



## Research Article

## Open access

# Molecular evolutionary analysis of the SARS-CoV-2 through the mutation analysis of Spike, Envelope and RdRp proteins

Dženita Omerkić Dautović<sup>1</sup>, Amra Džuhović<sup>2\*</sup>, Adna Ašić<sup>1,2</sup>

<sup>1</sup>International Burch University, Sarajevo, Bosnia and Herzegovina

<sup>2</sup>Verlab Research Institute for Biomedical Engineering, Medical Devices, and Artificial Intelligence, Sarajevo, Bosnia and Herzegovina

DOI: 10.31383/ga.vol8iss1ga03

## Abstract

### \*Correspondence

E-mail:  
amra@verlabinstitute.com

### Received

April, 2024

### Accepted

May, 2024

### Published

June, 2024

**Copyright:** ©2024 Genetics & Applications, The Official Publication of the Institute for Genetic Engineering and Biotechnology, University of Sarajevo

## Keywords

*Bioinformatics,  
COVID-19, mutation,  
phylogenetic analysis,  
SARS-CoV-2, Spike  
protein*

COVID-19 pandemic, caused by the Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2), was declared in 2020 by the World Health Organization. New mutations have been identified, leading to various variants of this virus, including Alpha, Beta, Gamma, Delta, and Omicron, which are classified as variants of concern (VOCs) and have raised considerable concerns for global public health. Such constant spread and changes in the genome of the virus require continuous monitoring. This research focuses on the evolution of SARS-CoV-2 through a detailed presentation of the viral genome, protein structure and interpretation, with the presentation of phylogenetic characteristics and patterns. We obtained the sequence data from the European region focusing on the S, E, and RdRp proteins from the publicly available NCBI database. We next used the MEGA11 package to generate the multiple sequence alignments and create phylogenetic trees. The SWISS-MODEL server was connected to the Protein Data Bank to obtain tertiary structure images of all the proteins presented in the paper. Stability studies of obtained mutations were performed via MUpro online tool. The results indicate a substantial impact of the Omicron variant relative to others, particularly concerning the alterations and mutations observed in the spike (S) protein, which is crucial in the infection process.

## Introduction

At the end of 2019, Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) rapidly spread globally, resulting in two million deaths within the first 12 months. During this period, a significant health crisis emerged, leading to widespread adverse consequences on a global scale (Kapoor et al., 2023; Yang and Rao, 2021).

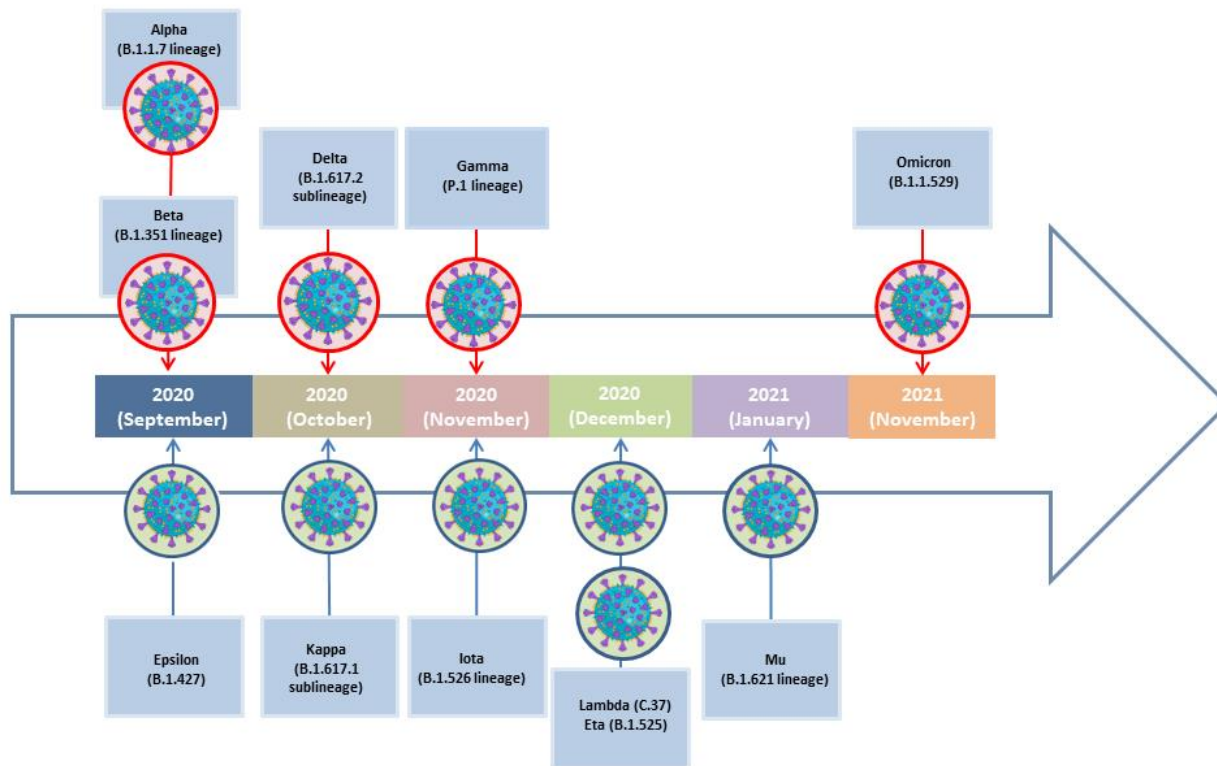
In the last twenty years, three coronaviruses (CoVs) have been identified, and the SARS-CoV-2 pandemic has had especially detrimental consequences (Sallard et al., 2021). The reservoir for the coronavirus evolution is a group of host mammals consisting of about 500 species (Cong et al., 2020; Sallard et al., 2021; Zhao et al., 2022). Important for the molecular evolution of these viruses, it is worth noting that the genomes of all members of the Coronavirus family are relatively long (Gomari et al., 2023).

The first sequence of the SARS-CoV-2 virus was reported in GenBank on 11th January 2020 (MN908947.3) (Rahimi et al., 2021). GenBank (<https://www.ncbi.nlm.nih.gov/sars-cov-2/>) has collected 8,723,875 nucleotide sequences (accessed 8th February 2024). Genome sequencing has been systematically implemented since the onset of the pandemic and remains critically important due to the continuous emergence of new variants and strains. Consistent with this, proper data collection, processing, and adequate genome analysis methods are required (Kandeel et al., 2021).

The phylogenetic approach enables the differentiation and identification of a range of phenomena, including different viral variants (Attwood et al., 2022). The classification of SARS-CoV-2 into Variants of Interest (VOIs) and Variants of Concern (VOCs) provides a

straightforward method for differentiation. VOIs are primarily characterized by mutations that alter the virus's properties (Figure 1). VOCs are generally monitored more closely because they significantly change the severity of the virus, its spread, and, therefore, the effectiveness of vaccines (Souza et al., 2022). The evolution and progression of a pandemic are influenced by the interplay between the pathogen, the host, and the environment in which they (co-)exist (Pinato et al., 2022).

In this research, we studied the SARS-CoV-2 describing the entire genome with the analysis of structural, non-structural and accessory proteins using *in silico* methods and detailed literature search. For phylogenetic analysis, we selected two structural proteins, Spike and Envelope proteins, as well as RdRp, a non-structural protein with a key role in virus replication. We chose the Spike protein because it has the highest mutation rate, resulting in the questionable effectiveness of vaccines that was discussed in the scientific community repeatedly with each new VOC emergence (Akkız, 2022; Hamdy et al., 2023). RdRp was chosen because many PCR tests target this gene, so changes can be reflected in the occurrence of false negative tests (Valadan et al., 2022). Studies related to viral pathogenesis, infection and function of E protein are scarce, when compared to other proteins that are in more detail researched and explained (Zhou et al., 2023). Additionally, this gene is not known for a high mutation rate, thus representing the lower end of the mutation rate spectrum in the SARS-CoV-2 genome. This comprehensive genomic approach can enhance our understanding of the evolutionary processes driven by various mutations, as such knowledge is applicable to pandemic monitoring, the design of detection tests, and pharmacological applications.



**Figure 1.** SARS-CoV-2 Variants of Concern and Variants of Interest (created with Smart, 2023). Alpha variant or B.1.1.7 was reported for the first time in September 2020 in the United Kingdom, followed by Beta (B.1.351; October 2020; South Africa), then Gamma (P.1; January 2020; Brazil), Delta (B.1.617.2; December 2020, India), and Omicron (BA.1 and BA.2; November 2021; Africa).

## Material and methods

The data filters used in the NCBI Virus database search were set as: the names of proteins (S, E and RdRp), geographic region: Europe, isolation source: oronasopharynx, and nucleotide completeness: protein. For the Pango lineage filter, we chose VOCs, with the addition of XBB.1.5 (nicknamed Kraken) that was contributing to increased number of reported cases, causing 49.1% of COVID-19 cases in January 2023 (Abdolreza et al., 2024; Scarpa et al., 2023), as well as BQ.1 or Cerberus, the BA.5 sublineage,

due to its wide prevalence, especially in January 2023 (Scarpa et al., 2023). Sample collection date was considered as the reporting date of the genome. Samples were chosen from the European region from the three periods:

1st September 2020 – 1st April 2021

31st August 2021 – 1st April 2022

31st August 2022 – 1st April 2023

We originally designed these periods because of the dynamics of spread of VOCs and peak of infections, in addition to the autumn and winter periods also being characteristic for flu and other respiratory infections.

A total of 121 nucleotide sequences were obtained for E protein, including the RefSeq Wuhan sequence, while for the analysis of S protein, 138 nucleotide sequences (including the RefSeq from Wuhan) were used. In collecting data on the RdRp molecule, due to specificities of the polypeptide processing and generation of mature protein, we used amino acid sequences and obtained 42 available samples, including the RefSeq entry. Genomes or sequences downloaded in FASTA file format were used in the Molecular Evolutionary Genetics Analysis Version 11 (MEGA11) package for alignment using default parameters (Kumar et al., 2018) for MSA with the ClustalW implementation (Gap Opening penalty 10.00 and Gap Extension Penalty 0.20 for multiple alignment). Protein sequences were aligned with the reference SARS-CoV-2 protein sequence from Wuhan (Wuhan-Hu-1 genome, GenBank accession NC\_045512), using the E protein information on [https://www.ncbi.nlm.nih.gov/protein/YP\\_009724392.1](https://www.ncbi.nlm.nih.gov/protein/YP_009724392.1), and S protein on <https://www.ncbi.nlm.nih.gov/protein/179631859>. Considering the manual selection of sequences and the absence of frequent or longer gaps in the viral gene or protein sequences, changes in alignment parameters did not produce significant differences in terms of genetic distances calculated for the MSA. When designing the phylogenetic tree, we set the bootstrap value to 1000 (Hall, 2013), using the neighbor-joining (NJ) method to design a cladogram. We have deployed UPGMA tree building method, as another distance-based alternative to the NJ method and no significant differences in tree topology were observed. In addition, more computationally intensive methods, including maximum parsimony and maximum likelihood were implemented, producing congruent topology to the one presented here. Combined with bootstrap values over 70%, these results confirmed well-supported NJ tree

topology, which remained the method of choice considering its simplicity, speed and low computational intensity. The tree in Newick format from the MEGA11 program was further edited and visualized by iTOL (<https://itol.embl.de/>) (Letunic and Bork, 2016).

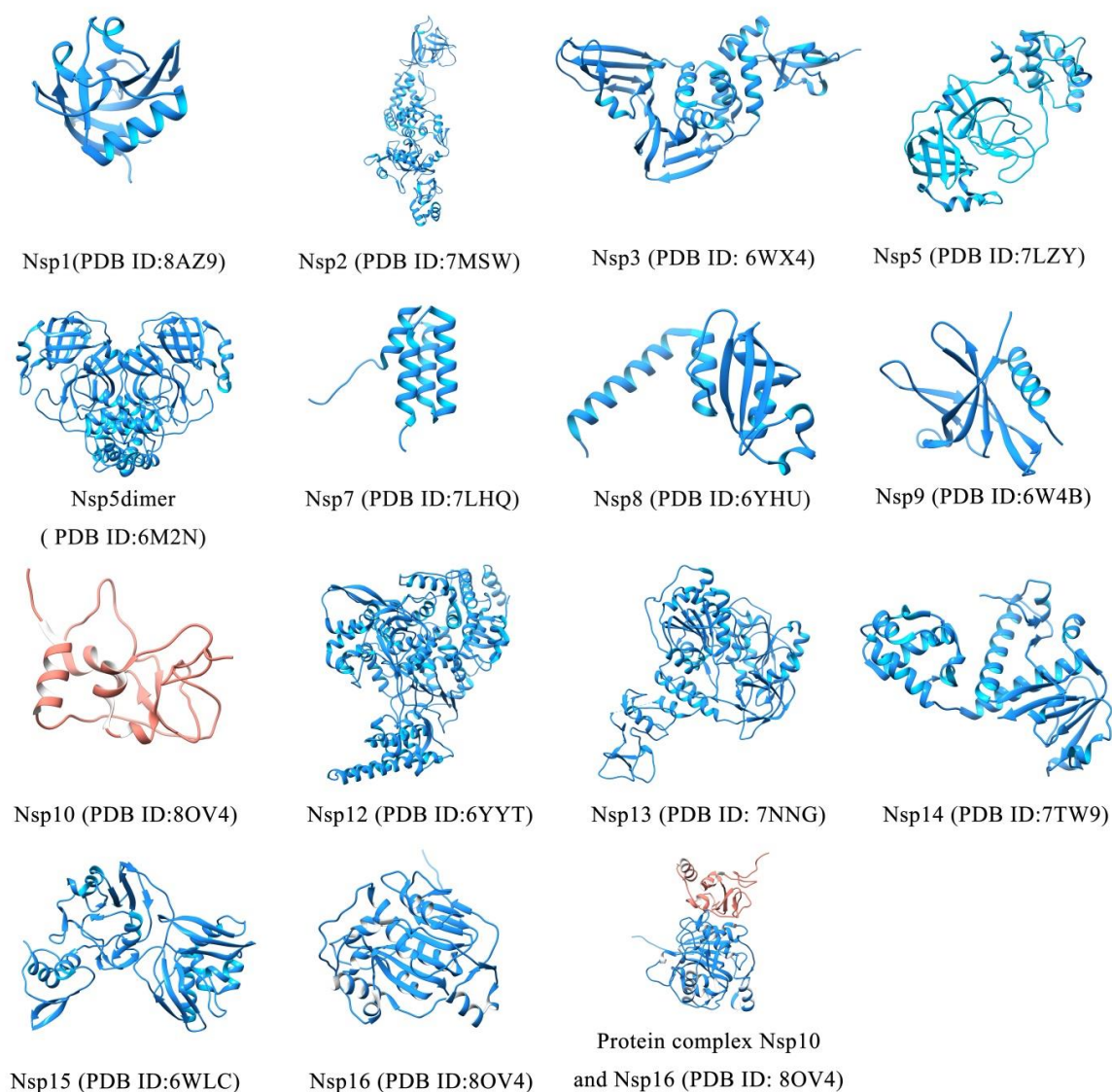
In order to perform protein structure visualization, the coordinates were downloaded directly from the Protein Data Bank (PDB), connected to the SWISS-MODEL server instead of performing *de novo* modeling of the protein structure, considering widely available high-quality imaging and annotation data for the SARS-CoV-2 proteome. We used UCSF Chimera to visualize selected molecules to show the structure of the proteins that make up the virus. The following structures of SARS-CoV-2 three-dimensional molecules with their PDB numbers were retrieved from the RCSB Protein Data Bank (<https://www.rcsb.org/>) in PDB format (Berman et al., 2000): Envelope protein (PDB ID: 7K3G), Spike protein (PDB ID: 7QUS), Membrane protein (PDB ID: 7VGS), N protein (PDB ID: 6YUN), Nsp1 (PDB ID: 8AZ9), Nsp2 (PDB ID: 7MSW), Nsp3 (PDB ID: 6WX4), Nsp4 (PDB ID: 7LZY), Nsp5 dimer (PDB ID: 6M2N), Nsp7 (PDB ID: 7LHQ), Nsp8 (PDB ID: 6YHU), Nsp9 (PDB ID: 6W4B), Nsp10 (PDB ID: 8OV4), Nsp12 (PDB ID: 6YYT), Nsp13 (PDB ID: 7NNG), Nsp14 (PDB ID: 7TW9), Nsp15 (PDB ID: 6WLC), Nsp16 (PDB ID: 8OV4), and Protein complex Nsp10 and Nsp16 (PDB ID: 8OV4). All of them were visualized by cartoon presentation using UCSF Chimera 1.16 (Pettersen et al., 2004). In order to obtain information on the effect of detected genetic variants on the protein stability, we used online tool MUpro (Cheng et al., 2006), whereby plain reference protein sequence of each S, E and RdRp protein was entered. We indicating all previously detected variable positions and the exact amino acid change, and obtained the results on the effect of specified changes.

## Results and Discussion

### *ORF1ab and its polyproteins*

The nonstructural proteins (Figure 2) of the virus (Nsp1-16) are created after proteolysis of the polypeptides pp1a and pp1ab by viral protease (Frazier et al., 2022).

Non-structural proteins of SARS-CoV-2 play a pivotal role by being involved in viral replication processes and replication transcription complex (Yan et al., 2020; [Supplementary Table 1](#)), and through their role in suppression of immune system or feedback of the host organism (Prasada et al. 2023).



**Figure 2.** Non-structural proteins of SARS-CoV-2 were obtained through the SWISS-MODEL project, whereby we utilized protein sequences and monitored the GMQE score. Visualization was performed using UCSF Chimera. The PDB ID of the protein in the image corresponds to the closest template used for modeling by connecting the SWISS-MODEL work directly to PDB.

### Structural proteins of SARS-CoV-2

Genes that encode for structural proteins (Figure 3) are found on the 3'-UTR, along with eight accessory genes (Naqvi et al., 2020).

SARS-CoV-2 Spike protein is in the group of fusion viral proteins (Santopolo et al., 2021) that has a vital role in the infection phase (Brintnell et al., 2021) and consists of two functional units labeled S1 and S2 (Kumavath et al., 2021). This protein plays a key role in viral entry into the host cells (Ghorbani et al., 2021). Glycosylation, phosphorylation, and acylation are some of the post-translational modifications (PTM) usually found on the Spike protein, following successful host infection (Cheng et al., 2023).

ACE2 is the primary receptor for Spike protein; their joining enables endocytosis or virus fusion inside the cell. In addition to the ACE2 receptor, other host proteins can serve as a receptor for coupling S proteins to the host, such as Neuropilin-1 (Kumavath et al., 2021; Yevsieieva et al., 2023). The RBD located on the Spike protein enables the recognition of hACE2, ensuring the penetration of the virus into the host (Yevsieieva et al., 2023). Within the RBD is a characteristic N501Y mutation that facilitates binding to the host, as well as E484K that was identified in Europe and America (Akkız, 2022). Changes in the Spike protein that mainly affect the coupling with the host ACE2 receptor are changes of K417N, N440K, G446S, S477N, T478K, E484A, Q493K, G496S, Q498R, N501Y, and Y505H in the S1 region of RBD (Jung et al., 2022). RBD mutations cause questionable effectiveness of therapy and vaccines (Han et al., 2022).

*ORF4* gene encodes for E protein (Yoshimoto, 2020), that contains from 76 to 109 amino acids,

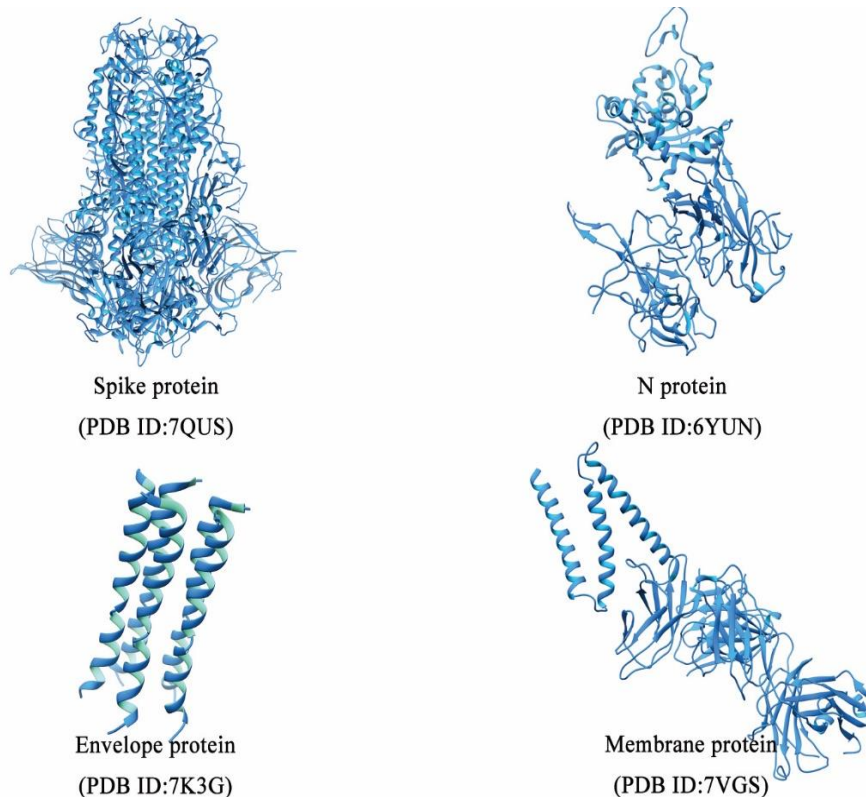
ranging from 8.4 to 12 kDa in size (Schoeman and Fielding, 2019; Zhou et al., 2023). E protein is an integral and short hydrophilic protein (Schoeman and Fielding, 2019), involved in the life cycle of SARS-CoV-2 from several perspectives (Cao et al., 2021). E protein belongs to the group of viroporins, that are regulating microenvironment of the host cell, including the pH and ion concentration (Bai et al., 2022; Cao et al., 2021; Schoeman and Fielding, 2019). It is also involved in viral assembly, release and pathogenesis (Yoshimoto, 2020). Through experimental efforts, it was reported that E protein contains parts subject to glycosylation (N48 and N66 parts) and it is known that through this process, the Spike protein deftly resists “obstacles” and thereby preserves infectivity.

The M protein shapes the viral appearance and is the most abundant viral protein (Schoeman and Fielding, 2019). It is encoded by *ORF5* gene (Yoshimoto et al., 2020), contains 222 amino acids (Bai et al. 2022), and plays a role in binding to the nucleocapsid (Khan et al., 2020) and determining the shape of the viral envelope (Bai et al., 2022).

Viral particle assembly is enhanced after succinylated N and M proteins. *ORF9b* gene encodes for Nucleocapsid phosphoprotein (N protein) (Cheng et al., 2023; Redondo et al., 2021), which is crucial for capsid formation (Schoeman and Fielding, 2019), in addition to being an antigen for virus detection (Rak et al., 2023). It comprises 419 amino acids (Bai et al., 2022) and is an antagonist of the host immune response (Yang and Rao, 2021). Back in 2020, it was already known that N parts might be detected in infected cells. This protein has a decisive role in the transfer of material to the progeny of the virus (Cong et al., 2020), but also promotes hyperinflammation and protects the viral genome (Bai et al., 2022) packed into a helical

ribonucleoprotein formed by the N protein (Low et al., 2022). The N protein is less prone to mutations and its higher sequence conservation make it an excellent diagnostic target (Yan et al., 2022). Beta and Omicron strains contain a mutation on the N protein, a P13L/S substitution, which is unique to these strains. It is believed that this mutation, as well as P13L, ERS31-33del, D63G, R203K/M, G204R, and D377Y occurred because the virus escaped the immune response of the host (Rak et al., 2023). Stabilizing mutations characteristic of the N protein are S194L, D103Y, P13L, S197L, M234I, and S188L (Jacob et al., 2021).

The last third of the viral genome codes for structural and accessory proteins (Wilson et al., 2022), involved in viral pathogenesis (Redondo et al., 2021). The accessory proteins include ORF3a, ORF3b, ORF6, ORF7a, ORF7b, ORF8, ORF9b, ORF9c, and ORF10 (Zhou et al., 2022), and are mostly scarcely researched (Giovanetti et al., 2021) with poor sequence conservation (van de Leemput and Han, 2021). Many accessory proteins affect the work of mitochondria and the dysregulation of processes signaled by this organelle (Prasada et al., 2023).



**Figure 3.** Structural proteins of SARS-CoV-2 were generated using the SWISS-MODEL, whereby the PDB templates with the best GMQE score were used for modeling.

#### *Mutations, stability, and phylogenetic analysis*

Alpha (B.1.1.7 lineage), Beta (B.1.351 lineage), Gamma (P.1), Delta (B.1.617.2 lineage), and Omicron (BA.1; BA.2) characterized the

coronavirus pandemic as VOCs, and their dynamics of spread, infectivity, and number of deaths are continuously reported, especially for Delta and Omicron variants which expanded multiregionally (Jung et al., 2022).

In the group of VOIs, there are Epsilon (B.1.427 and B.1.429), Zeta (P.2), Eta (B.1.525), Theta (P.3), Iota (B.1.526), Kappa (B.1.617.1), Lambda (C.37), and Mu (B.1.621) (Varghese Edwin and Stanislaus Antony, 2023; Wolf et al., 2023). By 21st January, 2024, there were 774,394,829 confirmed cases of COVID-19 (<https://covid19.who.int/>), compared to 27th January 2023, when 752,517,552 cases of COVID-19 were confirmed, alluding that the changes are still present and that the statistics is of great importance in monitoring the virus even after the official declaration of the end of the pandemic.

One of the reasons for such accelerated viral mutation accumulation is the lack of the viral genome proofreading activity of the virus, as well

as potential effects on the overall protein stability following the appearance of the initial mutations (Papanikolaou et al., 2022). In this study, we analyzed the SARS-CoV-2 sequences of the S, E, and RdRp proteins, observed mutations and their phylogenetic relationships, as well as their effect on the protein stability.

The results indicate evolutionary changes and provide information relevant for future monitoring and related research. MUPro analysis of mutation effect showed that the majority of observed changes act towards decreasing the protein stability (Table 1). Mutations were detected using the MEGA11 package, and their correctness was verified through literature review, followed by comparisons and commentary as appropriate.

**Table 1.** Results of MUPro protein stability analysis for S, E and RdRp proteins based on sequence information.

Mutation	Lineage	Effect
<b>S protein</b>		
P26S	P.1	Decreases stability of protein structure
W64R	BA.2.9.3	Increases the stability of protein structure
T19I	BA.2.9.3, BA.2.49, BA.2.9, BA.2	Decreases the stability of protein structure
T19R	B.1.617.2	Decreases the stability of protein structure
A27S	BA.2, BA.2.9.3, BA.2.49, BA.2.9, B.1.1.529 XBB.1, XBB.1.5, XBB.11.5.14, BF.7, BA.4.6, BN.1.2, BQ.1.1.18, BN.1.2, BQ.1.1, BQ.1, BA.5.2.1, BA.5.2, BA.5, BA.5.2	Decreases the stability of protein structure
W64R	BA.2.93	Increases the stability of protein structure
S71F	BQ.1	Decreases the stability of protein structure
D80A	B.1.351	Decreases the stability of protein structure
T95I	BA.1.17, BA.1.18, BA.1.1, BA.1.1.1, B.1.1.529, B.1.617.2, BA.2, BA.1	Decreases the stability of protein structure
Q134H	BQ.1.1.18	Decreases the stability of protein structure
D138Y	P.1	Decreases the stability of protein structure
G142D	BQ.1	Decreases the stability of protein structure
H146Q	XBB.1, XBB.1.5, XBB.1.5.14	Decreases the stability of protein structure
K147E	BA.5, BN.1.2	Decreases the stability of protein structure



E156G	B.1.617.2	Decreases the stability of protein structure
Q183E	XBB.1, XBB.1.5, XBB.1.5.14	Decreases the stability of protein structure
Q183L	BA.5.2.1	Increases the stability of protein structure
G184V	BA.5.2	Decreases the stability of protein structure
R190S	P.1	Decreases the stability of protein structure
T208M	BA.1	Decreases the stability of protein structure
V213G	BA.2.9.3, BA.2.49, BA.2.9, BN.1.2, BQ.1.1.18, BQ.1.1, BQ.1, BA.5, BA.5.2.1, BA.5.2	Decreases the stability of protein structure
L455R	Delta, BF.7, BA.4.6, BQ.1.1, BQ.1.1.18	Decreases the stability of protein structure
Y508H	Omicron	Decreases the stability of protein structure
T719I	Alpha	Decreases the stability of protein structure
P812S	Alpha	Decreases the stability of protein structure
Q957H	Gamma, Omicron	Decreases the stability of protein structure
<b>E protein</b>		
T9I	B.1.1.529, B.1.1.52, BA.1, BA.2, BA.1.1, BA.5, BA.5.1, BA.5.2, BA.5.2.1, BQ.1, BQ.1.1, BQ.1.1.18, BA.1.17.2, BA.1.17, BA.1.18, BA.1.1.1, BA.2.3, BA.2.9, BA.2.9.3, BA.2.23	Increases the stability of protein structure
S50I	BA.1.1.1	Increases the stability of protein structure
V62F	B.1.617.2	Decreases the stability of protein structure
P71L	B.1.351	Decreases the stability of protein structure
<b>RdRp protein</b>		
M196R	WBE23697.1	Decreases the stability of protein structure
P323L	B.1 lineage, including IDs WBE23697.1, WBE23695.1, WBE23684.1, WBE23682.1, WBE23680.1, WBE23676.1, WBE23672.1	Increases the stability of protein structure
C645G	WBE23697.1	Decreases the stability of protein structure
G671S	WBE23676.1	Decreases the stability of protein structure

### *Spike protein*

In this study, we detected 95 Spike glycoprotein mutations ([Supplementary Table 2](#)). The available evidence suggests that significant Spike protein mutations include E484K, K417T/N, N501Y, and D614G (Giovanetti et al. 2021; Mansbach et al.

2021). The N501Y mutation, located on the RBD, enhances the coupling of the ACE2 receptor and undoubtedly facilitates its entry into the host (John et al., 2021). This mutation is present in all VOCs except the Delta variant (Aleem, 2022), reducing vaccine efficacy (Saifi et al., 2022). Further

evidence supports that N501Y and D614G mutations combined increase viral replication rate (Akkız, 2022). N501Y also corroborates the highest binding affinity of RBD in all VOCs, followed by E484K (Shahhosseini et al., 2021).

Convergent substitutions (R346T, K444T, L452R, N460K, and F486V) are characteristic of the BQ.1 variant (Ito et al., 2023). We found the mutation K444T in BQ.1, BQ.1.1, BQ.1.1.18, as well as R346T and L452R in Delta samples (Supplementary Table 2).

The S375F mutation characterizes the Omicron BA.1 variant and thus affects the speed of Omicron spread due to increased binding affinity to hACE2. Mutations S371L, S373P, and S375F were not detected in earlier variants before the appearance of BA.1 (Kimura et al. 2022). In this study, we detected S371L in Omicron samples, in addition to the rare mutation Q493R which we found in Omicron and Gamma S protein sequences. This position is critical because of its role in interaction with ACE2, that is, increased binding affinity to hACE2 in case of mutant amino acid being present (Alkhatib et al., 2022). Specific mutants circulating worldwide (such as V367F) may affect splicing (Gusev et al. 2022; however, we did not detect this mutation).

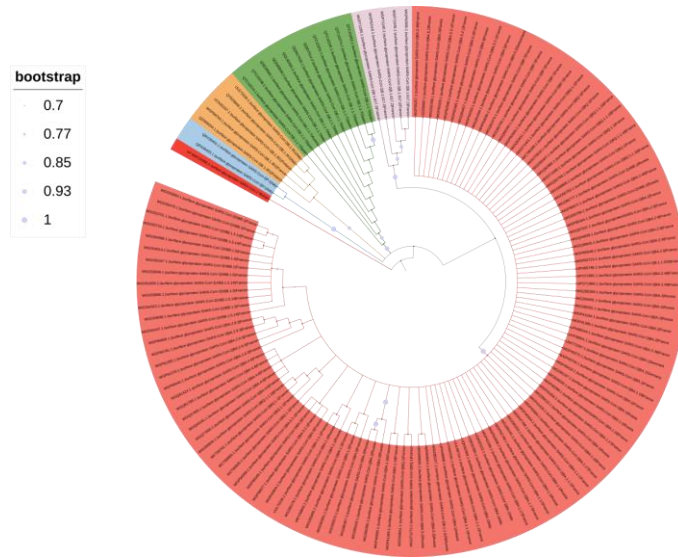
According to Varghese and Stanislaus (2023), N501Y, K417N, D614G, and T478K are the base mutations characteristic of S protein Omicron variants. Our results indicated similarity compared to their findings regarding Omicron mutations characteristic for S protein. Delta Spike mutations we detected, in line with the literature, are L452R and D614G, along with the Gamma mutations K417T, E484K, N501Y, and D614G. In addition, the P681H characteristic for the Alpha variant, was found in the Gamma variant in our research. Deletions that affect the viral dynamics by

increased infectivity are primarily characteristic of the S1 non-RBD region, namely:  $\Delta N69/\Delta V70/\Delta I156/\Delta I157/\Delta Y144/\Delta L242/\Delta A243/\Delta L244$  (Papanikolaou et al., 2022). Our evidence supports that mentioned mutations are present in our data except for  $\Delta L242$  and  $\Delta A243$ . Further evidence (see [Supplementary Table 2](#)), supports that Y144del was detected in Omicron lineages. The analysis included protein sequences of the S protein from 137 samples, along with the reference genome from Wuhan (n=138) downloaded from the NCBI database, whereby the phylogenetic tree was rooted at the RefSeq from Wuhan.

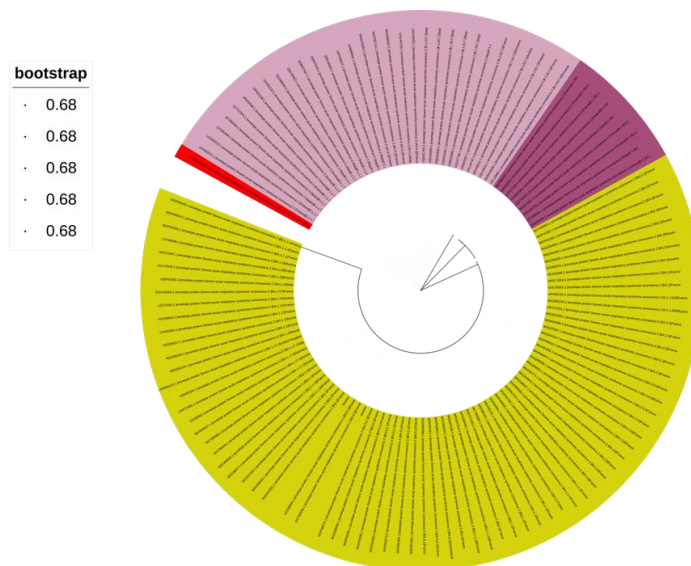
Our goal was to highlight the differences between the VOCs and their distribution within the phylogenetic tree, revealing clusters that emerged due to various mutations, particularly numerous in the S protein. Notably, Omicron forms a monophyletic clade among the other variants in this analysis. The genomic sequence of the SARS-CoV-2 S protein is of significant interest due to its functional importance and its high susceptibility to mutations, which greatly influence the mosaic properties of the virus.

#### *Envelope protein*

We reported four mutations ([Supplementary Table 3](#)) in the sequences included in the analysis. Variant P71L was present in a total of nine samples of the Beta lineage, while V62F was observed in the B.1.617.2 lineage. According to the global research, T9I and P71L mutations were identified as the most frequent E protein changes, followed by V62F, L21F/V and V58F (Abavisani et al. 2022). In our study, we did not detect L21F/V and V58F, while S50I was detected in our results, despite being not so often cited as an important E protein mutation.



**Figure 4.** Distance-based NJ phylogenetic tree of the Spike protein sequences composed of all VOCs representing the evolutionary course of SARS-CoV-2 spread. Each of the variants is represented as a subclade. Tree was rooted at the RefSeq from Wuhan highlighted as red branch. The following branches are presenting different VOCs: Alpha (green), Beta (pale orange), Gamma (light blue), Delta (lilac) and Omicron (red). Final tree was visualized using iTOL (Letunic & Bork, 2016).



**Figure 5.** Phylogenetic tree of the Envelope protein constructed by the NJ method with 1000 bootstrap iterations, colored by the encompassing VOCs, namely Alpha, Gamma and Delta in dark purple, Beta in lilac, and Omicron in yellow. The final tree was visualized using iTOL (Letunic & Bork, 2016).

The phylogenetic analysis of the E protein revealed that the Omicron variant has the highest total number of mutations compared to the other VOCs.

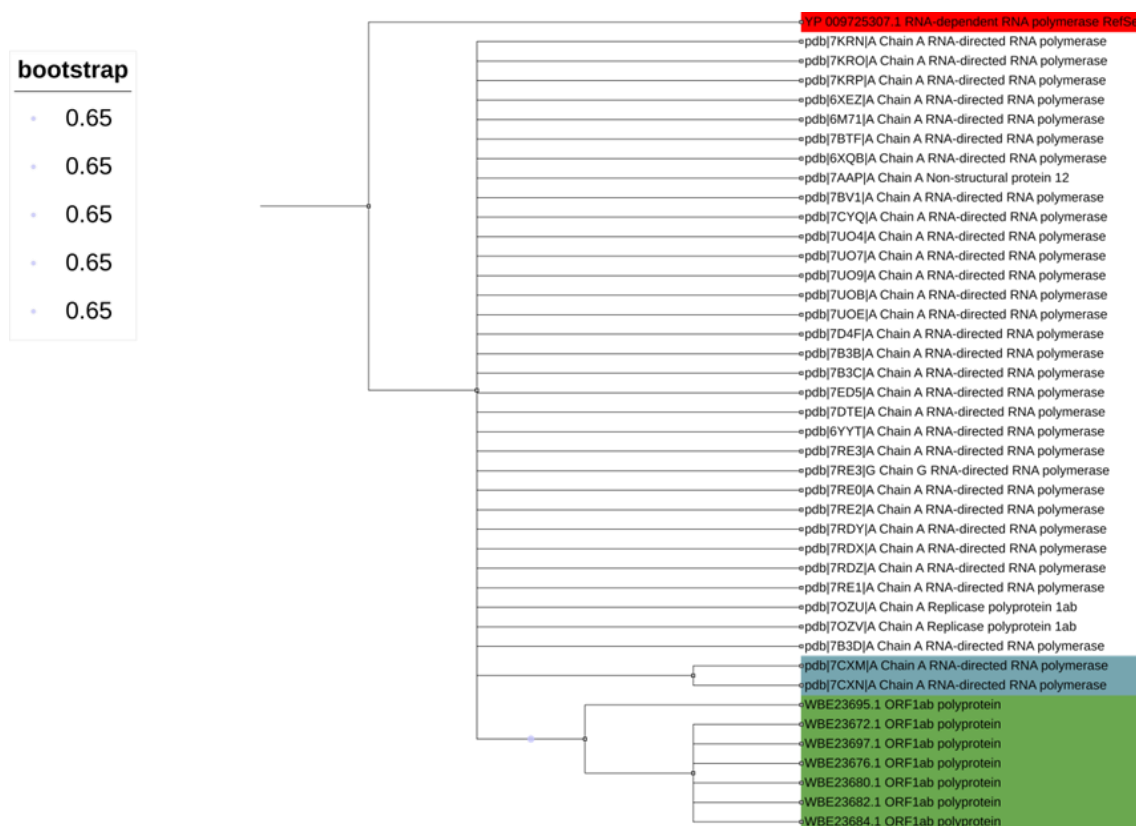
### RdRp

The leading and characteristic mutations of RdRp reported for Omicron are A1892T, I189V, P314L, K38R, T492I, and V57V (Bansal and Kumar, 2022). Subsequent changes in RNA editing and RNA replication errors are the main reasons for the emergence of further mutations (Wang et al., 2023), such as P323L caused by the 14408C>T mutation on RdRp, which results in high mutation rates throughout the rest of the viral genome,

despite being predicted to increase the protein stability.

In addition, mutations A97V and A185V can alter the secondary structure of a protein (Rahimi et al., 2021). It was also reported that P323L and A185V could eventually affect the secondary structure of the Nsp12 protein (Vilar and Isom, 2021).

The critical mutation for the B.1 lineage is P323L (P4715L) (Jacob et al., 2021), which is also the only RdRp mutation frequently reported in the literature which we also identified in this study. Three other mutation that we identified (Figure 6, [Supplementary Table 4](#)) are not reported in literature, but are predicted to decrease the protein stability (Table 1).



**Figure 6.** NJ-based phylogenetic tree of RdRp, supported by 1000 bootstrap iterations. The tree is rooted with RefSeq entry (red highlight). The regions colored in green and blue indicate the samples where most mutations were found, including M196R, P323L, G671S, and C645G. Final tree was visualized using iTOL (Letunic and Bork, 2016).

## Discussion

Evolutionary changes and the overall development of the virus are monitored by the viral genome research (Souza et al., 2022). Omicron (B.1.1.529) contains a major part of its modifying mutations in the RBD region of S glycoprotein (Han et al., 2022). The Omicron variant was reported to include 32 known mutations on the Spike protein, such as 69–70 del, K417N, T478K, N501Y, and P681R (Akkız, 2022); 15 substitutions are also reported in another work (Han et al., 2022); and this variant is otherwise known for containing increased number of mutations when compared to previous variants (Akkız, 2022).

The appearance of the XBB.1.5 'Kraken' variant, suppressed previously common variants, mainly due to its increased affinity towards the hACE2 receptor, marked by the F486P mutation on the Spike protein (Parums, 2023). However, the results of the binding analysis of the XBB.1.5 variant showed that it is not a highly threatening variant for global health, compared to the previous variants (Scarpa et al., 2023). EG.5 (Eris) was first reported by the WHO on 17th February 2023, and was declared VOI on 8th August 2023. Eris is a subvariant of Omicron, and during July and August of the same year, a more significant rise in the number of new infections was recorded mainly due to this variant.

The uniqueness of the Eris subvariant is reflected in the F456L mutation (Parums, 2023). In addition, new subvariant of Eris, EG.5.1 was described with new S protein mutation Q52H (Bilasy et al., 2024). Later, in September 2023, a new variant, BA.2.86 (or Pirola), was recommended for monitoring and was described as more resistant to monoclonal antibodies (mAbs) having ~35 mutations distinct from XBB.1.5 (Qu et al., 2023). BA.2.86 was also described as stable mutational variant, with higher affinity for receptor interactions, giving it a

potential for increased severity and transmissibility (Roy, 2023).

It was noted that the combination of N501Y, E484K, and K417N/T mutations could significantly affect the resistance and sensitivity to the currently designed therapy to control the infection (Akkız, 2022). All future approaches in the research of the new mutations are important because any failure can affect the research of new variants, leading to misinterpreting the findings or encountering unwanted challenges (Burns et al., 2022). Based on protein research, scientists can design novel drugs or proceed towards targeted therapy (van de Leemput and Han, 2021). Sharing data in science is crucial, especially during pandemics, as it forms the foundation for all health-related measures both locally and globally (Pan et al., 2021).

S, E, and RdRp are essential proteins; their mutations have high information content, and our work aimed to present the diversity of identified mutations. E protein is known to be highly conserved in terms of its nucleotide and protein sequence; contrary, the S protein accumulates mutations much faster, while RdRp is an essential and complex protein. We did not include the Nsps in the analysis because the mutations are of low frequency and are persistent; they do not offer much information in terms of drug design, epidemics, and immunology. Also, accessory proteins need to be more researched, since available information is still speculative and under further testing and validation.

Although E and RdRp proteins are known to mutate poorly, their mutations and the detection of new mutations significantly affects the detection of the SARS-CoV-2 virus since these two genes are common targets of PCR diagnostic tests. A rapid and accurate diagnostic strategy is the key link for managing the pandemic and monitoring the

movement of the virus, which is a very important implication for future pandemics as well (Bilasy et al., 2024).

## Conclusion

The results from our study provide evidence of the rapid evolution of the SARS-CoV-2 virus, particularly in certain genomic regions, which is significant for predicting and comparing virus variants. This approach is also used in forecasting the course of a pandemic, developing optimal immunization and treatment strategies, and monitoring animal reservoirs of the virus (Cheng et al., 2023). Our findings confirm that during the vaccination period, selective pressure favored the viral evolution, representing an adaptive response (Duerr et al., 2023; Rando et al., 2023). This led to a significant increase in the number of new variants and mutations detected. Consequently, while symptoms have become less severe, the virus continues to modify its genetic code to maintain its presence in the host population for an extended period.

## Authors' contributions

**DŽ. Omerkić Dautović:** conceptualization, formal analysis, investigation, visualization, original draft.

**A. Džuh:** review & editing, investigation, data curation.

**A. Ašić:** methodology, supervision, validation, review & editing.

## Conflict of interest

Authors declare no conflict of interest.

## References

Abavisani M, Rahimian K, Mahdavi B, Tokhanbigli S, Mollapour Siasakht M, Farhadi A, Kodori M, Mahmanzar M, Meshkat Z (2022) Mutations in SARS-

CoV-2 structural proteins: a global analysis. *Virology* 19(1), 220.

Abdolreza E, Fereshteh E, Armin JM, Amir S (2024) EG.5 (Eris) and BA.2.86 (Pirola) two new subvariants of SARS-CoV-2: a new face of old COVID-19. *Infection* 52(2):337-343

Akkız H (2022) The Biological Functions and Clinical Significance of SARS-CoV-2 Variants of Concern. *Frontiers in medicine*, 9, 849217.

Aleem A, Akbar Samad A B, Slenker A K (2022) Emerging Variants of SARS-CoV-2 And Novel Therapeutics Against Coronavirus (COVID-19). In *StatPearls*. StatPearls Publishing.

Alkhatib M, Salpini R, Carioti L, Ambrosio FA, D'Anna S, Duca L, Costa G, Bellocchi MC, Piermatteo L, Artese A, Santoro MM, Alcaro S, Svicher V, Ceccherini-Silberstein F (2022) Update on SARS-CoV-2 Omicron Variant of Concern and Its Peculiar Mutational Profile. *Microbiology spectrum*, 10(2), e0273221.

Attwood SW, Hill SC, Aanensen DM, Connor TR, Pybus OG (2022) Phylogenetic and phylodynamic approaches to understanding and combating the early SARS-CoV-2 pandemic. *Nature reviews. Genetics*, 23(9), 547–562.

Bai C, Zhong Q, Gao, GF (2022) Overview of SARS-CoV-2 genome-encoded proteins. *Science China. Life sciences*, 65(2), 280–294.

Bansal K, Kumar S (2022) Mutational cascade of SARS-CoV-2 leading to evolution and emergence of omicron variant. *Virus research*, 315, 198765.

Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE (2000) The protein data bank. *Nucleic Acids Res*, 28(1):235–242.

Bilasy SE, Wahyuni TS, Ibrahim M, El-Shamy A (2024) What SARS-CoV-2 Variants Have Taught Us: Evolutionary Challenges of RNA Viruses. *Viruses*, 16(1), 139.

Brintnell E, Gupta M, Anderson DW (2021) Phylogenetic and Ancestral Sequence Reconstruction of SARS-CoV-2

- Reveals Latent Capacity to Bind Human ACE2 Receptor. *Journal of molecular evolution*, 89(9-10), 656–664.
- Burns BL, Moody D, Tu ZJ, Nakitandwe J, Brock JE, Bosler D, Mitchell SL, Loeffelholz MJ, Rhoads DD (2022) Design and Implementation of Improved SARS-CoV-2 Diagnostic Assays To Mitigate the Impact of Genomic Mutations on Target Failure: the Xpert Xpress SARS-CoV-2 Experience. *Microbiology spectrum*, 10(6), e0135522.
- Cao Y, Yang R, Lee I, Zhang W, Sun J, Wang W, Meng X (2021) Characterization of the SARS-CoV-2 E Protein: Sequence, Structure, Viroporin, and Inhibitors. *Protein science : a publication of the Protein Society*, 30(6), 1114–1130.
- Cheng J, Randall A, Baldi P (2006) Prediction of protein stability changes for single-site mutations using support vector machines. *Proteins*, 62(4), 1125–1132.
- Cheng N, Liu M, Li W, Sun B, Liu D, Wang G, Shi J, Li L (2023) Protein post-translational modification in SARS-CoV-2 and host interaction. *Frontiers in immunology*, 13, 1068449.
- Cong Y, Ulasli M, Schepers H, Mauthe M, V'kovski P, Kriegenburg F, Thiel V, de Haan CAM, Reggiori F (2020) Nucleocapsid Protein Recruitment to Replication-Transcription Complexes Plays a Crucial Role in Coronaviral Life Cycle. *Journal of virology*, 94(4), e01925-19.
- Duerr R, Dimartino D, Marier C, Zappile P, Wang G, François F, Ortigoza M B, Iturrate E, Samanovic MI, Mulligan MJ, Heguy A (2023) Selective adaptation of SARS-CoV-2 Omicron under booster vaccine pressure: a multicentre observational study. *EBioMedicine*, 97, 104843.
- Frazier MN, Riccio AA, Wilson IM, Copeland WC, Stanley RE (2022) Recent insights into the structure and function of coronavirus ribonucleases. *FEBS open bio*, 12(9), 1567–1583.
- Ghorbani A, Samarfard S, Eskandarzade N, Afsharifar A, Eskandari MH, Niazi A, Izadpanah K, Karbanowicz TP (2021) Comparative phylogenetic analysis of SARS-CoV-2 spike protein-possibility effect on virus spillover. *Briefings in bioinformatics*, 22(5), bbab144.
- Giovanetti M, Benedetti F, Campisi G, Ciccozzi A, Fabris S, Ceccarelli G, Tambone V, Caruso A, Angeletti S, Zella D, Ciccozzi M (2021) Evolution patterns of SARS-CoV-2: Snapshot on its genome variants. *Biochemical and biophysical research communications*, 538, 88–91.
- Gomari MM, Tarighi P, Choupani E, Abkhiz S, Mohamadzadeh M, Rostami N, Sadroddiny E, Baammi S, Uversky VN, Dokholyan NV (2023) Structural evolution of Delta lineage of SARS-CoV-2. *International journal of biological macromolecules*, 226, 1116–1140.
- Gusev E, Sarapultsev A, Solomatina L, Chereshev V (2022) SARS-CoV-2-Specific Immune Response and the Pathogenesis of COVID-19. *International journal of molecular sciences*, 23(3), 1716.
- Hall BG (2013) Building phylogenetic trees from molecular data with MEGA. *Molecular biology and evolution*, 30(5), 1229–1235.
- Hamdy M E, El Deeb A H, Hagag N M, Shahein M A, Alaidi O, Hussein H A (2023) Interspecies transmission of SARS CoV-2 with special emphasis on viral mutations and ACE-2 receptor homology roles. *International journal of veterinary science and medicine*, 11(1), 55–86.
- Han P, Li L, Liu S, Wang Q, Zhang D, Xu Z, Han P, Li X, Peng Q, Su C, Huang B, Li D, Zhang R, Tian M, Fu L, Gao Y, Zhao X, Liu K, Qi J, Gao G F, Wang P (2022) Receptor binding and complex structures of human ACE2 to spike RBD from omicron and delta SARS-CoV-2. *Cell*, 185(4), 630–640.e10.
- Ito J, Suzuki R, Uriu K, Itakura Y, Zahradnik J, Kimura KT, Deguchi S, Wang L, Lytras S, Tamura T, Kida I, Nasser H, Shofa M, Begum MM, Tsuda M, Oda Y, Suzuki T, Sasaki J, Sasaki-Tabata K, Fujita S, Sato K (2023) Convergent evolution of SARS-CoV-2 Omicron subvariants leading to the emergence of BQ.1.1 variant. *Nature communications*, 14(1), 2671.
- Jacob JJ, Vasudevan K, Pragasam AK, Gunasekaran K, Veeraraghavan B, Mutreja A (2021) Evolutionary

- Tracking of SARS-CoV-2 Genetic Variants Highlights an Intricate Balance of Stabilizing and Destabilizing Mutations. *mBio*. 31;12(4):e0118821.
- John G, Sahajpal NS, Mondal AK, Ananth S, Williams C, Chaubey A, Rojiani AM, Kolhe R (2021) Next-Generation Sequencing (NGS) in COVID-19: A Tool for SARS-CoV-2 Diagnosis, Monitoring New Strains and Phylodynamic Modeling in Molecular Epidemiology. *Current issues in molecular biology*, 43(2), 845–867.
- Kumar S, Stecher G, Li M, Knyaz C, Tamura K (2018).MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. *Molecular biology and evolution*, 35(6), 1547–1549.
- Jung C, Kmiec D, Koepke L, Zech F, Jacob T, Sparrer KMJ, Kirchhoff F (2022) Omicron: What Makes the Latest SARS-CoV-2 Variant of Concern So Concerning?. *Journal of virology*, 96(6), e0207721.
- Kandeel M, Mohamed M E M, Abd El-Lateef H M, Venugopala K N, El-Beltagi H S (2022) Omicron variant genome evolution and phylogenetics. *Journal of medical virology*, 94(4), 1627–1632.
- Kapoor S, Singh A, Gupta V (2023) In silico evaluation of potential intervention against SARS-CoV-2 RNA-dependent RNA polymerase. *Physics and chemistry of the earth* (2002), 129, 103350.
- Khan A, Khan M, Saleem S, Babar Z, Ali A, Khan AA, Sardar Z, Hamayun F, Ali SS, Wei DQ (2020) Phylogenetic Analysis and Structural Perspectives of RNA-Dependent RNA-Polymerase Inhibition from SARs-CoV-2 with Natural Products. *Interdisciplinary sciences, computational life sciences*, 12(3), 335–348.
- Kimura I, Yamasoba D, Nasser H, Zahradnik J, Kosugi Y, Wu J, Nagata K, Uriu K, Tanaka YL, Ito J, Shimizu R, Tan TS, Butlertanaka EP, Asakura H, Sadamasu K, Yoshimura K, Ueno T, Takaori-Kondo A, Schreiber G, Genotype to Phenotype Japan (G2P-Japan) Consortium, Sato K (2022) The SARS-CoV-2 spike S375F mutation characterizes the Omicron BA.1 variant. *iScience*, 25(12), 105720.
- Kumavath R, Barh D, Andrade BS, Imchen M, Aburjaile FF, Ch A, Rodrigues DL, Tiwari S, Alzahrani KJ, Góes-Neto A, Weener ME, Ghosh P, Azevedo V (2021) The spike of SARS-COV-2: Uniqueness and applications. *Frontiers in Immunology*, 12.
- Letunic I, Bork P (2016) Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Research*. 2016;44(W1):W242–W245.
- Low ZY, Zabidi NZ, Yip AJW, Puniyamurti A, Chow VTK, Lal SK (2022) SARS-CoV-2 Non-Structural Proteins and Their Roles in Host Immune Evasion. *Viruses*, 14(9), 1991.
- Mansbach RA, Chakraborty S, Nguyen K, Montefiori DC, Korber B, Gnanakaran S (2021) The SARS-CoV-2 Spike variant D614G favors an open conformational state. *Science advances*, 7(16), eabf3671.
- Naqvi AAT, Fatima K, Mohammad T, Fatima U, Singh IK, Singh A, Atif SM, Hariprasad G, Hasan GM, Hassan MI (2020) Insights into SARS-CoV-2 genome, structure, evolution, pathogenesis and therapies: Structural genomics approach. *Biochimica et biophysica acta. Molecular basis of disease*, 1866(10), 165878.
- Pan B, Ji Z, Sakkiah S, Guo W, Liu J, Patterson TA, Hong H (2021) Identification of Epidemiological Traits by Analysis of SARS-CoV-2 Sequences. *Viruses*, 13(5), 764.
- Papanikolaou V, Chrysovergis A, Ragos V, Tsiambas E, Katsinis S, Manoli A, Papouliakos S, Roukas D, Mastronikolis S, Peschos D, Batistatou A, Kyrodimos E, Mastronikolis N (2022) From delta to Omicron: S1-RBD/S2 mutation/deletion equilibrium in SARS-CoV-2 defined variants. *Gene*, 814, 146134.
- Parums DV (2023) Editorial: The XBB.1.5 ('Kraken') Subvariant of Omicron SARS-CoV-2 and its Rapid Global Spread. *Medical science monitor : international medical journal of experimental and clinical research*, 29, e939580.



- Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE (2004) UCSF chimera—a visualization system for exploratory research and analysis. *J Comput Chem* 25(13):1605–1612.
- Pinato DJ, Aguilar-Company J, Ferrante D, Hanbury G, Bower M, Salazar R, Mirallas O, Sureda A, Plaja A, Cucurull M, Mesia R, Townsend S, Jackson A, Dalla Pria A, Newsom-Davis T, Handford J, Sita-Lumsden A, Aphorpe E, Vincenzi B, Bertuzzi A, OnCovid study group (2022) Outcomes of the SARS-CoV-2 omicron (B.1.1.529) variant outbreak among vaccinated and unvaccinated patients with cancer in Europe: results from the retrospective, multicentre, OnCovid registry study. *The Lancet. Oncology*, 23(7), 865–875.
- Prasada Kabekkodu S, Chakrabarty S, Jayaram P, Mallya S, Thangaraj K, Singh KK, Satyamoorthy K (2023) Severe acute respiratory syndrome coronaviruses contributing to mitochondrial dysfunction: Implications for post-COVID complications. *Mitochondrion*, 69, 43–56. Advance online publication.
- Qu P, Xu K, Faraone JN, Goodarzi N, Zheng YM, Carlin C, Bednash JS, Horowitz JC, Mallampalli RK, Saif LJ, Oltz EM, Jones D, Gumina RJ, Liu SL (2023) Immune Evasion, Infectivity, and Fusogenicity of SARS-CoV-2 Omicron BA.2.86 and FLip Variants. *bioRxiv* : the preprint server for biology, 2023.09.11.557206.
- Rahimi A, Mirzazadeh A, Tavakolpour S (2021) Genetics and genomics of SARS-CoV-2: A review of the literature with the special focus on genetic diversity and SARS-CoV-2 genome detection. *Genomics* vol. 113,1 Pt 2 (2021): 1221-1232.
- Rak A, Gorbunov N, Kostevich V, Sokolov A, Prokopenko P, Rudenko L, Isakova-Sivak I (2023) Assessment of Immunogenic and Antigenic Properties of Recombinant Nucleocapsid Proteins of Five SARS-CoV-2 Variants in a Mouse Model. *Viruses* 13;15(1):230.
- Rando HM, Lordan R, Lee AJ, Naik A, Wellhausen N, Sell E, Kolla L, COVID-19 Review Consortium, Gitter A, Greene C S (2023) Application of Traditional Vaccine Development Strategies to SARS-CoV-2. *mSystems*, 8(2), e0092722.
- Redondo N, Zaldívar-López S, Garrido JJ, Montoya M (2021) SARS-CoV-2 Accessory Proteins in Viral Pathogenesis: Knowns and Unknowns. *Frontiers in immunology*, 12, 708264.
- Roy U (2023) Computational Investigation of Selected Spike Protein Mutations in SARS-CoV-2: Delta, Omicron, and Some Circulating Subvariants. *Pathogens* (Basel, Switzerland), 13(1), 10.
- Saifi S, Ravi V, Sharma S, Swaminathan A, Chauhan NS, Pandey R (2022) SARS-CoV-2 VOCs, Mutational diversity and clinical outcome: Are they modulating drug efficacy by altered binding strength?. *Genomics*, 114(5), 110466.
- Sallard E, Halloy J, Casane D, Decroly E, van Helden J (2021) Tracing the origins of SARS-COV-2 in coronavirus phylogenies: a review. *Environmental chemistry letters*, 19(2), 769–785.
- Santopolo S, Riccio A, Santoro MG (2021) The biogenesis of SARS-CoV-2 spike glycoprotein: multiple targets for host-directed antiviral therapy. *Biochemical and biophysical research communications*, 538, 80–87.
- Scarpa F, Azzena I, Locci C, Casu M, Fiori PL, Ciccozzi A, Angeletti S, Imperia E, Giovanetti M, Maruotti A, Borsetti A, Cauda R, Cassone A, Via A, Pascarella S, Sanna D, Ciccozzi M (2023) Molecular In-Depth on the Epidemiological Expansion of SARS-CoV-2 XBB.1.5. *Microorganisms*, 11(4), 912.
- Schoeman D, Fielding BC (2019) Coronavirus envelope protein: current knowledge. *Virology journal*, 16(1), 69.
- Shahhosseini N, Babuadze GG, Wong G, Kobinger GP (2021) Mutation Signatures and In Silico Docking of Novel SARS-CoV-2 Variants of Concern. *Microorganisms*, 9(5), 926.
- Smart. Servier Medical Art. (2023, April 25). Retrieved July 27, 2023, from <https://smart.servier.com/>
- Souza PFN, Mesquita FP, Amaral JL, Landim PGC, Lima KRP, Costa MB, Farias IR, Belém MO, Pinto YO,

- Moreira HHT, Magalhaes ICL, Castelo-Branco DSCM, Montenegro RC, de Andrade CR (2022) 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. *International journal of biological macromolecules*, 208, 105–125.
- Valadan R, Golchin S, Alizadeh-Navaei R, Haghshenas M, Zargari M, Mousavi T, Ghamati M (2022) Differential gene expression analysis of common target genes for the detection of SARS-CoV-2 using real time-PCR. *AMB Express*, 12(1), 112.
- van de Leemput J, Han Z (2021) Understanding Individual SARS-CoV-2 Proteins for Targeted Drug Development against COVID-19. *Molecular and cellular biology*, 41(9), e0018521.
- Varghese Edwin H, Stanislaus Antony C (2023) An update on COVID-19: SARS-CoV-2 variants, antiviral drugs, and vaccines. *Heliyon*, e13952. Advance online publication.
- Vilar S, Isom DG (2021) One Year of SARS-CoV-2: How Much Has the Virus Changed?. *Biology*, 10(2), 91.
- Wang J, Wu L, Pu X, Liu B, Cao M (2023) Evidence Supporting That C-to-U RNA Editing Is the Major Force That Drives SARS-CoV-2 Evolution. *Journal of molecular evolution*, 1–11. Advance online publication.
- Wilson IM, Frazier MN, Li JL, Randall TA, Stanley RE (2022) Biochemical Characterization of Emerging SARS-CoV-2 Nsp15 Endoribonuclease Variants. *Journal of molecular biology*, 434(20), 167796.
- Wolf JM, Wolf LM, Bello GL, Maccari JG, Nasi LA (2023) Molecular evolution of SARS-CoV-2 from December 2019 to August 2022. *Journal of medical virology*, 95(1), e28366.
- Yan L, Zhang Y, Ge J, Zheng L, Gao Y, Wang T, Jia Z, Wang H, Huang Y, Li M, Wang Q, Rao Z, Lou Z (2020) Architecture of a SARS-CoV-2 mini replication and transcription complex. *Nature communications*, 11(1), 5874.
- Yang H, Rao Z (2021) Structural biology of SARS-CoV-2 and implications for therapeutic development. *Nature reviews. Microbiology*, 19(11), 685–700.
- Yevsieieva LV, Lohachova KO, Kyrychenko A, Kovalenko SM, Ivanov VV, Kalugin ON (2023) Main and papain-like proteases as prospective targets for pharmacological treatment of coronavirus SARS-CoV-2. *RSC advances*, 13(50), 35500–35524.
- Yoshimoto FK (2020) The Proteins of Severe Acute Respiratory Syndrome Coronavirus-2 (SARS CoV-2 or n-COV19), the Cause of COVID-19. *The protein journal*, 39(3), 198–216.
- Zhao N, Zhou N, Fan H, Ding J, Xu X, Dong X, Dong X, Xu D, Min X, Yu Y, Gong H, Mao L, He M (2022) Mutations and Phylogenetic Analyses of SARS-CoV-2 Among Imported COVID-19 From Abroad in Nanjing, China. *Frontiers in microbiology*, 13, 851323.
- Zhou Q, Chen Y, Wang R, Jia F, He F, Yuan F (2022) Advances of CRISPR-Cas13 system in COVID-19 diagnosis and treatment. *Genes & diseases*, 10.1016/j.gendis.2022.11.016. Advance online publication.
- Zhou S, Lv P, Li M, Chen Z, Xin H, Reilly S, Zhang X (2023) SARS-CoV-2 E protein: Pathogenesis and potential therapeutic development. *Biomedicine & pharmacotherapy = Biomedecine & pharmacotherapie*, 159, 114242.