DOI: 10.24996/ijs.2026.67.3.15

ISSN: 0067-2904

Characterization of Mutational Clusters in the N-Terminal Domain of SARS-CoV-2 Spike Protein

Noor Saber[1]*, Nuha J Kandala[2]
[1]*National Center of Hematology, Mustansiriyah University, Iraq*
[2]*Department of Biotechnology, College of Science, University of Baghdad, Baghdad, Iraq*

Received: 31/10/2024 Accepted: 20/3/2025 Published: xx

Abstract

Since the first report of SARS-CoV-2 in Wuhan, China, the virus has undergone multiple mutations leading to the emergence of variants of concern. While mutations in the receptor binding domain (RBD) of the spike gene have been extensively studied, mutations and indels in the N-terminal domain (NTD) have received less attention. In this study, we use a simple and cost-effective Sanger sequencing approach by targeting a 757bp region in the NTD of the spike gene. Analysis results showed the dominancy of G142D mutation resulted from single nucleotide polymorphism (SNP). Furthermore, our results showed the prevalence of two distinctive mutational clusters, the first pattern includes the 69H/70V, 143V, 144Y, L212I, and 211N deletions, insertion of 214EPE and T95I substitution. The second cluster includes one amino acid change of V213G. Subsequent analysis revealed the persistence of these genetic changes in current global circulating strains and emergence of new low circulating mutations within studied viral isolates represented by W258G. Our findings highlight the genetic changes shared among the studied isolates, providing a mutational catalogue of the NTD region and establishing a reference database for further local studies. Additionally, this study introduces a novel primer pair, designed to enable a rapid and cost-effective approach for SARS-CoV-2 lineages tracking.

Keywords: SARS-CoV-2, N-Terminal Domain, Mutational Clusters, Sanger Sequencing, Spike.

توصيف مجاميع الطفرات في المجال الطرفي الاميني للجين الشوكي في فايروس *SARS-CoV-2*

نور صابر [1]*, نهى جوزيف قندلا[2]
[1]التقنيات الاحيائية, كلية العلوم, جامعة بغداد, بغداد, العراق
[2]المركز الوطني لبحوث وعلاج امراض الدم، الجامعة المستنصرية، بغداد، العراق

الخلاصة

منذ التقرير الاول لـ SARS-CoV-2 في ووهان، الصين، خضع الفايروس لطفرات متعددة ادت الى ظهور المتغيرات المثيرة للقلق. في حين ان الطفرات في منطقة ارتباط المُستقبل (RBD) للجين الشوكي قد تمت

* Email: noor.saber@uomustansiriyah.edu.iq

دراستها على نطاق واسع، الا ان الطفرات والـ indels قد حضيت بأهتمام اقل. فهم التغيرات الوراثية في منطقة NTD امر بالغ الأهمية لتحديد كيفية تطور الفيروس للتهرب من الجهاز المناعي وزيادة الإمراضية. في هذه الدراسة، قمنا باجراء تحليل التتابع الوراثي بطريقة سانجر لقُطع من الـNTD بحجم 757 زوج قاعدي من 88 عينة وأجرينا تحليلات Insilco لتحديد الطفرات والإدراج والحذف في المحتوى الوراثي. وجدنا أن 97٪ من العينات التي تحتوي على حذف الـ V70 / H69 كانت مصحوبة بتغييرات جينية محددة ، بما في ذلك T95I و V143 و Y144 و N211 و L212I وإدخال 214 EPE. وكشف التحليل اللاحق عن استمرار هذه التغيرات الجينية في السلالات العالمية المنتشرة الحالية. تسلط النتائج التي توصلنا إليها الضوء على التغيرات الجينية المشتركة بين العينات العراقية في منطقة الـNTD و توفر دليل طفرات شامل يعمل كقاعدة بيانات مرجعية للدراسات المحلية المستقبلية. بالإضافة إلى ذلك، تقدم هذه الدراسة زوجًا جديدًا من البادئات، مصممًا لتمكين نهج سريع وفعال من حيث التكلفة لتتبع سلالات SARS-CoV-2.

1. Introduction

Severe acute respiratory syndrome coronavirus 2 is a highly contagious virus that is spread via inhalational or direct contact routes and caused the recent coronavirus disease of 2019 (COVID-19) pandemic. In Iraq, COVID-19 was reported for the first time in Al-Najaf city south of Baghdad on 24th of February 2020 [1-4]. Rapid implementation of epidemiological monitoring has led to the development of many protocols, software, and databases to track emerging variants and their genetic signatures within infected patients [5,6]. In the context of SARS-CoV-2 viral tracking and monitoring, next-generation sequencing (NGS) technologies are highly accurate and reliable. Nonetheless, Sanger sequencing remains a reasonable choice for targeted mutation analysis due to its rapid and cost-effective nature in resource-limited settings. This technique has successfully identified mutations within the spike gene, which can threaten public health by reducing the effectiveness of available vaccines and therapeutics, as well as the accuracy of standard diagnostic tests [7-10]. The NTD is a hotspot region that harbored many mutations and indels that participated in emergence of new circulating lineages and sublineages [11,12]. The NTD polypeptide encompasses epitopes that are identified by neutralizing antibodies of host adaptive immune response and mediates interactions with glycosyl groups in the cellular glycoenvironment and is actively implicated in host cell surface attachment [13,15]. Recently, it has been discovered that the SARS-CoV-2 Spike protein binds specifically to the tyrosine kinase receptor UFO (AXL). AXL is a host receptor that is involved in interaction with the NTD of the Spike protein. This interaction will enhance viral entry to the targeted cells, particularly human bronchial and pulmonary epithelial cells. However, acquired mutations in the NTD may induce conformational changes, that could reduce binding affinity or enhance viral fitness by reducing the neutralizing activity as the antibody-mediated protection depends on the target antigen's structure. Furthermore, such mutations could compromise the efficacy of existing vaccines and therapeutics [16-19]. The NTD encompasses different regions known as N1 (14-26 residues), N2 (67-79 residues), N3 (141-156 residues), N4 (177-186 residues), and N5 (246-260 residues). It contains a specific β-hairpin structure covered by four N-linked glycans including four aspartic amino acids at codons 17, 75, 122, and 149 [20-21]. In this study, we aim to investigate the genetic variations of NTD of SARS-CoV-2 spike protein region through employing a targeted Sanger sequencing approach. Using a specific primer set derived from the ARTICV3 primer pool, we were able to synthesize an amplicon fragment spanning the NTD region in local viral isolates.

2. Methodology

2.1 Sampling Strategy

A total of 88 nasopharyngeal swabs were obtained from Baghdad Central Public Health Laboratory (CPHL) during the fourth wave of the pandemic between January and March 2022.

All specimens were referred originally from different public health laboratories in Iraqi governorates. Viral RNA was automatically extracted using the ExiPrep™ 96 Lite (A-5250, BIONEER) with the ExiPrep™ Viral DNA/RNA extraction kit (K-4614, BIONEER), following the manufacturer's instructions. RT-PCR was performed using the TaqPath™ COVID-19 Combo Kit (Thermo Fisher) on the Applied Biosystems 7500 Fast Real-Time PCR. All specimens with cycle threshold (Ct) values ranging from 14.9 to 21.6 were selected to avoid amplicon abundance dropout, which is typically observed in low viral load samples [22,23].

2.2 Primers Design

The synthesized cDNA served as a template for amplifying a 757bp NTD fragment using the SubA_21587F: `5-CCACTAGTCTCTAGTCAGTGTGTT-3` and SubA_22344R: `5-CCAGCTGTCCAACCTGAAGA3` ARTICV3 primers sourced from the ARTIC Network. ARTIC V3 is a primer pool designed for next-generation sequencing applications which amplify the whole genome of SARS-CoV-2. In this study, one primer from two different sets (one forward and one reverse) was used to cover the NTD region. The primers aligned successfully to the genome Wuhan-Hu-1 parental strain (NCBI Reference Sequence: NC_045512.2) and the WIV04 spike sequence of Global Initiative on Sharing All Influenza Data (GISAID), accession number: hCoV-19/Wuhan/WIV04/2019 The GISAID reference sequence share 100% identity to the early SARS-CoV-2 outbreak sequences with high-quality data and great coverage and fully identical to Whuan-Hu-1 sequence [24-26]. The designed primers are intended to cover the NTD region from 13 to 304 and the trimer interface at codon ranges from 38 to 282.

2.3 PCR Assay

Complementary DNA (cDNA) synthesis was performed using the Promega GoScript™ Reverse Transcription Mix with Random Primers system (A2800) following manufacturer recommended protocols. The quality of the generated cDNA was assessed using Quantus fluorometer. GoTaq® Master Mix, Promega (M7122) was used for the PCR amplification, and the program consisted of an initial denaturation step at 95°C for 5 minutes, followed by 40 cycles of denaturation at 95°C for 30 seconds, annealing at 60°C for 30 seconds, and extension at 72°C for 1 minute. A final extension step was performed at 72°C for 7 minutes, followed by holding at 10°C for 10 minutes. The resulting fragments were assessed using gel electrophoresis and visualized using a gel documentation system.

2.4 Sanger Sequencing and Data Analysis

The sequencing protocol was carried out using an ABI3730XL automated DNA sequencer at Macrogen, South Korea. Data analysis was performed using Basic Local Alignment Tool (BLASTn)/NCBI for primary identification, and further analysis was handled mainly using Geneious Prime software and Mega X. The annotation of the Wuhan-Hu-1 parental strain gained from the NCBI, RefSeq database, all fragments aligned using Clustal Omega, aligned fragments mapped to the reference parental genome using map to reference assembly tool on Geneious prime. All mutations were identified using SNPs/Variation identifier tool in Geneious prime. Nextclade on Galaxy (https://usegalaxy.org/) was used to verify the amino acids variations and to identify SARS-CoV-2 variants. A phylogenetic tree was constructed using the Auspice platform (https://auspice.us/) powered by real SARS-CoV-2 genomic data from the GISAID repository.

The curated fragments were translated into amino acid sequences using Geneious Prime and then modelled using the Phyre2 platform. Further protein visualization and variant labelling were performed using PyMOL. CoVsurver (v1.22.06) was used to visualize these changes

across the entire Spike protein, and EpiCoV™ was used to track variations in currently circulating SARS-CoV-2 strains

2.5 Data Availability

All adopted methodology is deposited online and accessible from the following published dx.doi.org/10.17504/protocols.io.x54v9d63zg3e/v1. All analysed sequenced fragments are published online at NCBI Genbank and on GISAID. Fragments accession numbers on both repositories are available in supplementary file 1.

3. Results

The TaqPath assay discriminated 88 SARS-CoV-2 included cases into 36 possessing the 69H/70V deletion and 44 without it. The deletion was identified through spike gene target failure (SGTF) which was used in this study as a proxy marker for investigating circulating SARS-CoV-2 variants and as a prescreening step prior to sequencing. The NTD FASTA files were trimmed using Geneious Prime software to remove poor-quality reads, generating fragments of varying lengths. A total of 88 targeted sequences were analyzed, spanning codon 27 to codon 258 within the NTD of the spike gene. The current reads covered a big portion of the NTD, within which most of the crucial mutations fall within this codon range. NCBI BLASTn results showed that 49 sequences were 100% identical to Genbank deposited sequences, while 39 sequences had disagreements at different loci with identity percentages of 98% and 99%, mismatches were present and represented by 34 transversions, 19 transitions, 2 insertions, and 11 deletions as depicted in supplementary file 2. Upon mapping the NTD reads to the reference sequence, all mutational events within the targeted region were identified. In silico analysis revealed a total of 78 polymorphisms, including six deletions, four insertions, 25 transitions, 37 transversions, and eight substitutions, the results fully listed in supplementary file 3. Among these variations, the nucleotide changes at codon 142 were dominant, with a frequency of 100% due to a SNP that changed the wild glycine to aspartic acid (G142D). Another frequently occurring mutation was observed at codon 213, where thymine was replaced by guanine, resulting in the substitution of valine with glycine (V213G), with a frequency of 60%. The percentages of transitions and transversions detected are presented in Figure 1.

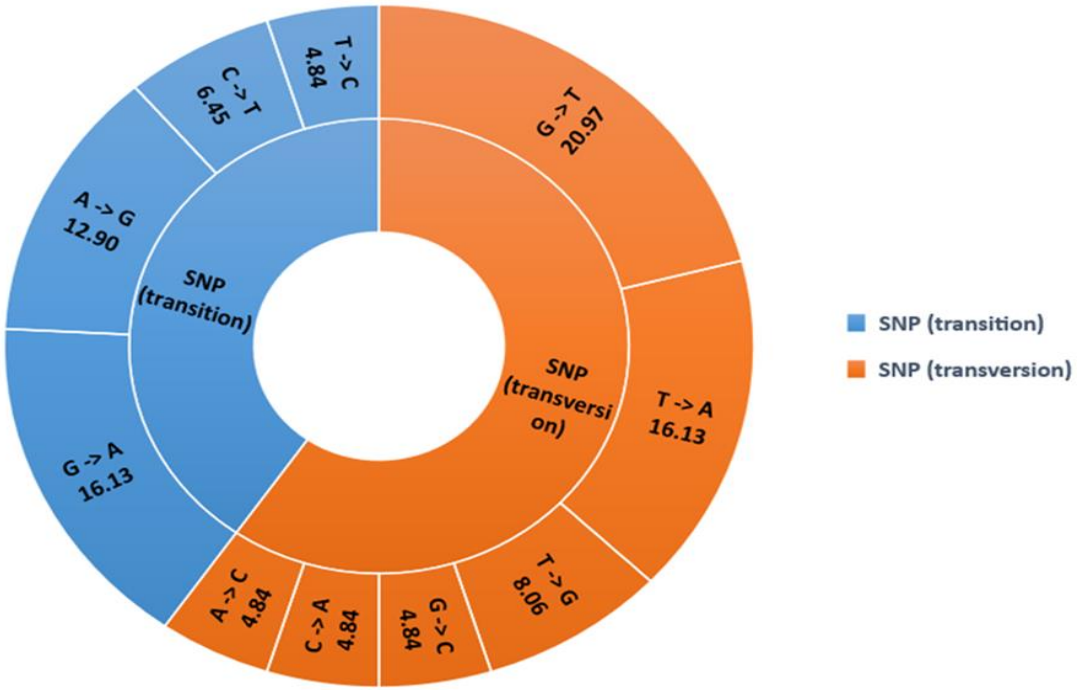

Figure 1: Hierarchical concentric ring representation of transition/transversion SNPs using sunburst chart, 48.39% of listed substitution reduce GC% content within NTD fragments, highest mutation frequency rate was G/T transversions, and G/A transitions with a percentage of 20.97% and 16.13% respectively.

To identify conserved areas in the NTD, we determined the minimum and maximum lengths of nucleotide disagreements resembling the length of each variation on the entire investigated region from nucleotide 82 to 792 in the NTD region of the spike gene. Such regions are instrumental in designing universal primers capable of detecting all SARS-CoV-2 variants. The results showed that N2 and N5 regions were highly variable. On other hand, the downstream area of the N2 was conserved along with other areas distributed in the area between the upstream regions of N3 and N4. These variations are represented in a scatter plot, as depicted in Figure 2.

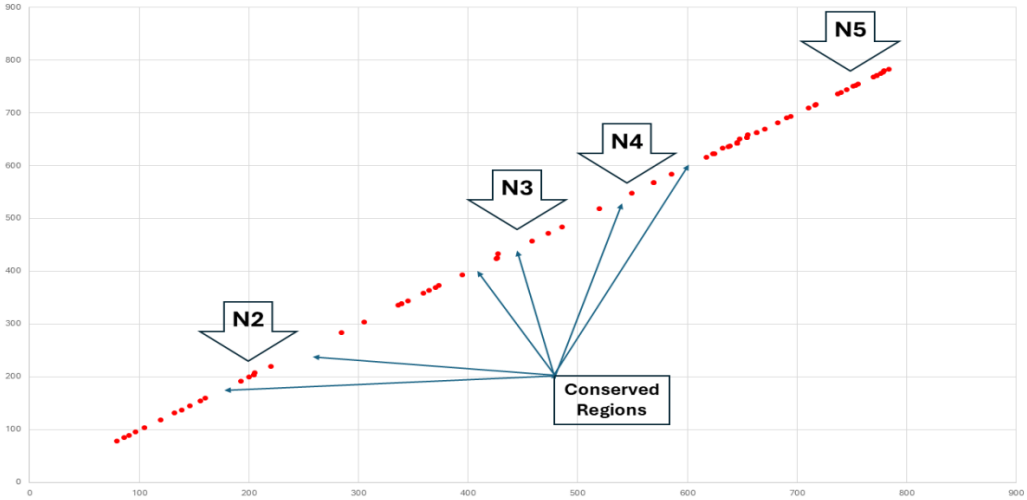

Figure 2: Scatter chart illustration highlighting conserved regions in NTD fragments, each dot represents a mutated locus across the studied region of NTD fragments. The x axis represents the minimum length, the Y axis represents the maximum length of each genetic variation.

Bioinformatics analysis showed mutational clustering shared between 69/70 deletion, T95I, 143V, 144Y, N211 deletions, L212I, and the insertion of 214 EPE within 97% of analyzed fragments. Further analysis conducted using the GISAID EpiCov platforms confirms the clusters, The results showed 34 different patterns built upon the mutational content of each sequenced fragment. Mutational clusters could be broadly categorized into two types: patterns with and without the 69/70 deletion. Table 1 provides a detailed breakdown of these patterns.

Table 1: Statistical representation of mutational patterns shared among NTD fragments; all SNPs, insertion (ins) and deletion (del) are clustered within revealed patterns.

ID	Amino Acid Substitutions within NTD fragments of Spike gene	Occurrences	Percentage
1	A67V, L54M	1	1.14%
2	D228H, G232C, G142D, V213G, E224K, I231V, R246K, R237T, S247I	1	1.14%
3	G142D	1	1.14%
4	G142D, K206T	1	1.14%
5	G142D, V213G	31	35.23%
6	G142D, V213G, A27S	1	1.14%
7	G142D, V213G, A27S, T29I	1	1.14%
8	G142D, V213G, N74D	1	1.14%
9	G142D, V213G, P251T, A260S	1	1.14%
10	G142D, V213G, Q52R	1	1.14%
11	G142D, V213G, T29I	1	1.14%
12	G142D, V213G, W258G	3	3.41%
13	G257R, G142D, V213G	1	1.14%
14	H69 deletion, V70 deletion, A67V, G142D, V213G	1	1.14%
15	H69del, T95I, A67V, E132Q, K113N, V70del	1	1.14%
16	H69del, T95I, A67V, L212I, D40G, N211del, G142D, V70del, V143del, Y145del, Y144del, 214EPE insertion	1	1.14%
17	H69del, T95I, A67V, L212I, ins214EPQ, N211del, G142D, Q218H, V70del, V143del, Y145del, Y144del, R190K	1	1.14%
18	H69del, T95I, A67V, L212I, N211del, G142D, M153K, V70del, V143del, Spike_Y145del, Y144del, ins214EPE	1	1.14%
19	H69del, T95I, A67V, L212I, N211del, G142D, R102K, V70del, V143del, Y145del, Y144del, ins214EPE	1	1.14%
20	H69del, T95I, A67V, L212I, N211del, G142D, T208M, V70del, V143del, Y145del, Y144del, ins214EPE	1	1.14%
21	H69del, T95I, L212I, N211del, G142D, T259K, V70del, V143del, Y145del, Y144del, W258G, ins214EPE	2	2.27%
22	H69del, T95I, A67V, L212I, N211del, G142D, V70del, V143del, Y145del, Y144del, ins214EPE	21	23.86%
23	H69del, T95I, A67V, L212I, N211del, G142D, V70del, V143del, Y145del, Y144del, W258G, ins214EPE	2	2.27%
24	H69del, T95I, A67V, Q173H, L212I, N211del, G142D, R246K, V70del, V143del, Y145del, Y144del, ins214EPE	1	1.14%
25	H69del, T95I, A67V, Q173H, R190T, G142D, S162T, R158K, V70del, W64C, V143del, Y145del, Y144del	1	1.14%
26	ins214EPK, H69del, T95I, A67V, L212I, N211del, G142D, V70del, V143del, Y145del, Y144del	1	1.14%
27	N122F, F32L, N125F, V120A, T124A	1	1.14%
28	Q173H, G142D, V213W	1	1.14%
29	Q173H, G252A, G142D, V213G, H49P, Q218H	1	1.14%
30	Q183H, G142D, K195N	1	1.14%
31	Q239R, G142D, V213G	1	1.14%
32	T95I	1	1.14%
33	T95I, K113N, Q115H	1	1.14%
34	A27S, G142D, Q239R, S247I, V213G	1	1.14%
	TOTAL	88	100.00%

The Nextclade, Galaxy online platform used to predict variants based on mutations among processed fragments, all analyzed mutations were acquired by Omicron variants, specifically, Omicron B.1.1.529 and its sub-lineage, the BA.2. Variant and related clades confirmation of carried on through phylogenetic tree using Auspice, the results showed that all observed mutations were consistent with those previously reported in Omicron which confirm variant prediction. Furthermore, the results showed that those mutations are related to the main clades in the global phylogeny of SARS-CoV-2, those which includes 21K and 21 L omicron clades, two fragments furtherly characterized within 22A and 22B branches of the 21L clades. The phylogenetic tree is illustrated in Figure 3.

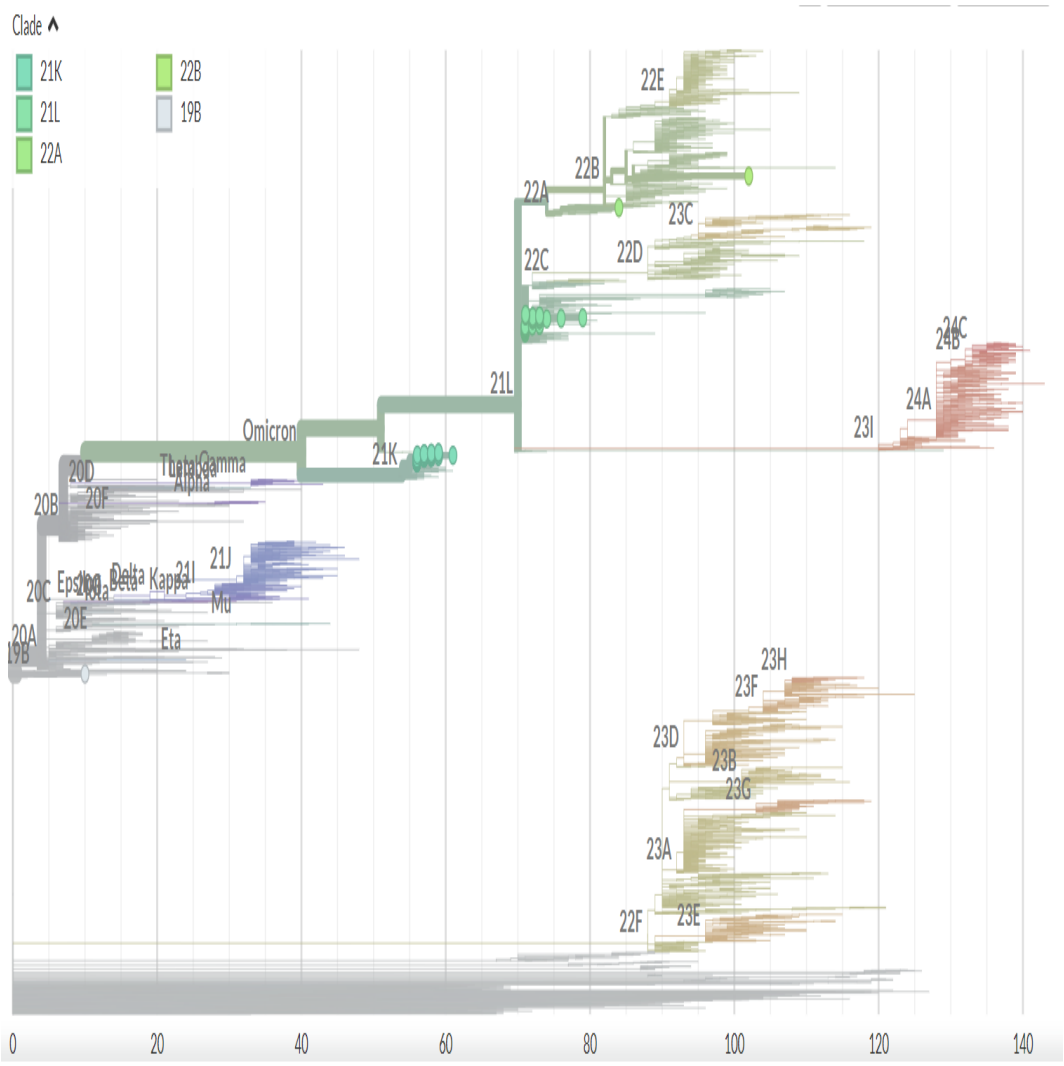

Figure 3: Phylogenetic analysis of NTD fragments demonstrates consistent mutations within Omicron, specifically in the 21L and 21K clades. This analysis highlights the relationship between local isolates and globally reported lineages, providing insights into their evolutionary connections.

The mutational clusters highlighted the presence of a highly variable site in the codon range of 214 – 216 in the upstream region of N5. This variation resulted from a deletion event out of the proximal reading frame at codon 211/212 and is accompanied by an insertion of EPE amino acids in codon 214, as shown in Figure 4.

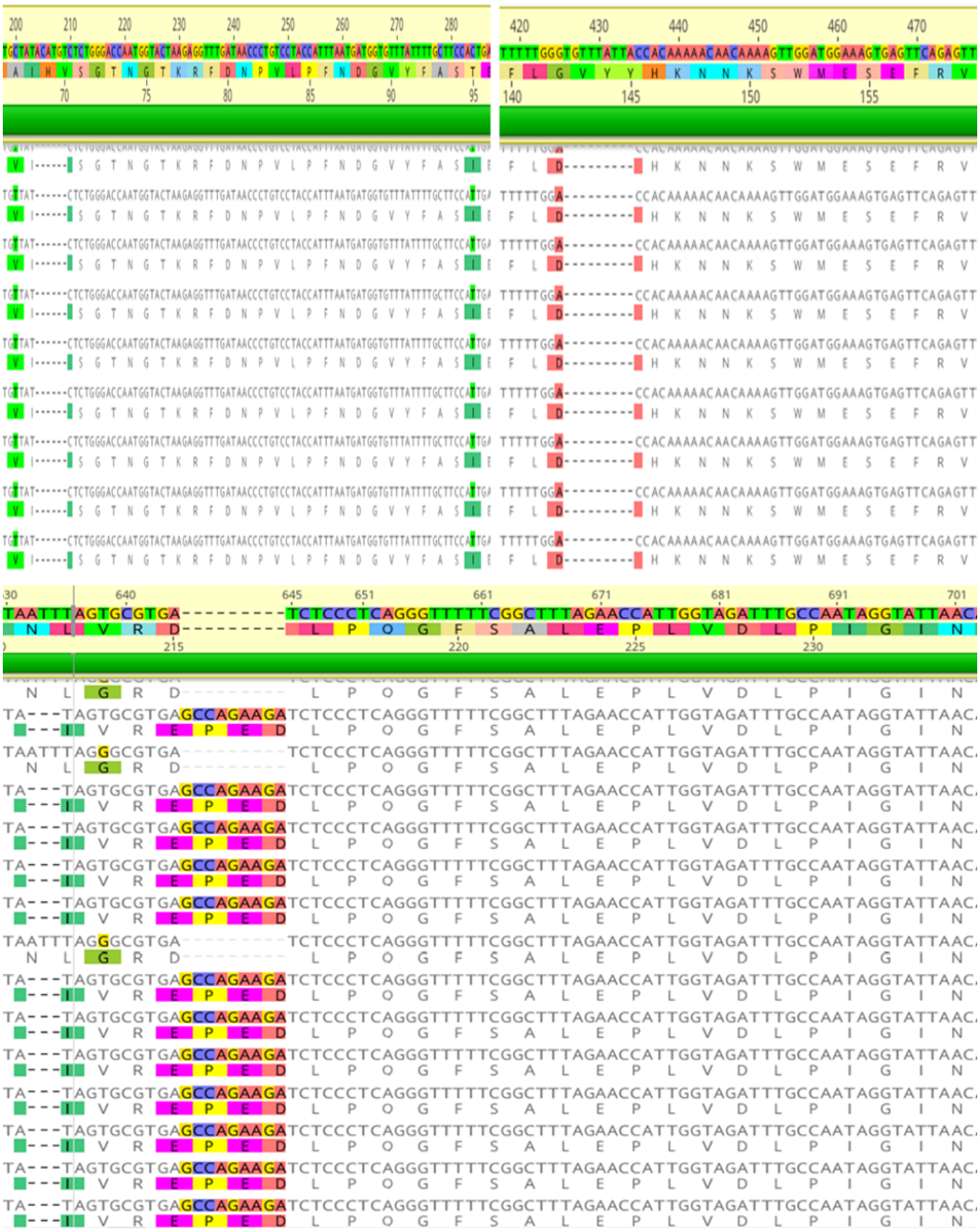

Figure 4: NTD fragments mapped to Wuhan reference Spike gene sequence showing A67V, 69H/70Vdeletion, G142D, 143V/144Y deletion, 211N deletion, L212I, and 214 EPE insertion as one cluster of amino acids changes, the second cluster includes only one unique V213G mutation.

To label these changes in corresponding structural peptide, Phyre2 was used for the structure modelling, and the protein model visualized and inspected using Pymol. The results showed a tendency toward mutations in the loops of the NTD rather than α-helices or β-sheets, those changes are listed in Figure 5.

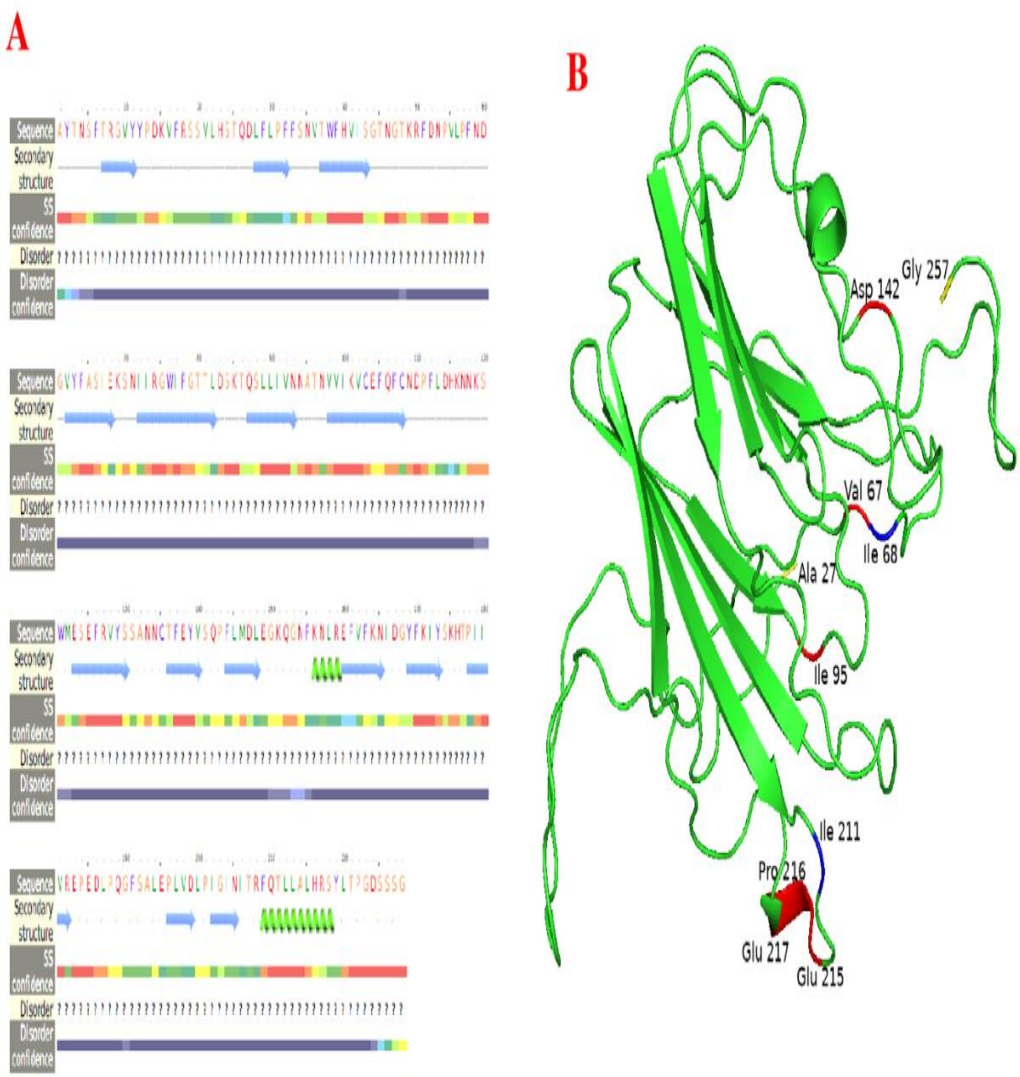

Figure 5: A model structural representation of SARS-CoV-2 NTD fragments, based on the 8UPX PDB template with 100% confidence. A: Secondary structure prediction of studied NTD fragment, blue arrows represent the beta sheets while green spiral label stands for alpha helices, B: representation of detected amino acid changes within the NTD fragments, the green arrows represent the beta sheets, spring structures represent alpha helices, while wired structure stands for loops, the fragment starts with alanine at codon 27 and end with glycine at codon 257 which labelled with yellow colour.

The GISAID CoVSurver tool was used to analyze the epidemiological status, frequency and persistence of these amino acid changes among different SARS-CoV-2 strains across worldwide. The results showed that these changes persist in newer emerging variants like JN.1. Furthermore, amino acid changes represented by A67V, 69/70 deletion, T95I, G142D, and 143/144 deletion show high dominance across different SARS-CoV-2 lineages and sublineages, as depicted in Figure 6.

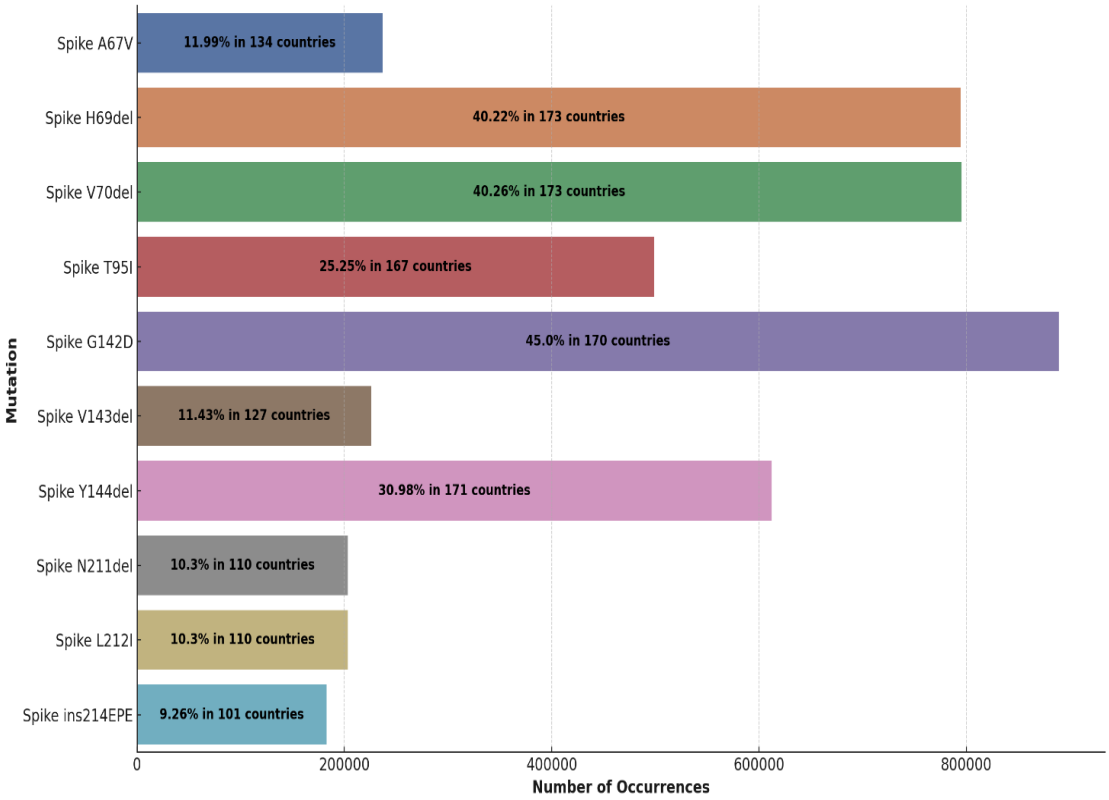

Figure 6: Dominancy of the detected changes in rSARS-CoV-2 isolates reported worldwide.

4. Discussion

Genomic surveillance systems have become crucial in monitoring the evolution of SARS-CoV-2 during the pandemic. Viral widespread presence in the human population has facilitated the emergence of several problematic variants through acquiring a large number of non-synonymous mutations. One of the methods for tracking variants is the detection of 69/70 deletion within the NTD of the Spike gene. The H69/V70 deletion first emerged in the Alpha B1.1.7 variant and emerged again in Omicron variants, although it is missing from the BA.2, BA.2.12.1, and BA.2.75 [24-28]. The H69/V70 deletion falls within the NTD N2 region, which is a focal point in monitoring SARS-CoV-2 variants of concerns, as it increases the efficiency of infectivity, enhances Spike cleavage, and boosts the syncytium formation process [29,30]. The results of the current study indicate a high proportion of SNPs resulting from both transition and transversion events, these types of SNPs are linked with a decrease in the overall GC%. content of the viral genome. This phenomenon could be attributed to a reduction in the energy required for replication, resulting in faster reproduction. Alternatively, it could facilitate the emergence of new variants that are better adapted to the host environment, which induces a positive fitness effect, especially in the human lung [31-33]. Current analysis revealed that G142D was dominant across all the analysed fragments with 100% frequency. This mutation was first reported in March 2020 and is present in 45% of all sequences in GISAID and persists among Delta and Omicron variants [34,35]. Among 36 NTD fragments with the 69/70 deletion, a transition of cytosine to thymine at codon 95 led to a change of threonine to isoleucine. It has been reported T95I enhance the effect of G142D, and these mutations are linked to an increased viral load, which is expected in COVID-19 cases with low Ct value [36]. Additionally, NTD surface topography in the antibody binding area is likely altered by G142D and T95I, so it may be linked to host immune evasion. The etiological explanation is not definitive, but it may arise from the conversion of a β-strand to α-helix polypeptide structure around 159-167 amino acids range and 183-190 amino acid range [36]. In general, two sets of amino acid changes were

dominant. The first set of changes harboured the 69/70, 143-145, and codon 211 deletions, T95I and L212I mutations, and the insertion of 214 EPE. Spike 214 EPE insertion was frequently reported among Omicron in different lineages [37,38]. In the second group of NTD sequences, all mutations clustered within the first pattern are missing. However, it possesses one unique change of V213G resulting from a SNP transversion of thymine to guanine. One NTD fragment was found to contain both the 69/70 deletion and V213G mutations, but it lacked the other mutations noted in in the first pattern. The prevalence of G142D and V213G in the NTD may indicate that these mutations confer an adaptive advantage, primarily by enhancing immune evasion through the reduction of neutralizing antibody activity [39]. Structural modelling of these mutations suggests alterations in the Spike protein's topology, which could enhance its affinity for host receptors or shield critical epitopes from immune recognition. For instance, the G142D mutation, positioned near the RBD, could theoretically alter spike's structural characteristics, facilitating a more favourable interaction with human receptors [40,43]. Furthermore, the observed tendency toward maintaining mutations in the loops of the NTD can be attributed to the fact that the loops are more flexible, less constrained than α-helices and β-sheets, and less likely to disrupt Spike integrity. Moreover, loops tend to be exposed on the protein surface, which resembles a primary target for neutralizing antibodies. This can justify the cause of harbouring the mutational hotspots area in the loops rather than other than α-helices and β-sheets. These factors collectively contribute to the high frequency of mutations observed in the loops of the NTD [11,44-46]. Several low-frequency mutations were also identified within the studied region. These include A27S, T29I, F32L, H49P, Q52R, N74D, Q173H, Q183H, K195N, K206T, V213W, Q218H, E224K, D228H, I231V, G232C, R237T, Q239R, R246K, S247I, P251T, G252A, and G257R. While the exact role of these mutations is not yet clear, further surveillance may reveal their persistence or displacement among circulating variants. These low-frequency mutations, in addition to the dominant V213G mutation, may also be of interest to researchers studying the evolution and transmission of SARS-CoV-2. We also identified a mutation signature represented by a transition of thymine to guanine at codon 258 in both clusters, resulting in the amino acid change from tryptophan to glycine. The amino acid change W258G has only been reported 20 times in GISAID Spike sequences, indicating a low circulating frequency. The exact induced effect of this mutation is unknown, highlighting the need for more local studies.

Conclusions

While NGS platforms are renowned for their high throughput and extensive data collection capabilities, the Sanger sequencing method employed in our study proves invaluable tool for precise mutation identification, especially in resource-limited settings. Our findings underscore the high prevalence of two mutational clusters in local isolates, which illustrate the rapid evolutionary adaptations of SARS-CoV-2. Furthermore, a tendency toward changing amino acids within loops structures of NTD polypeptide which conger additional flexibility and enrollment in the host immune evasion process. These adaptations toward neutralizing antibodies challenge existing public health strategies and should bring more attention to the NTD harbored mutations.

Acknowledgment

The authors would like to acknowledge all researchers and contributors who submit their data to the GISIAD repository. Furthermore, the authors thank National Influenza Center, Central Public Health Laboratory for their support.

Ethical Approval

The research topic was approved by the ethics committee at the College of Science, University of Baghdad (CSEC/0222/0155) on January 28th, 2022.

Conflict of Interest

The authors declare that they have no conflicts of interest.

Funding

The current work didn't receive funds or grants from public or private agencies.

References

[1] D. K. Sabir, "Analysis of SARS-COV2 spike protein variants among Iraqi isolates," *Gene Reports*, vol. 26, p. 101420, 2022. doi: 10.1016/j.genrep.2021.101420.

[2] R. Alsayed, A. A. Ali, R. Makia, M. Kadhom, R. Raheem, O. Al-Obaidi, A. Hadi, D. K. Sabir, and E. Yousif, "Challenges facing Iraq to tackle the spread of COVID-19: An overview," *Journal of University of Anbar for Pure Science*, vol. 2020, no. 2, pp. 22–27, 2020. doi: 10.37652/JUAPS.2020.14.2.5.

[3] R. Z. Jasim, "Biochemical Action of Vaccines in Iraqi Patients with COVID-19 Infection," *Baghdad Science Journal*, vol. 20, pp. 1469–1479, 2023, doi: 10.21123/bsj.2023.8750.

[4] S. K. Mohammed, M. M. Taha, E. M. Taha, M. Najim, and A. Mohammad, "Cluster Analysis of Biochemical Markers as Predictor of COVID-19 Severity," *Baghdad Science Journal*, vol. 19, no. 6,p. 1423, 2022. doi: 10.21123/bsj.2022.4242.

[5] C. Lam, K. Gray, M. Gall, R. Sadsad, A. Arnott, J. Johnson-Mackinnon, W. Fong, K. Basile, J. Kok, D. E. Dwyer, V. Sintchenko, and R. J. Rockett, "SARS-CoV-2 genome sequencing methods differ in their abilities to detect variants from low-viral-load samples," *Journal of Clinical Microbiology*, vol. 59, no. 11, p. e0104621, 2021. doi: 10.1128/JCM.01046-21.

[6] C. Lam, K. Gray, M. Gall, R. Sadsad, A. Arnott, J. Johnson-Mackinnon, W. Fong, K. Basile, J. Kok, D. E. Dwyer, V. Sintchenko, and R. J. Rockett, "SARS-CoV-2 genome sequencing methods differ in their abilities to detect variants from low-viral-load samples," *Journal of Clinical Microbiology*, vol. 59, no. 11, p. e0104621, 2021. doi: 10.1128/JCM.01046-21.

[7] S. H. Lee, "A routine Sanger sequencing target specific mutation assay for SARS-CoV-2 variants of concern and interest," *Viruses*, vol. 13, no. 12, p. 2386, 2021. doi: 10.3390/v13122386.

[8] M. F. Bezerra, L. C. Machado, V. D. C. V. De Carvalho, C. Docena, S. P. Brandão-Filho, C. F. J. Ayres, M. H. S. Paiva, and G. L. Wallau, "A Sanger-based approach for scaling up screening of SARS-CoV-2 variants of interest and concern," *Infection, Genetics and Evolution: Journal of Molecular Epidemiology and Evolutionary Genetics in Infectious Diseases*, vol. 92, p. 104910, 2021. doi: 10.1016/j.meegid.2021.104910.

[9] A. S. Lauring and E. B. Hodcroft, "Genetic variants of SARS-CoV-2-What do they mean?," *Journal of American Medical Association*, vol. 325, no. 6, pp. 529–531, 2021. doi: 10.1001/jama.2020.27124.

[10] A. Grimaldi, F. Panariello, P. Annunziata, T. Giuliano, M. Daniele, B. Pierri, C. Colantuono, M. Salvi, V. Bouché, A. Manfredi, M. C. Cuomo, D. Di Concilio, C. Tiberio, M. Fiorenza, G. Portella, I. Cimmino, A. Sorrentino, G. Fusco, M. R. Granata, P. Cerino, and D. Cacchiarelli, "Improved SARS-CoV-2 sequencing surveillance allows the identification of new variants and signatures in infected patients," *Genome Medicine*, vol. 14, no. 1, p. 90, 2022. doi: 10.1186/s13073-022-01098-8.

[11] M. McCallum, A. De Marco, F. A. Lempp, M. A. Tortorici, D. Pinto, A. C. Walls, M. Beltramello, A. Chen, Z. Liu, F. Zatta, S. Zepeda, J. di Iulio, J. E. Bowen, M. Montiel-Ruiz, J. Zhou, L. E. Rosen, S. Bianchi, B. Guarino, C. S. Fregni, R. Abdelnabi, and D. Veesler, "N-terminal domain antigenic mapping reveals a site of vulnerability for SARS-CoV-2," *Cell*, vol. 184, no. 9, pp. 2332–2347.e16, 2021. doi: 10.1016/j.cell.2021.03.028.

[12] P. C. Resende, F. G. Naveca, R. D. Lins, F. Z. Dezordi, M. V. F. Ferraz, E. G. Moreira, D. F. Coêlho, F. C. Motta, A. C. D. Paixão, L. Appolinario, R. S. Lopes, A. C. D. F. Mendonça, A. S. B. da Rocha, V. Nascimento, V. Souza, G. Silva, F. Nascimento, L. G. L. Neto, F. V. da Silva, I.

Riediger, M. D. C. Debur, A. B. Leite, T. Mattos, C. F. da Costa, F. M. Pereira, C. A. Dos Santos, D. B. Rovaris, S. B. Fernandes, A. Abbud, C. Sacchi, R. Khouri, A. F. L. Bernardes, E. Delatorre, T. Gräf, M. M. Siqueira, G. Bello, and G. L. Wallau, "The ongoing evolution of variants of concern and interest of SARS-CoV-2 in Brazil revealed by convergent indels in the amino (N)-terminal domain of the spike protein," *Virus Evolution*, vol. 7, no. 2, p. veab069, 2021. doi: 10.1093/ve/veab069.

[13] A. Klinakis, Z. Cournia, and T. Rampias, "N-terminal domain mutations of the spike protein are structurally implicated in epitope recognition in emerging SARS-CoV-2 strains," *Computational and Structural Biotechnology Journal*, vol. 19, pp. 5556-5567, 2021. doi: 10.1016/j.csbj.2021.10.004.

[14] X. Chi, R. Yan, J. Zhang, G. Zhang, Y. Zhang, M. Hao, Z. Zhang, P. Fan, Y. Dong, Y. Yang, Z. Chen, Y. Guo, J. Zhang, Y. Li, X. Song, Y. Chen, L. Xia, L. Fu, L. Hou, J. Xu, C. Yu, J. Li, Q. Zhou, and W. Chen, "A neutralizing human antibody binds to the N-terminal domain of the Spike protein of SARS-CoV-2," *Science*, vol. 369, no. 6504, pp. 650-655, 2020. doi: 10.1126/science.abc6952.

[15] M. Awasthi, S. Gulati, D. P. Sarkar, S. Tiwari, S. Kateriya, P. Ranjan, and S. K. Verma, "The sialoside-binding pocket of SARS-CoV-2 Spike glycoprotein structurally resembles MERS-CoV," *Viruses*, vol. 12, no. 9, p. 909, 2020. doi: 10.3390/v12090909.

[16] W. S, Qiu Z, Hou Y, Deng X, Xu W, Zheng T, Wu P, Xie S, Bian W, Zhang C, Sun Z, Liu K, Shan C, Lin A, Jiang S, Xie Y, Zhou Q, Lu L, Huang J, and Li X, "AXL is a candidate receptor for SARS-CoV-2 that promotes infection of pulmonary and bronchial epithelial cells," *Cell Research*, vol. 31, no. 2, pp. 126-140, 2021. doi: 10.1038/s41422-020-00460-y.

[17] Q. Li, J. Wu, J. Nie, L. Zhang, H. Hao, S. Liu, C. Zhao, Q. Zhang, H. Liu, L. Nie, H. Qin, M. Wang, Q. Lu, X. Li, Q. Sun, J. Liu, L. Zhang, X. Li, W. Huang, and Y. Wang, "The impact of mutations in SARS-CoV-2 Spike on viral infectivity and antigenicity," *Cell*, vol. 182, no. 5, pp. 1284-1294.e9, 2020. doi: 10.1016/j.cell.2020.07.012.

[18] A. Kwarteng, E. Asiedu, A. A. Sylverken, A. Larbi, S. A. Sakyi, and S. O. Asiedu, "Molecular characterization of interactions between the D614G variant of SARS-CoV-2 S-protein and neutralizing antibodies: A computational approach," *Infection, Genetics and Evolution*, vol. 91, 2021, doi: 10.1016/j.meegid.2021.104815.

[19] Y. Singh, N. K. Fuloria, S. Fuloria, V. Subramaniyan, D. U. Meenakshi, S. Chakravarthi, U. Kumari, N. Joshi, and G. Gupta, "N-terminal domain of SARS CoV-2 spike protein mutation associated reduction in effectivity of neutralizing antibody with vaccinated individuals," *Journal of Medical Virology*, vol. 93, no. 10, pp. 5726-5728, 2021. doi: 10.1002/jmv.27181.

[20] J. Zhang, T. Xiao, Y. Cai, and B. Chen, "Structure of SARS-CoV-2 spike protein," *Current Opinion in Virology*, vol. 50, pp. 173-182, 2021. doi: 10.1016/j.coviro.2021.08.010.

[21] P. Tong, A. Gautam, I. W. Windsor, M. Travers, Y. Chen, N. Garcia, N. B. Whiteman, L. G. A. McKay, N. Storm, L. E. Malsick, A. N. Honko, F. J. N. Lelis, S. Habibi, S. Jenni, Y. Cai, L. J. Rennick, W. P. Duprex, K. R. McCarthy, C. L. Lavine, T. Zuo, J. Lin, A. Zuiani, J. Feldman, E. A. MacDonald, B. M. Hauser, A. Griffths, M. S. Seaman, A. G. Schmidt, B. Chen, D. Neuberg, G. Bajic, S. C. Harrison, and D. R. Wesemann, "Memory B cell repertoire for recognition of evolving SARS-CoV-2 spike," *Cell*, vol. 184, no. 19, pp. 4969-4980.e15, 2021. doi: 10.1016/j.cell.2021.07.025.

[22] European Centre for Disease Prevention and Control, "Guidance for representative and targeted genomic SARS-CoV-2 monitoring – 2021," ECDC, Stockholm, 2021. [Online].

[23] K. Itokawa, T. Sekizuka, M. Hashino, R. Tanaka, and M. Kuroda, "Disentangling primer interactions improves SARS-CoV-2 genome sequencing by multiplex tiling PCR," *PLoS One*, vol. 15, no. 9, 2020. doi: 10.1371/journal.pone.0239403.

[24] P. Okada, R. Buathong, S. Phuygun, T. Thanadachakul, S. Parnmen, W. Wongboot, S. Waicharoen, S. Wacharapluesadee, S. Uttayamakul, A. Vachiraphan, M. Chittaganpitch, N. Mekha, N. Janejai, S. Iamsirithaworn, R. T. Lee, and S. Maurer-Stroh, "Early transmission patterns of coronavirus disease 2019 (COVID-19) in travellers from Wuhan to Thailand, January 2020," *Euro Surveillance*, vol. 25, no. 8, p. 2000097, 2020. doi: 10.2807/1560-7917.ES.2020.25.8.2000097.

[25] "Genomic sequencing of SARS-CoV-2: a guide to implementation for maximum impact on public health," *World Health Organization*, Geneva, 2021.

[26] G. Dudas, S. L. Hong, B. I. Potter, S. Calvignac-Spencer, F. S. Niatou-Singa, T. B. Tombolomako, T. Fuh-Neba, U. Vickos, M. Ulrich, F. H. Leendertz, K. Khan, C. Huber, A. Watts, I. Olendraitė, J. Snijder, K. N. Wijnant, A. M. J. J. Bonvin, P. Martres, S. Behillil, A. Ayouba, M. F. Maidadi, D. M. Djomsi, C. Godwe, C. Butel, A. Šimaitis, M. Gabrielaitė, M. Katėnaitė, R. Norvilas, L. Raugaitė, G. W. Koyaweda, J. K. Kandou, R. Jonikas, I. Nasvytienė, Ž. Žemeckienė, D. Gečys, K. Tamušauskaitė, M. Norkienė, E. Vasiliūnaitė, D. Žiogienė, A. Timinskas, M. Šukys, M. Šarauskas, G. Alzbutas, A. A. Aziza, E. K. Lusamaki, J. M. Cigolo, F. M. Mawete, E. L. Lofiko, P. M. Kingebeni, J. M. Tamfum, M. R. D. Belizaire, R. G. Essomba, M. C. O. Assoumou, A. B. Mboringong, A. B. Dieng, D. Juozapaitė, S. Hosch, J. Obama, M. O. Ayekaba, D. Naumovas, A. Pautienius, C. D. Rafaï, A. Vitkauskienė, R. Ugenskienė, A. Gedvilaitė, D. Čereškevičius, V. Lesauskaitė, L. Žemaitis, L. Griškevičius, and G. Baele, "Emergence and spread of SARS-CoV-2 lineage B.1.620 with variant of concern-like mutations and deletions," *Nature Communications*, vol. 12, no. 1, p. 5769, 2021. doi: 10.1038/s41467-021-26055-8.

[27] E. Volz, S. Mishra, M. Chand, J. C. Barrett, R. Johnson, L. Geidelberg, W. R. Hinsley, D. J. Laydon, G. Dabrera, Á. O'Toole, R. Amato, M. Ragonnet-Cronin, I. Harrison, B. Jackson, C. V. Ariani, O. Boyd, N. J. Loman, J. T. McCrone, S. Gonçalves, D. Jorgensen, R. Myers, V. Hill, D. K. Jackson, K. Gaythorpe, N. Groves, J. Sillitoe, D. P. Kwiatkowski, S. Flaxman, O. Ratmann, S. Bhatt, S. Hopkins, A. Gandy, A. Rambaut, and N. M. Ferguson, "Assessing transmissibility of SARS-CoV-2 lineage B.1.1.7 in England," *Nature*, vol. 593, no. 7858, pp. 266-269, 2021. doi: 10.1038/s41586-021-03470-x.

[28] M. Kidd, A. Richter, A. Best, N. Cumley, J. Mirza, B. Percival, M. Mayhew, O. Megram, F. Ashford, T. White, E. Moles-Garcia, L. Crawford, A. Bosworth, S. F. Atabani, T. Plant, and A. McNally, "S-Variant SARS-CoV-2 lineage B1.1.7 is associated with significantly higher viral load in samples tested by TaqPath polymerase chain reaction," *Journal of Infectious Diseases*, vol. 223, no. 10, pp. 1666-1670, 2021. doi: 10.1093/infdis/jiab082.

[29] B. Meng, S. A. Kemp, G. Papa, R. Datir, I. A. T. M. Ferreira, S. Marelli, W. T. Harvey, S. Lytras, A. Mohamed, G. Gallo, N. Thakur, D. A. Collier, P. Mlcochova, L. M. Duncan, A. M. Carabelli, J. C. Kenyon, A. M. Lever, A. De Marco, C. Saliba, K. Culap, E. Cameroni, N. J. Matheson, L. Piccoli, D. Corti, L. C. James, D. L. Robertson, D. Bailey, and R. K. Gupta, "Recurrent emergence of SARS-CoV-2 spike deletion H69/V70 and its role in the Alpha variant B.1.1.7," *Cell Reports*, vol. 35, no. 13, p. 109292, 2021. doi: 10.1016/j.celrep.2021.109292.

[30] J. Wise, "Covid-19: New coronavirus variant is identified in UK," *British medical journal*, vol. 371, p. m4857, 2020. doi: 10.1136/bmj.m4857.

[31] A. Kumar, N. Goyal, N. Saranathan, S. Dhamija, S. Saraswat, M. B. Menon, and P. Vivekanandan, "The slowing rate of CpG depletion in SARS-CoV-2 genomes is consistent with adaptations to the human host," *Molecular Biology and Evolution*, vol. 39, no. 3, p. msac029, 2022. doi: 10.1093/molbev/msac029.

[32] A. Vandelli, M. Monti, E. Milanetti, A. Armaos, J. Rupert, E. Zacco, E. Bechara, R. Delli Ponti, and G. G. Tartaglia, "Structural analysis of SARS-CoV-2 genome and predictions of the human interactome," *Nucleic Acids Research*, vol. 48, no. 20, pp. 11270-11283, 2020. doi: 10.1093/nar/gkaa864.

[33] Y. Wang, J. M. Mao, G. D. Wang, Z. P. Luo, L. Yang, Q. Yao, and K. P. Chen, "Human SARS-CoV-2 has evolved to reduce CG dinucleotide in its open reading frames," *Scientific Reports*, vol. 10, no. 1, p. 12331, 2020. doi: 10.1038/s41598-020-69342-y.

[34] H. Y. Cai and A. Cai, "SARS-CoV2 spike protein gene variants with N501T and G142D mutation-dominated infections in mink in the United States," *Journal of Veterinary Diagnostic Investigation*, vol. 33, no. 5, pp. 939-942, 2021. doi: 10.1177/10406387211023481.

[35] Wu, and Q. Zhang, "Tracking SARS-CoV-2 Omicron diverse spike gene mutations identifies multiple inter-variant recombination events," *Signal Transduction and Targeted Therapy*, vol. 7, no. 1, p. 138, 2022. doi: 10.1038/s41392-022-00992-2.

[36] R. Muttineni, B. R. N, K. Putty, K. Marapakala, S. K. P, J. Panyam, A. Vemula, S. M. Singh, S. Balachandran, V. R. S. T, and A. K. Kondapi, "SARS-CoV-2 variants and spike mutations

involved in second wave of COVID-19 pandemic in India," *Transboundary and Emerging Diseases*, vol. 69, no. 5, pp. e1721–e1733, 2022. doi: 10.1111/tbed.14508.

[37] M. Gerdol, K. Dishnica, and A. Giorgetti, "Emergence of a recurrent insertion in the N-terminal domain of the SARS-CoV-2 spike glycoprotein," *Virus Research*, vol. 310, p. 198674, 2022. doi: 10.1016/j.virusres.2022.198674.

[38] P. T. Skidmore, E. A. Kaelin, L. A. Holland, R. Maqsood, L. I. Wu, N. J. Mellor, J. M. Blain, V. Harris, J. LaBaer, V. Murugan, and E. S. Lim, "Genomic sequencing of SARS-CoV-2 E484K variant B.1.243.1, Arizona, USA," *Emerging Infectious Diseases*, vol. 27, no. 10, pp. 2718-2720, 2021. doi: 10.3201/eid2710.211189..

[39] X. Wang, M. Hu, B. Liu, H. Xu, Y. Jin, B. Wang, Y. Zhao, J. Wu, J. Yue, and H. Ren, "Evaluating the effect of SARS-CoV-2 spike mutations with a linear doubly robust learner," *Frontiers in Cellular and Infection Microbiology*, vol. 13, p. 1161445, 2023. doi: 10.3389/fcimb.2023.1161445.

[40] D. J. Benton, A. G. Wrobel, P. Xu, C. Roustan, S. R. Martin, P. B. Rosenthal, J. J. Skehel, and S. J. Gamblin, "Receptor binding and priming of the spike protein of SARS-CoV-2 for membrane fusion," *Nature*, vol. 588, no. 7837, pp. 327-330, 2020. doi: 10.1038/s41586-020-2772-0.

[41] J. Lan, J. Ge, J. Yu, S. Shan, H. Zhou, S. Fan, Q. Zhang, X. Shi, Q. Wang, L. Zhang, and X. Wang, "Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor," *Nature*, vol. 581, no. 7807, pp. 215-220, 2020. doi: 10.1038/s41586-020-2180-5.

[42] D. Wrapp, N. Wang, K. S. Corbett, J. A. Goldsmith, C. L. Hsieh, O. Abiona, B. S. Graham, and J. S. McLellan, "Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation," *Science*, vol. 367, no. 6483, pp. 1260-1263, 2020. doi: 10.1126/science.abb2507.

[43] T. N. Starr, A. J. Greaney, S. K. Hilton, D. Ellis, K. H. D. Crawford, A. S. Dingens, M. J. Navarro, J. E. Bowen, M. A. Tortorici, A. C. Walls, N. P. King, D. Veesler, and J. D. Bloom, "Deep mutational scanning of SARS-CoV-2 receptor binding domain reveals constraints on folding and ACE2 binding," *Cell*, vol. 182, no. 5, pp. 1295-1310.e20, 2020. doi: 10.1016/j.cell.2020.08.012.

[45] G. K. Veeramachaneni, V. B. S. C. Thunuguntla, J. Bobbillapati, and J. S. Bondili, "Structural and simulation analysis of hotspot residues interactions of SARS-CoV 2 with human ACE2 receptor," *Journal of Biomolecular Structure and Dynamics*, vol. 39, no. 11, pp. 4015-4025, 2021. doi: 10.1080/07391102.2020.1773318.

[46] D. Pal, "Spike protein fusion loop controls SARS-CoV-2 fusogenicity and infectivity," *Journal of Structural Biology*, vol. 213, no. 2, 2021, doi: 10.1016/j.jsb.2021.107713.

[47] A. G. Wrobel, D. J. Benton, P. Xu, L. J. Calder, A. Borg, C. Roustan, S. R. Martin, P. B. Rosenthal, J. J. Skehel, and S. J. Gamblin, "Structure and binding properties of Pangolin-CoV spike glycoprotein inform the evolution of SARS-CoV-2," *Nature Communications*, vol. 12, no. 1, p. 837, 2021. doi: 10.1038/s41467-021-21006-9.