Cellular and Molecular Biology
E-ISSN : 1165-158X / P-ISSN : 0145-5680
www.cellmolbiol.org

Genomic analysis of SARS-CoV-2 omicron sublineage BA.5.2.1 in Erbil/ Iraq

Sevan Omer Majed[1,2], Paywast Jamal Jalal[3], Mohammed Hassan Fatah[4], Karzan Khawaraham Karim[5], Abdulkarim Yasin Karim[6,7], Monika Miasko [8], Sahar Hassannajad[9], Suhad Asaad Mustafa[10*]

[1] Biology Department, College of Education, Salahaddin University-Erbil, Erbil, 44001, Kurdistan Region, Iraq
[2] Department of Pathological Analysis, College of Science, Knowledge University, Erbil, 44001, Kurdistan Region, Iraq
[3] Biology Department, College of Science, University of Garmian, Sulaymaniyah, 46001, Kurdistan Region, Iraq
[4] Medical Lab, Technology Department, Kalar Technical College, Garmian Polytechnic University, Kalar, 46021, Kurdistan Region, Iraq
[5] The American University of Kurdistan, Duhok, 42001, Kurdistan Region, Iraq
[6] Department of Biology, College of Science, Salahaddin University-Erbil, Erbil, 44001, Kurdistan Region, Iraq
[7] Department of Community Health Nursing, College of Health Technology, Cihan University-Erbil, 44001, Kurdistan Region, Iraq
[8] Medical Analysis Department, Faculty of Applied Science, Tishk International University, Erbil, 44001, Kurdistan Region, Iraq
[9] Department of Medical Laboratory Science, College of Science, Knowledge University, Erbil, 44001, Kurdistan Region, Iraq
[10] General Directorate of Scientific Research Center, Salahaddin University-Erbil, Erbil,44001, Kurdistan Region, Iraq

ARTICLE INFO

Original paper

Article history:
Received: June 25, 2023
Accepted: September 01, 2023
Published: November 15, 2023

Keywords:

BA.1 Sublineage, Omicron, Subvariant BA.5.2.1, Variant of Concern (VOC)

ABSTRACT

Due to several mutations in its genomic sequence, particularly in the spike protein region, the recently-discovered SARS-CoV-2 variant B.5.2.1 has alarmed health policy authorities worldwide. The World Health Organization (WHO) has labelled it "Omicron" and classified it as a worldwide variant of concern (VOC). Following the appearance of Omicron in Iraq, new cases were also detected and analyzed in Kurdistan regions. Two hundred patients were recruited in this study from Erbil/Iraq. The RNA genome samples were extracted, the qRT-PCR performed, and 10 samples were sequenced. The sample sequence was published (EPI ISL 15921492) in the GISAID international gene bank for COVID-19. When compared to the BA.1 Omicron sublineage, 17 new mutations and five deletions in the Omicron subvariant BA.5.2.1 sequence were detected. The spike region includes eight of these variations and one deletion. Overall, 30 substitutions were shared between those previously seen in the BA.1 sublineage and the newly-detected BA.5.2.1 Omicron subvariant. We detected eight new substitutions in our BA.5.2.1 subvariants (T112I, A27S, V213G, T376F, D405N, R408S, L452R, F486V), which were not mentioned previously, should be cause for concern and may be related to immune escape or viral oligomerization. Omicron might be more immune-escape-capable than the current VOCs/VOIs. However, the predicted mutational research shows no conclusive evidence that the Omicron variant may be more virulent or fatal than other variations, including Delta. The greater capacity for immunological evasion may cause the current increase in Omicron cases in Erbil/Iraq.

Doi: http://dx.doi.org/10.14715/cmb/2023.69.11.10

Introduction

On March 11, 2020, the World Health Organization (WHO) classified the prevalence of Coronavirus disease 2019 (Covid-19) as a global pandemic crisis (1). Despite the restricted measures of lockdown and quarantine, SARS-COV-2, like other RNA viruses, we have experienced multiple mutations and emerging variants (2). Omicron was announced as the first variant in South Africa on November 24, 2021(3, 4). It spread worldwide and became a dominant variant (2). Accordingly, multiple subvariants of Omicron have emerged: BA.2, BA.3, BA.4, and BA.5 (5). In addition, WHO has recorded B.1.1.529 as a Variant of Concern (VOC) (WHO, 2021) involving four variants: Alpha (lineage B.1.1.7, Beta (lineage B.1.351), Gamma (lineage P.1), and Delta (lineage B.1.617.2) (4).

The SARS-CoV-2 genome is between 29.8 kb and 29.9 kb and consists of spike protein (S), envelope (E), membrane (M), and nucleocapsid (N) proteins (6). The Omicron variant had several mutations, including 37 nonsynonymous changes in the spike protein, 11 in the amino-terminal domain (N-terminal domain [NTD]), and 15 in the receptor-binding domain (RBD) (7).

The BA.5 subvariant has been characterized by spike mutations, L452R and F486V (the BA.4 sequence is identical to the BA.5 sequence for the spike protein) (7, 8). Omicron BA.4 shares all mutations/deletions with the BA.2 lineage except NSP4: L438F, which is reverted to WT (wild type); S: 69/70 deletion, L452R, F486V, Q493 (WT); ORF 6: D61 (WT); ORF 7b: L11F; N: P151S. Ultimately, Omicron, BA.5 shares the same mutations/deletions as BA.4 except M: D3N; ORF7b: L11 (WT); N: P151 (WT); synonymous SNPs: A27038G, and C27889T (8).

Omicron variants have become a real challenge to the efficacy of COVID-19 vaccines (9). The BA.4 and BA.5 variants have mutations that make reinfection in vaccinated individuals possible (8). The Omicron variant also spread among Kurdistan people (10). The current study aimed to analyze the genomic structure of Omicron sub-

* Corresponding author. Email: suhad.mustafa@su.edu.krd

lineage BA5.2.1 isolated in Erbil, Kurdistan Region, Iraq, to summarize the evolutionary mechanisms through which these variants are emerging and outline the phylogeny of other related variants to show the risk of infection and its impact on the flow of immunity and vaccination.

Materials and Methods

Ethical statement and consent to cases

All methods performed in this research study, including human cases and clinical characteristics, were tracked and permitted by the Local Human Research Ethics Committee (HERC) at Salaheddin University-Erbil (Reference No. 3c/234). During the sample collection period, written informed consent and permission were obtained from the research cases. In addition, all procedures while conducting the research were carried out following the 1964 Helsinki Declaration.

Sample Collection

Oro-nasopharyngeal swabs from 200 cases at the BIO laboratory were taken to demonstrate the performance of the RT-qPCR for detecting SARS-CoV-2 variants. The samples were immediately utilized for RNA extraction. Aside from the swabs taken, all clinical features were gathered via a questionnaire [swabs taken showed that all swabs tested positive for SARS-CoV-2 sub-lineages.

RNA extraction and RT-qPCR for detection of coronavirus strains

From each 200 samples 200microliters was used , to isolate the viral genome, as mentioned in the Nucleic Acid Extraction Kit (Magnetic Bead Method, Reference No.: B200-32). A Zybio EXM3000 Nucleic Acid Isolation System (Catalogue No.: ZBI-EXM3000, Thailand) was utilized to detect the SARS-CoV-2 genome.

The PowerChekTM SARS-CoV-2-S-gene Mutation Detection Kit Version 3.0 (Cat. No. R6922Q) identified SARS-CoV-2 variations in these samples. The total reaction volume for preparation of an RT-qPCR reaction was 20µ, including 5µl of sample RNA, 5µl of each Primer/ Probe mix (FastPlex™ 1 Step SARS-CoV-2 Detection Kit/Cat. No.: 02.01.1020), and 10 µL of a 2X RT-PCR Master Mix. The company's instructions for executing RT-qPCR were followed in applying ZYbio. The first strand of cDNA was synthesized from the template RNA at 50°C for 20 minutes using reverse transcriptase and a specific primer. The RT-qPCR program was set up by starting the initial denaturation step at 95°C for 5 minutes. Then, there were 35 cycles of denaturation at 95 °C for 10 seconds, followed by annealing at 60 °C for 30 seconds and elongation at 72 °C for 30 seconds. The instrument adjusts each Kit's threshold lines by the CoA-specified Ct values. Finally, fluorophore curves on FAM, JOE, ROX, cy5, and Quasar were assessed for data analysis.

Whole-genome sequencing workflow for identification of coronavirus strains

Sample preparation

Whole-genome sequencing was performed for 10 out of the 200 samples with Ct< 30 to identify the coronavirus strains. The clinical characterization of these included Sex (5 males and 5 females); age (50< 10 cases); BMI (Normal weight=3 cases and overweight=7 cases); symptoms (Fever= 10 cases, chills= 7 cases, stomach ache=7 cases, cough=10 cases); vaccination (vaccinated= 2 cases, and unvaccinated=8 cases); hospitalization (non-hospitalized= 10 cases). Total RNA molecules were extracted from each sample (as described in QIAGEN QIAmp Viral Mini Kit, PN 52904).

RNA Library Construction

This study used 200ng of the sample RNA to construct an RNA library. Ten RNA libraries were generated from reduced RNA for ten samples by utilizing the Kit of TruSeq Stranded Total RNA (Illumina, Cat no. 20020599) and the IDT for Illumina TruSeq RNA UD Indexes (96 indexes, 96 wells, Illumina, Cat no. 20022371). Human ribosomal RNA (rRNA) was then eliminated using the Ribo-Zero Gold rRNA depletion protocol (Illumina, 96 samples, Cat no. 20020599). After removing rRNA molecules, first- and second-strand cDNA molecules were synthesized from the remaining RNA molecules using DNA polymerase I and RNase H, as explained in the TruSeq Stranded Total RNA procedure. Following this, adenylation was performed by adding adenine (A) to the blunt double-stranded (ds)-cDNA fragments to prevent them from ligating to each other during the adapter ligation reaction. After that, adapter ligation was carried out to prepare the ds-cDNA for hybridization onto a flow cell. To amplify the prepared ds-cDby using the Illumina NextSeq 500 method with a 150-cycle high-output kit (v2.5) to generate 75-bp paired-end reads, DNA Polymerase I utilized in the assay did not combine past dUTP.

Consequently, the second strand of cDNA was effectively quenched during amplification. The amplifications were then enriched with PCR and purified by KAPA Library Quantification Kit (Kapa Biosystems, Roche Diagnostics Corporation, USA) to generate the final cDNA library. The products were saved for further study (Nextera) in the future. After the Bcl files were converted to Fastq, the CLC genomics workbench version 11.0 was used to measure them (CLC, QIAGEN, Germany). Genetic mutations were confirmed and shown with the BAM files, applying Geneious software.

Data analysis

This study used FAST-QC to perform read quality control, as explained in (11). The Fastp tool version 0.19 was applied to trim adapter sequences and low-quality bases, as mentioned in (12). The CD-HIT-DUP v4.6.8 tool was utilized to filter out low-complexity reads and duplicates with fewer than forty bases, as mentioned in (13). Using Bowtie2 v2.3.4.3 and the SILVA database as a reference compared to the human genome version GRCh38. p13, off-target readings were eliminated (14).

De novo assembly was performed using mapped reads with the SPAdes v3.14.0 tool. Viral genomes were assembled from the input by comparing the trimmed readings to the Wuhan-Hu-1 reference genome sequence (MN908947). The reads collected for each sample were compared to the reference genome using the Bowtie2 v2.3.4.3 program. The majority threshold criterion was used to generate consensus genome sequences. Only aligned reads with a coverage level higher than 80% and an average depth of 8 were used in the experiments.

The genome sequence was sent to GISAID (Global Initiative on Sharing All Influenza Data) on 2022 November

11 to receive an admission number. The complete GISAID database personnel validated the genome; accession number EPI ISL 15921492 was assigned. The whole genome is now accessible to all GISAID researchers.

Phylogenetic analysis

To find the closest relationship between our sample (virus name: hCoV-19/Iraq/SOM/2022; Accession ID: EPI ISL 15921492) and other SARS-CoV-2 variants in Iraq, the whole-genomic sequence of Omicron variants across the globe was collected in GISAID (https://gisaid.org/phylodynamics/global/nextstrain/). The primary local alignment search (BLAST) tool on the GISAID website was searched for closely-related sequences against all sequences in the database (https://github.com/nextstrain/ncov/). Approximately 2276 Omicron genome sequences from the GISAID in different countries were retrieved and arranged to find the closest relationship between our genome and these genomes. Then, Molecular Evolutionary Genetics Analysis Version 11 (MEGA11) software was used to construct a phylogenetic tree.

Spike protein secondary structure

The secondary structure in different views was designed to determine the spike mutations using CCP4 software, version 8.0 (https://www.ccp4.ac.uk/download/#os=windows).

Results

Patients Characteristics

Out of 200 samples (Table 1), approximately 60% were male, and 40% were female. Most patients were older than 50 years old (65%). The weight of the patients was also considered in this study; 32.5% ranged in standard categories (18.5-24.9), 11% were considered underweight, 25% were overweight, and 31.5% were obese (BMI ≥ 30). Cough and runny nose were the most significant symptoms among 97.5% and 94.5%, respectively, followed by fever (72.5%), chills (47.5%), stomach ache (42.5%) and, to a lesser extent, diarrhoea; found in 22.5%. Most of the patients were not hospitalized, and 72.5% were non-vaccinated.

Characteristics of the related genomes

The 2276 Omicron genome sequences retrieved from GISAID were mainly isolated from the United States (>70%), followed by strains from Canada and Japan, at 20% (Figure 1A). The set of genomes was collected over three months (Figure 1B). The isolated genome sequence was compared to the related genomes to define the Omicron variant in Erbil-Iraq. The sequenced results show that the isolated sequence had the highest sequence similarity with the lineage of BA. 5.2.1. Figure 1C showed that the most frequent and related lineage was BA.5.2.1 and, to a lesser extent, the BA.5 lineage. The highest number of the recorded genomes by distance was closely related to our genome (Figure 1D).

Phylogeny of the query sequences and related genomes

The isolated strain from Erbil/ Iraq recorded from the GISAID database (accession number: EPI_ISL_15921492) was found to belong to the lineage of BA.5.2.1 (Figure 2). Based on the GISAID BLAST tool, the sequence similarity of the isolated whole genome was identified, with 61 complete genome sequences from nine countries. A phylogenetic tree was constructed based on the whole related genome of Omicron variants. It showed a very close

Table 1. Characteristics of cases with SARS-CoV-2 Omicron variant (BA.5.2.1).

Characteristic	No. of cases (%)	OR (95% C1)	Significance
Sex			
Male	120 (60)	6.61 (5.10-10.50)	**
Female	80 (40)	7.23 (5.11-9.34)	
Age			
50<	70 (35)	6.53 (5.06-10.12)	***
50≥	130 (65)	7.32 (5.43-9.76)	
BMI Categories:			
Underweight ≤18.5	22 (11)	4.45 (3-6.98)	
Normal weight = 18.5–24.9	65 (32.5)	7.89 (3.91-12)	**
Overweight = 25–29.9	50 (25)	5.45 (4.45-9.64)	
Obesity = BMI of 30 or greater	63 (31.5)	5.01 (4.21-6.54)	
Common Symptoms			
Fever	145 (72.5)	5.34 (4-9.40)	*
Chills	95 (47.5)	6.11 (5.67-8.43)	**ns**
Cough	190 (97.5)	5.21 (4.18-11.32)	**ns**
Stomach ache	85 (42.5)	5.09 (4.44-8.06)	**ns**
Runny nose	189 (94.5)	6.32 (5.09-8.26)	*
Diarrhea	45 (22.5)	7.45 (6.09-10.25)	
Vaccination			
Vaccinated	55 (27.5)	5.84 (3.20-7.55)	**
Non-vaccinated	145 (72.5)	7.43 (4.97-11.57)	
Hospitalized			
Non-hospitalized	175 (87.5)	10.22 (8.43-12.07)	**
Hospitalized	25 (12.5)	5.65 (4.50-8.82)	

*p<0.05; **p<0.01; ***p<0.001; ns=Non-significant; BMI: Body Mass Index

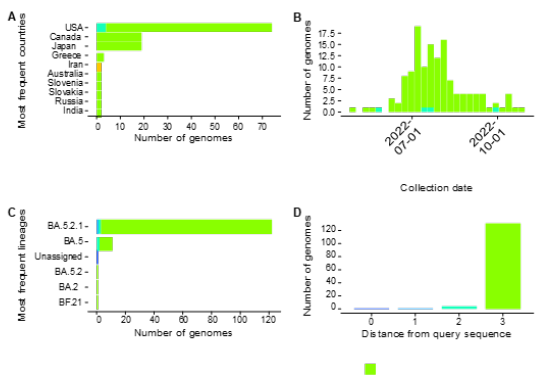

Figure 1. The related genomes are colored by distance to the nearest query sequence. The charts show the number of associated genomes from each of the most frequent countries (A), the number of related genomes by collection date (B), the number of associated genomes assigned to each of the most frequent lineages (C), and the number of related genomes by distance from the closest query sequence (D).

relationshipbetween the isolated strain and the strain from Iran (EPI_ISL_15885942) (Fig.2A). All the genomic sequences from different countries in the last six months of 2022 were isolated and closely related (Fig.2B). Figure 2B shows that most US residents were infected with Omicron sub-lineages. Figure 2C compares many mutated amino acids in spike proteins across different countries. Several charged amino acids were found to be shared between them, including S. D796Y, S. N764K, S. N969K, S. P25del, S. P26del, S. T19I, and S. V213G. The retrieved sequences from the GISAID database were also compared in their spike regions, and the analysis showed that most of the mutations are unique to the Omicron variant. The isolated BA.5.2.1 Omicron variant from this study had a higher close relationship with the Omicron variants found in Iran, the US, and Denmark (Table 2).

Genomic analysis of BA.5.2.1 strain

The sequence analysis of the newly-isolated strain from the Erbil/Kurdistan region of Iraq showed that it belongs to the BA.5.2.1 sub-lineage. The genome consists of 28,955 nucleotides (Figure 3). Many gelation and amino acid substitutions were detected in most of the genome; some are novel. The others were also previously seen in strain from Erbil/ Iraq, belonging to the BA.1—sub-lineage (15). The analysis showed that these novel substitutions in the BA.5.2.1 sub-lineage from the Erbil/ Iraq strain were detected, which are T24I, L264F, T327I, P132H (ORF1a), R392C, A509V, I42V (ORF1b), T112I, A27S, V213G, T376F, D405N, R408S, L452R, F486V (spike), S314R, A398V (nucleocapsid). Moreover, amino acids were also deleted in the ORF1b (4 amino acids) and spike (1 amino acid) regions which are T19I, L24, P25, P26 and G142D, respectively.

On the other hand, other amino acids substituted in the BA.5.2.1 sub-lineage were also detected in the previously-isolated BA.1. sub-lineage. In ORF1a, both amino acids K384R and T492I were substituted. In ORF1b, P323L is substituted. Approximately 20 amino acids are substituted in the.2.1; the same is found in AB.1. sub-lineage and they are G339D, S371F, S373P, S375F, K417N, N440K, S477N, T478K, E484A, Q498R, N501Y, Y505H, D614G, H655Y, N679K, P681H, N764K, D796Y, Q954H, N969K. Amino acids in the envelope (T9I), membrane (D3N, Q19E, A63T) and nucleocapsid (P13L, R203K, G204R) were detected as well.

Furthermore, deleted amino acids in AB.5.2.1. were observed as well, which were previously detected in the isolated strain AB.1. from Erbil/Iraq; these are L105, S106, G107 (ORF1a); H69, V70, N144 (ORF1b); E31, R32, S33 (nucleocapsid).

Spike glycoprotein structure and mutation determination

The alignment specifically for the spike region showed that in the total of 1276 amino acids, there are 30 amino acid substitutions (Figure 4) which are: T20I, T112I, A27S, V213G, G339D, S371F, S373P, S375F, T376F, D405N, R408S, K417N, N440K, L452R, S477N, T478K, E484A,

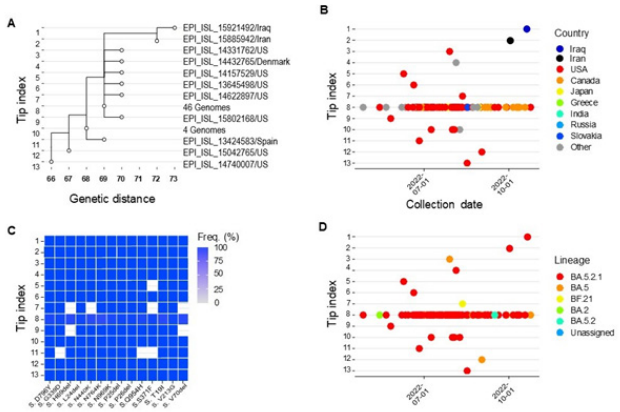

Figure 2. Maximum-likelihood tree of the query sequences, related genomes, and an outgroup (the WIV04 reference sequence from Wuhan). The figure shows (A) the tree, with branch lengths given by the genetic distance (the estimated number of mutations) [A yellow background indicates the query sequences]. The Accession ID, or the number of genomes in the group label tips. (B) The collection date of the sequences within each tip group is coloured by location. (C) The frequency of occurrence of spike mutations within each tip group. (D) The collection date of the sequences within each tip group, coloured by lineage.

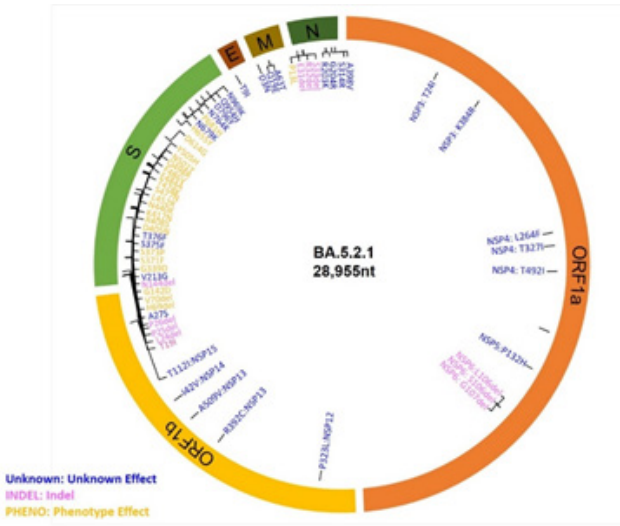

Figure 3. Genome sequence analysis of BA.5.2.1. The genome was about 28.955 nucleotides. The mutation loci on the genome were determined by colour; blue colour indicated the unknown effect of mutated amino acids, pink indicated indels, and green showed phenotype effects.

Table 2. Sixty-one related genomes to this genomic sequence (EPI_ISL_15921492).

Accession ID	Tip	Collection date	Submit date	Lineage	Clade	Distance	Quality	Country / State
EPI_ISL_15885942	1	2022-06	2022-11-23	BA.5.2.1	GRA	1	0.954	Iran / Hormozgan
EPI_ISL_15410331	8	2022-09-29	2022-10-18	BA.5	GRA	2	0.961	USA / North Carolina
EPI_ISL_15825602	8	2022-07-21	2022-11-19	BA.5.2.1	GRA	2	0.935	USA / Washington
EPI_ISL_14105014	8	2022-07-11	2022-07-28	BA.5	GRA	2	0.922	USA / New Hampshire
EPI_ISL_15802168	9	2022-05-26	2022-11-17	BA.5.2.1	GRA	2	0.907	USA / Kentucky
EPI_ISL_14331762	3	2022-07-29	2022-08-09	BA.5	GRA	3	0.968	USA / South Carolina
EPI_ISL_14432765	4	2022-08-05	2022-08-12	BA.5.2.1	GRA	3	0.966	Denmark / Sjaelland
EPI_ISL_14157529	5	2022-06-09	2022-07-30	BA.5.2.1	GRA	3	0.949	USA / Arizona / Maricopa County
EPI_ISL_13645498	6	2022-06-20	2022-07-05	BA.5.2.1	GRA	3	0.939	USA / Illinois
EPI_ISL_14622897	7	2022-08-12	2022-08-25	BF.21	GRA	3	0.917	USA / North Carolina / Pitt County
EPI_ISL_14010757	8	2022-07-09	2022-07-23	BA.5.2.1	GRA	3	0.978	USA / Massachusetts
EPI_ISL_14653864	8	2022-08-09	2022-08-26	BA.5.2.1	GRA	3	0.973	USA / Connecticut
EPI_ISL_13717494	8	2022-06-19	2022-07-08	BA.5.2.1	GRA	3	0.971	USA / Washington
EPI_ISL_14913088	8	2022-08-28	2022-09-11	BA.5.2.1	GRA	3	0.971	Australia / Sydney
EPI_ISL_14913084	8	2022-08-24	2022-09-11	BA.5.2.1	GRA	3	0.971	Australia / Sydney
EPI_ISL_13915909	8	2022-07-05	2022-07-19	BA.5.2.1	GRA	3	0.970	USA / Washington
EPI_ISL_15824267	8	2022-10-19	2022-11-18	BA.5.2.1	GRA	3	0.969	Canada / Alberta
EPI_ISL_13644645	8	2022-06-14	2022-07-05	BA.5.2.1	GRA	3	0.968	USA / California
EPI_ISL_15178941	8	2022-09-16	2022-09-29	BA.5.2	GRA	3	0.968	USA / Indiana
EPI_ISL_15279104	8	2022-09-09	2022-10-06	BA.5.2.1	GRA	3	0.967	Canada / Alberta
EPI_ISL_12973031	8	2022-05-21	2022-05-28	BA.5.2.1	GRA	3	0.967	Belgium / Brussels Capital Region
EPI_ISL_15667100	8	2022-10-10	2022-11-04	BA.5.2.1	GRA	3	0.967	Canada / Alberta
EPI_ISL_15278918	8	2022-09-05	2022-10-06	BA.5.2.1	GRA	3	0.967	Canada / Alberta
EPI_ISL_15666799	8	2022-10-13	2022-11-04	BA.5.2.1	GRA	3	0.967	Canada / Alberta
EPI_ISL_15666783	8	2022-10-12	2022-11-04	BA.5.2.1	GRA	3	0.967	Canada / Alberta
EPI_ISL_15599660	8	2022-10-06	2022-10-31	BA.5.2.1	GRA	3	0.967	Canada / Alberta
EPI_ISL_15272936	8	2022-08-07	2022-10-06	BA.5.2.1	GRA	3	0.967	USA / New York / Niagara County
EPI_ISL_14964710	8	2022-08-04	2022-09-14	BA.5.2.1	GRA	3	0.967	USA / Alaska / Anchorage
EPI_ISL_14513141	8	2022-08-08	2022-08-18	BA.5.2.1	GRA	3	0.967	USA / Massachusetts
EPI_ISL_15200098	8	2022-09-12	2022-09-29	BA.5.2.1	GRA	3	0.967	Canada / Manitoba
EPI_ISL_15376486	8	2022-09-12	2022-10-14	BA.5.2.1	GRA	3	0.967	Canada / Alberta
EPI_ISL_15376411	8	2022-09-06	2022-10-14	BA.5.2.1	GRA	3	0.967	Canada / Alberta
EPI_ISL_14501973	8	2022-08-03	2022-08-17	BA.5.2.1	GRA	3	0.967	USA / Connecticut
EPI_ISL_14294139	8	2022-07-24	2022-08-08	BA.5.2.1	GRA	3	0.967	USA / District of Columbia
EPI_ISL_14263558	8	2022-07-18	2022-08-05	BA.5.2.1	GRA	3	0.967	USA / Washington
EPI_ISL_15666949	8	2022-10-10	2022-11-04	BA.5.2.1	GRA	3	0.967	Canada / Alberta
EPI_ISL_14054153	8	2022-07-10	2022-07-26	BA.5.2.1	GRA	3	0.967	USA / Illinois
EPI_ISL_13807435	8	2022-06-26	2022-07-13	BA.5.2.1	GRA	3	0.966	USA / Arizona
EPI_ISL_14651763	8	2022-08-06	2022-08-26	BA.5.2.1	GRA	3	0.966	USA / Arizona
EPI_ISL_14502434	8	2022-08-04	2022-08-17	BA.5.2.1	GRA	3	0.966	USA / Connecticut
EPI_ISL_14556062	8	2022-07-31	2022-08-19	BA.5.2.1	GRA	3	0.966	USA / Maryland
EPI_ISL_13424583	8	2022-06-07	2022-06-23	BA.5.2.1	GRA	3	0.966	Spain / Catalunya
EPI_ISL_13931508	8	2022-06-29	2022-07-19	BA.5.2.1	GRA	3	0.966	Canada / Ontario
EPI_ISL_14958676	8	2022-07-21	2022-09-14	BA.5.2.1	GRA	3	0.966	USA / Alaska / Anchorage
EPI_ISL_14191964	8	2022-07-09	2022-08-02	BA.5.2.1	GRA	3	0.966	USA / Ohio
EPI_ISL_14191830	8	2022-07-09	2022-08-02	BA.5.2.1	GRA	3	0.966	USA / Ohio
EPI_ISL_14271578	8	2022-07-11	2022-08-05	BA.5.2.1	GRA	3	0.966	USA / Maryland
EPI_ISL_14271608	8	2022-07-11	2022-08-05	BA.5.2.1	GRA	3	0.966	USA / Minnesota
EPI_ISL_14280262	8	2022-07-12	2022-08-05	BA.5.2.1	GRA	3	0.966	USA / California
EPI_ISL_14641646	8	2022-07-27	2022-08-26	BA.5.2.1	GRA	3	0.966	Japan / Fukushima
EPI_ISL_14726224	8	2022-07-19	2022-08-30	BA.5.2.1	GRA	3	0.966	Japan / Tokyo
EPI_ISL_14732215	8	2022-07-19	2022-08-30	BA.5.2.1	GRA	3	0.966	Japan / Tokyo
EPI_ISL_14856744	8	2022-07-08	2022-09-08	BA.5.2.1	GRA	3	0.966	Japan / Tokyo
EPI_ISL_14856747	8	2022-07-08	2022-09-08	BA.5.2.1	GRA	3	0.966	Japan / Tokyo
EPI_ISL_14861132	8	2022-07-19	2022-09-08	BA.5.2.1	GRA	3	0.966	Japan / Tokyo
EPI_ISL_14861140	8	2022-07-19	2022-09-08	BA.5.2.1	GRA	3	0.966	Japan / Tokyo
EPI_ISL_14989962	8	2022-07-20	2022-09-16	BA.5.2.1	GRA	3	0.966	Japan / Tokyo
EPI_ISL_14990279	8	2022-07-20	2022-09-16	BA.5.2.1	GRA	3	0.966	Japan / Tokyo
EPI_ISL_13760462	8	2022-06-20	2022-07-11	BA.5.2.1	GRA	3	0.966	Netherlands / Overijssel
EPI_ISL_15069070	8	2022-08-30	2022-09-22	BA.5.2.1	GRA	3	0.966	Japan / Kanagawa

F486V, Q498R, N501Y, Y505H, D614G, H655Y, N674K, N679K, P681H, N764K, D796Y, Q954H, N969K. The eight deletions (P24, P25, P26, H69, V69, V70, G142D, N144 (Y144)) were detected in the C-terminal region of the spike region. Most of the polymorphisms were located in the S1 region and S1-S2 junction (Figure 5). These were considered to impact the binding process to the ACE2 receptor, the binding of amino acids with the ACE2 receptor and the amino acid substitutions. The analysis also showed that amino acids from 339-417 can form a complex and bind to ACE2. Eight amino acids were shown to be substituted: G339D, S371F, S373P, S375F, N408S, D405N, K417N, N440K, T478K, Q498R. Six of these were also detected in the previously-isolated strain of the AB.1 sublineage. However, two of these substitutions, N408S and D405N, are only found in the AB.5.2.1 sublineage.

Phylogenetic tree analysis of the Omicron strains

To analyze the samples from this study with the published omicron variants, 2275 sequences were retrieved from the GISAID database. The samples were isolated from their appearance from November 2021 until December 2022. Interestingly, all the samples created eight clades; including 21K, 21L, 22A, 22B, 22C, 22D, 22E, and 22F (Figure 5a). The BA.5.2.1 Omicron variants from Erbil/Iraq belong to Clade 22B, which includes the most transmitted variants (Figure 5b).

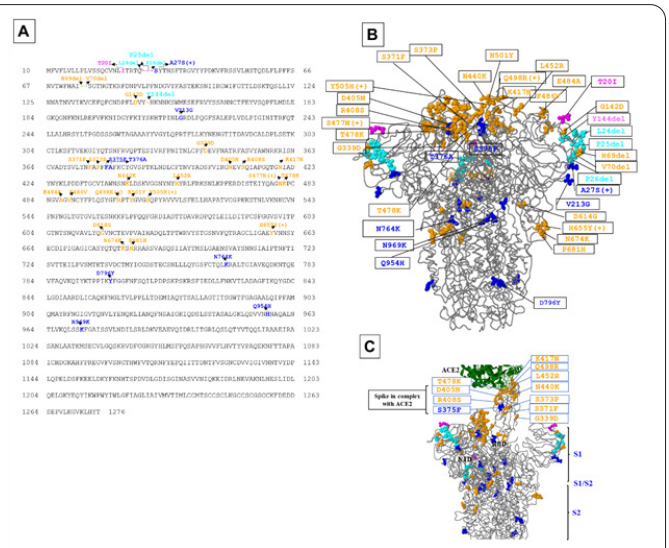

Figure 4. Spike glycoprotein sequence with mutation determination. A. Spike Sequence with 35 mutations, Length = 1278 aa. B. Structure profile of the spike and mutation determination. (C) Spike-ACE2 interaction and Mutation determination.

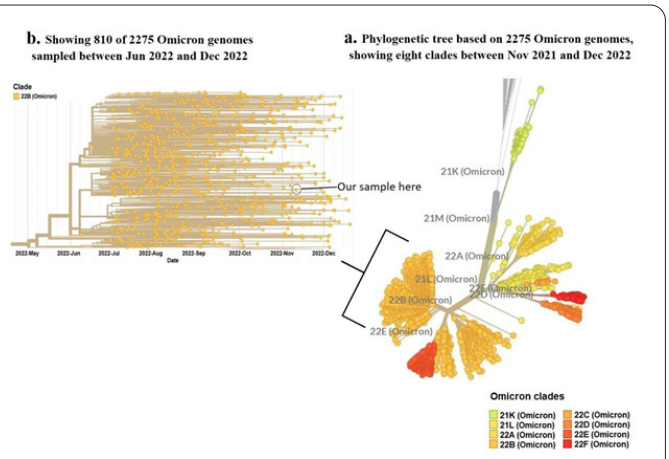

Figure 5. Analysis of Phylogenetic tree. **a.** Unrooted tree was built up based on 2275 Omicron samples. They show eight clades of Omicron detected between Nov 2021 and Dec 2022. **b.** The rectangular tree, including 810 of 2275 Omicron samples, was zoomed in to show our sample in the 22B clade of Omicron.

Discussion

In July 2022, the discovery of COVID-19 subvariant Omicron BA.5.2.1 was officially announced in Shanghai. Last October and September, a new wave of severe infection with COVID-19-like symptoms dominated Iraq's Kurdistan region. The Omicron subvariant known as BA.5 is the most contagious subvariant to date and has evolved into the predominant variant circulating in the United States and much of the world. In many nations, BA.5 is gradually replacing the early BA.1 and BA.2 Omicron subvariants of SARS-CoV-2 (16). Many subvariants of the Omicron strain have been discovered worldwide since its appearance. A considerable number of variations were detected in the analysis of subvariants, which generally impacted the ACE2 binding site, immune escape and viral oligomerization (17-19).

More than any other SARS-CoV-2 variant before, Omicron has roughly 50 mutations, 32 of which were in the spike protein, which most vaccines work to neutralize (20). The computational tools analyzed the number of mutations among different Omicron subvariants and showed 39 variations in BA.1, 40 in BA.1.1 and 31 in BA.3 21. When compared to the BA.1 Omicron sublineage, 17 new mutations and five deletions in the entire sequence of Omicron subvariant BA.5.2.1 were detected in this study. The spike region included eight of these variations and one deletion. Overall, 30 substitutions were shared between the previously-detected BA.1 sublineage and the newly-detected BA.5.2.1 Omicron subvariant.

Compared to wild-type (WT), Omicron and sub-variants demonstrated a greater affinity for human ACE2 and the potential for more transmission (21). The Omicron variation binds receptors less effectively than the WT, according to the N-terminal domain (NTD) of the spike protein's negative electrostatic potential (NEP) value (22). The electrostatic potential of NTD negative value might be one of the reasons why the Omicron form is considered less hazardous to the lower respiratory tract, given that at least one receptor is strongly expressed in lung and bronchial cells (22). The transmission potential of the Omicron sub-lineages BA.2 and BA.3 is more potent than that of BA.1 and BA.1.1. Previously, it was revealed that new salt bridges and hydrogen bonds formed by altered residues in BA.1.1 (K478), BA.2 (R400, R490, and R495), and BA.3 (R397 and H499) (23). The subvariant BA.5.2.1 also has T478K variation in the spike NTD, which is thought to increase the binding affinity to hACE2.

The amino acid variations at the Receptor-binding Motif (RBM) residues such as T478K, Q498R, N501Y and Y505H all considerably impact effective binding affinity for human ACE2, which may decline due to the presence of variation at 417 residues (K417N) (24). A previous analysis showed that G339D substitution increased RBD expression and affected antibody binding (25). The three substitutions at S371F, S373P, and S375F have a role in escaping from immunity through recognizing antibodies, impacting viral oligomerisation, and increasing human ACE2 binding (26).

Both amino acids H69 and V70 were deleted in the BA.5.2.1 Omicron subvariant spike region, which is believed to impact the infectivity compared to the wild type, specifically in causing chronic inflammation and viral excretion because of the effect on host interactions and specifically on immune escape (27). Moreover, other variations at the E484K, N501Y, and P681H were related to immune escape and inflammation (27, 28). Both variations, P323L, in the RNA-dependent-RNA Polymerase and D614G in the spike region, were revealed to be related to viral replication and the severity of the viruses (29).

Even though many studies suggest that an Omicron infection is less severe than a Delta infection, the rapid increase in cases will lead to more hospitalizations, imposing stress on healthcare systems as they try to treat patients with COVID-19 and other diseases simultaneously. Despite the high estimated transmissibility of Omicron subvariants, we believe that future COVID-19 waves may be prevented by updated vaccines, reduced immunization inequalities, enhanced antiviral therapy, and preventive measures adopted by susceptible populations.

In conclusion, the emergence of the Omicron subvariant BA.5.2.1 as the most contagious COVID-19 variant is a cause for concern. The subvariant is replacing earlier subvariants in many countries, including the Kurdistan region of Iraq, where it is the most frequent and related

lineage. The fact that vaccinated individuals may still be infected with this subvariant highlights the possibility of antigenic drift.

Particular amino acid variations, such as those in the RBM, can affect the binding affinity for human ACE2, potentially leading to immune escape and inflammation. Although Omicron infections may be less severe than Delta infections, the rapid increase in cases could still strain healthcare systems.

Continuous surveillance and assessment are crucial to effectively respond to the Omicron subvariant and any future emerging variants. This could include monitoring the subvariant's spread and impact, evaluating current vaccines' effectiveness, and developing new vaccines if necessary. Overall, it is essential to remain vigilant and take appropriate measures to contain the spread of the virus and mitigate its impact on public health.

Conclusions

In conclusion, the emergence of the Omicron subvariant BA.5.2.1 as the most contagious COVID-19 variant is a cause for concern. The subvariant is replacing earlier subvariants in many countries, including the Kurdistan region of Iraq, where it is the most frequent and related lineage. The fact that vaccinated individuals may still be infected with this subvariant highlights the possibility of antigenic drift. Specific amino acid variations, such as those found in the RBM, can affect the binding affinity for human ACE2, potentially leading to immune escape and inflammation. Although Omicron infections may be less severe than Delta infections, the rapid increase in cases could still strain healthcare systems. Continuous surveillance and assessment are crucial to effectively respond to the Omicron subvariant and any future emerging variants. This could include monitoring the subvariant's spread and impact, evaluating current vaccines' effectiveness, and developing new vaccines if necessary. Overall, it is essential to remain vigilant and take appropriate measures to contain the spread of the virus and mitigate its impact on public health.

Author Contributions

Sevan Omer Majed: data acquisition and analysis. Paywast Jamal Jalal: wrote the main manuscript. Mohammed Hassan Fatah, Monika Henryka Misako, and Sahar Hassannejad: manuscript drafting and revision. Karzan Khawaraham Karim and Abdulkarim Yasin Karim participated in data acquisition, analysis, and manuscript drafting. Suhad Asad Mustafa : Editing, and final draft. All authors read and approved the final manuscript.

Funding

This research received no external funding.

Institutional Review Board Statement: All methods performed in this research study, including human cases and clinical characteristics, were tracked and permitted by the Local Human Research Ethics Committee (HERC) at Salaheddin University-Erbil (Reference No. 3c/234).

Informed Consent Statement:
During the sample collection period, written informed consent and permission were obtained from the research cases. In addition, all procedures while conducting the research were carried out following the 1964 Helsinki Declaration.

Data Availability Statement:
The authors confirm that the data supporting the findings of this study are available within the article.

Acknowledgments

We thank the staff and management team of Rojawa Emergency Hospital.

Conflicts of Interest

The authors declare no conflict of interest.

References

1. Bartleson JM, Radenkovic D, Covarrubias AJ, Furman D, Winer DA, and Verdin E, SARS-CoV-2, COVID-19 and the Ageing Immune System. Nat Aging 2021; 1(9): p. 769-782. doi: 10.1038/s43587-021-00114-7.
2. Vitiello A, Ferrara F, Auti AM, Di Domenico M, and Boccellino M, Advances in the Omicron variant development. J Intern Med 2022; 292(1): p. 81-90. doi: 10.1111/joim.13478.
3. Gowrisankar A, Priyanka TMC, and Banerjee S, Omicron: a mysterious variant of concern. Eur Phys J Plus 2022; 137(1): p. 100. doi: 10.1140/epjp/s13360-021-02321-y.
4. Quarleri J, Galvan V, and Delpino MV, Omicron variant of the SARS-CoV-2: a quest to define the consequences of its high mutational load. Geroscience 2022; 44(1): p. 53-56. doi: 10.1007/s11357-021-00500-4.
5. Yao L, Zhu KL, Jiang XL, Wang XJ, Zhan BD, Gao HX, Geng XY, Duan LJ, Dai EH, and Ma MJ, Omicron subvariants escape antibodies elicited by vaccination and BA.2.2 infection. Lancet Infect Dis 2022; 22(8): p. 1116-1117. doi: 10.1016/S1473-3099(22)00410-8.
6. Khailany RA, Safdar M, and Ozaslan M, Genomic characterization of a novel SARS-CoV-2. Gene Rep 2020; 19: p. 100682. doi: 10.1016/j.genrep.2020.100682.
7. Ren SY, Wang WB, Gao RD, and Zhou AM, Omicron variant (B.1.1.529) of SARS-CoV-2: Mutation, infectivity, transmission, and vaccine resistance. World J Clin Cases 2022; 10(1): p. 1-11. doi: 10.12998/wjcc.v10.i1.1.
8. Cao Y, Yisimayi A, Jian F, Song W, Xiao T, Wang L, Du S, Wang J, Li Q, Chen X, Yu Y, Wang P, Zhang Z, Liu P, An R, Hao X, Wang Y, Wang J, Feng R, Sun H, Zhao L, Zhang W, Zhao D, Zheng J, Yu L, Li C, Zhang N, Wang R, Niu X, Yang S, Song X, Chai Y, Hu Y, Shi Y, Zheng L, Li Z, Gu Q, Shao F, Huang W, Jin R, Shen Z, Wang Y, Wang X, Xiao J, and Xie XS, BA.2.12.1, BA.4 and BA.5 escape antibodies elicited by Omicron infection. Nature 2022; 608(7923): p. 593-602. doi: 10.1038/s41586-022-04980-y.
9. Cele S, Jackson L, Khoury DS, Khan K, Moyo-Gwete T, Tegally H, San JE, Cromer D, Scheepers C, Amoako DG, Karim F, Bernstein M, Lustig G, Archary D, Smith M, Ganga Y, Jule Z, Reedoy K, Hwa S-H, Giandhari J, Blackburn JM, Gosnell BI, Abdool Karim SS, Hanekom W, Davies M-A, Hsiao M, Martin D, Mlisana K, Wibmer CK, Williamson C, York D, Harrichandparsad R, Herbst K, Jeena P, Khoza T, Kløverpris H, Leslie A, Madansein R, Magula N, Manickchund N, Marakalala M, Mazibuko M, Moshabela M, Mthabela N, Naidoo K, Ndhlovu Z, Ndung'u T, Ngcobo N, Nyamande K, Patel V, Smit T, Steyn A, Wong E, von Gottberg A, Bhiman JN, Lessells RJ, Moosa M-YS, Davenport MP, de Oliveira T, Moore PL, Sigal A, Ngs SA, and Team C-K, Omicron extensively but incompletely escapes Pfizer BNT162b2 neutralization. Nature 2022; 602(7898): p. 654-656. doi: 10.1038/

s41586-021-04387-1.

10. Merza MA and Almufty HB, The Omicron Variant is Striking Iraqi Kurdistan in January 2022: Would Preventive Measures Contain the new Wave? Disaster Med Public Health Prep 2022: p. 1-2. doi: 10.1017/dmp.2022.156.
11. Andrews S, FastQC: A Quality Control Tool for High Throughput Sequence Data. 2010.
12. Chen S, Zhou Y, Chen Y, and Gu J, fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics 2018; 34(17): p. i884-i890. doi: 10.1093/bioinformatics/bty560.
13. Fu L, Niu B, Zhu Z, Wu S, and Li W, CD-HIT: accelerated for clustering the next-generation sequencing data. Bioinformatics 2012; 28(23): p. 3150-2. doi: 10.1093/bioinformatics/bts565.
14. Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, Peplies J, and Glockner FO, The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Res 2013; 41(Database issue): p. D590-6. doi: 10.1093/nar/gks1219.
15. Majed SO, Mustafa SA, Jalal PJ, Fatah MH, Miasko M, Jawhar Z, and Karim AY, SARS-CoV-2 Omicron Variant Genomic and Phylogenetic Analysis in Iraqi Kurdistan Region. Genes 2022; 14(1): p. 413-26. doi: 10.3390/genes1030413.
16. Ong CP, Ye ZW, Tang K, Liang R, Xie Y, Zhang H, Qin Z, Sun H, Wang TY, Cheng Y, Chu H, Chan JF, Jin DY, and Yuan S, Comparative analysis of SARS-CoV-2 Omicron BA.2.12.1 and BA.5.2 variants. J Med Virol 2023; 95(1): p. e28326. doi: 10.1002/jmv.28326.
17. Duan LJ, Jiang WG, Wang ZY, Yao L, Zhu KL, Meng QC, Wang BS, Li LB, Wang GL, and Ma MJ, Neutralizing immunity against SARS-CoV-2 Omicron BA.1 by infection and vaccination. iScience 2022; 25(9): p. 104886. doi: 10.1016/j.isci.2022.104886.
18. Evans JP, Zeng C, Qu P, Faraone J, Zheng YM, Carlin C, Bednash JS, Zhou T, Lozanski G, Mallampalli R, Saif LJ, Oltz EM, Mohler PJ, Xu K, Gumina RJ, and Liu SL, Neutralization of SARS-CoV-2 Omicron sub-lineages BA.1, BA.1.1, and BA.2. Cell Host Microbe 2022. doi: 10.1016/j.chom.2022.04.014.
19. Liang H, Nian X, Wu J, Liu D, Feng L, Lu J, Peng Y, Zhou Z, Deng T, Liu J, Ji D, Qiu R, Lin L, Zeng Y, Xia F, Hu Y, Li T, Duan K, Li X, Wang Z, Zhang Y, Zhang H, Zhu C, Wang S, Wu X, Wang X, Li Y, Huang S, Mao M, Guo H, Yang Y, Jia R, Xufang J, Wang X, Liang S, Qiu Z, Zhang J, Ding Y, Li C, Zhang J, Fu D, He Y, Zhou D, Li C, Zhang J, Yu D, and Yang XM, COVID-19 vaccination boosts the potency and breadth of the immune response against SARS-CoV-2 among recovered patients in Wuhan. Cell Discov 2022; 8(1): p. 131. doi: 10.1038/s41421-022-00496-x.
20. Khandia R, Singhal S, Alqahtani T, Kamal MA, El-Shall NA, Nainu F, Desingu PA, and Dhama K, Emergence of SARS-CoV-2 Omicron (B.1.1.529) variant, salient features, high global health concerns and strategies to counter it amid ongoing COVID-19 pandemic. Environ Res 2022; 209: p. 112816. doi: 10.1016/j.envres.2022.112816.
21. Li M, Ren Y, Aw ZQ, Chen B, Yang Z, Lei Y, Cheng L, Liang Q, Hong J, Yang Y, Chen J, Wong YH, Wei J, Shan S, Zhang S, Ge J, Wang R, Dong JZ, Chen Y, Shi X, Zhang Q, Zhang Z, Chu JJH, Wang X, and Zhang L, Broadly neutralizing and protective nanobodies against SARS-CoV-2 Omicron subvariants BA.1, BA.2, and BA.4/5 and diverse sarbecoviruses. Nat Commun 2022; 13(1): p. 7957. doi: 10.1038/s41467-022-35642-2.
22. Wang X, Chen X, Tan J, Yue S, Zhou R, Xu Y, Lin Y, Yang Y, Zhou Y, Deng K, Chen Z, Ye L, and Zhu Y, 35B5 antibody potently neutralizes SARS-CoV-2 Omicron by disrupting the N-glycan switch via a conserved spike epitope. Cell Host Microbe 2022. doi: 10.1016/j.chom.2022.03.035.
23. Kumar S, Karuppanan K, and Subramaniam G, Omicron (BA.1) and sub-variants (BA.1.1, BA.2, and BA.3) of SARS-CoV-2 spike infectivity and pathogenicity: A comparative sequence and structural-based computational assessment. J Med Virol 2022; 94(10): p. 4780-4791. doi: 10.1002/jmv.27927.
24. Miller NL, Clark T, Raman R, and Sasisekharan R, Insights on the mutational landscape of the SARS-CoV-2 Omicron variant receptor-binding domain. Cell Rep Med 2022; 3(2): p. 100527. doi: 10.1016/j.xcrm.2022.100527.
25. Starr TN, Greaney AJ, Hilton SK, Ellis D, Crawford KHD, Dingens AS, Navarro MJ, Bowen JE, Tortorici MA, Walls AC, King NP, Veesler D, and Bloom JD, Deep Mutational Scanning of SARS-CoV-2 Receptor Binding Domain Reveals Constraints on Folding and ACE2 Binding. Cell 2020; 182(5): p. 1295-1310.e20. doi: 10.1016/j.cell.2020.08.012.
26. Starr TN, Greaney AJ, Addetia A, Hannon WW, Choudhary MC, Dingens AS, Li JZ, and Bloom JD, Prospective mapping of viral mutations that escape antibodies used to treat COVID-19. Science 2021; 371(6531): p. 850-854. doi: 10.1126/science.abf9302.
27. Takano K, Watanabe Y, Hariu M, and Seki M, Detection of Representative Mutant Strains and a Case of Prolonged Infection by SARS-CoV-2 with Spike 69/70 Deletion in Japan. Infect Drug Resist 2021; 14: p. 2579-2581. doi: 10.2147/idr.s320658.
28. Brejová B, Boršová K, Hodorová V, Čabanová V, Reizigová L, Paul ED, Čekan P, Klempa B, Nosek J, and Vinař T, A SARS-CoV-2 mutant from B.1.258 lineage with ΔH69/ΔV70 deletion in the Spike protein circulating in Central Europe in the fall 2020. Virus Genes 2021: p. 1-5. doi: 10.1007/s11262-021-01866-5.
29. Biswas SK and Mudi SR, Spike protein D614G and RdRp P323L: the SARS-CoV-2 mutations associated with severity of COVID-19. Genomics Inform 2020; 18(4): p. e44. doi: 10.5808/GI.2020.18.4.e44.