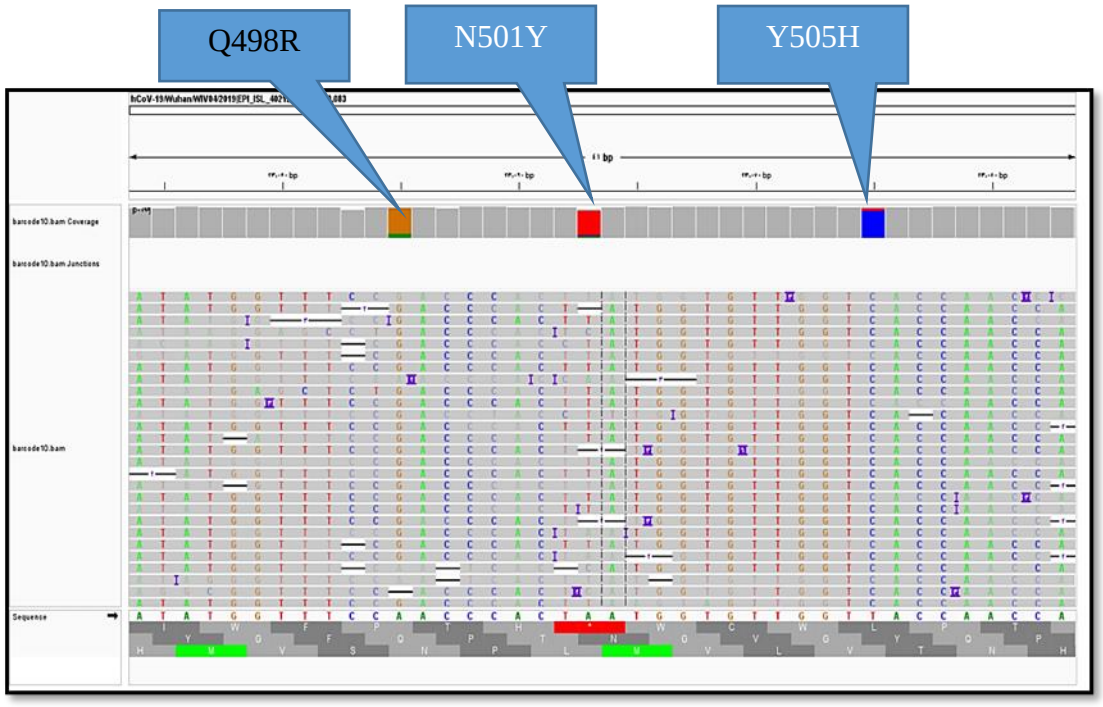


Figure 19 The mutations (Q498R, N501Y, Y505H) in hCoV-19/Syria/CPHL-SU-3237S/2023|EPI_ISL _18234342 by IGV

To study the effectivity of the mutations we apply a model for the RBD region to a building structure pdb because there are no experimental results of X-ray crystallography or NMR for samples, to perform a structure by homology modeling. download the sample from GISAID (CoVserver) with the accession number EPI_ISL_18810712 which has

96.544 % AA identity with the reference hCoV-19/Wuhan/WIV04/2019|EPI ISL 402124.
This sample genome contains 23 mutations in their RBD region without deletion or insertion. the RBD region containing RBM position 430–522 is a subset of the positions in the RBD of the spike amino acid sequence. Determine the RBD structure using MEGA X, I align the S protein structure reference downloaded from the Uniprot database in accession number P0DTC2 with spike protein of the sample using MEGA X, then trim the RBD structure that ranges (319- 541) residue.

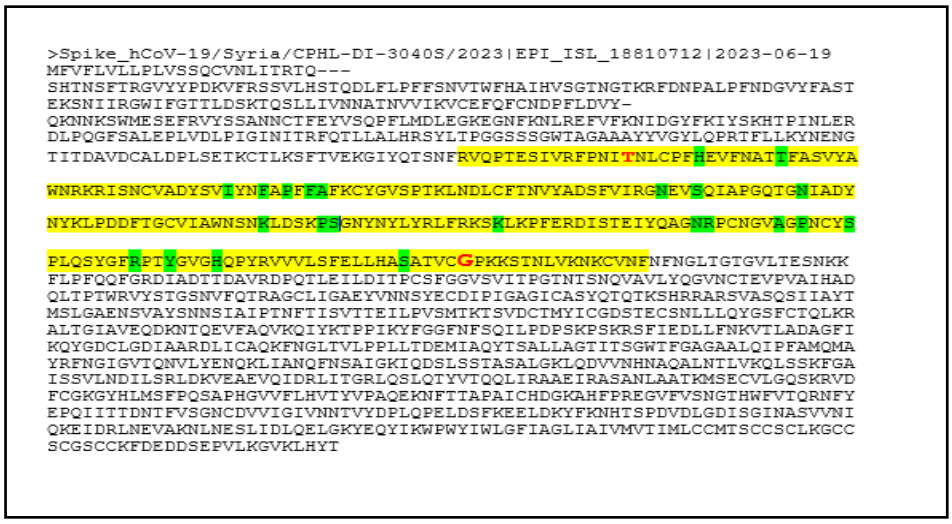


Figure 20 The spike protein sequence for hcov-19/Syria/CPHL-DI-3040/2023 Omicron sample indicates the initiate and the end residue of the RBD region with red color and substitutions with green color using Word Microsoft.

To Study homology, I have used Alpha Fold provided by Google Colab Notebook, allowing access to the Alpha Fold2 codebase from the official GitHub repository. After obtaining the model determine the substitutions of amino acids in the model using the Chimera X program. we align the RBD region of our sample with the wild-type RBD reference a part of (PDB: 6M0J) protein structure and visualize this alignment.
The alignment shows that the value of RMSD (Root Mean Squared Distance) is 1.089 Angstroms indicating better alignment. [85] Vaccines are based on the composition of RBD, which is considered an epitope due to its association with ACE2. Completing molecular docking and molecular dynamics may allow the study of medicines that suit Syrian society and support the study.

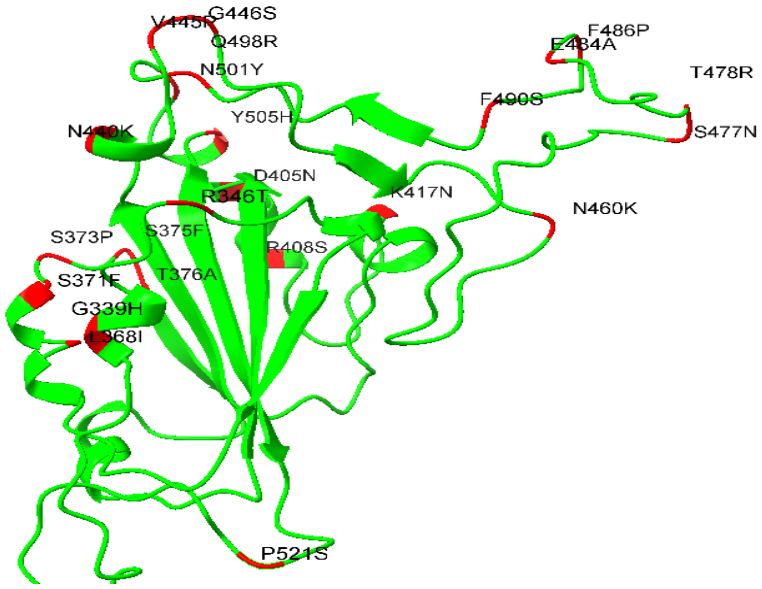


Figure 21 3D homology modeling of RBD omicron for Syrian sample (hcov-19/syria/CPHL-DI-3040/2023) by Alpha fold collab visualize mutations by Chimera X v1.8.

Pango results

The Pangolineage website used in this study generates lineages of strain on patients. Pangolin identified four recombinant XBB lineage as follow: XBB.2.3.11(16.67 %) exist in 4 strains, XBB.1.9.1(4.17%) in one strain, FL.12(XBB.1.9.1.12) (4.17%) in one strain, and XBB.1.42.1(75%) in 18 strains coverage all districts. All samples sequenced in Pangolin are confirmed to be omicron strains that the lineage XBB.1.42.1 is designed as Syrian Lineage.

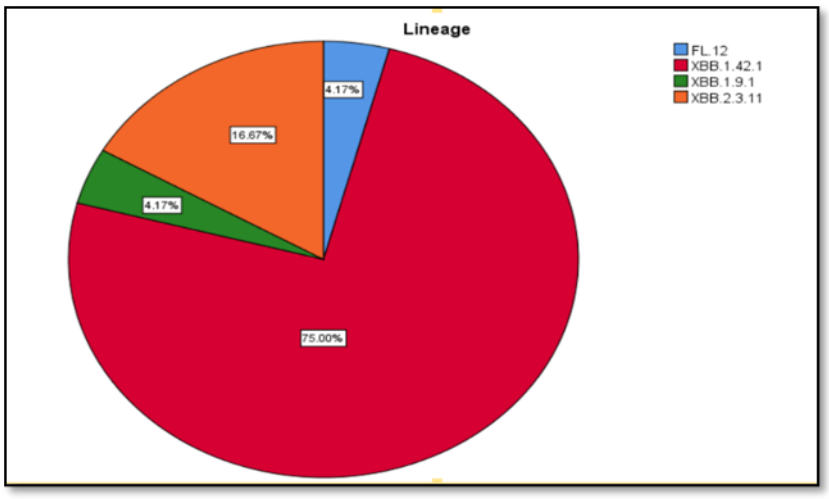


Figure 22 Lineages of SARS-CoV-2detected by Pangolin

According to https://outbreak.info/ XBB.1.42.1 which was detected on 21-04-2023 located mostly in the United States of America at 17.5%, Canada at 11.0%, France at 9.0%, Spain at 9.0%, United Kingdom at 6.0%.
The prevalence rate of XBB. 1.42.1 in Syria is 10% relative to the total number of samples (122) estimated 10% (95% CI: 5%-16%) risk of SARS-CoV-2 followed by Lebanon, Iraq, Ukraine, and Hungary.

Nextclade results:

Upload fasta sequences onto the Nextclade website we choose SARS-CoV-2 protein-mature as a reference to obtain the cluster of samples that classify it in three clades 22F, 23D, 23E, while GISAID has a special classification of clades, defines EPI_ISL_18234343 as GR clade and 18 submitted samples are classified as GRA clade. According to https://outbreak.info/situation-reports and CoVariant https://covariants.org/ enabled by data from GISAID, XBB Omicron (22F) a recombinant variant, was first detected probably in South Asia in mid-2022. 22F clade formed concern because of the role in increasing the transmission and immune evasion of this variant family due to large numbers of mutations in the spike gene including the RBD region and N-terminal domain. During January 2023, the (22F) clade became dominant until decreasing in February 2023.

sample	Clade/ Nextclade	Pango Lineage	WHO Name	Date of exposure in the world	Location of exposure
hCoV-19/Syria/CPHL-HA-1S/2023	22F	XBB 1.42.1	Omicron	7/Jun/ 2023	India USA Indonesia Singapore
hCoV-19/Syria/CPHL-LA-2S/2023	22F	XBB 1.42.1	Omicron		
hCoV-19/Syria/CPHL-LA-3S/2023	22F	XBB 1.42.1	Omicron		
hCoV-19/Syria/CPHL-TA-4S/2023	22F	XBB 1.42.1	Omicron		
hCoV-19/Syria/CPHL-HM-5S/2023	22F	XBB 1.42.1	Omicron		
hCoV-19/Syria/CPHL-SU-3103S/2023|EPI_ISL_18869832	22F	XBB 1.42.1	Omicron		

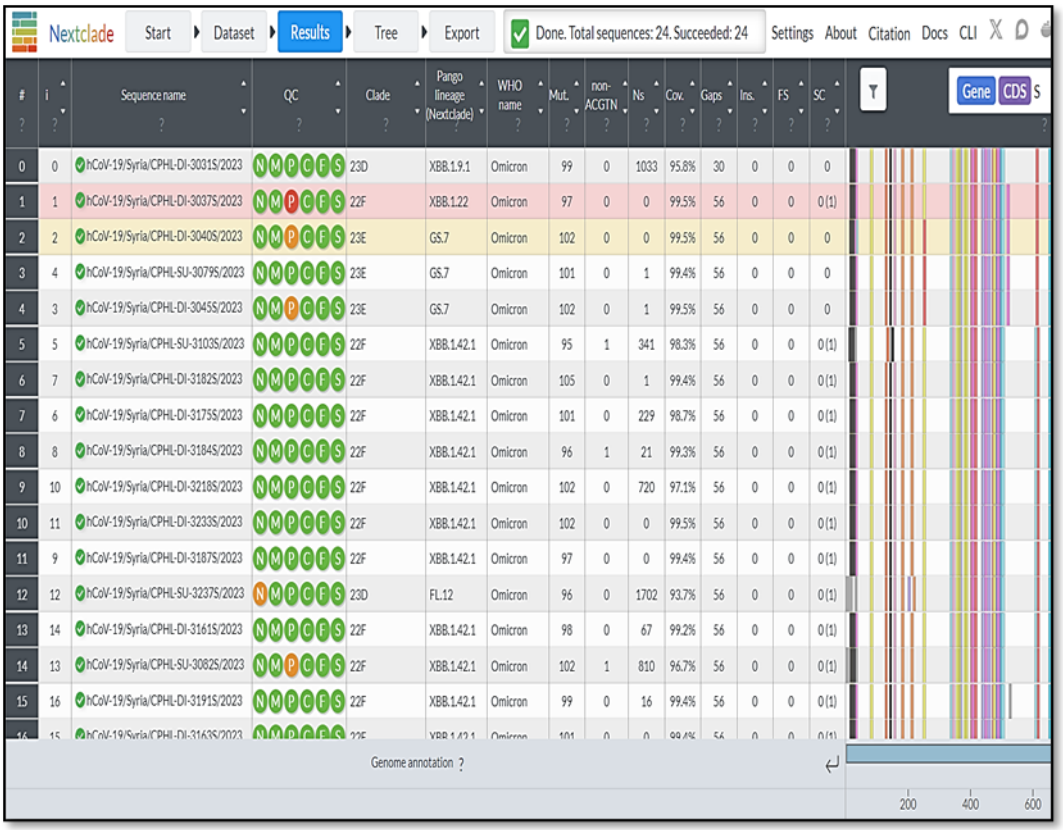
hCoV-19/Syria/CPHL-DI-3182S/2023|EPI_ISL_18234343	22F	XBB 1.42.1	Omicron		
hCoV-19/Syria/CPHL-DI-3175S/2023|EPI_ISL_18869833	22F	XBB 1.42.1	Omicron		
hCoV-19/Syria/CPHL-DI-3184S/2023|EPI_ISL_18869834	22F	XBB 1.42.1	Omicron		
hCoV-19/Syria/CPHL-DI-3218S/2023|EPI_ISL_18869835	22F	XBB 1.42.1	Omicron		
hCoV-19/Syria/CPHL-SU-3082S/2023|EPI_ISL_18869837	22F	XBB 1.42.1	Omicron		
hCoV-19/Syria/CPHL-DI-3161S/2023|EPI_ISL_18869838	22F	XBB 1.42.1	Omicron		
hCoV-19/Syria/CPHL-DI-3163S/2023|EPI_ISL_18869839	22F	XBB 1.42.1	Omicron		
hCoV-19/Syria/CPHL-DI-3215S/2023|EPI_ISL_18869840	22F	XBB 1.42.1	Omicron		
hCoV-19/Syria/CPHL-DI-3187S/2023|EPI_ISL_18869841	22F	XBB 1.42.1	Omicron		
hCoV-19/Syria/CPHL-DI-3191S/2023|EPI_ISL_18869842	22F	XBB 1.42.1	Omicron		
hCoV-19/Syria/CPHL-DI-3252S/2023|EPI_ISL_18869843	22F	XBB 1.42.1	Omicron		
hCoV-19/Syria/CPHL-DI-3233S/2023|EPI_ISL_18946673	22F	XBB 1.42.1	Omicron		
hCoV-19/Syria/CPHL-DI-3031S/2023|EPI_ISL_18869829	23D	XBB.1.9.1	Omicron	11/Dec/2022	Indonesia/France India
how-19/Syria/CPHL-SU-3237S/2023|EPI_ISL_18869836	23D	FL.12 XBB.1.9.1.12	Omicron	18/Feb/2023	
hCoV-19/Syria/CPHL-DI-3037S/2023|EPI_ISL_18234342	23E	XBB.2.3.11 GS.7	Omicron	8/Jan/2023	India/ USA
hCoV-19/Syria/CPHL-DI-3040S/2023|EPI_ISL_18810712	23E	XBB.2.3.11 GS.7	Omicron		
hCoV-19/Syria/CPHL-DI-3045S/2023|EPI_ISL_18869830	23E	XBB.2.3.11 GS.7	Omicron		

hCoV-19/Syria/CPHL-SU-3079S/2023|EPI_ISL_18869831	23E	XBB.2.3.11 GS.7	Omicron		

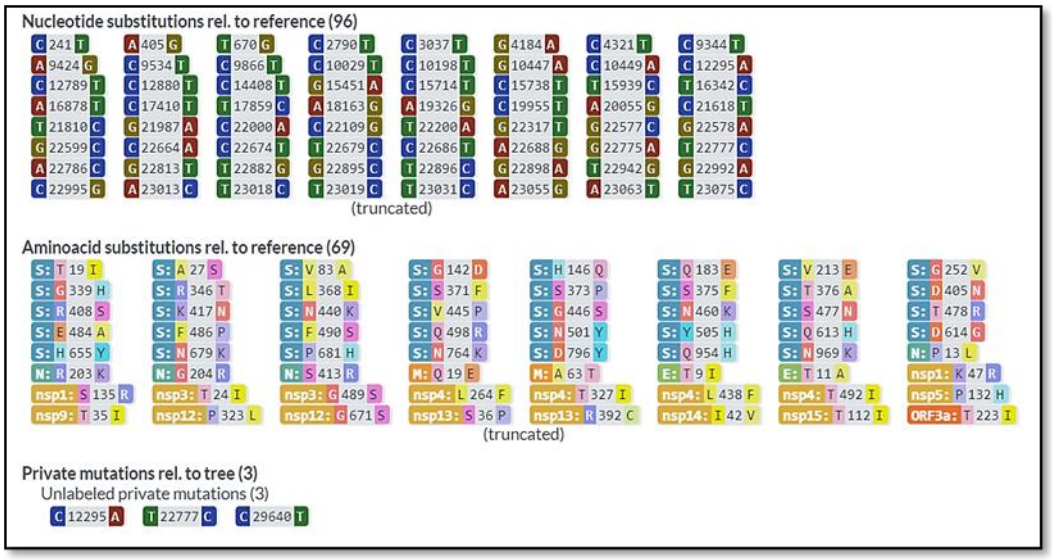
Table 6 Lineages and clusters of samples as provided by Pangolin, Nextclade, and WHO

Nextclade provides a variety of information about sequences and is easy. The interface of the results page provides the substitution in nucleotide and protein to every sequence with information's about gaps, deletion, number of mutations, quality control for regions in the genome, and name of lineage in the WHO and Pangolin. For example, hcov-19/Syria/CPHL-DI-3184/2023 sample we can see mutations in nucleotide and amino acids. In addition to classifying samples in clades visualize in tree explaining genetic relatedness between sequences and the important branches supported by the Next strain as we see in Figure 23.

(A)



(B)



(C)

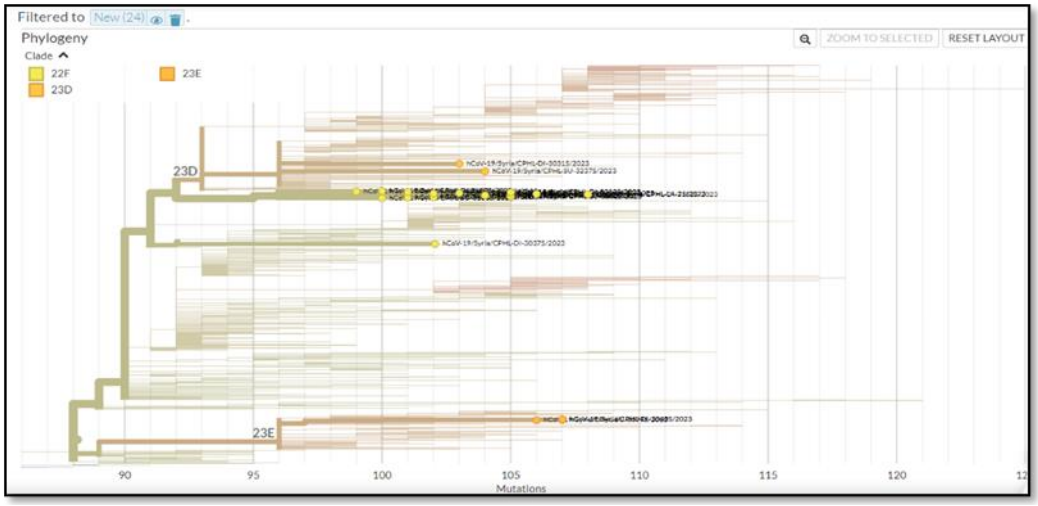


Figure 23 Outputs of Nextclade (A) The interface of Nextclade results of 24 samples. (B) mutation substitution in nucleotide and amino acids of hcov-19/Syria/CPHL-DI-3184/2023 sample. (C) The Nextclade provides phylogenetic Tree show the cluster of 24 samples

GISAID results:

From 24 individual samples, only 19 sample sequences were successfully submitted to GISAID. We fill in the Metadata for every sample that shares a uniform suffix name in all entries (hcov-19/Syria/CPHL-city name + type of sample/2023). It can be obtained by searching the EpiCoV database gives every sample a unique accession number. We can

search by country Syria. Clade definition in GISAID are informed by the statistical distribution of genome distances in phylogenetic clusters.
So 18 Samples were classified as GRA clusters while one sample was classified as a GR cluster.

Accession ID	AA identity %	Collection Date	Length(nt)	Clade	aa changes
EPI_ISL_18946673	96.544	14-08-2023	29690	GRA	44
EPI_ISL_18869843	96.701	17-08-2023	29675	GRA	42
EPI_ISL_18869842	96.544	06-08-2023	29674	GRA	44
EPI_ISL_18869841	96.544	06-08-2023	29675	GRA	44
EPI_ISL_18869840	97.251	09-08-2023	29594	GRA	35
EPI_ISL_18869839	96.465	31-07-2023	29676	GRA	45
EPI_ISL_18869838	96.544	31-07-2023	29676	GRA	44
EPI_ISL_18869837	96.779	26-06-2023	29675	GRA	41
EPI_ISL_18869836	97.172	09-08-2023	29673	GRA	36
EPI_ISL_18869835	96.544	12-08-2023	29687	GRA	44
EPI_ISL_18869834	96.622	06-08-2023	29674	GRA	43
EPI_ISL_18869833	96.544	01-08-2023	29674	GRA	44
EPI_ISL_18869832	96.858	06-07-2023	29674	GRA	40
EPI_ISL_18869831	96.622	26-06-2023	29673	GRA	43
EPI_ISL_18869830	96.622	20-06-2023	29689	GRA	43
EPI_ISL_18869829	96.701	15-06-2023	29659	GRA	42
EPI_ISL_18810712	96.544	19-06-2023	29692	GRA	44
EPI_ISL_18234343	97.879	05-08-2023	28605	GR	27
EPI_ISL_18234342	95.915	19-06-2023	29654	GRA	52

Table 7 Accession numbers and clade provided by GISAID for 19 samples.

We applied the Blast tool (v1.2.2) provided by the GISAID database which found similarities between the input sequences and the sequence submitted in the GISAID database, providing alignment information such as the identity, scores, gaps, coverage, and collection date. For the input (hcov-19/Syria/CPHL-DI-3163/2023|EPI_ISL_18869839) shows the top ten sequences between the query and outputs. We chose ten top alignment outputs, Blast found similarities with 7 sequences from the USA, two sequences from JAPAN, and one sequence from England. with 99.9 % similarity.

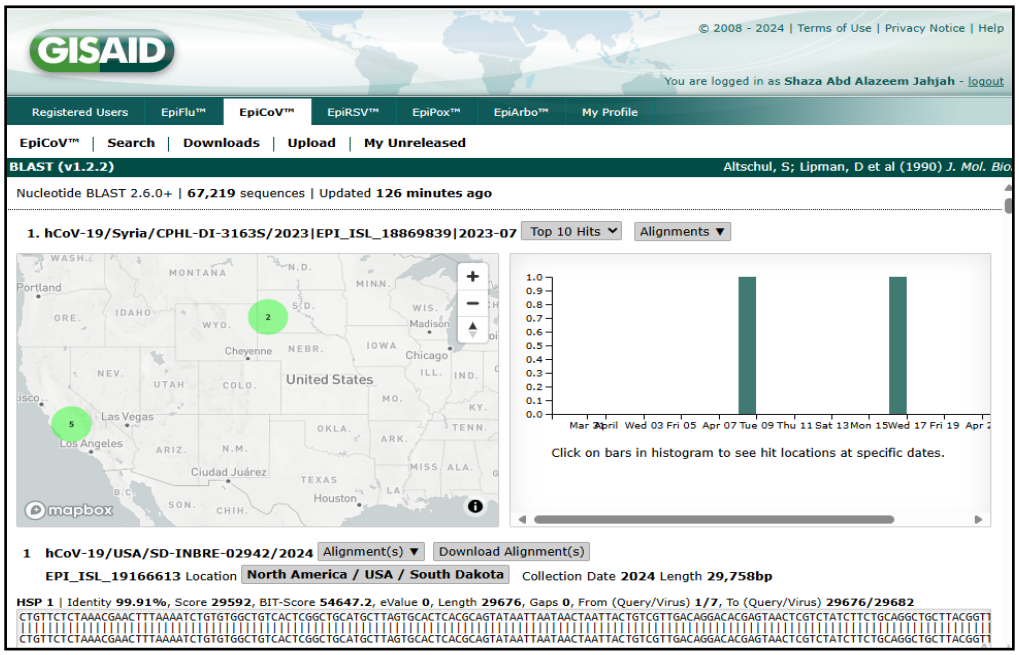


Figure 24 Blast output for a query EPI_ISL_18869839

For phylogenetic analysis of our sequences, I chose 140 sequences from 9 countries, downloaded from the GISAID database and 10 samples from Syria/Damascus (EPI ISL 18234342, EPI ISL 18810712, EPI ISL 18869830, EPI ISL 18869833, EPI ISL 18869838, EPI ISL 18869839, EPI ISL 18869841, EPI ISL 18869842, EPI ISL 18869843, EPI ISL 18946673). These 10 selected sequences from this study from Damascus aligned with 130 sequences from countries in the Middle East (Lebanon, Palestine, Iraq, Jordan, Saud Arabia, Kuwait, Iran, Bahrain, and Turkey).

The date of sample collection was chosen from April to September for all sequences and All was Omicron variant. Then aligned with reference hCoV-19/Wuhan/WIV04/2019|EPI ISL 402124 using MAFF tool provide URL for results to visualize this alignment in Jalview, input results to IQ-TREE 2 using fast model selection via ModelFinder (GTR+F+R3) and ultrafast bootstrap replicates 1000 to generate a Maximum likelihood phylogenetic tree (ML).

Figure 25 Jalview visualization for 140 sequences from GISAID

For visualization use the ITOL phylogenetic tree. the phylogenetic tree of 141 genomes of the Omicron variant with reference shows 7 full-length genomes of this study making a cluster with two sequences from Iran and one sequence from Jordan. these genomes are closely related to a cluster containing two sequences from Jordan. While the other three full-length genomes from this study were among a cluster of 23 leaves of sequences with near sequences from Kuwait and Saudi Arabia and far sequences from Palestine and Bahrain. Notice that these three genomes are closely related to a cluster containing two sequences from Iran and one sequence from Lebanon. While the sequences of Turkey, Bahrain, and Palestine remained evolutionarily far from the sequences in Syria.

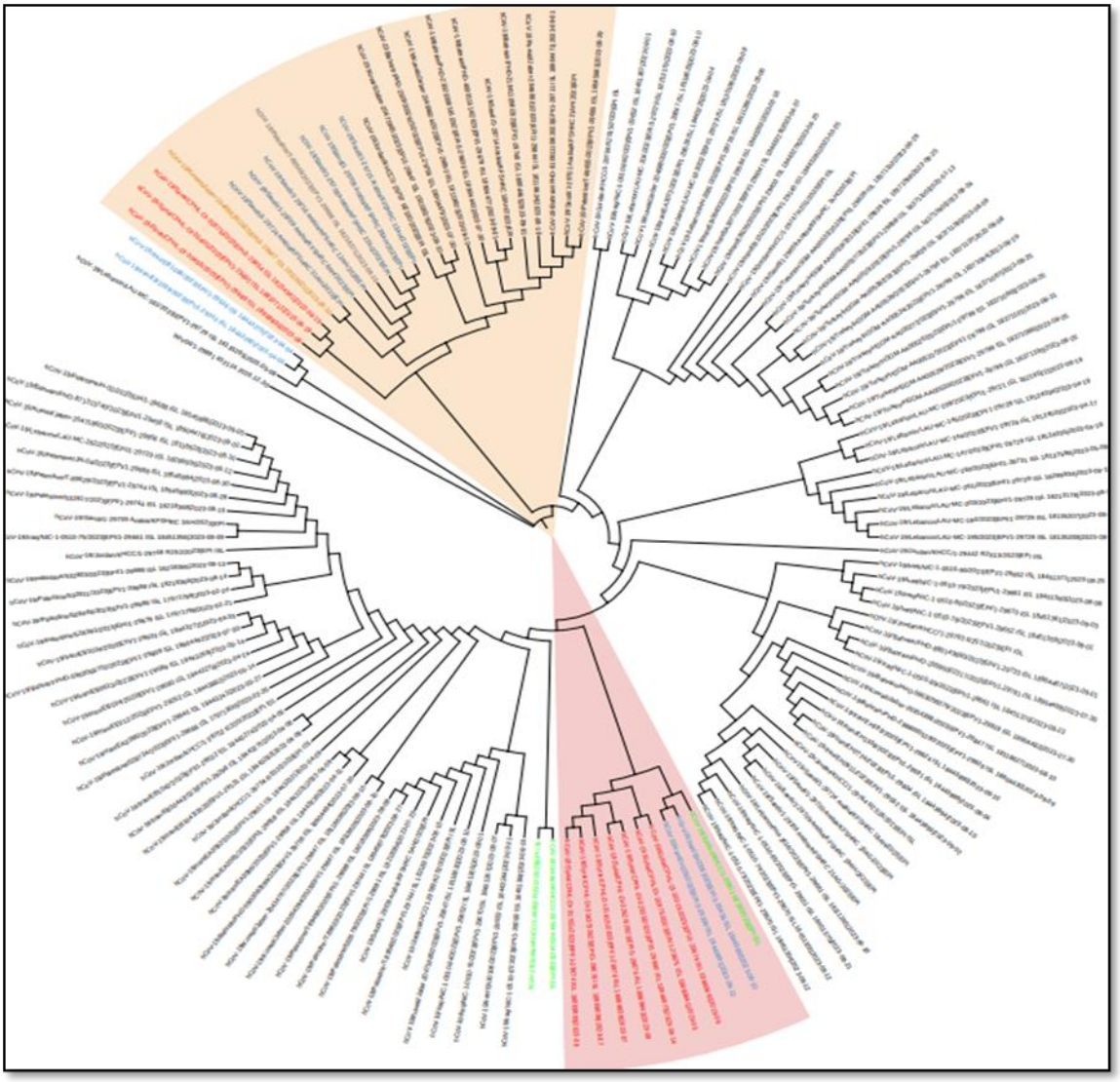


Figure 26 Phylogenetic tree for 10 Syrian samples

Discussion

This study provides a detailed analysis of 24 viral isolations of hospitalized patients with Covid-19. List and characteristics description of samples collected from major cities in Syria using Oxford Nanopore Sequencing MinION which became available in the Emerging Disease laboratory in CPHL mid-2023. Prepare the results and process them through many steps to generate consensus sequences. To identify the mutations we apply Coronaapp software to provide mutations in samples studying their prevalence and effect. Some of them are highly frequent

and others are rare. Noticing that substitution mutations in S protein are the most frequent followed by NSP4 and NSP3. S protein has been the main target of diagnostics and therapeutics. This importance was gained through the presence of the RBD region in the S1 subunit, it is responsible for entering the cell and causing pathogenesis, and that make it the most protein exposure to mutagenesis. [39][28]
PredictSNP results show four mutations considered as deleterious mutations that effect the function of the protein, and have an important effect on the binding affinity for ACE2. Noting the Y505H reduces the binding affinity for ACE2.
Among the distinctive mutations that exist in all isolates of samples that play an important role in the infection, transmissibility, and persistence of the virus, which was added to its distinctive location in the S protein.
K417N (Lysine to Asparagine amino acid) was first reported in South Africa mutation found in Beta, Gamma, and Delta VOCs, located in the S-RBD region, in a specific epitope can escape from neutralization by various monoclonal antibodies mAbs elicited by vaccines Pfizer-BioNtech BNT162b2 and Moderna mRNA-1273. [45] [86]
K417N with the mutations in isolates **S371L, S373P, S375F, G446S, E484A, and Y505H**, decrease the binding affinity of the RBD region with ACE2 in addition to the **Y144-** deletion and **G252V** which decreased pseudovirus infectivity while **N440K, S477N, T478K, N501Y** increase the binding affinity for ACE2, also **V83A** increase pseudovirus infectivity. [53] [52][60]

The **S: N501Y** (asparagine to tyrosine amino acid) was first detected in Alpha, Beta, and Gamma. associated with other mutations namely N439K, H96-/ V70- deletion. N501Y located in the RBD region, appears to play a critical role in the transmission and virulence of SARS-CoV-2, increasing the affinity to the ACE2 that reduces the probability of interaction with antibodies. [78] [87] K417N and N501Y these two mutations are a source of concern cause contribute to immune escape and higher infectivity.
S: S477N: Existing in all strains and major circulating in Omicron VOCs increased the binding affinity with the hACE2 receptor and improved transmissibility. location in an epitope makes it targeted by large variety of monoclonal antibodies. [45]

S477N and Q498R found in our isolates that have H-bonds and salt bridges that give more positive to the RBD surface increasing the interaction with negatively charged ACE2. [86]

E484A: a new substitution in Omicron instead of the previous mutation E484K that was found in the Beta and Gamma variants, founded in all strains. It plays a role in changing the interaction between the RBD and the ACE2 and it is more involved in escaping from 2B04 and 1B07 monoclonal antibodies. [59]

S: D614G (from Aspartic acid 614 to GLY) is the most prominent missense mutation found located in the C-terminal region of the S1 domain of spike protein. Shown in all waves of a pandemic, It was detected in late 2020 in Germany- China found that all our sequences play a significant role in the transmission as the instability of D416 impacts the integrity of the local membrane and post-translational modification of the S protein, as it eliminates hydrogen bonds between S1. S2 units, this physicochemical change increased the protein chain flexibility and transmission of SARS-CoV-2 up to approximately 50% Studies have not proven that this mutation may change the effectiveness of Monoclonal antibody treatments. [86][88][45]

N679K: New mutation in Omicron near the cleavage site Here function adds a positively charged Lysine residue, making the neighbor satisfactory for furin cleavage. N679K incorporates variant **P681H (**proline to histidine) providing efficient S1 and S2 cleavage, enhancing replication fitness raising the transmissibility of S protein. [24][51]
We found the accessory proteins ORF3a and ORF8 displayed the most mutations relative to other ORF proteins that are more conserved. But analysis of the ORF8 protein there are two deletion mutations D119- and F120-, which contribute with the structural instability, [89] and recognition of a significant Nonsense mutation (**G8*)** as an SNP-stop codon, found in 18 of 24 samples. This mutation associated The distribution of ORF8 knockouts was in clade 20I Alpha, which had Q27* and K68* then G8* mutation appears in clade 22F (lineage XBB) and clade 23A (XBB.1.5). According to CoVsurver on GISAID, the first detection of this aa change was hcov-19/Zambia/ZMB-159/2020 and still occurred recently in strain hcov-19/Japan/FR062-0102/2024 and submitted in 110 countries. ORF8 (Nonstructural protein NS8) belongs to

accessory proteins on the genome, its length is 366 bp, and 121 amino acid protein. ORF8 protein has been associated with accurate diagnosis of COVID-19 due to many functions related to induced early antibody response, inhibiting type I INF production and causing cytokine storms by activating IL-17. [34][90] A study showed that there is a small effect of the ORF8 protein on the gene expression of MHC-I cells, but not on ACE2 cells and the S protein in CoV2 This may be beneficial, but the ORF8 intracellular is not beneficial due to the induction of endoplasmic reticulum (ER) stress.[91] ORF8 gene is considered one of the most rapidly evolving in the betacoronavirus family and with these characteristics ORF8 is a therapeutic target by inactivation viral replication and host immune downregulation. [34][90]

XBB lineage has higher fitness through recombination and shows resistance to the antiviral humoral by gaining a pair of immune escape associated and infectivity as F486S and N460K on the 3′ fragment and Y144- and V83A on the 5′ fragment. [60] Pangolin results showed that XBB.1.42.1 was designed as Syrian lineages located in 18 samples. in the pangolin, Scorpio-call indicates another lineage for the sample EPI_ISL_18234343, this strain is a mother lineage (BA.3) Omicron that was last detected in the world was in 29 Jun 2022.
This sample was classified by GISAID in cluster GR, it was the lowest sample containing mutations, with 27 mutations only. A study indicates. This requires continuous monitoring of changes in SARS-CoV-2.
The results of the phylogenetic tree and www.outbreak.info showed a match in terms that the lineage XBB.1.42.1 is prevalent in countries in the Middle East that XBB.1.42.1 estimated 10% (95% CI: 5%-16%) risk of SARS-CoV-2 in Syria followed by Lebanon, Iraq, Ukraine, and Hungary. This is likely to spread to three countries in the Middle East beginning in 2023. sequence analysis was consistent with the global distribution. XBB.1.42.1 (22F), XBB.2.3.11(23E), and XBB.1.9.1(23D) are the most remain clades still circulating all of the year 2023.

The phylogenetic tree generated by ITOL of 140 isolates of whole genome sequences includes 10 isolates from Syria. All belong to the Omicron variant representing the evolutionary relationships among sequences based on the reference, the close related through evolution between sequences from Jordan, and Iran help indicate the origins of the virus and understand the trace of the infection through the population.

Conclusion

This study is the first descriptive genotype study in Syria for the whole genome of SARS-CoV-2. Applying a variety of bioinformatic tools to allow the analysis of variants, identify mutations, and determine lineage circulating in strains putting Syria on the genome map of the world, sharing information in the GISAID database. This study consists of a basis and introduction to other research regarding treatments and vaccines that are compatible with the strains spread in Syria. The continuity of circulating in the world increases the changes in spike protein. Remain using vaccines and monoclonal antibodies as an effective and successful therapeutic alternative in reducing infections, limiting their spread, and alleviating symptoms.

Limitation

The limited number of samples, may not reflect the extent of the spread of the SARS-CoV-2 virus in Syria especially since there is no genomic data during the early stages of the epidemic. Study a specific period of 2023, showed the spread of a recombinant variant XBB Omicron with increasing in the RBD Region mutations.

Funding

Laboratory materials, reagents, and supplies were funded solely by the World Health Organization in Syria for National Emergence Diseases Laboratory EDL. The research was conducted personally after getting the approval to obtain data and the search was completed without funding from any health authority.

Reference

[1] H. Lu, F. Giordano, and Z. Ning, “Oxford Nanopore MinION Sequencing and Genome Assembly,” *Genomics, Proteomics Bioinforma.*, vol. 14, no. 5, pp. 265–279, 2016, doi: 10.1016/j.gpb.2016.05.004.

[2] F. Wu *et al.*, “A new coronavirus associated with human respiratory disease in China,” *Nature*, vol. 579, no. 7798, pp. 265–269, 2020, doi: 10.1038/s41586-020-2008-3.

[3] Y. Z. Zhang and E. C. Holmes, “A Genomic Perspective on the Origin and Emergence of SARS-CoV-2,” *Cell*, vol. 181, no. 2, pp. 223–227, 2020, doi: 10.1016/j.cell.2020.03.035.

[4] K. Athanasopoulou, M. A. Boti, P. G. Adamopoulos, P. C. Skourou, and A. Scorilas, “Third-Generation Sequencing: The Spearhead towards the Radical Transformation of Modern Genomics,” *Life*, vol. 12, no. 1, 2022, doi: 10.3390/life12010030.

[5] G. John *et al.*, “Next-generation sequencing (Ngs) in covid-19: A tool for sars-cov-2 diagnosis, monitoring new strains and phylodynamic modeling in molecular epidemiology,” *Curr. Issues Mol. Biol.*, vol. 43, no. 2, pp. 845–867, 2021, doi: 10.3390/cimb43020061.

[6] K. Athanasopoulou, M. A. Boti, P. G. Adamopoulos, P. C. Skourou, and A. Scorilas, “Third-generation sequencing: The spearhead towards the radical transformation of modern genomics,” *Life*, vol. 12, no. 1, 2022, doi: 10.3390/life12010030.

[7] A. M. Giani, G. R. Gallo, L. Gianfranceschi, and G. Formenti, “Long walk to genomics: History and current approaches to genome sequencing and assembly,” *Comput. Struct. Biotechnol. J.*, vol. 18, pp. 9–19, 2020, doi: 10.1016/j.csbj.2019.11.002.

[8] D. Bolognini, R. Semeraro, and A. Magi, “Versatile Quality Control Methods for Nanopore Sequencing.,” *Evol. Bioinform. Online*, vol. 15, p. 1176934319863068, 2019, doi: 10.1177/1176934319863068.

[9] R. M. Wadapurkar and R. Vyas, “Computational analysis of next generation sequencing data and its applications in clinical oncology,” *Informatics Med. Unlocked*, vol. 11, no. May, pp. 75–82, 2018, doi: 10.1016/j.imu.2018.05.003.

[10] Y. Wang, Y. Zhao, A. Bollas, Y. Wang, and K. F. Au, “Nanopore sequencing technology, bioinformatics and applications,” *Nat.*

Biotechnol., vol. 39, no. 11, pp. 1348–1365, 2021, doi: 10.1038/s41587-021-01108-x.

[11] J. Zeng, H. Cai, H. Peng, H. Wang, and Y. Zhang, "Causalcall : Nanopore Basecalling Using a Temporal Convolutional Network," vol. 10, no. January, pp. 1–11, 2020, doi: 10.3389/fgene.2019.01332.

[12] G. John *et al.*, "Next-Generation Sequencing (NGS) in COVID-19: A Tool for SARS-CoV-2 Diagnosis, Monitoring New Strains and Phylodynamic Modeling in Molecular Epidemiology," *Curr. Issues Mol. Biol.*, vol. 43, no. 2, pp. 845–867, 2021, doi: 10.3390/cimb43020061.

[13] S. Devarakonda, D. Morgensztern, and R. Govindan, "Role of Next-Generation Sequencing Technologies," *Target. Ther. Oncol.*, vol. 122, no. 1, pp. 1–15, 2013, doi: 10.1201/b15944-6.

[14] M. Jain, H. E. Olsen, B. Paten, and M. Akeson, "The Oxford Nanopore MinION: delivery of nanopore sequencing to the genomics community," *Genome Biol.*, vol. 17, no. 1, pp. 1–11, 2016, doi: 10.1186/s13059-016-1103-0.

[15] I. D. Capraru *et al.*, "Identification of Genomic Variants of SARS-CoV-2 Using Nanopore Sequencing," *Med.*, vol. 58, no. 12, 2022, doi: 10.3390/medicina58121841.

[16] Z. Abdelrahman, M. Li, and X. Wang, "Comparative Review of SARS-CoV-2 , SARS-CoV , MERS-CoV , and Influenza A Respiratory Viruses," vol. 11, no. July 2003, 2020, doi: 10.3389/fimmu.2020.552909.

[17] M. H. Shum *et al.*, "Binding affinity between coronavirus spike protein and human ACE2 receptor," *Comput. Struct. Biotechnol. J.*, vol. 23, no. September 2023, pp. 759–770, 2024, doi: 10.1016/j.csbj.2024.01.009.

[18] F. Li and L. Du, "MERS Coronavirus : An Emerging Zoonotic Virus," vol. 4, pp. 1–6, 2019.

[19] A. N. Alshukairi *et al.*, "High Prevalence of MERS-CoV Infection in Camel Workers in Saudi Arabia," *MBio*, vol. 9, no. 5, pp. 10.1128/mbio.01985-18, 2018, doi: 10.1128/mbio.01985-18.

[20] A. Mittal, K. Manjunath, R. K. Ranjan, S. Kaushik, S. Kumar, and V. Verma, "COVID-19 pandemic: Insights into structure, function, and hACE2 receptor recognition by SARS-CoV-2," *PLOS Pathog.*,

vol. 16, no. 8, p. e1008762, Aug. 2020, [Online]. Available: https://doi.org/10.1371/journal.ppat.1008762

[21] S. Veit, S. Jany, R. Fux, and G. Sutter, "CD8 + T Cells Responding to the Middle East Respiratory Syndrome Coronavirus Nucleocapsid Protein Delivered by Vaccinia Virus MVA in Mice," 2018, doi: 10.3390/v10120718.

[22] X. Chen *et al.*, "Next-Generation Sequencing Reveals the Progression of COVID-19," *Front. Cell. Infect. Microbiol.*, vol. 11, no. March, pp. 1–14, 2021, doi: 10.3389/fcimb.2021.632490.

[23] A. Mittal, K. Manjunath, R. Kumar, and R. Id, "COVID-19 pandemic : Insights into structure , function , and hACE2 receptor recognition by," pp. 1–19, 2020, doi: 10.1371/journal.ppat.1008762.

[24] P. F. N. Souza *et al.*, "The spike glycoprotein of SARS-CoV-2: A review of how mutations of spike glycoproteins have driven the emergence of variants with high transmissibility and immune escape," *Int. J. Biol. Macromol.*, vol. 208, pp. 105–125, 2022, doi: https://doi.org/10.1016/j.ijbiomac.2022.03.058.

[25] C. Li, Q. He, H. Qian, and J. Liu, "Overview of the pathogenesis of COVID-19 (Review)," *Exp. Ther. Med.*, vol. 22, no. 3, pp. 1–10, 2021, doi: 10.3892/etm.2021.10444.

[26] Y.-C. Liu, R.-L. Kuo, and S.-R. Shih, "COVID-19: The first documented coronavirus pandemic in history," *Biomed. J.*, vol. 43, no. 4, pp. 328–333, 2020, doi: https://doi.org/10.1016/j.bj.2020.04.007.

[27] Y. M. Bar-On, A. Flamholz, R. Phillips, and R. Milo, "Science Forum: SARS-CoV-2 (COVID-19) by the numbers," *Elife*, vol. 9, p. e57309, 2020, doi: 10.7554/eLife.57309.

[28] T. N. Starr *et al.*, "Deep Mutational Scanning of SARS-CoV-2 Receptor Binding Domain Reveals Constraints on Folding and ACE2 Binding," *Cell*, vol. 182, no. 5, pp. 1295-1310.e20, 2020, doi: 10.1016/j.cell.2020.08.012.

[29] Y. Huang, C. Yang, X. feng Xu, W. Xu, and S. wen Liu, "Structural and functional properties of SARS-CoV-2 spike protein: potential antivirus drug development for COVID-19," *Acta Pharmacol. Sin.*, vol. 41, no. 9, pp. 1141–1149, 2020, doi: 10.1038/s41401-020-0485-4.

[30] Y. Wan, J. Shang, R. Graham, R. S. Baric, and F. Li, “Receptor Recognition by the Novel Coronavirus from Wuhan: an Analysis Based on Decade-Long Structural Studies of SARS Coronavirus.,” *J. Virol.*, vol. 94, no. 7, Mar. 2020, doi: 10.1128/JVI.00127-20.

[31] D. A. Jamison *et al.*, “A comprehensive SARS-CoV-2 and COVID-19 review, Part 1: Intracellular overdrive for SARS-CoV-2 infection,” *Eur. J. Hum. Genet.*, vol. 30, no. 8, pp. 889–898, 2022, doi: 10.1038/s41431-022-01108-8.

[32] Z. Y. Low, N. Z. Zabidi, A. J. W. Yip, A. Puniyamurti, V. T. K. Chow, and S. K. Lal, “SARS-CoV-2 Non-Structural Proteins and Their Roles in Host Immune Evasion,” *Viruses*, vol. 14, no. 9, 2022, doi: 10.3390/v14091991.

[33] E. Hartenian, D. Nandakumar, A. Lari, M. Ly, J. M. Tucker, and B. A. Glaunsinger, “The molecular virology of coronaviruses,” *J. Biol. Chem.*, vol. 295, no. 37, pp. 12910–12934, 2020, doi: 10.1074/jbc.REV120.013930.

[34] J. J. Garrido, “SARS-CoV-2 Accessory Proteins in Viral Pathogenesis : Knowns and Unknowns,” vol. 12, no. July, pp. 1–8, 2021, doi: 10.3389/fimmu.2021.708264.

[35] R. Gorkhali, P. Koirala, S. Rijal, A. Mainali, A. Baral, and H. K. Bhattarai, “Structure and Function of Major SARS-CoV-2 and SARS-CoV Proteins,” *Bioinform. Biol. Insights*, vol. 15, no. Figure 1, 2021, doi: 10.1177/11779322211025876.

[36] M. H.-H. Shum *et al.*, “Binding affinity between coronavirus spike protein and human ACE2 receptor,” *Comput. Struct. Biotechnol. J.*, vol. 23, pp. 759–770, 2024, doi: https://doi.org/10.1016/j.csbj.2024.01.009.

[37] O. Bel, K. R. P. Lima, M. B. Costa, and I. R. Farias, “Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID- 19 . The COVID-19 resource centre is hosted on Elsevier Connect , the company ’ s public news and information ,” no. January, 2020.

[38] B. Cosar, Z. Yagmur, S. Unal, and A. Turan, “Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID- 19 . The COVID-19 resource centre is hosted on Elsevier Connect , the company ’ s public news and information ,” no.

January, 2020.

[39] S. Xia *et al.*, “A pan-coronavirus fusion inhibitor targeting the HR1 domain of human coronavirus spike,” *Sci. Adv.*, vol. 5, no. 4, p. eaav4580, 2019, doi: 10.1126/sciadv.aav4580.

[40] S. Lukassen *et al.*, “SARS\&\#x2010;CoV\&\#x2010;2 receptor ACE2 and TMPRSS2 are primarily expressed in bronchial transient secretory cells,” *EMBO J.*, vol. 39, no. 10, p. e105114, 2020, doi: https://doi.org/10.15252/embj.20105114.

[41] S. Dong, J. Sun, Z. Mao, L. Wang, Y.-L. Lu, and J. Li, “A guideline for homology modeling of the proteins from newly discovered betacoronavirus, 2019 novel coronavirus (2019-nCoV).,” *J. Med. Virol.*, vol. 92, no. 9, pp. 1542–1548, Sep. 2020, doi: 10.1002/jmv.25768.

[42] M. A. Martin, D. VanInsberghe, and K. Koelle, “Insights from SARS-CoV-2 sequences,” *Science (80-.).*, vol. 371, no. 6528, pp. 466–467, 2021, doi: 10.1126/science.abf3995.

[43] S. Nelson-Sathi *et al.*, “Mutational landscape and in silico structure models of SARS-CoV-2 spike receptor binding domain reveal key molecular determinants for virus-host interaction,” *BMC Mol. Cell Biol.*, vol. 23, no. 1, p. 2, 2022, doi: 10.1186/s12860-021-00403-4.

[44] B. Cosar *et al.*, “SARS-CoV-2 Mutations and their Viral Variants.,” *Cytokine Growth Factor Rev.*, vol. 63, pp. 10–22, Feb. 2022, doi: 10.1016/j.cytogfr.2021.06.001.

[45] M. Alkhatib *et al.*, “SARS-CoV-2 Variants and Their Relevant Mutational Profiles: Update Summer 2021,” *Microbiol. Spectr.*, vol. 9, no. 3, 2021, doi: 10.1128/spectrum.01096-21.

[46] J. Y. Choi and D. M. Smith, “SARS-CoV-2 Variants of Concern,” *Yonsei Med J*, vol. 62, no. 11, pp. 961–968, Nov. 2021, [Online]. Available: https://doi.org/10.3349/ymj.2021.62.11.961

[47] S. S. A. Karim and Q. A. Karim, “Omicron SARS-CoV-2 variant: a new chapter in the COVID-19 pandemic,” *Lancet*, vol. 398, no. 10317, pp. 2126–2128, 2021, doi: 10.1016/S0140-6736(21)02758-6.

[48] K. Dhama *et al.*, “Global emerging Omicron variant of SARS-CoV-2: Impacts, challenges and strategies,” *J. Infect. Public Health*, vol. 16, no. 1, pp. 4–14, 2023, doi: 10.1016/j.jiph.2022.11.024.

[49] S. Chatterjee, M. Bhattacharya, S. Nag, K. Dhama, and C.

Chakraborty, “A Detailed Overview of SARS-CoV-2 Omicron: Its Sub-Variants, Mutations and Pathophysiology, Clinical Characteristics, Immunological Landscape, Immune Escape, and Therapies,” *Viruses*, vol. 15, no. 1, 2023, doi: 10.3390/v15010167.

[50] S. J. R. da Silva, A. Kohl, L. Pena, and K. Pardee, “Recent insights into SARS-CoV-2 omicron variant,” *Rev. Med. Virol.*, vol. 33, no. 1, 2023, doi: 10.1002/rmv.2373.

[51] C. Jung *et al.*, “Omicron: What Makes the Latest SARS-CoV-2 Variant of Concern So Concerning?,” *J. Virol.*, vol. 96, no. 6, 2022, doi: 10.1128/jvi.02077-21.

[52] L. Wu *et al.*, “SARS-CoV-2 Omicron RBD shows weaker binding affinity than the currently dominant Delta variant to human ACE2,” *Signal Transduct. Target. Ther.*, vol. 7, no. 1, p. 8, 2022, doi: 10.1038/s41392-021-00863-2.

[53] D. Mannar *et al.*, “SARS-CoV-2 Omicron variant: Antibody evasion and cryo-EM structure of spike protein–ACE2 complex,” *Science (80-.).*, vol. 375, no. 6582, pp. 760–764, 2022, doi: 10.1126/science.abn7760.

[54] R. Hirose *et al.*, “Differences in environmental stability among SARS-CoV-2 variants of concern: both omicron BA.1 and BA.2 have higher stability.,” *Clin. Microbiol. Infect. Off. Publ. Eur. Soc. Clin. Microbiol. Infect. Dis.*, vol. 28, no. 11, pp. 1486–1491, Nov. 2022, doi: 10.1016/j.cmi.2022.05.020.

[55] K. S. Sandhya, A. A. Kishore, A. Unni, P. Sunitha, C. V Sajithra, and A. S. Nair, “Interaction analysis of SARS-CoV-2 omicron BA1 and BA2 of RBD with fifty monoclonal antibodies: Molecular dynamics approach,” *J. Mol. Graph. Model.*, vol. 128, p. 108719, 2024, doi: https://doi.org/10.1016/j.jmgm.2024.108719.

[56] K. Dhama *et al.*, “Global emerging Omicron variant of SARS-CoV-2: Impacts, challenges and strategies.,” *J. Infect. Public Health*, vol. 16, no. 1, pp. 4–14, Jan. 2023, doi: 10.1016/j.jiph.2022.11.024.

[57] Y. Cao *et al.*, “Omicron escapes the majority of existing SARS-CoV-2 neutralizing antibodies,” *Nature*, vol. 602, no. 7898, pp. 657–663, 2022, doi: 10.1038/s41586-021-04385-3.

[58] C. S. Lupala, Y. Ye, H. Chen, X.-D. Su, and H. Liu, “Mutations on RBD of SARS-CoV-2 Omicron variant result in stronger binding to human ACE2 receptor,” *Biochem. Biophys. Res. Commun.*, vol. 590, pp. 34–41, 2022, doi:

https://doi.org/10.1016/j.bbrc.2021.12.079.

[59] S. Chatterjee, M. Bhattacharya, S. Nag, K. Dhama, and C. Chakraborty, “A Detailed Overview of SARS-CoV-2 Omicron: Its Sub-Variants, Mutations and Pathophysiology, Clinical Characteristics, Immunological Landscape, Immune Escape, and Therapies,” *Viruses*, vol. 15, no. 1, pp. 1–27, 2023, doi: 10.3390/v15010167.

[60] T. Tamura *et al.*, “Virological characteristics of the SARS-CoV-2 XBB variant derived from recombination of two Omicron subvariants,” *Nat. Commun.*, vol. 14, no. 1, pp. 13–15, 2023, doi: 10.1038/s41467-023-38435-3.

[61] R. You *et al.*, “Systematic Genomic Surveillance of SARS-CoV-2 at Xiamen International Airport and the Port of Xiamen Reveals the Importance of Incoming Travelers in Lineage Diversity.,” *Viruses*, vol. 16, no. 1, Jan. 2024, doi: 10.3390/v16010132.

[62] J. Guo, J. Ge, and Y. Guo, “Recent advances in methods for the diagnosis of Corona Virus Disease 2019.,” *J. Clin. Lab. Anal.*, vol. 36, no. 1, p. e24178, Jan. 2022, doi: 10.1002/jcla.24178.

[63] G. Rong, Y. Zheng, Y. Chen, Y. Zhang, P. Zhu, and M. Sawan, “COVID-19 Diagnostic Methods and Detection Techniques.,” *Encyclopedia of Sensors and Biosensors*. pp. 17–32, 2023. doi: 10.1016/B978-0-12-822548-6.00080-7.

[64] A. C. Shi and P. Ren, “SARS-CoV-2 serology testing: Progress and challenges.,” *J. Immunol. Methods*, vol. 494, p. 113060, Jul. 2021, doi: 10.1016/j.jim.2021.113060.

[65] L. Ning, H. B. Abagna, Q. Jiang, S. Liu, and J. Huang, “Development and application of therapeutic antibodies against COVID-19,” *Int J Biol Sci*, vol. 17, pp. 1486–1496, 2021, doi: 10.7150/ijbs.59149.

[66] A. Valcarcel, A. Bensussen, E. R. Álvarez-buylla, J. Díaz, and S. P. Smieszek, “Structural Analysis of SARS-CoV-2 ORF8 Protein : Pathogenic and Therapeutic Implications,” vol. 12, no. September, pp. 1–8, 2021, doi: 10.3389/fgene.2021.693227.

[67] M. Ejemel *et al.*, “A cross-reactive human IgA monoclonal antibody blocks SARS-CoV-2 spike-ACE2 interaction.,” *Nat. Commun.*, vol. 11, no. 1, p. 4198, Aug. 2020, doi: 10.1038/s41467-020-18058-8.

[68] A. Saleh, S. Qamar, A. Tekin, R. Singh, and R. Kashyap, “Vaccine Development Throughout History.,” *Cureus*, vol. 13, no. 7, p. e16635, Jul. 2021, doi: 10.7759/cureus.16635.

[69] V. Prasad and A. Haslam, “COVID-19 vaccines: history of the pandemic’s great scientific success and flawed policy implementation,” *Monash Bioeth. Rev.*, 2024, doi: 10.1007/s40592-024-00189-z.

[70] A. Y.-K. Thye, J. W.-F. Law, P. Pusparajah, V. Letchumanan, K.-G. Chan, and L.-H. Lee, “Emerging SARS-CoV-2 Variants of Concern (VOCs): An Impending Global Crisis.,” *Biomedicines*, vol. 9, no. 10, Sep. 2021, doi: 10.3390/biomedicines9101303.

[71] A. D. Iuliano *et al.*, “Trends in Disease Severity and Health Care Utilization During the Early Omicron Variant Period Compared with Previous SARS-CoV-2 High Transmission Periods — United States, December 2020–January 2022,” *MMWR Recomm. Reports*, vol. 71, no. 4, pp. 146–152, 2022, doi: 10.15585/mmwr.mm7104e4.

[72] H. Li, “Minimap2: pairwise alignment for nucleotide sequences,” *Bioinformatics*, vol. 34, no. 18, pp. 3094–3100, Sep. 2018, doi: 10.1093/bioinformatics/bty191.

[73] T. Seemann, “Prokka: rapid prokaryotic genome annotation,” *Bioinformatics*, vol. 30, no. 14, pp. 2068–2069, 2014, doi: 10.1093/bioinformatics/btu153.

[74] K. Smith, “Genetic Polymorphism and SNPs,” [Online]. Available: https://api.semanticscholar.org/CorpusID:40212128

[75] D. Mercatelli, L. Triboli, E. Fornasari, F. Ray, and F. M. Giorgi, “Coronapp: A web application to annotate and monitor SARS-CoV-2 mutations.,” *J. Med. Virol.*, vol. 93, no. 5, pp. 3238–3245, May 2021, doi: 10.1002/jmv.26678.

[76] J. Bendl *et al.*, “PredictSNP: robust and accurate consensus classifier for prediction of disease-related mutations.,” *PLoS Comput. Biol.*, vol. 10, no. 1, p. e1003440, Jan. 2014, doi: 10.1371/journal.pcbi.1003440.

[77] H. Thorvaldsdóttir, J. T. Robinson, and J. P. Mesirov, “Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration,” *Brief. Bioinform.*, vol. 14, no. 2, pp. 178–192, Mar. 2013, doi: 10.1093/bib/bbs017.

[78] O. Ovchynnykova *et al.*, “Homology Modeling and Molecular

Dynamics-Driven Search for Natural Inhibitors That Universally Target Receptor-Binding Domain of Spike Glycoprotein in SARS-CoV-2 Variants.," *Molecules*, vol. 27, no. 21, Oct. 2022, doi: 10.3390/molecules27217336.

[79] E. F. Pettersen *et al.*, "UCSF ChimeraX: Structure visualization for researchers, educators, and developers.," *Protein Sci.*, vol. 30, no. 1, pp. 70–82, Jan. 2021, doi: 10.1002/pro.3943.

[80] Á. O'Toole, O. G. Pybus, M. E. Abram, E. J. Kelly, and A. Rambaut, "Pango lineage designation and assignment using SARS-CoV-2 spike gene nucleotide sequences," *BMC Genomics*, vol. 23, no. 1, 2022, doi: 10.1186/s12864-022-08358-2.

[81] Á. O'Toole *et al.*, "Assignment of epidemiological lineages in an emerging pandemic using the pangolin tool.," *Virus Evol.*, vol. 7, no. 2, p. veab064, 2021, doi: 10.1093/ve/veab064.

[82] S. Khare *et al.*, "GISAID's Role in Pandemic Response.," *China CDC Wkly.*, vol. 3, no. 49, pp. 1049–1051, Dec. 2021, doi: 10.46234/ccdcw2021.255.

[83] K. Katoh and D. M. Standley, "MAFFT multiple sequence alignment software version 7: improvements in performance and usability.," *Mol. Biol. Evol.*, vol. 30, no. 4, pp. 772–780, Apr. 2013, doi: 10.1093/molbev/mst010.

[84] B. Q. Minh *et al.*, "IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era," *Mol. Biol. Evol.*, vol. 37, no. 5, pp. 1530–1534, May 2020, doi: 10.1093/molbev/msaa015.

[85] I. Kufareva and R. Abagyan, "Methods of protein structure comparison.," *Methods Mol. Biol.*, vol. 857, pp. 231–257, 2012, doi: 10.1007/978-1-61779-588-6_10.

[86] R. A. A. Pondé, "Physicochemical effect of the N501Y, E484K/Q, K417N/T, L452R and T478K mutations on the SARS-CoV-2 spike protein RBD and its influence on agent fitness and on attributes developed by emerging variants of concern," *Virology*, vol. 572, pp. 44–54, 2022, doi: https://doi.org/10.1016/j.virol.2022.05.003.

[87] R. Kumar *et al.*, "Understanding Mutations in Human SARS-CoV-2 Spike Glycoprotein: A Systematic Review & Meta-Analysis," *Viruses*, vol. 15, no. 4, 2023, doi: 10.3390/v15040856.

[88] L. Zhang *et al.*, "SARS-CoV-2 spike-protein D614G mutation

increases virion spike density and infectivity," *Nat. Commun.*, vol. 11, no. 1, p. 6013, 2020, doi: 10.1038/s41467-020-19808-4.

[89] F. A. Alsuwairi *et al.*, "Genomic Surveillance and Mutation Analysis of SARS-CoV-2 Variants among Patients in Saudi Arabia," 2024.

[90] C. Wagner *et al.*, "Positive selection underlies repeated knockout of ORF8 in SARS-CoV-2 evolution," *Nat. Commun.*, vol. 15, no. 1, pp. 1–12, 2024, doi: 10.1038/s41467-024-47599-5.

[91] K. Matsuoka *et al.*, "RESEARCH ARTICLE EDITORS ' PICK SARS-CoV-2 accessory protein ORF8 is secreted extracellularly as a glycoprotein homodimer," *J. Biol. Chem.*, vol. 298, no. 3, p. 101724, 2022, doi: 10.1016/j.jbc.2022.101724.