JOURNAL OF
BIOLOGICAL
INSIGHTS

Vol. 1 | Issue 1 | pp:1-29 | 2025 |
DOI: 10.4197/jbi.1-1.1

Sharing Knowledge, Shaping the Future

Published by Scientific Publishing Center, King Abdulaziz University

Research Article

Received: 2025/10/04 | **Accepted**: 2025/11/12 | **Published**: 2025/12/29 |

Handling Editor: Irfan A. Rather | Department of Biological Sciences | King Abdulaziz University | Jeddah | Saudi Arabia

Phylogenomic Characterization of SARS-CoV-2 Variants in Jeddah, Saudi Arabia: Insights into Viral Evolution and Transmission Dynamics

Saleh Ali Arishi[1], Wael A. Fathy[1,2,*], Mohammed Al-matary[1], Ahmed Bahieldin[1], Sherif Edris[1], Ahmed M. Ramadan[1,*]

[1]Department of Biological Sciences, Faculty of Science, King Abdulaziz University, 21589 Jeddah, Saudi Arabia.
[2]Botany and Microbiology Department, Faculty of Science, Beni-Suef University, Beni-Suef, 62511, Egypt.

*Corresponding Authors: aamara@kau.edu.sa (A.M.R) | wfathy@stu.kau.edu.sa (W.A.F)

Abstract

Objective: The purpose of the study was to describe the genomic profile of SARS-CoV-2 variants in circulation in Jeddah, Saudi Arabia, as the country is a key pilgrimage destination and central hub in the Middle East, to guide the national health policies and learn about the evolution of the virus in the region. **Methods**: A total of 94 SARS-CoV-2 samples were collected in the city of Jeddah from January to December 2022. The whole-genome sequencing was done via the Illumina CovidSeq platform with the help of phylogenetic reconstruction (maximum likelihood), sequence alignment (Mauve), lineage assignment (Pangolin, Nextclade), and mutation profiling (with the focus on the spike gene) being conducted. **Results**: Phylogenetic analysis indicated that there were several clusters of closely related isolates, which showed that there was limited genetic diversity, and these were extensively transmitted locally. The BA.2 variant prevailed (59.57) and BA.1 (26.60), and small BA.2 sublineages were signs of continued viral evolution. Huge mutational hotspots were found in the spike gene (42.19%) and the ORF1ab region (35.94%), with single-nucleotide variants representing 92.31% of mutations. The most important spike protein mutations, including D614G (98.94%), P681H/N679K (100%), were almost fixed with multiple receptor-binding domain alterations involved in viral entry and immune evasion. **Conclusion**: BA.2 and its offshoots were the most dominant in Jeddah, Saudi Arabia, in 2022, and mutation patterns indicated adaptive evolution due to immune selection. The presence of concentrated mutations in key genome regions was an indication of continued adaptation to enhance viral transmissibility and immune evasion. The results provide the necessary background information on genomic surveillance and evidence-based Aramex policies in Saudi Arabia and other Middle East countries.

Keywords: SARS-CoV-2 | Saudi Arabia | Genomic investigation | Omicron variants | Viral mutations | Spike protein

1. Introduction

The SARS-CoV-2 pandemic requires a large-scale monitoring of genomes to monitor the evolution of the virus, record the dynamics of transmission, and make informed decisions on the strategies of population health [1]. Since the virus continues in the evolutionary processes and generates variants of concern, detailed phylogenetic and mutational studies have become key instruments for explaining the regional epidemiological trends and genetic variation [2, 3]. The Kingdom of Saudi Arabia is in a unique position regarding the global SARS-CoV-2 surveillance because it is considered one of the main pilgrimage destinations [4], whereas the country hosts millions of Muslim pilgrims annually to Hajj and Umrah, which creates the possibility of hotspots of viral spread and development of variations [5, 6]. This geographic and demographic importance, along with the fact that Saudi Arabia has a substantial burden of COVID-19 of more than 816,000 cases and is in a strategic location in the Middle East, underscores the importance of genomic characterization of viral strains that circulate in a detailed manner [7, 8].

Earlier studies conducted in the Middle East have revealed a high degree of genetic diversity in the SARS-CoV-2 populations [9], and phylogenetic analyses have revealed a complex set of transmission processes and a set of regional-specific mutations [10]. Additionally, studies conducted in the Middle Eastern countries have revealed more than 2200 different genome variants, with a myriad of missense variants, frame shifts, and deletions, which together promote viral adaptation and immunity escape [11]. The fact that the D614G mutation is common in the entire region and is on the verge of fixation is a demonstration of the selective advantage of specific replacements that increase the viral transmissibility [12, 13]. Furthermore, the incidence of various variants and subvariants has been documented through genomic surveillance efforts across the globe and particularly in the Middle East, with a focus on Omicron lineages that were predominant in the year 2022 and whose trends were consistent with those of the global epidemiology [14, 15].

The integration of whole-genome sequencing technologies and bioinformatics pipelines has also enabled the characterization of viral populations in detail on a scale never before achieved [16, 17]. The application of phylogenetic reconstruction using the maximum likelihood and Bayesian approaches provides information about the evolutionary patterns, routes of transmission, and the temporal patterns of variant appearance [18]. Also, multiple sequence alignment kits such as Mauve usually aim to compare genomic studies, consequently revealing conserved regions and detecting hotspots of mutations that influence viral phenotypes [19, 20]. Moreover, lineage classification systems, such as the Pango nomenclature, and analytical tools such as Nextclade, can be used to standardise the identification of variants and enable real-time tracking of strains in circulation [21, 22]. The current study is a genomic study of 94 samples of SARS-CoV-2 collected in Jeddah, Saudi Arabia, in 2022 and analyzes these samples using a range of complementary techniques to provide insight into the diversity of the virus and its evolution. The study seeks to offer information on the molecular epidemiology of SARS-CoV-2 in KSA. The focus of this investigation is the mutations in the spike protein, which are credited with their critical role in the entry mechanisms of the virus,

immune system evasion, as well as classification of viral variants. Overall, this study contributes to the understanding of the evolutionary processes of SARS-CoV-2 in the Middle Eastern region, as it evaluates the prevalence of mutations in different genomic areas and functional domains. In addition, it offers vital information upon which ongoing surveillance activities and development of policy on the health of the population rely.

2. Materials and Methods

2.1 Sample collection

Consecutive sampling collection was conducted in 12 months (January-December 2022) when samples that were positive on reverse transcription polymerase chain reaction included in the study became available. The distribution of the collected specimens by time showed that they varied over the months and reflected natural changes in test volume, positivity rate, and specimen quality at various points of the 2022 pandemic. Sample acquisition to perform whole-genome sequencing (WGS) was performed at Alborg Laboratories, Jeddah, Saudi Arabia (Figure 1) simultaneously. Potential inclusion was screened with residual nasopharyngeal swab samples of patients who were previously tested with real-time RT-PCR to test SARS-CoV-2. Eligible specimens were confirmed to be RT-PCR-positive with a cycle threshold (Ct) value ≤30 to ensure sufficient viral RNA output for sequencing. Sample collection was limited to a maximum of seven days of symptom onset in symptomatic cases to maximize the RNA integrity. Also, participants were required to be either residents or visitors of Jeddah during sampling. The specimens were eliminated when they were found to have a Ct value greater than 30, indicated RNA degradation, missing clinical or epidemiological metadata, or duplication of the same individual. The outcome of this selection was the identification of 94 specimens that met all the inclusion criteria to be included in sequencing. The remaining swab material was subjected to standardized laboratory conditions of viral RNA extraction to preserve the integrity of nucleic acids. Whole genome sequencing was then conducted using the Illumina CovidSeq Kit (Illumina, San Diego, CA, USA) according to the protocol, and thereby high-quality genomic data were generated that were to be used in downstream analyses.

Figure 1. Shows the geographical representation of Saudi Arabia, and the specific location of Jeddah city as given by Hussein et al. [23]. The exact position of the Alborg main branch (21°37'22.5"N 39°09'17.1" E) is indicated in this map, and it was used to collect the samples.

2.2 RNA extraction and sequencing

RNA was extracted using the automated Abbott M1000 nucleic acid extraction system along with QIAamp Nucleic Acid Extraction Kit (Qiagen, Hilden, Germany) as per the manufacturer's instructions. The eluted RNA was further subjected to whole-genome sequencing by using the amplicon-based Illumina CovidSeq standard protocol, which allowed thorough amplification of the SARS-CoV-2 genome. The ready sequencing libraries were then loaded onto the Illumina MiSeq platform and sequenced using the 600-cycle v3 MiSeq Reagent Kit, following the recommended workflow provided by the manufacturer to ensure high-quality and high-coverage genome data that could be used to do downstream analysis.

2.3 Bioinformatics and data analysis

Initial processing of the raw sequencing data was done using CLC Genomics Workbench v20.0.2 (QIAGEN) and DNASTAR Lasergene v18 (DNASTAR Inc., USA) to perform base calling, quality trimming, and initial read alignment. Poor reads with a Phred score below 30 or less than 50 base pairs were filtered out in order to reduce sequencing artifacts. The derivation of consensus sequences was after alignment with the SARS-CoV-2 reference genome (Wuhan-Hu-1, GenBank accession: NC_045512.2). Unclear base calls were sorted by the manual separation of chromatograms of Sanger sequencing data or read pile-ups of next-generation sequencing (NGS) data. The Mauve software was used to perform multiple sequence alignment. Mutations were annotated at both the nucleotide and amino acid levels, with particular focus given to substitutions, additions, and deletions relevant to the variants of concern (VOCs). To classify lineages and clades, the Pangolin COVID-19 Lineage Assigner was

used to designate Pango lineages, with Nextclade v2.5 doing the clade assignment, mutation profiling, and general metrics of sequence quality. The Maximum Likelihood analysis was used to explain phylogenetic relationships. This end-to-end analytical pipeline supported the effective identification of SARS-CoV-2 mutations, accurate lineage and clade assignment of the analysis, and a comprehensive analysis of the evolutionary relationship among the strains of SARS-CoV-2 in circulation in Jeddah, and enhanced continuing genomic surveillance and monitoring of viral evolution. The dimensional structural mapping of the mutations that were identified was performed with the GISAID CoVsurver platform, which is a platform that combines the cryo-electron microscopy (cryo-EM) structures to enable visualization of mutations on the existing protein data bank structures. In this analysis, two complementary structural models were used: the prefusion spike trimer (PDB: 6acc) resolved at 3.6 A by cryo-EM, which was presented in the receptor-binding domain down conformation, and the spike-ACE2 complex (PDB: 6acj) resolved at 4.2 A by cryo-EM, which was presented in the open receptor-engaged state.

3. Results

3.1 Phylogenetic relationship

As depicted in Figure 2, the circular phylogeny of the 94 SARS-CoV-2 genomes exhibits multiple clusters of closely related isolates, indicative of recent shared ancestry within the sampled population. While many isolates are interconnected by short branches, this suggests a high degree of sequence similarity and low genetic diversity amongst these isolates. Additionally, several smaller clusters can be identified, implying the presence of closely related subgroups or potential transmission chains. These clusters can be indicative of local outbreaks, rapid spread events, or a shared source of infection. The isolates in a cluster will typically have limited genetic variations when compared to other isolates in the tree, while those located on single long branch segments indicate outlier mutations or unrepresented diversity. A few tip sequences appear to differ from the remainder of the tree by being substantially longer than the others and therefore, include more mutations, potential for unique genomics through insertions/deletions or amino acid substitutions, or an introduction into the tree from separate transmission pathways. Outlier sequences are of particular interest because they may signify individual introductions of the virus, error(s) in the sequence data collection process, or unusual evolutionary changes.

Another interpretation of the alignment results (Figure 3) is that most of the samples are characterized by a large sequence similarity, i.e., the large stretches of green horizontal bars with very few breaks. This consistency shows that there are conserved regions in the viral genome across the entire population that was studied. The vertical lines and gaps of different colors indicate the positions of the mutations or genomic variations, where red, yellow, and blue vertical lines are the types of different substitutions or structural changes that occurred in the 94 sequences. The presence of distinct horizontal gaps or missing regions indicates deletions, lack of coverage, or unclear sequence data in some samples. There are no uniform distributions of mutations, and multiple loci show concentrated vertical activity, especially in

the 5,000-10,000 bp region and around 25,000 bp. The distribution of differentiating mutations is concentrated in the middle (between 5,000-10,000 bp) and the end (about 25,000 bp) of the genome, which suggests that there are genomic regions in which the viral population experiences more variation. Minor variations are also scattered at other loci, which shows evidence of sporadic point mutations.

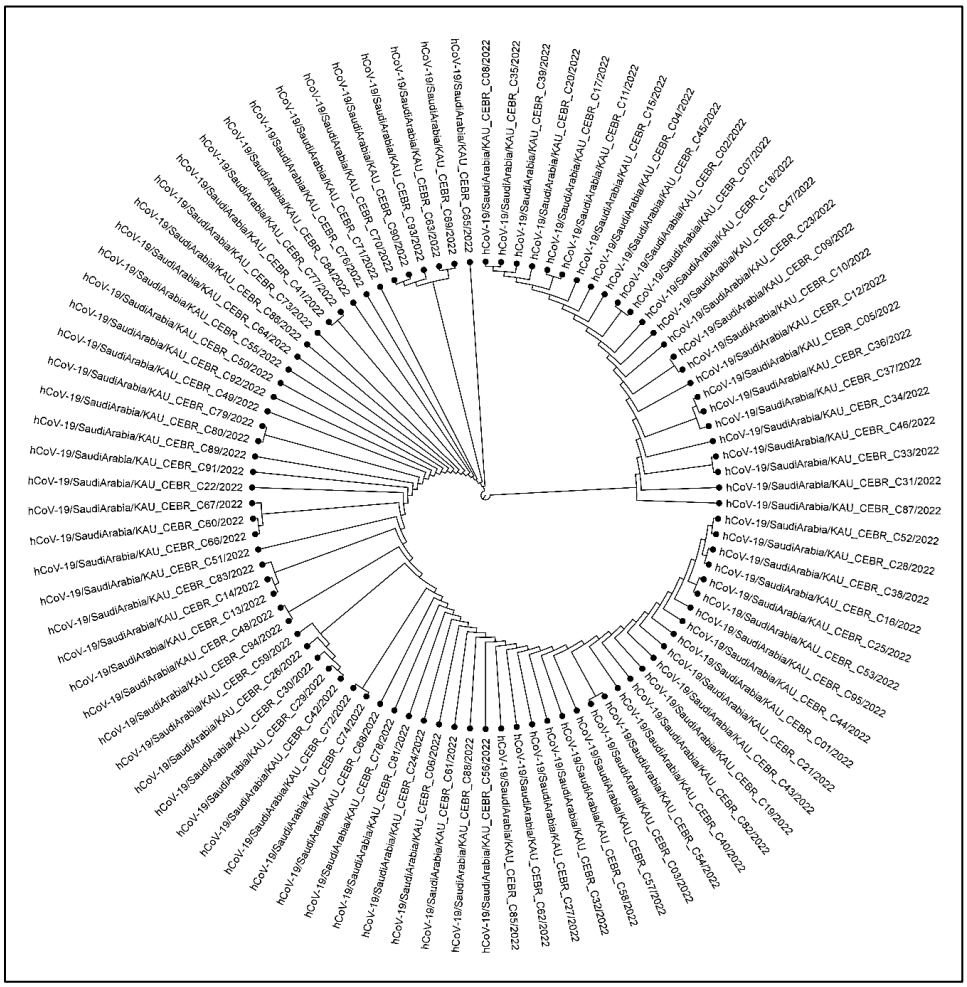

Figure 2. Shows the phylogenetic tree results that interpret the relationships among SARS-CoV-2 genome samples.

While the Nextclade Pango Lineage (as illustrated in Figure 4) gives a detailed phylogenomic and mutational study of 94 SARS-CoV-2 genomes using Nextclade and Pango lineage classification, in two panels: a circular phylogenetic tree that shows the lineage diversity and a mutation frequency landscape aligned with viral genes. The top panel contains a radial phylogenetic tree with the names of SARS-CoV-2 variants and Pango lineage subcategories. The most common genomes belong to Omicron sublineages, specifically BA.2 and its variants (BA.2.3, BA.2.10, BA.2.32, BA.2.5, BA.2.57, and BA.2.64), with the most diverse ones being

depicted in blue, purple, and yellow to emphasize their prevalence. The other smaller peripheral cluster is an expression of other variants of concern, including BA.1. The complex branching patterns of the Omicron lineages testify to its ongoing diversification and rapid evolution throughout the sampling time, and outgroup sequences provide the rooted evolutionary structure of the tree. The lower panel represents the frequency of mutation in the approximate 29800 base SARS-CoV-2 genome that is further divided into functional gene units such as ORF1a, ORF1b, and the spike gene. The frequency of mutation is not evenly distributed, with multiple hotspots that represent genetic hotspots. The presence of clusters in the ORF1a and ORF1b suggests functional diversification or selection sweeps related to replication and transcription, but the spike gene has noticeable peaks, which highlight its critical role in viral entry and evasion of the immune system. On the other hand, some areas have fewer mutations, implying the presence of functionally constrained segments of the genome that remain conserved. The analysis overall clarifies why the Omicron variant, in particular, BA.2 and its sublineages, were prevailing in the population under evaluation, with a specific point towards a rapid local diversification and transmission. The mutation clusters and the evolutionary pressures on the replication machinery and the spike protein are emphasized by the phylogenetic structures, driving the evolution of viruses and leading to changes that can change the ability to replicate, polymerase fidelity, receptor binding, and antigenic properties.

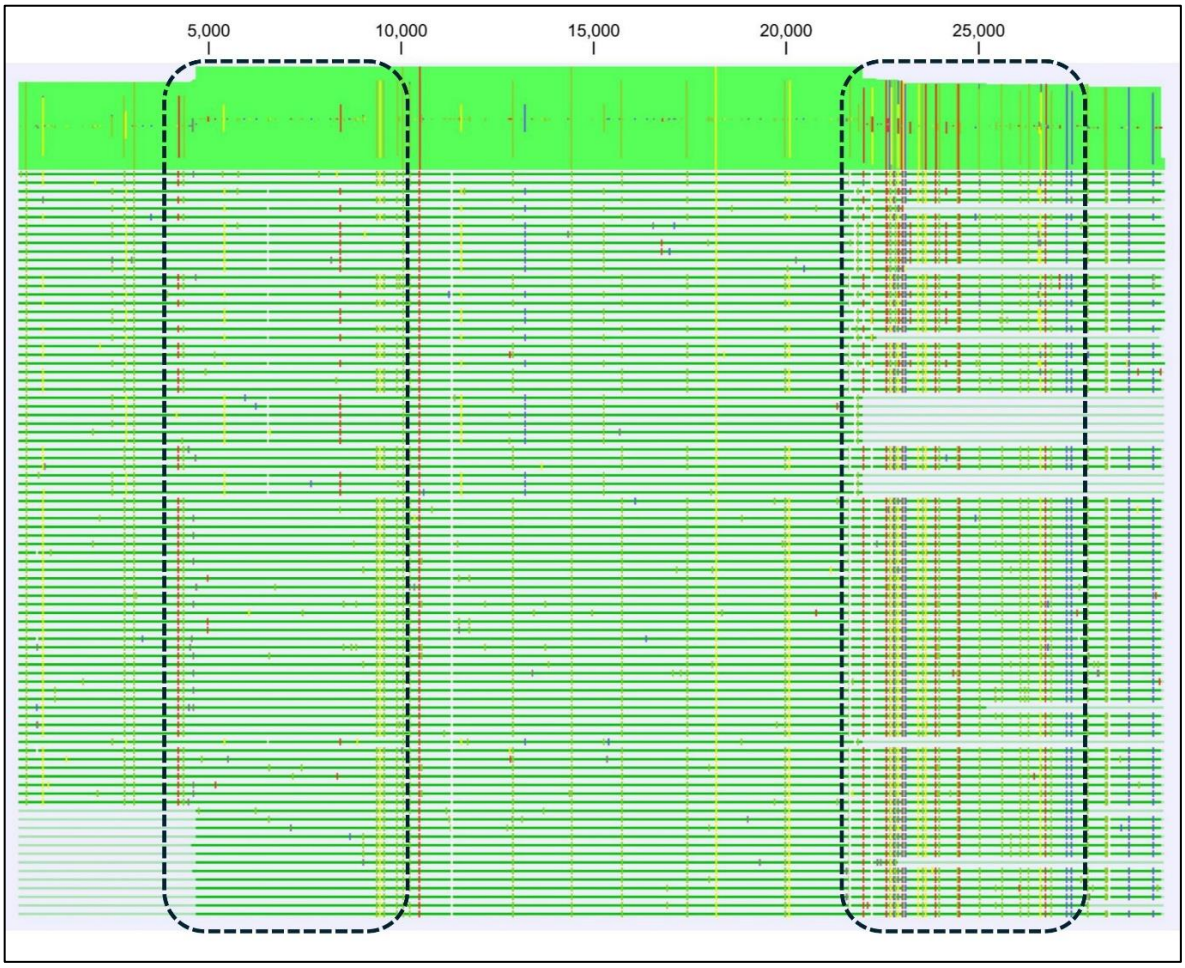

Figure 3. Multiple sequence alignment of the 94 sequences, revealing the presence of highly conserved regions (indicated in green) interspersed with mutational hotspots.

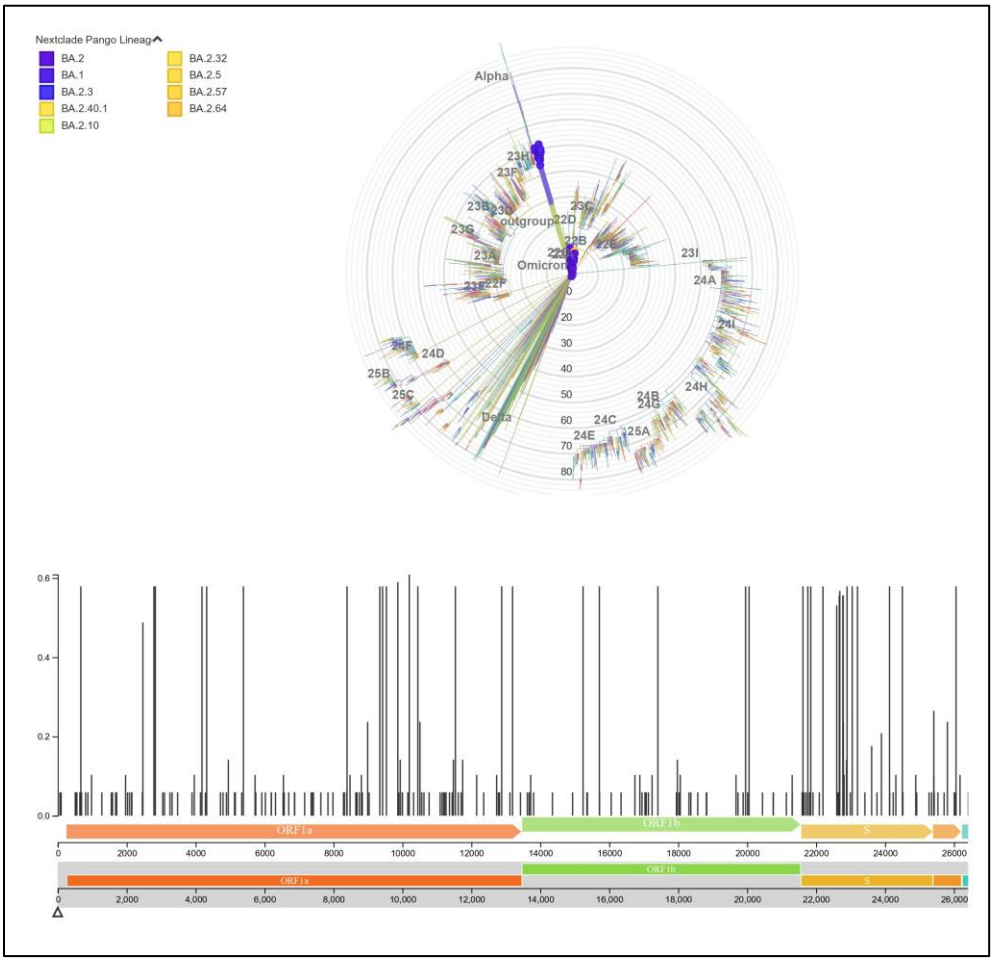

Figure 4. Shows the Nextclade Phylogenomic tree and the mutation frequency of 94 SARS-CoV-2 genomes, highlighting the predominance of the Omicron BA.2 variant and the identification of mutation hotspots within ORF1a, ORF1b, and spike regions.

3.2 Indel detection

The complete list of mutations found in the SARS-CoV-2 genomes under analysis is provided by an inclusive analysis, as shown in Table 1. The mutations have primarily included single-nucleotide variants (SNVs), with a few cases of deletions and multi-nucleotide variants (MNVs), as examples of viral genetic changes. Most of these variants are very high frequency, with some of them being close to or even exceeding 100 percent, which is a strong indication that these mutations are fixed or almost fixed in this viral population. These mutations are classic genetic changes associated with dominant viral lineages or clades. These high-frequency mutations would probably act as lineage-defining markers that arose early during the viral growth and have remained because of selective benefits or founder effects. SNVs on the other hand refer to point mutations whereby a nucleotide base is replaced by another in a distinctive way (e.g. C to T) with a distribution that cuts across the whole genome and contains vital viral genes essential to replication, transcription and interaction with the host. Of

particular interest are SNVs that are highly frequent in the spike protein, e.g., 23403 (A to G), 23063 (A to T), and 22992 (G to A), potentially affecting the viral infectivity, transmission, and immune evasion. These SNVs and other mutations in the spike protein cause alterations in the receptor-binding domain (RBD) or other critical functional sites, and thus alter the binding affinity of the virus to host ACE2 receptors or its antibody recognition. Whereas deletions that involve more than one base pair, including the 9-base pair deletions at 11288-11296, 21633-21641, and 28362-28370, involve deletions that incur the deletion of sequence stretches that may translate into protein structure or protein expression changes. Whereas, deletion at 28362-28370 affects the N gene, which encodes the nucleocapsid protein, which plays a very important role in the packaging and replication of viral RNA. These deletions affect the viral phenotype by altering protein functionality or immunity, and are commonly linked with new variants of concern. Moreover, MNVs occur in the form of multiple replacements in a group of bases rather than in the form of single-nucleotide changes, which reflect the existence of complicated mutational processes. These MNVs suggest that there can be localized hypermutation or evolutionary hotspots, and can occur because of host-mediated RNA editing or replication errors. The above-mentioned MNVs (i.e., the triple substitution at positions 28881-28883) are very important because they were linked with the changes in antigenic characteristics and the fitness of the viruses. The mutation frequencies of 60% to 90% indicate polymorphic sites under active selection or transitional forms of a sampled population. These rates imply continual viral development, where some variants develop, survive, or compete among host groups or geographical areas. These loci are important in epidemiological surveillance, which can refer to adaptive mutations that give selective advantages or protection against antiviral agents or immune reactions.

The identification of insertions and deletions is presented in Figure 5. This value classifies the nature of mutations identified in the studied SARS-CoV-2 genomes with a focus on the dominance and roles of multiple mutation forms. These mutations are SNVs, which account for the overwhelming majority (92.31%). Where SNVs take place when one base in the viral RNA genome is substituted with another, they are the main cause of genetic variation in SARS-CoV-2. These mutations typically result in changes to the viral proteins that may influence features like transmissibility and pathogenicity. As an example, SNVs in the spike protein that are considered critical have been associated with an increased binding affinity with human ACE2 receptors or decreased neutralization by antibodies. The other deletions (4.62% of the mutations identified) involve the loss of one or more nucleotides of the viral genome. Although less frequent than SNVs, deletions can cause substantial structural changes in viral proteins, which sometimes change the viral infectivity or enable immune evasion. Such deletions have been linked with emerging variants that exhibit altered antigenic properties or replication dynamics. The last mutation type identified was MNVs, which represent 3.08% of the mutations, wherein multiple adjacent nucleotides are substituted simultaneously. These complex mutations are either a result of errors during replication or a result of editing by hosts, which results in localized hypermutation. MNVs can simultaneously target and have the potential to multiplex their effects on larger regions of viral proteins and potentially combine their effect on viral functions, potentially speeding up the process of adaptation.

Table 1. Mutations of high frequency observed in SARS-CoV-2 genomes.

Region	Type	Reference	Allele	Length	Frequency
241	SNV	C	T	1	100.00
670	SNV	T	G	1	67.61
2790	SNV	C	T	1	69.01
3037	SNV	C	T	1	100.00
4184	SNV	G	A	1	68.06
4321	SNV	C	T	1	68.06
9344	SNV	C	T	1	73.56
9424	SNV	A	G	1	73.56
9534	SNV	C	T	1	73.56
9866	SNV	C	T	1	71.59
10029	SNV	C	T	1	100.00
10198	SNV	C	T	1	68.89
10447	SNV	G	A	1	73.91
10449	SNV	C	A	1	98.91
11288..11296	Deletion	TCTGGTTTT	-	9	74.19
12880	SNV	C	T	1	74.73
14408	SNV	C	T	1	100.00
15714	SNV	C	T	1	73.91
17410	SNV	C	T	1	73.91
18163	SNV	A	G	1	100.00
19955	SNV	C	T	1	74.16
20055	SNV	A	G	1	74.16
21618	SNV	C	T	1	73.56
21633..21641	Deletion	TACCCCCTG	-	9	73.56
21987	SNV	G	A	1	81.82
22200	SNV	T	G	1	83.12
22578	SNV	G	A	1	100.00
22674	SNV	C	T	1	100.00
22679	SNV	T	C	1	100.00
22686	SNV	C	T	1	100.00
22688	SNV	A	G	1	84.00
22775	SNV	G	A	1	84.00
22786	SNV	A	C	1	84.00
22813	SNV	G	T	1	100.00
22882	SNV	T	G	1	74.32
22992	SNV	G	A	1	100.00
22995	SNV	C	A	1	100.00
23013	SNV	A	C	1	100.00
23040	SNV	A	G	1	100.00
23055	SNV	A	G	1	100.00
23063	SNV	A	T	1	100.00
23075	SNV	T	C	1	100.00
23403	SNV	A	G	1	100.00
23525	SNV	C	T	1	100.00
23599	SNV	T	G	1	100.00
23604	SNV	C	A	1	100.00
23854	SNV	C	A	1	100.00
23948	SNV	G	T	1	100.00
24424	SNV	A	T	1	100.00
24469	SNV	T	A	1	100.00
25000	SNV	C	T	1	94.37
25584	SNV	C	T	1	100.00
26060	SNV	C	T	1	85.71
26270	SNV	C	T	1	100.00
26577	SNV	C	G	1	94.12
26709	SNV	G	A	1	100.00
26858	SNV	C	T	1	86.76
27259	SNV	A	C	1	100.00
27382..27384	MNV	GAT	CTC	3	86.57
27807	SNV	C	T	1	100.00
28271	SNV	A	T	1	100.00
28311	SNV	C	T	1	100.00
28362..28370	Deletion	GAGAACGCA	-	9	100.00
28881..28883	MNV	GGG	AAC	3	100.00
29510	SNV	A	C	1	85.48

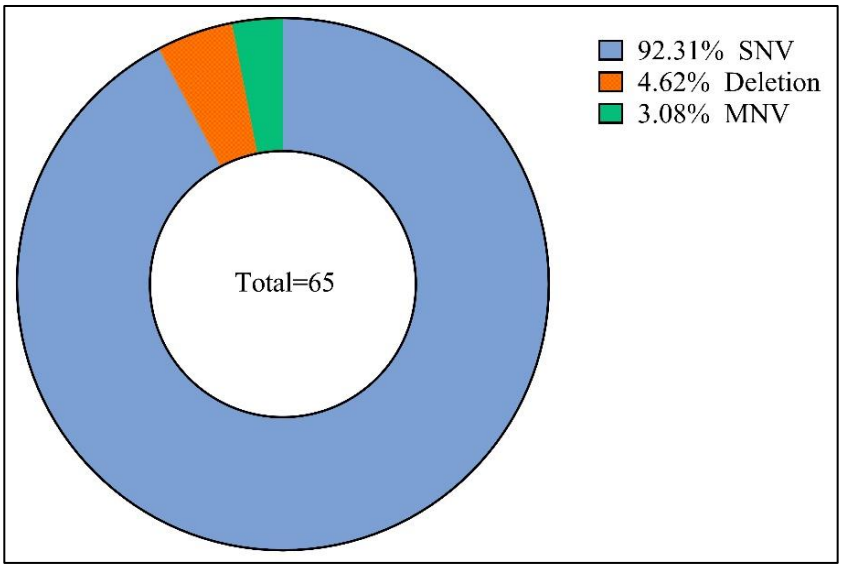

Figure 5. Present the distribution of indels within SARS-CoV-2 genomes of the 94 samples.

3.3 SARS-CoV-2 detected variants

The results from this study clarify the occurrence and distribution of types of SARS-CoV-2 and its lineages (Figure 6). This chart shows how 94 SARS-CoV-2 genomes were classified following the Pango nomenclature system, with the proportions and diversity of the circulating variants of the sampled population emphasized. Whereas the most common lineage, BA.2, has 59.57% of the total samples, which translates to 56 genomes. This over-representation of BA.2 supports the fact that it is the main viral strain that circulated during the period of the study. Then, the second most common lineage, BA.1, contains 26.60% or 25 genomes, indicating that it is an important but declining subvariant compared to BA.2. Nevertheless, a number of minor BA.2 sublineages are associated with a smaller but significant percentage of the population, showing that the Omicron variant continues to diversify virally. It is important to note that BA.2.3 represents 6.38% and it is a significant occurrence that can be connected to local outbreaks or new clusters. Other sublineages include BA.2.10, BA.2.32, BA.2.5, BA.2.57, and BA.2.64, which are single lineages of about 1.06, and number to about one genome each. They are all rare, but when summed up, they constitute about 5.3% of the dataset, highlighting the evolution of viruses into various distinct sublineages. The most noticeable of these minor sublineages is BA.2.40.1, with 2.13% prevalence, which should be considered one of the variants that may emerge. Comparatively, the share of BA.1 lineage (26.60%) is less than two times that of BA.2, which means that the prevalence of BA.2 and its offshoots shifted. This trend is aligned with the currently known epidemiological trend where BA.2 became more transmissible and able to evade the immune system, leading to the replacement of BA.1 in many parts of the world. The data also show that there are many small sublineages that have been formed out of BA.2, which serves to demonstrate the current

adaptive evolution of SARS-CoV-2 to various selective pressures, such as host immunity and vaccination. Their identification at low frequencies across these different lineages provides good information on the viral population structure, revealing hidden diversity that could trigger future outbreaks or influence epidemic pathways.

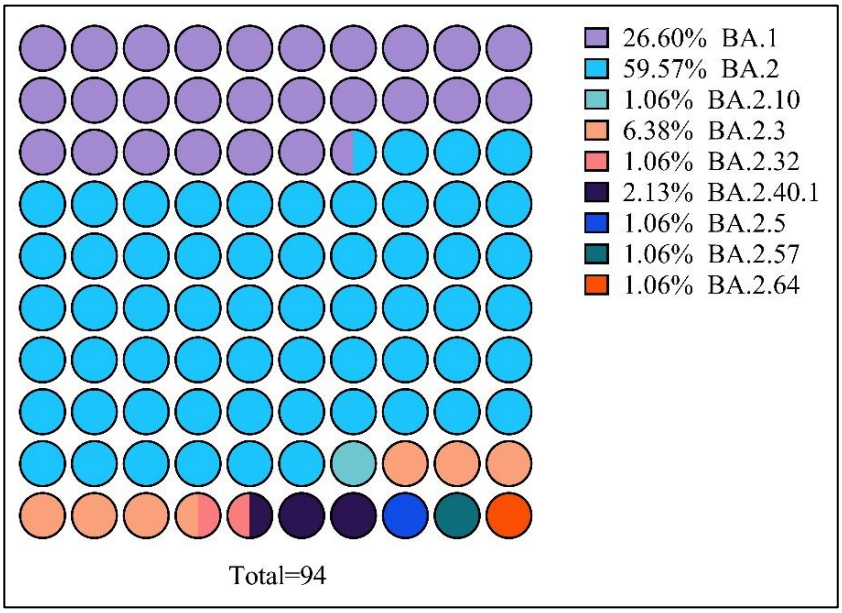

Figure 6. The presence and distribution of SARS-CoV-2 lineages among 94 genomes, as classified by the Pango classification system.

Furthermore, the mutational prevalence analysis of the genomic regions of SARS-CoV-2 (Figure 7) shows that genetic variations are unevenly distributed, and specific differences between the rates of mutations are observed in the entire genome of the virus. It is worth noting that the spike (S) gene has the highest level of mutational prevalence of 42.19%, and it is very important in the adaptation and evolution of the virus. Such a high rate of mutations is also in line with the role of the spike protein in helping the virus to gain entry into the host cell by binding to the ACE2 receptor and its status as a major immune antigen. Spike gene mutations have the potential to directly affect the viral transmissibility, infectivity, and immune evasion, thereby affecting the epidemiological properties of the virus and impacting vaccine efficacy. The second most prevalent mutation at 35.94% occurs in the ORF1ab region immediately after the spike gene. It is a large genomic region that encodes the replicase polyprotein that is cleaved to form a number of non-structural proteins that are essential in the viral synthesis of RNA, replication, and transcription regulation. The mutations in this region can influence viral replication fitness and host-virus interactions and, therefore, enable the virus to adapt to a wide range of cellular environments or antiviral pressures. The comparatively high mutation rate in ORF1ab suggests that these crucial functional areas are still under selection or genetic drift, and can have an effect on viral replication efficiency or pathogenicity. The nucleocapsid (N) gene exhibits a mutation prevalence of 7.81% which is a

medium genetic variation. The N protein is part of the viral RNA packaging and genome stability, and mutations in this region affect the viral assembly and replication or immune surveillance. There is a relatively low rate of mutation in other accessory proteins and structural genes: membrane (M), ORF3a, and ORF6 have a prevalence of 3.13%, and the envelope (E), ORF10, and ORF7b regions have a prevalence of 1.56%. The mutation of these regions is less common, but they can undergo functional constraints, which do not compromise protein integrity but can allow infrequent mutational events that could affect viral fitness or host response. The nonhomogeneous distribution of mutations is the reflection of the evolutionary forces that have acted on SARS-CoV-2 to increase transmissibility and fitness without causing the organism to become functionally constrained. Areas with high mutation rates, including the spike and ORF1ab genes, are probably evidence of adaptive evolution hotspots, driven by selection pressures such as host immune responses, therapeutic interventions, and transmission dynamics. Conversely, regions with low-mutational prevalence suggest purifying selection to preserve vital viral functions or structural integrity.

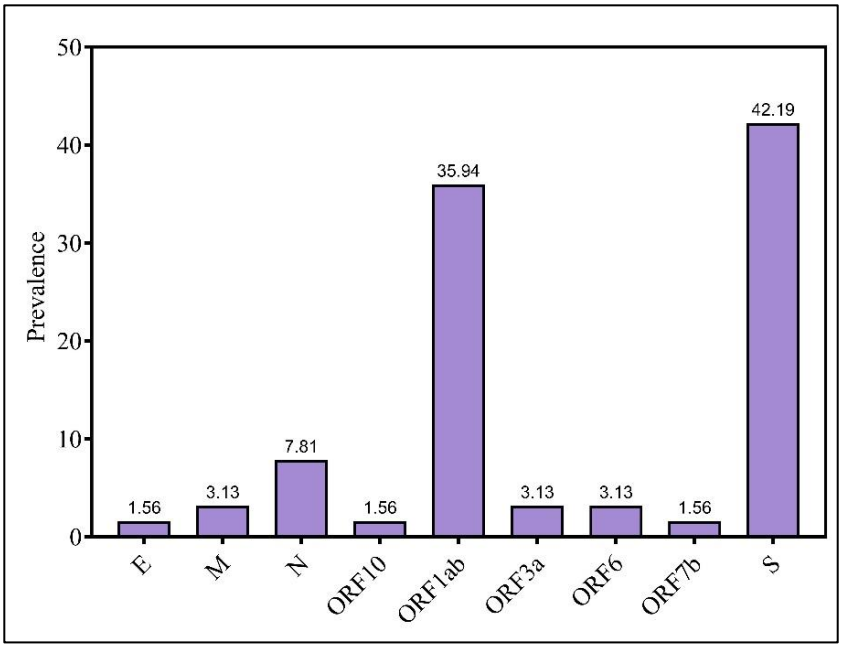

Figure 7. Presents an analysis of the prevalence of mutations across the 94 genomic regions of SARS-CoV-2.

To determine the frequency of amino acid substitutions occurring at the SARS-CoV-2 spike protein loci in 94 genomes (Figure 8) and identify a major grouping of near fixed substitutions along with many of the mid-to-low frequency substitutions occurring at cleavage sites, the receptor-binding domain (RBD), and immune-escape sites, the majority of the mutations have high-prevalence spike protein substitutions that correlate with the Omicron lineage structure, while there is also additional diversity caused by mid-and low-frequency variations in the RBD and N-terminal domains. There were some core substitutions with high prevalence rates; D614G (98.94%), along with other S2/proximal substitutions having >95%

prevalence rates: N969K (100%), Q954H (100%), D796Y (97.87%), N764K (96.81%), and H655Y (100%), all contributing to the dominant haplotype architecture. Mutations in the cleavage-site area presented with S1/S2 polybasic cluster mutations P681H and N679K with a 100% prevalence rate. The A684T mutation was found at low frequency (4.26%), indicating a small sublineage with an alternative substitution to the furin site. High prevalence rates for substitutions in the RBD were seen with the following: Q498R (95.74%), Q493R (94.68%), N501Y (95.74%), E484A (94.68%), S477N (94.68%), T478K (94.68%), Y505H (94.68%), and S375F (92.55%). Other RBD substitutions were seen with variable prevalence; N440K (70.21%), D405N (75.53%), R408S (75.53%), and S371F (75.53%) were moderately frequent, while G446S (21.28%), G496S (21.28%), and R346K (22.34%) represented smaller subvariant-specific antigenic trajectories. Immune-evasion and variant of concern-associated genomic locations were often present, including K417N (86.17%) and the previously mentioned RBD substitutions. Escape-associated substitutions of moderate to low prevalence were also observed, including S373P (92.55%), T376A (74.47%), and S371L (18.09%), which indicate the existence of antigenically distinct clades. In addition to the previously mentioned substitutions, the N-terminal and S2 areas contain additional substitutions that form subdomains; V213G (70.21%) and T19I (70.21%) were notably common. Conversely, substitutions like A67V, T95I, T547K, N856K, and L981F occurred in 26.60% of isolates, which represent minority lineages that include these markers in conjunction with the primary framework. Substitutions of H49Y, L48V, D178N, A892V, and E1262D were rarely observed (1.06-2.13%), thus suggesting a limited propagation of the lineages defined by these changes. From a functional perspective, the cleavage-site pair P681H/N679K, along with numerous RBD substitutions (Q498R, Q493R, N501Y, E484A, S477N, T478K, Y505H), were found in no less than 89 genomes and therefore define a high-prevalence axis that defines the primary phenotype of currently circulating viral strains. Intermediate prevalence (70-76%) markers that contribute to variability include N440K, D405N, R408S, S371F, T376A, V213G, and T19I, in addition to low-frequency substitutions, including G446S, G496S, R346K, S371L, and A684T, which define subvariants and their respective antigenic properties. From a functional standpoint, the simultaneous occurrence of paired mutations in the cleavage site, along with a localized concentration of RBD substitutions, suggests an integrated selective pressure favoring increased viral entry efficiency and evasion of antibody-mediated immunity, while minority alternatives may define parallel antigenic modifications in the same epidemiologic setting.

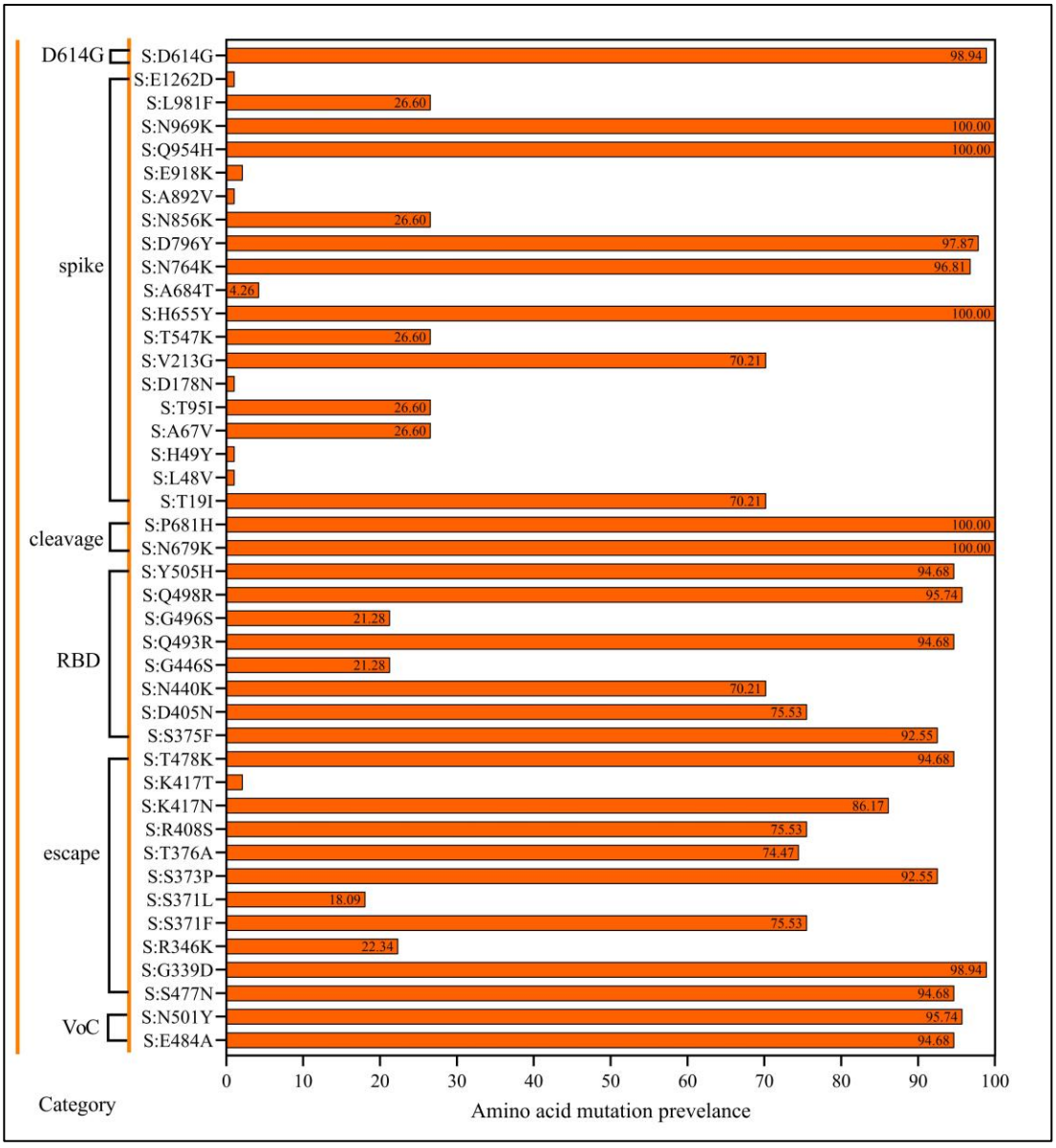

Figure 8. Distribution of amino acid mutations among spike protein sites in 94 SARS-CoV-2 genomes.

Analysis of the SARS-CoV-2 spike glycoprotein using dimensional structural analysis provides insight into the spatial organization and functional impact of identified amino acid substitutions found in our viral samples. The three-dimensional structure of the spike trimer in the pre-fusion, receptor-binding-domain conformation (PDB: 6acc) contains an extensive number of sites with mutations located throughout its architecture (Figure 9). Each color-coded sphere represents an individual mutation found during genomic analysis. Each color corresponds to specific functional or structural categories. Orange colored mutations, including G142D, V143del, Y144del, Y145del, D178N, L212I, V213G, G339D, R346K, S371L, S371F, S373P, S375F, and T376A, are spread out over the NTD and RBD and indicate areas of

the virus that have been subjected to constant evolutionary pressure. High frequency mutations (>95%) and intermediate frequency mutations (70-92%) include both high and low frequency substitutions that reflect the range of viral adaptation within this population. Blue mutations, such as T95I, D405N, R408S, N440K, G446S, K417N, N501Y, Q493R, Q498R, E484A, S477N, T478K, Y505H, G496S, and E1262D, are found in regions related to receptor binding and conformational changes and represent the functional core of viral infectivity. Cyan mutations, including L24del, P25del, P26del, H69del, and V70del, are found in the NTD antigenic supersite and represent peripheral regions involved in immune recognition and antibody evasion. Magenta markers highlight positions such as L48V, H49Y, and A67V, which occur at intermediate frequencies and represent emerging "hot-spots" or sub-lineage-defining variants that affect spike assembly and conformational transitions. While the two structures illustrated are very different, the open ACE2-bound conformation (PDB: 6acj) illustrates how mutations affect receptor interaction and viral entry by showing the spike protein in complex with the host receptor ACE2 (green ribbon). The Blue mutations cluster around the receptor binding interface and nearly fixed substitutions such as Q498R (95.74%), Q493R (94.68%), N501Y (95.74%), E484A (94.68%), S477N (94.68%), T478K (94.68%), and Y505H (94.68%) are found in close proximity to ACE2 contact residues. This indicates that there has been strong selective pressure on maintaining or finely tuning receptor affinity. Intermediate frequency RBD mutations N440K (70.21%), D405N (75.53%), R408S (75.53%), and S371F (75.53%) are spatially clustered within the ACE2-binding interface and represent multiple co-evolving lineages with differing receptor-binding strategies. Orange mutations, such as G339D, R346K, S375F, and T376A, span the RBD and adjacent domains and can affect either direct contact residues or structural components that regulate binding access and conformational stability. Cyan mutations, including T19I (70.21%) and V213G (70.21%), are localized to peripheral NTD positions and facilitate the transition between the closed and open states necessary for receptor interaction. Of the 56 identified amino acid substitutions, they are non-randomly distributed and localized in functionally significant regions. The NTD is home to several deletions, including L24del, P25del, P26del, and the H69del-V70del deletion pattern (73.56%-74.19%) that have been linked to immune evasion through neutralizing antibodies targeting the antigenic supersite. These deletions have now become established as lineage-defining markers within the BA.2-dominant population. Blue mutations, including K417N (86.17%), S477N (94.68%), T478K (94.68%), E484A (94.68%), Q493R (94.68%), Q498R (95.74%), and N501Y (95.74%), are localized to the RBD and are directly adjacent to ACE2 interface residues identified in previous structural studies. Near fixation (>94%) among these substitutions indicates strong purifying selection maintaining these contact residues in this viral population. Substitutions in the S2 subunit of the spike protein that affect membrane fusion represent all color categories. The D614G substitution (98.94% prevalence, orange) is found ubiquitously throughout the spike protein and appears in both conformations. It is thought to enhance spike stability and transmissibility. Orange and magenta marked substitutions S371F (92.55%), S373P (92.55%), and S375F (92.55%) are found near the furin cleavage site (P681H/N679K, 100% prevalence). Polybasic site mutations, represented by blue, may be regulating protease activation specificity and efficiency. Additional substitutions such as G339D and R346K, represented by orange, are

localized to areas that could affect the conformational flexibility and energy transitions required for membrane fusion.

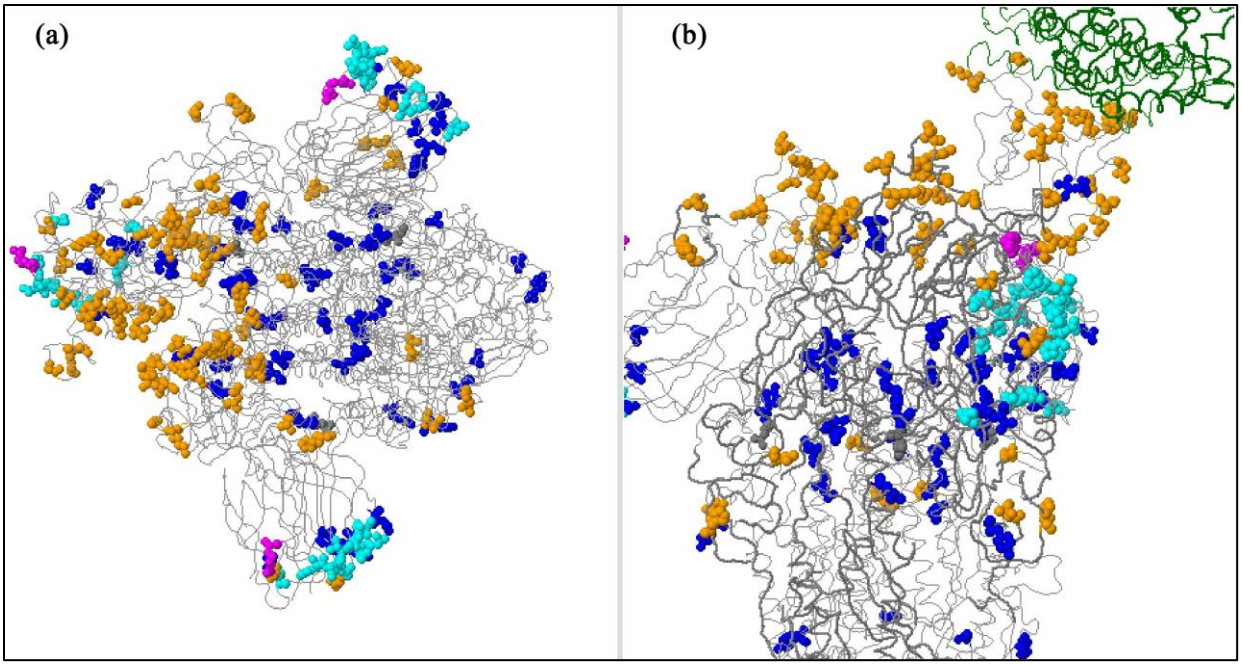

Figure 9. Show the 3D structural mapping of SARS-CoV-2 spike protein mutations identified. (a) Spike glycoprotein (PDB: 6acc; cryo-EM, 3.6 Å resolution) in the RBD-down conformation. (b) Spike glycoprotein (PDB: 6acj; cryo-EM, 4.2 Å resolution) in complex with the host cell receptor ACE2 (green ribbon).

4. Discussion

4.1 Phylogenetic relationships and transmission dynamics

The results of the phylogenetic reconstruction demonstrate the peculiarities of a well-developed viral population with apparent extensive chains of local transmissions, which are consistent with the results of other studies of SARS-CoV-2 surveillance in the region. The high proportion of short branches linking closely related isolates indicates a fast viral spread in a temporally and geographically limited population, which is indicative of the epidemiological situation of a lasting community transmission throughout the Omicron-dominated era [24]. This distribution is consistent with known principles of viral phylogenetics, in which related sequences clustering on short branches tend to indicate recent common ancestry and chains of transmission within local populations are active [25]. Contrary to the fact that the recognition of many different clusters in the phylogeny suggests the existence of many different events of introduction or the establishment of parallel transmission chains, which is a common

occurrence in cosmopolitan regions with high connectivity and movement between populations [26]. These clusters are likely to refer to unique epidemiological units that could be different social networks, geographical areas, or institutional outbreaks in Saudi Arabia. The presence of outlier sequences on long branches indicates that there were discrete viral infections of the external origin, the development of specific mutations as a result of long-term situations of infection, or the sampling of divergent viral branches that evolved under distinct selection pressures [27]. Moreover, the limited genetic diversity observed in most of the samples can be attributed to recent phenomena of viral growth, which is consistent with the date of the advent of Omicron and its subsequent spread across the world [28]. This finding aligns with observations of other national and regional SARS-CoV-2 surveillance programs, in which most dominant variant replacement events often lead to populations of very similar sequences, which in turn diversify with time as a result of mutation accumulation and selection [29].

4.2 Genomic architecture and mutational landscape

The results of the multiple sequence alignment show that the genetic variation is highly heterogeneous across the whole SARS-CoV-2 genome, with certain mutational hotspots coinciding with functional areas of the genome [30]. The clustering of mutations in the 5,000-10,000 base pair region that is the locations of the ORF1ab that encode key elements of the replication machinery illustrates the continued evolution of viral fitness under the pressure of selection related to replication efficiency and host adaptation [31, 32]. These findings are consistent with trends in the global development of SARS-CoV-2, with mutations in ORF1ab and especially in nonstructural proteins associated with RNA synthesis and processing making a significant contribution to viral adaptation [33]. Moreover, the accumulation of mutations near the 25,000 base pair location, which is the location of the structural and accessory gene regions, indicates selection forces related to the assembly of viruses as well as their adherence to the host [34, 35]. This distribution pattern reflects the sophisticated evolutionary limitations on SARS-CoV-2, in which different genomic regions are exposed to different selection forces depending on their functional roles and interactions with host immune functions [36, 37]. The conservation of large areas of conserved areas with localized concentrations of variation is representative of the balance between the maintenance of key viral functions and environmental adaptation [38].

4.2 Genomic characterization and evolutionary analysis

4.2.1 **Lineage dynamics and micron predominance**

The complete lineage analysis has clarified a striking prevalence of the Omicron lineage BA.2 and its sublineages BA.2.3, BA.2.10, BA.2.32, BA.2.5, BA.2.57, and BA.2.64 in the Jeddah, Saudi Arabian setting, which includes most of the spreading viral strains [39, 40]. This supremacy fits the global epidemiological trend in 2022, where BA.2 was found to be more transmissible than

other Omicron subvariants, thus becoming the dominant strain worldwide [41]. The variety of BA.2 derivatives is indicative of the continued viral diversification in the presence of local selection pressures, and thus, it is reflective of continued evolutionary events in the context of the predominant lineage. Within the context of phylogenomics, the combination of mutation frequency landscapes illustrates the focalization of evolutionary stress in the areas of functional value, especially ORF1ab and the spike protein. The fact that the spike protein is the most noticeable in the mutation landscape indicates that it is a bifunctional protein with the main function of neutralizing antibodies and the primary one in terms of receptor binding. The combination of these two functions makes it a key target of adaptive evolution under immune selection pressure [42, 43]. The high frequency of mutations in ORF1ab, especially in the nsp3 and RNA-dependent RNA polymerase regions, can be considered as adaptations that might contribute to better viral replication, an increase in the host range, or resistance to antiviral agents. Conversely, the mutation rate observed in structural genes (M) and those involved in accessory genes (ORF3a, ORF6, ORF7b, ORF10) is low, indicating that there are functional constraints that are essential to both virion assembly and budding processes [44]. The mutational landscape depicts dynamic processes of evolution where both variants are reaching fixation and new adaptive changes are propagating throughout the population. This patchy distribution is suggestive of continued selective forces, with mutation, natural selection, and genetic drift all acting together in creating the apparent genetic diversity [45]. The unique mutational patterns of different Omicron sublineages offer useful information on evolutionary paths leading to viral heterogeneity that illustrate not only the historical selection pressures that defined the Omicron variant but also the current evolutionary processes that are taking place in particular epidemiological settings.

4.3 Mutational architecture: SNVs, indels, and functional significance

The large number of single-nucleotide variants, as compared to deletions and multi-nucleotide variants, represents the largest component of the mutational spectrum in SARS-CoV-2, as these account for 92.31% of all genetic alterations; deletions account for 4.62%; and multi-nucleotide variants account for 3.08%. This pattern of distributions is consistent with previously described paradigms in the field of SARS-CoV-2 genomic surveillance, in which the majority of genetic heterogeneity is due to point mutations [46]; whereas deletions produce an exaggerated effect on phenotypic variation, as they result in the simultaneous removal or substitution of many amino acids [47-49]. The high frequency of SNVs is attributed to both replication errors caused by the virus's RNA-dependent RNA polymerase, as well as the host-mediated RNA editing activity of enzymes such as APOBEC and ADAR, which catalyze C→T and A→G nucleotide substitutions, respectively [50, 51]. High frequency deletions (74% up to 100%) at three highly conserved locations: 9 bp at nsp6 (11288-11296); the S1/S2 junction (21633-21641); and the N gene (28362-28370) indicate early and robust selective pressure. The

deletion at nsp6 disrupts the host cell's ability to carry out autophagy and could potentially allow the virus to increase its replication rate through altered membrane dynamics. The deletion of the S gene at the S1/S2 junction may affect the efficiency of the cleavage of the S protein and thus facilitate entry into the host cell [52, 53]; while the deletion of the N gene results in the removal of the carboxyl terminal portion of the N protein that influences RNA packaging, as well as the ability of the N protein to suppress the host cell's interferon response and induce antigenic responses through altered antigenic properties. While MNVs make up a small percentage (3.08%), they are evidence of localized hypermutation "hotspots". An example of this complexity is the triple nucleotide substitution at positions 28881-28883 (GGG→AAC) in the N gene, resulting in two amino acid changes (R203K and G204R) that enhance the production of subgenomic RNAs and increase viral replication. As MNVs occur in response to host-mediated nucleotide editing during the course of acute infection, they demonstrate the dynamic relationship between viral replication fidelity and host defense mechanisms. The frequency of indel occurrence provides insight into the selection processes occurring within viral populations. High frequency deletions (>70%) are indicative of either strong positive selection or founder effects, where early circulating viral strains were selected over other viral strains. Deletions found at intermediate frequencies (40-70%) are representative of transitional strains that are in a state of partial fixation. Low-frequency deletions (<40%, are representative of new events that possess the potential to expand based upon future selective pressures. Functionally, indels located within the spike and nucleocapsid proteins, which are key targets of the host immune system, have the potential to alter the epitope recognition by antibodies and T cells. Altered epitope recognition has significant implications for vaccine efficacy and the sensitivity of diagnostic tests for detecting viral infections, emphasizing the need for ongoing monitoring of indel dynamics [54, 55].

4.4 Spike protein mutations and mechanistic adaptations

The results revealed that 42.2% of the identified genomic mutations have been implicated with the gene encoding the spike glycoprotein, highlighting its key role as the primary target of immune system reactions and as a factor in viral infection of host cells [56]. This is further adjusted by the selective forces of neutralizing antibodies and antiviral interventions that hasten adaptive genomic changes. An evolutionary analysis based on the stratification approach indicates different evolutionary consequences in different categories of mutation frequencies. Core Nearly Fixed Mutations are those with a prevalence rate of more than 94% and include D614G, P681H, N679K, Q498R, Q493R, N501Y, E484A, and S477N. These mutations form the key Omicron scaffold, which enhances the transmission efficacy and host tropism via convergent evolutionary mechanisms, thus optimizing the conformation of spike protein in response to the immunological pressures [57]. Intermediate-Frequency Mutations, those with prevalence rates of 70-76, including N440K, D405N, R408S, S371F, T376A, V213G,

and T19I, define points of divergence in the BA.2 lineage. Such mutations can affect sublineage-specific properties, such as divergent antibody sensitivity or a change in cellular tropism. Low-Frequency Variations (Less than 25%): G446S, G496S, R346K, S371L, A684T are nascent variants with localized antigenic changes that require close monitoring since in the past, such mutations have been able to gain dominance in circumstances that cause selective advantages [36, 58]. The overlap of mutations at the sites of cleavage (P681H/N69K) and the widespread changes to the RBD highlight the dual nature of the virus in terms of maximizing spike protein activation and simultaneously avoiding the immune response of the host [59]. P681H mutation, which is located next to the furin cleavage site, promotes increased priming of the spike protein by host proteases, thus increasing the rate of viral entry [60]. Structural studies reveal that the P681H mutation gives the furin cleavage site greater conformational flexibility, thereby increasing recognition and proteolytic processing by host furin proteases [61]. At the same time, the adjacent mutation (N679K) alters the local electrostatic environment, thereby changing the geometry of the substrate to one more favorable to the cleavage site accessibility [62, 63]. At the same time, the N679K mutation, which is nearby, changes the local electrostatic environment to allow the substrate geometry that allows the best access to the cleavage site to occur. The P681H and N679K mutations that are found in this group with a frequency of 100 percent are located at the S1/S2 polybasic cleavage site and are key genetic changes in the Omicron strains. These mutations work together to maximize the efficiency of cleavage of the spikes, thereby promoting the fusion of the membranes and the direct person-to-person viral transmission, which explains the significant transmissibility benefit of the Omicron variants [64]. Cell-based cleavage experiments indicate enhanced S1/S2 separation compared to the wild-type spike, and in vivo experiments using hamster and mouse models find these mutations to enhance the propagation of the virus in alignment with the near fixation of the mutations in circulating isolates [65]. These mutations have been identified to be prevalent in the dataset at 100 percent, which is an indicator of high positive selection pressure to promote better viral entry and immune evasion. D614G mutation with an almost universal fixation of 98.94 has been widely studied using multiple experimental approaches. Experiments involving the use of pseudoviruses have reported a 4-8 fold rise in titer of the virus in cultured cells [66]. Further studies have shown that D614G alters S1-S2 domain orientation, exposing more spike protein and destabilizing the open conformational form to enhance ACE2 accessibility.

Structural studies showed that D614G changes the orientation of the S1-S2 domain, exposing the spike protein and destabilizing the open form of conformation, making it more accessible to ACE2 [67]. The results of experiments on animal models point to the idea that D614G has a survival benefit in transgenic mice with ACE2 expression, which firmly supports its status as a key mutation that increases transmissibility [68, 69]. The presence of mutations in the RBD (Q498R-95.74%), Q493R (94.68%), N501Y (95.74%), E484A (94.68%), S477N (94.68%),

and T478K (94.68%) can be attributed to adaptive evolution under the influence of the selection pressure that preserves and strengthens the ACE2-binding ability and helps to avoid antibody-mediated neutralization. The Alpha and Omicron mutation N501Y is linked to the maintenance or increase in affinity of ACE2 binding and a massive decrease in recognition by some monoclonal antibodies, and thus an example of an immune evasion mechanism [70]. The E484A protein helps escape such escape by disrupting important interactions with Class 1 and Class 2 neutralizing antibodies with minimal effects on ACE2 binding [71]. The location of S477N and T478K, which are part of the secondary antibody epitopes, suggests a cumulative strategy of immune evasion by serial mutations, which make the protein less vulnerable to polyclonal antibodies. Taken together, these RBD mutations are adaptive response mechanisms that maintain viral fitness through ACE2 interaction and have a high ability to evade host immune responses [72, 73].

The nine-base pair deletions, which are found at positions 11288-11296 in the nsp6, 21633-21641 in the spike gene, and 28362-28370 in the nucleocapsid gene, are found with frequencies ranging between 74-100 percent and with known functional implications. The nsp6 gene deletion is experimentally linked with the changes in the modulation of autophagy and the development of replication organelles, evidenced through the confocal microscopy and membrane traffic experiments [74]. The efficiency of furin cleavage and the following processing of the spike protein is also influenced by the deletion of the S gene at the cleavage site of S1/S2 as shown by the analysis of Western blot of cleavage products of the spike protein [52, 53]. Also, the deletion in the nucleocapsid gene affects the carboxy-terminal domain, which in turn influences the efficiency of RNA packaging and oligomerization of the nucleocapsid as explained under biochemical characterization experiments [75].

5. Conclusions

The genomic characterization of SARS-CoV-2 transmission in Jeddah, Saudi Arabia, in 2022 provides important insight into viral evolution in a region of high global epidemiological relevance. Analysis of 94 viral genomes showed widespread community transmission and local diversification, with short phylogenetic branch lengths reflecting rapid dissemination during the Omicron-dominant phase. The strong predomination of the Omicron BA.2 variant and its certain sublineages, constitute about 60% of the samples, mirrors global trends and underscores continued adaptive evolution under local selection pressures. This is particularly significant considering that Jeddah is a major international travel and pilgrimage hub, which facilitates interaction among diverse viral populations. Mutational analysis indentified concentrated hotspots in the spike (42.19%) and ORF1ab (35.94%) regions, indicating strong selective pressures on the viral entry and replication mechanisms. Fixed or nearly fixed mutations, such as D614G, P681H/N679K, and key receptor-binding domain substitutions,

exhibit convergent adaptation that increases transmissibility and immune evasion. The high rate of single-nucleotide variation (2.31%) and frequent deletions in the nsp6, spike, and nucleocapsid genes are also likely reflect lineage-defining adaptive advantages.

These results have direct implications for genomic-surveillance policy and public health preparedness. Saudi Arabia should integrate real-time genomic surveillance into the national health surveillance system to enable rapid detection of emerging variants and linkage of genomic data with clinical outcomes. Targeted sequencing campaigns during Hajj and Umrah, when millions of international pilgrims arrive, would provide early warning of variant introduction and inform containment strategies. Establishing routine genomic screening at airports and other points of entry, with timely data sharing through local and international databases such as Weqaya and GISAID, would further strengthen regional and global situational awareness.

With the high mutational rate in immune-targeted areas, vaccine strategy updates should reply on domestic genomic data to ensure booster formulations account for locally prevalent sublineages. The systematic combination of genomic, clinical, and epidemiological datasets into national systems of disease surveillance will enhance to predict the development of variants and strengthen evidence-based policymaking.

Future research studies should include longitudinal genomic monitoring of breakthrough and severe infections, along with functional characterization of spike and ORF1ab mutations, to clarify their phenotypic impacts on transmissibility and immune escape mechanisms. Integrating genomic surveillance into regular infectious disease monitoring, especially in international transit centers such as Jeddah, will strengthen Saudi Arabia's capacity for early detection of variants, international data collaboration, and pandemics preparedness. Such a system would reinforce national health security while contributing meaningfully to global efforts to predict and prevent future SARS-CoV-2 like threats.

Author Contributions

A.B. and S. E. performed the conceptualization, S. A methodology design, W. A. F. and M. A. formal analysis, A. M. R. wrote the original draft. S. E. writing-review and editing.

Funding

No external funding was received for this study

Institutional Review Board Statement

The study was approved by the National Committee of Biomedical Ethics at King Abdul-Aziz City for Science and Technology, Saudi Arabia, approval number [HAP-02-J-099], dated [13/12/2022].

Informed Consent Statement

Not applicable

Data Availability Statement

Data are available from the corresponding author upon reasonable request.

Conflicts of Interest

The authors declare no conflicts of interest.

References

[1] Ling-Hu T, Rios-Guzman E, Lorenzo-Redondo R, Ozer EA, Hultquist JF. Challenges and Opportunities for Global Genomic Surveillance Strategies in the COVID-19 Era. *Viruses*. 2022;14(11):2532.10.3390/v14112532

[2] Ciccozzi M, Lai A, Zehender G, Borsetti A, Cella E, Ciotti M, et al. The phylogenetic approach for viral infectious disease evolution and epidemiology: An updating review. *Journal of Medical Virology*. 2019;91(10):1707-24

[3] Ogunlaja A, Oginni O. Molecular epidemiology and phylodynamics of emerging and re-emerging viral infections: Implications for surveillance and control. *Inter J Res Publicat Rev*. 2022;6(2):2924-41

[4] Alharbi AM. Coronavirus and co-infections: A Saudi Arabian perspective. *Saudi J Biol Sci*. 2023;30(9):103739.10.1016/j.sjbs.2023.103739

[5] Khan A, Alsofayan Y, Alahmari A, Alowais J, Algwizani A, Alserehi H, et al. COVID-19 in Saudi Arabia: the national health response. *East Mediterr Health J*. 2021;27(11):1114-24.10.26719/emhj.21.048

[6] Al Khawaja S. The Impact Of COVID-19 On The Importance And Use Of Public Parks In Saudi Arabia (Dammam City As A Case Study). 2022

[7] Alshalwi AM. Reported Incidence and Case-Fatality Rates of COVID-19 Among Physicians in Riyadh Region, Saudi Arabia: A Cross-Sectional Retrospective Study, 2021-2022: *Alfaisal University (Saudi Arabia)*; 2025

[8] Badreldin HA, Al-jedai A, Alghnam S, Nakshabandi Z, Alharbi M, Alzahrani A, et al. Sustainability and Resilience in the Saudi Arabian Health System. 2025

[9] Smatti MK, Al-Sarraj YA, Albagha O, Yassine HM. Host Genetic Variants Potentially Associated With SARS-CoV-2: A Multi-Population Analysis. *Front Genet*. 2020;11:578523.10.3389/fgene.2020.578523
[10] Chakraborty C, Sharma AR, Bhattacharya M, Agoramoorthy G, Lee SS. Evolution, Mode of Transmission, and Mutational Landscape of Newly Emerging SARS-CoV-2 Variants. *mBio*. 2021;12(4):e0114021.10.1128/mBio.01140-21
[11] Bindayna KM, Crinion S. Variant analysis of SARS-CoV-2 genomes in the Middle East. *Microb Pathog*. 2021;153:104741.10.1016/j.micpath.2021.104741
[12] Plante JA, Liu Y, Liu J, Xia H, Johnson BA, Lokugamage KG, et al. Author Correction: Spike mutation D614G alters SARS-CoV-2 fitness. *Nature*. 2021;595(7865):E1.10.1038/s41586-021-03657-2
[13] Weissman D, Alameh MG, de Silva T, Collini P, Hornsby H, Brown R, et al. D614G Spike Mutation Increases SARS CoV-2 Susceptibility to Neutralization. *Cell Host Microbe*. 2021;29(1):23-31 e4.10.1016/j.chom.2020.11.012
[14] Alkhamis MA, Hussain A, Al-Therban F. Comparative Evolutionary Epidemiology of SARS-CoV-2 Delta and Omicron Variants in Kuwait. *Viruses*. 2024;16(12):1872.10.3390/v16121872
[15] Khasawneh AI, Himsawi NM, Abu-Raideh JA, Sammour A, Abu Safieh H, Obeidat A, et al. Prevalence of SARS-COV-2 and other respiratory pathogens among a Jordanian subpopulation during Delta-to-Omicron transition: Winter 2021/2022. *PLoS One*. 2023;18(3):e0283804.10.1371/journal.pone.0283804
[16] Elrashedy A, Mousa W, Nayel M, Salama A, Zaghawa A, Elsify A, et al. Advances in bioinformatics and multi-omics integration: transforming viral infectious disease research in veterinary medicine. *Virol J*. 2025;22(1):22.10.1186/s12985-025-02640-x
[17] Maljkovic Berry I, Melendrez MC, Bishop-Lilly KA, Rutvisuttinunt W, Pollett S, Talundzic E, et al. Next Generation Sequencing and Bioinformatics Methodologies for Infectious Disease Research and Public Health: Approaches, Applications, and Considerations for Development of Laboratory Capacity. *J Infect Dis*. 2020;221(Suppl 3):S292-S307.10.1093/infdis/jiz286
[18] Featherstone LA, Zhang JM, Vaughan TG, Duchene S. Epidemiological inference from pathogen genomes: A review of phylodynamic models and applications. *Virus Evol*. 2022;8(1):veac045.10.1093/ve/veac045
[19] Mostefai F. Big Data approaches to viral genomics and evolutionary dynamics of SARS-CoV-2. 2025
[20] Singhvi N, Talwar C, Mahanta U, Kaur J, Mondal K, Ahmad N, et al. Comparative genomics and integrated system biology approach unveiled undirected phylogeny patterns, mutational hotspots, functional patterns, and molecule repurposing for monkeypox virus. *Funct Integr Genomics*. 2023;23(3):231.10.1007/s10142-023-01168-z
[21] McBroome J, de Bernardi Schneider A, Roemer C, Wolfinger MT, Hinrichs AS, O'Toole AN, et al. A framework for automated scalable designation of viral pathogen lineages from genomic data. *Nat Microbiol*. 2024;9(2):550-60.10.1038/s41564-023-01587-5
[22] Codda G. Next Generation Sequencing-based detection and characterization of microbial pathogens causing invasive infections and outbreaks in ICU: Towards improved management of the high-risk patient. 2024
[23] Hussein MH, Magram SF. Domestic Water Quality in Jeddah, Saudi Arabia. *Journal of King Abdulaziz University: Engineering Sciences*. 2012;23(1)
[24] Xie H, Huang Y, Dick D, Lee S. Although the primary and secondary vaccination rates in Korea account for over 75% of the total population, confirmed cases of COVID-19 are dramatically increasing due to immune waning and the Omicron variant. Therefore, it is urgent to evaluate the effectiveness of booster vaccination strategies for living with COVID-19. In this work, we have

developed an age-specific mathematical model with eight age groups. *Evolution of SARS-CoV-2: impact of variants on hosts, COVID-19 vaccines and countermeasures*. 2024:21

[25] Pereson MJ, Mojsiejczuk L, Martinez AP, Flichman DM, Garcia GH, Di Lello FA. Phylogenetic analysis of SARS-CoV-2 in the first few months since its emergence. *J Med Virol*. 2021;93(3):1722-31.10.1002/jmv.26545

[26] Sallard E, Halloy J, Casane D, Decroly E, van Helden J. Tracing the origins of SARS-COV-2 in coronavirus phylogenies: a review. *Environ Chem Lett*. 2021;19(2):769-85.10.1007/s10311-020-01151-1

[27] Eulenfeld T, Triebel S, Marz M. AnchoRNA: Full virus genome alignments through conserved anchor regions. *bioRxiv*. 2025:2025.01. 30.635689

[28] Holtz AM. From viral origins to global dispersion: exploring factors that drive and sustain viral diffusion over time and space: *Université Paris Cité*; 2023

[29] Han SM, Shiino T, Masuda S, Furuse Y, Yasaka T, Kanda S, et al. Phylogenetic Study of Local Patterns Influenza A(H3N2) Virus Transmission in a Semi-Isolated Population in a Remote Island in Japan Between 2011 and 2013. *Influenza Other Respir Viruses*. 2025;19(3):e70089.10.1111/irv.70089

[30] Islam MR, Hoque MN, Rahman MS, Alam A, Akther M, Puspo JA, et al. Genome-wide analysis of SARS-CoV-2 virus strains circulating worldwide implicates heterogeneity. *Sci Rep*. 2020;10(1):14004.10.1038/s41598-020-70812-6

[31] Mikhailovna SI. Regulation of DNA replication in the cell. Replication of various DNA, its regulation and repair Main enzymes of DNA replication.

[32] McLean M. The Novel Functions of the SARS CoV-2 Accessory Protein ORF8: *McGill University (Canada)*; 2022.

[33] Saha I, Ghosh N, Maity D, Sharma N, Mitra K. Inferring the genetic variability in Indian SARS-CoV-2 genomes using consensus of multiple sequence alignment techniques. *Infect Genet Evol*. 2020;85:104522.10.1016/j.meegid.2020.104522

[34] Banoun H. Evolution of SARS-CoV-2: Review of Mutations, Role of the Host Immune System. *Nephron*. 2021;145(4):392-403.10.1159/000515417

[35] Leider JM. Determination of mutation rates of RNA viruses and characterization of the receptor-destroying enzyme of bovine coronavirus: *City University of New York*; 1989.

[36] Telenti A, Hodcroft EB, Robertson DL. The Evolution and Biology of SARS-CoV-2 Variants. *Cold Spring Harb Perspect Med*. 2022;12(5):a041390.10.1101/cshperspect.a041390

[37] Hasan M. Evolution of Leading Mutations in Spike Glycoproteins of SARS-CoV-2: *Hong Kong University of Science and Technology (Hong Kong)*; 2024.

[38] Cobar O, Cóbar S. SARS-CoV-2 Orf1ab Genome Mutations, the Driving Force for Virus Pathogenicity. *Medical Research Archives*. 2023;11(11)

[39] Farahat RA, Abdelaal A, Umar TP, El-Sakka AA, Benmelouka AY, Albakri K, et al. The emergence of SARS-CoV-2 Omicron subvariants: current situation and future trends. *Infez Med*. 2022;30(4):480-94.10.53854/liim-3004-2

[40] Kimura I, Yamasoba D, Tamura T, Nao N, Suzuki T, Oda Y, et al. Virological characteristics of the SARS-CoV-2 Omicron BA.2 subvariants, including BA.4 and BA.5. *Cell*. 2022;185(21):3992-4007 e16.10.1016/j.cell.2022.09.018

[41] Chrysostomou AC, Vrancken B, Haralambous C, Alexandrou M, Gregoriou I, Ioannides M, et al. Unraveling the dynamics of Omicron (BA. 1, BA. 2, and BA. 5) waves and emergence of the deltacron variant: Genomic epidemiology of the SARS-CoV-2 epidemic in cyprus (Oct 2021–Oct 2022). *Viruses*. 2023;15(9):1933

[42] Mittal A, Khattri A, Verma V. Structural and antigenic variations in the spike protein of emerging SARS-CoV-2 variants. *PLoS Pathog*. 2022;18(2):e1010260.10.1371/journal.ppat.1010260

[43] Harvey WT, Carabelli AM, Jackson B, Gupta RK, Thomson EC, Harrison EM, et al. SARS-CoV-2 variants, spike mutations and immune escape. *Nat Rev Microbiol*. 2021;19(7):409-24.10.1038/s41579-021-00573-0
[44] Hassan SS, Choudhury PP, Roy B. Rare mutations in the accessory proteins ORF6, ORF7b, and ORF10 of the SARS-CoV-2 genomes. *Meta Gene*. 2021;28:100873
[45] Markov PV, Ghafari M, Beer M, Lythgoe K, Simmonds P, Stilianakis NI, et al. The evolution of SARS-CoV-2. *Nature Reviews Microbiology*. 2023;21(6):361-79
[46] Schmidt M, Arshad M, Bernhart SH, Hakobyan S, Arakelyan A, Loeffler-Wirth H, et al. The evolving faces of the SARS-CoV-2 genome. *Viruses*. 2021;13(9):1764
[47] Hu TL. Harnessing Genomic Surveillance to Characterize the Molecular Epidemiology and Uncover Antiviral Resistance Mechanisms of SARS-CoV-2: *Northwestern University*; 2024
[48] Jiang W, Ji W, Zhang Y, Xie Y, Chen S, Jin Y, et al. An Update on Detection Technologies for SARS-CoV-2 Variants of Concern. *Viruses*. 2022;14(11):2324.10.3390/v14112324
[49] Thakur S, Sasi S, Pillai SG, Nag A, Shukla D, Singhal R, et al. SARS-CoV-2 Mutations and Their Impact on Diagnostics, Therapeutics and Vaccines. *Front Med (Lausanne)*. 2022;9:815389.10.3389/fmed.2022.815389
[50] Su S. Investigating roles of RNA editing in virus host interaction 2024
[51] Domingo E. Molecular basis of genetic variation of viruses: Error-prone replication. *Virus as Populations*. 2019:35
[52] Yu S, Zheng X, Zhou B, Li J, Chen M, Deng R, et al. SARS-CoV-2 spike engagement of ACE2 primes S2' site cleavage and fusion initiation. *Proc Natl Acad Sci U S A*. 2022;119(1):e2111199119.10.1073/pnas.2111199119
[53] Xia X. Domains and Functions of Spike Protein in Sars-Cov-2 in the Context of Vaccine Design. *Viruses*. 2021;13(1):109.10.3390/v13010109
[54] Lee E, Sandgren K, Duette G, Stylianou VV, Khanna R, Eden JS, et al. Identification of SARS-CoV-2 Nucleocapsid and Spike T-Cell Epitopes for Assessing T-Cell Immunity. *J Virol*. 2021;95(6):10.1128/jvi. 02002-20.10.1128/JVI.02002-20
[55] Farhud DD, Mojahed N. SARS-COV-2 Notable Mutations and Variants: A Review Article. *Iran J Public Health*. 2022;51(7):1494-501.10.18502/ijph.v51i7.10083
[56] Salleh MZ, Derrick JP, Deris ZZ. Structural evaluation of the spike glycoprotein variants on SARS-CoV-2 transmission and immune evasion. *International journal of molecular sciences*. 2021;22(14):7425
[57] Parsons RJ, Acharya P. Evolution of the SARS-CoV-2 Omicron spike. *Cell Rep*. 2023;42(12):113444.10.1016/j.celrep.2023.113444
[58] Mohandas S, Yadav PD, Sapkal G, Shete AM, Deshpande G, Nyayanit DA, et al. Pathogenicity of SARS-CoV-2 Omicron (R346K) variant in Syrian hamsters and its cross-neutralization with different variants of concern. *EBioMedicine*. 2022;79:103997.10.1016/j.ebiom.2022.103997
[59] Ghimire D, Han Y, Lu M. Structural plasticity and immune evasion of SARS-CoV-2 spike variants. *Viruses*. 2022;14(6):1255
[60] Takeda M. Proteolytic activation of SARS-CoV-2 spike protein. *Microbiology and immunology*. 2022;66(1):15-23
[61] Cassari L, Pavan A, Zoia G, Chinellato M, Zeni E, Grinzato A, et al. SARS-CoV-2 S mutations: a lesson from the viral world to understand how human furin works. *International journal of molecular sciences*. 2023;24(5):4791
[62] Cui Z, Liu P, Wang N, Wang L, Fan K, Zhu Q, et al. Structural and functional characterizations of infectivity and immune evasion of SARS-CoV-2 Omicron. *Cell*. 2022;185(5):860-71 e13.10.1016/j.cell.2022.01.019

[63] Naveca F, Nascimento V, Souza V, Corado A, Nascimento F, Silva G, et al. Emergence and spread of SARS-CoV-2 P. 1 (Gamma) lineage variants carrying Spike mutations ?141-144, N679K or P681H during persistent viral circulation in Amazonas, Brazil. *image*. 2021;420(784):75.7
[64] Faraji N, Zeinali T, Joukar F, Aleali MS, Eslami N, Shenagari M, et al. Mutational dynamics of SARS-CoV-2: Impact on future COVID-19 vaccine strategies. *Heliyon*. 2024;10(9)
[65] Bashor L, Gagne RB, Bosco-Lauth AM, Bowen RA, Stenglein M, VandeWoude S. SARS-CoV-2 evolution in animals suggests mechanisms for rapid variant selection. *P Natl Acad Sci USA*. 2021;118(44):e2105253118.ARTN e2105253118
[66] Daniloski Z, Jordan TX, Ilmain JK, Guo X, Bhabha G, tenOever BR, et al. The Spike D614G mutation increases SARS-CoV-2 infection of multiple human cell types. *Elife*. 2021;10:e65365.10.7554/eLife.65365
[67] Gobeil SM, Janowska K, McDowell S, Mansouri K, Parks R, Manne K, et al. D614G Mutation Alters SARS-CoV-2 Spike Conformation and Enhances Protease Cleavage at the S1/S2 Junction. *Cell Rep*. 2021;34(2):108630.10.1016/j.celrep.2020.108630
[68] Santa Ardisson J, Vedovatti Monfardini Sagrillo M, Ramos Athaydes B, Corredor Vargas AM, Torezani R, Ribeiro-Rodrigues R, et al. Comparative spatial-temporal analysis of SARS-CoV-2 lineages B.1.1.33 and BQ.1.1 Omicron variant across pandemic phases. *Sci Rep*. 2025;15(1):10319.10.1038/s41598-025-95140-5
[69] FDA U. Fact Sheet for Health Care Providers. *Emergency Use Authorization (EUA) of Bamlanivimab, US Food and Drug Administration*. 2021
[70] Liu Y, Liu J, Plante KS, Plante JA, Xie X, Zhang X, et al. The N501Y spike substitution enhances SARS-CoV-2 infection and transmission. *Nature*. 2022;602(7896):294-9.10.1038/s41586-021-04245-0
[71] Chakraborty S. E484K and N501Y SARS-CoV 2 spike mutants Increase ACE2 recognition but reduce affinity for neutralizing antibody. *Int Immunopharmacol*. 2022;102:108424.10.1016/j.intimp.2021.108424
[72] Wadford DA, Baumrind N, Baylis EF, Bell JM, Bouchard EL, Crumpler M, et al. Implementation of California COVIDNet - a multi-sector collaboration for statewide SARS-CoV-2 genomic surveillance. *Front Public Health*. 2023;11:1249614.10.3389/fpubh.2023.1249614
[73] Suljic A, Zorec TM, Zakotnik S, Vlaj D, Kogoj R, Knap N, et al. Efficient SARS-CoV-2 variant detection and monitoring with Spike Screen next-generation sequencing. *Brief Bioinform*. 2024;25(4).10.1093/bib/bbae263
[74] Ricciardi S, Guarino AM, Giaquinto L, Polishchuk EV, Santoro M, Di Tullio G, et al. The role of NSP6 in the biogenesis of the SARS-CoV-2 replication organelle. *Nature*. 2022;606(7915):761-8.10.1038/s41586-022-04835-6
[75] Zhou R, Zeng R, von Brunn A, Lei J. Structural characterization of the C-terminal domain of SARS-CoV-2 nucleocapsid protein. *Mol Biomed*. 2020;1(1):2.10.1186/s43556-020-00001-4

How to cite this paper:

Saleh Ali Arishi, Wael A. Fathy, Mohammed Al-matary, Ahmed Bahieldin, Sherif Edris, Ahmed M. Ramadan. Phylogenomic Characterization of SARS-CoV-2 Variants in Jeddah, Saudi Arabia: Insights into Viral Evolution and Transmission Dynamics. *J Biol Insights*. 2025;1(1):1–29.

Arabic translation of title page

التوصيف الجينومي التطوري لسلالات فيروس SARS-CoV-2 في جدة بالمملكة العربية السعودية: رؤى حول تطور الفيروس وديناميكيات انتشاره

صالح علي العريشي[1]، وائل فتحي[2،1]*، محمد المطري[1]، أحمد باهي الدين[1]، شريف إدريس[1]، أحمد رمضان[1]*

[1]قسم العلوم البيولوجية، كلية العلوم، جامعة الملك عبدالعزيز، جدة، المملكة العربية السعودية.

[2]قسم النبات والأحياء الدقيقة، كلية العلوم، جامعة بني سويف، بني سويف، مصر.

المؤلفون المراسلون: aamara@kau.edu.sa | wfathy@stu.kau.edu.sa

الملخص

الهدف: هدفت هذه الدراسة إلى الوصف الجينومي لمتحورات فيروس SARS-CoV-2 في مدينة جدة بالمملكة العربية السعودية، باعتبارها وجهة رئيسية للحج والعمرة ومركزًا محوريًا في منطقة الشرق الأوسط، وذلك للمساهمة في توجيه السياسات الصحية الوطنية وفهم تطور الفيروس في المنطقة. **الطرق**: تم جُمِع ما مجموعه ٩٤ عينة من فيروس SARS-CoV-2 من مدينة جدة خلال الفترة من يناير إلى ديسمبر ٢٠٢٢. أُجري تسلسل الجينوم الكامل باستخدام Illumina CovidSeq، مع تطبيق إعادة البناء التطوري، ومحاذاة التسلسلات، وتحديد السلالات باستعمال Pangolin وNextclade، وتحليل الطفرات مع التركيز على جين البروتين الشوكي.(Spike) . **النتائج**: أظهر التحليل التطوري وجود عدة عناقيد ومجموعات من العزلات الفيروسية المتقاربة وراثيًا، مما يشير إلى محدودية التنوع الجيني وحدوث انتقال وانتشار محلي واسع النطاق. حيث كان المتحور BA.2 هو السائد بنسبة (٥٩,٥٧%) من العزلات ثم يليه BA.1 بنسبة (26.60%)، كما دل ظهور سلالات فرعية صغيرة من BA.2 على استمرار تطور الفيروس. كما لوحظ مناطق شائعة للطفرات بشكل كبير في جين البروتين الشوكي بلغت (٤٢,١٩%) ومنطقة(35.94%) ORF1ab ، حيث شكّلت الطفرات أحادية النوكليوتيد ٩٢,٣١% من إجمالي الطفرات. وكانت أهم طفرات البروتين الشوكي، بما في ذلك (98.94%) D614G و(100%) P681H/N679K، شبه ثابتة في معظم العزلات، إلى جانب عدة تغيّرات في نطاق ارتباط المستقبل (RBD) المرتبطة بدخول الفيروس والتهرب المناعي. **الاستنتاج**: كان المتحور BA.2 وتفرعاته الأكثر هيمنة في مدينة جدة خلال عام ٢٠٢٢، وأشارت أنماط الطفرات إلى تطور تكيفي ناتج عن الضغط المناعي. كما أن تركز الطفرات في مناطق جينومية رئيسية يعكس استمرار تكيف الفيروس لتعزيز قابليته للانتقال والتهرب من الاستجابة المناعية. وتوفر هذه النتائج خلفية علمية مهمة للرصد الجينومي وصياغة السياسات الصحية المبنية على الأدلة في المملكة العربية السعودية ودول الشرق الأوسط الأخرى.

الكلمات المفتاحية: فيروس SARS-CoV-2 | المملكة العربية السعودية | الفحص الجينومي | متحورات أوميكرون | الطفرات الفيروسية | بروتين الشوكي/ منطقة ارتباط المستقبل.