Article

Not peer-reviewed version

# Phylogenetic, Sequencing, and Mutations Analysis of the SARS-CoV-2 Omicron (BA.1) and Sub-variants (BA.1.1, BA.2) during the Fifth Wave of the Pandemic in the Iraqi Kurdistan Region

Sherzad Majeed Taher *, Jassim Mohammed Abdo, Muayad Aghali Merza

Posted Date: 16 August 2023

doi: 10.20944/preprints202308.1167.v1

Keywords: COVID-19; SARS-COV-2; Omicron Variants; Phylogenetic tree; Genome Sequence

*Article*

# Phylogenetic, Sequencing, and Mutations analysis of the SARS-CoV-2 Omicron (BA.1) and Sub-Variants (BA.1.1, BA.2) during the Fifth Wave of the Pandemic in the Iraqi Kurdistan Region

**Sherzad Majeed Taher [1,*], Jassim M Abdo [1] and Muayad A. Merza [2]**

[1] Department of Basic Science, College of Pharmacy University of Duhok, Duhok 42001, Kurdistan, Iraq
[2] Department of Internal Medicine, Azadi Teaching Hospital, College of Medicine, University of Duhok, Duhok 42001, Kurdistan, Iraq
* Correspondence: sherzad81@uod.ac; Tel.: +9647504709533

**Abstract:** Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) was first identified in humans in Wuhan City at the end of December 2019. Since then, it has spread to all the countries. Therefore, global interest has been focused on discovering treatments and developing successful vaccines. This study sequenced the complete genome of the SARS-COV-2 Omicron Omicron (BA.1) and sub-variants (BA.1.1, BA.2), which were isolated from 40 individuals in Duhok, Iraq. Ninety-five different mutations were identified when the complete genome of the SARS-COV-2 virus discovered in Wuhan, China (accession number: NC 045512.2) was matched to the virus sequence using sequencing technology (Illumina, USA). Sequence analysis revealed 38 mutations in spike glycoprotein (S), 30 of which were found in ORF1a. Additionally, 11 mutations were found in ORF1b, and 7,3,2,1 mutations were found in Nucleocapsid (N), membrane protein (M), Open Reading Frames 6 (ORF6), Open Reading Frames 9 (ORF9), and Envelope (E) genes, respectively. Phylogenetic analysis and transmission further confirmed that the isolates found in Iraq had distinct infection origins and were closely related to those from other countries and states. According to the findings of this study, a new vaccine can be developed based on identifying new Omicron variant mutations and sub-variants such as BA.2, which were identified for the first time in Iraq.

**Keywords:** COVID-19; SARS-COV-2; Omicron Variants; Phylogenetic tree; Genome Sequence

## 1. Introduction

Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) caused an unusual global outbreak in Wuhan, China, in December 2019, which was quickly identified as a pandemic and known as coronavirus disease of 2019 ( COVID-19).This outbreak suddenly gained international awareness and prompted a global campaign to develop treatment and vaccines. More than 5 million deaths and 260 million confirmed cases of COVID-19 were reported by the World Health Organization (WHO) by November 2021. SARS-CoV-2 is a single-stranded positive-sense Ribonucleic acid (RNA ) virus. SARS-CoV-2 has one of the longest genomes among all RNA virus families, measuring 30 kb and encoding nearly 9860 amino acids (1,2).

In consideration of how urgent the epidemic was, scientists have put various efforts into producing effective vaccines. However, similar to other infections, viruses are more likely to spread in open-air environments. From a health standpoint, the increase in visitors to crowded areas seemed absurd, but it was less shocking than anticipated. It is strenuous to accurately estimate the risk of self-behavior, and the length of the pandemic has challenged the limits of human patience and self-control (3). Full viral genome sequencing is essential for diagnosing infectious viral illnesses and designing vaccines. Viruses' virulence, pathogenicity, and evolutionary connections between their hosts can also be studied using genome sequencing (4). Rapid country-to-country transmission of SARS-CoV-

2 has raised questions regarding whether virus mutations are the reason for its outspread (5,6,7). Random mutations occur because SARS-CoV-2, an RNA virus, lacks a "proofreading" mechanism as it multiplies. These are the main strains of the virus that spread in numerous places and have multiplied slightly (8). Additionally, the virus is modified by nucleotide changes, which is one of the most significant indications of viral progression (9,10,11, 12,13,14,15).

Next-generation sequencing (NGS) technology is used to unravel the genome of the novel virus, and continues to provide fuel for SARS-CoV-2 sequencing projects worldwide (15). EpiCoV contains more than 12 million whole-genome sequences (WGS) of SARS-CoV-2, which can be accessed through the Global Initiative on Sharing All Influenza Data (GISAID). There are currently 11 clades in the GISAID nomenclature system based on shared marker mutations, with L and S forming early in the pandemic, before L splinters into V and G (16). Evolution and transmission observations of SARS-CoV-2 were possible owing to the unprecedented rate of genome generation, as viruses circulating in different regions started to diverge and form distinct lineages caused by mutation accumulation during viral genome replication and spread among susceptible individuals (17). A variety of nomenclature systems are currently being used to track SARS-CoV-2 genetic lineages at the local and global levels, including WHO labels (18), GISAIDs(19), NextStrains (20), and Phylogenetic Assignment of Named Global Outbreak Lineages (Pango lineages) (21).

The WHO recognized four variations of concern (VOCs) in sequence: the alpha variant (B.1.1.7) found in the UK; a Gamma (B.1.1.28) variation descendant, named P.1, first detected in South Africa in October 2020; and Beta (B.1.351) and Delta (B.1.617.2), originally reported in Brazil and India (22).

Iraqi Kurdistan has been affected by neighboring countries, such as Iran and Turkey, as their COVID-19 trends have increased. The first cases were reported in Iraq and Iraqi Kurdistan on February 24, 2020, and March 1, 2020, respectively (23,24). Iraqi Kurdistan was exposed to three waves of COVID-19: the first wave was from March 2020 to December 2020, the second wave lasted from January to June 2021, and the third wave lasted from July to December 2021(25). In the first two waves, Alpha and Beta variants were the most prevalent variants of concern (VOC); however, the delta strain was the most prevalent variant in the third wave (26,27). Overall, in Iraq, death rates in the first, second, and third waves were 2.15%, 0.58%, and 0.92%, respectively (28). The Delta variant was much more severe; however, the death rates in the third wave were lower than those in the first wave (29). This is attributable to the lack of experience in COVID-19 case management and inadequate healthcare facilities during the initial stage of the pandemic. At the beginning of January 2022, five cases of the Omicron variant were reported in a family in Duhok after a member returned from abroad (27,30).

In Iraq, virus sequencing has been limited in number since the first emergence of infection cases, resulting in only 1,409 complete genome sequences submitted to the GISAID database since the beginning of the pandemic. The first sequence of Iraqi patients available from the first wave showed the presence of a GH clade with D614G mutation(31).

To tailor public health measures at the regional or national level in the face of the potential epidemiological consequences of novel mutations, an ongoing genomic surveillance program is essential. We sequenced 40 SARS-CoV-2 genomes to identify dominant variants, clades, and lineages and to reveal the potential mutation patterns.

This study contributes to genomic surveillance efforts and provides a comprehensive and updated overview of COVID-19 epidemiology in Kurdistan by analyzing the genomic sequences of Omicron variants originating from Duhok City in the Global Initiative on Sharing All Influenza Data (GISAID) EpiCoV database from May 2022.

## 2. Materials and Methods

### *2.1. Setting*

Duhok COVID-19 Hospital is the main tertiary care referral hospital for COVID-19 cases in the Duhok Governorate. It consisted of 50 ward beds and 20 ICU beds. The primary goal of the hospital is to manage severe, critical, and complicated cases.

 doi:10.20944/preprints202308.1167.v1



The UoD COVID-19 Center for Research and Diagnosis is an independent center under the University of Duhok, which was officially launched in 2021 following endorsement by the Ministry of Higher Education and Scientific Research and the Ministry of Health of the Kurdistan Region of Iraq.

*2.2. Clinical sample and processing*

2.2.1. Sampling

**Viral RNA extraction and Real-Time PCR**

The specimens were investigated using the QIAprep&amp Viral RNA UM Kit (Qiagen), which utilizes RNA extraction and RT-PCR, according to the manufacturer's protocol. The cycle threshold ( Ct ) values of selected samples was (< 23). The genomes of the 40 samples were later dispatched to the Scripps Research Institute (TSRI), La Jolla, California, USA, for genome sequencing.

**Complete Genome Sequencing of SARS-COV-2 Omicron Variant**

The 40 genome samples were subjected to complete genome sequencing at the Scripps Research Institute (TSRI), California, USA, as described previously. The resultant sequences were later deposited by the (TSRI) into the GISAID database (http://www.gisaid.org/), where it was given the assigned accession numbers. EPI_ISL_12604438,EPI_ISL_12604442,,EPI_ISL_12604444,EPI_ISL_12604448,EPI_ISL_12604451,EPI_ISL_12604457,EPI_ISL_12604460,EPI_ISL_12604463,EPI_ISL_12604471,EPI_ISL_12604476,EPI_ISL_12604477, EPI_ISL_12604478,EPI_ISL_12604481,EPI_ISL_12604482,EPI_ISL_12604483,EPI_ISL_12604487,EPI_ISL_12604488,EPI_ISL_12604489,EPI_ISL_12604490,EPI_ISL_12604495,EPI_ISL_12604496,EPI_ISL_12604501 ,EPI_ISL_12604502,EPI_ISL_12604503,EPI_ISL_12604507,EPI_ISL_12604508,EPI_ISL_12604509, EPI_ISL_12604510,EPI_ISL_12604514,EPI_ISL_12604516,EPI_ISL_12604517,EPI_ISL_12604521 ,EPI_ISL_12604526,EPI_ISL_12604527,EPI_ISL_12604528,EPI_ISL_12604532,EPI_ISL_12604845, EPI_ISL_12604846 ,EPI_ISL_12604847, EPI_ISL_12604848 **( Data S1 ).**

**3. Genome Alignment, and Phylogenetic Analysis**

SARS-CoV-2 genome sequence alignment and mutation analysis were interpreted using the NextClade sequence analysis and the GISAID database platform tools (32,33). Lineage and clade assignments were identified using pangolins (version v.3.1.7) (34) (Table 1). Phylogenetic trees were constructed by NextClade software . We constructed a phylogenetic tree using the accession numbers of 40 sequences of hCoV-19/Iraq/KR compared to the closest SARS-CoV-2 genome sequences downloaded from the GISAID database for details of the accession numbers and other descriptions. (Figure 1)

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 16 August 2023 doi:10.20944/preprints202308.1167.v1



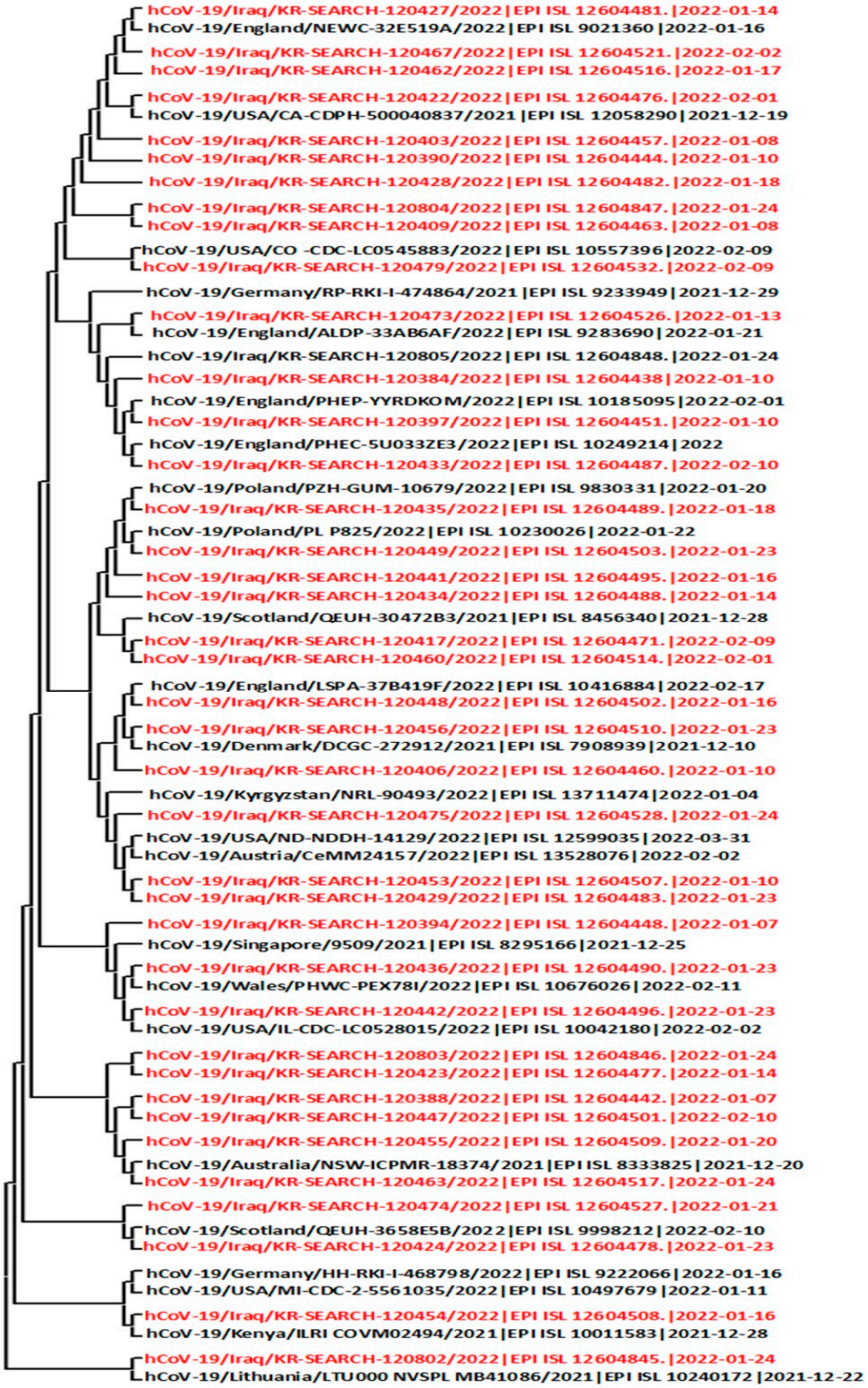


**Figure 1. Phylogenetic tree using SARS-COV-2 isolates from various regions and their almost complete gene sequences.** The sequences marked red colour were isolated from Duhok province, Iraqi-Kurdistan. The GISAID databases' websites were used to download the sequence. The MEGA 11 tree-building program employed the neighbor joining approach.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 16 August 2023 doi:10.20944/preprints202308.1167.v1



**Table 1.** Total mutations, effect on amino acid, and distribution on genomic positions of SARS-CoV-2 (Omicron variant) isolates of this study. The Entire genome sequences (40 Sequences) were aligned to the SARS-CoV-2 reference genome (NC 045512.2) using the program Nextclade version (Version 2.8.0).

| Mutation | Position | Nucleotide change | Code | Amino acid Change | Type of Mutation |
|---|---|---|---|---|---|
| **ORF1a (266...13468)** | | | | | |
| | 444 | GTT > GCT | V 60 A | Valin>Alanine | Non-synonymous SNV |
| | 593 | CAT > TAT | H 110 Y | Histidine>Tyrosine | Non-synonymous SNV |
| | 670 | AGT > AGG | S 135 R | Serine>Arginine | Non-synonymous SNV |
| | 1415 | CTT > TTT | L 384 F | Leucine>Phenylalanine | Non-synonymous SNV |
| | 2790 | ACT > ATT | T 842 I | Threonine>Isoleucine | Non-synonymous SNV |
| | 2832 | AAG > AGG | K 856 R | Lysine>Arginine | Non-synonymous SNV |
| | 2883 | TGT > TAT | C 873 Y | Cisteine>Tyrosine | Non-synonymous SNV |
| | 3896 | GTT > TTT | V 1211 F | Valine>Phenylalanine | Non-synonymous SNV |
| | 4184 | GGT > AGT | G 1307 S | Glycine>Serine | Non-synonymous SNV |
| | 4893 | ACA > ATA | T 1543 I | Threonin>Isoleucine | Non-synonymous SNV |
| | 5007 | ACG > ATG | T 1581 M | Threonin>Methionine | Non-synonymous SNV |
| | 510 – 518 | ATG > -TG | del82/84 | del82/84 | Non-frame shift deletion |
| | 519 | ATG > -TG | M 85 V | Methionine>Valine | Non-synonymous SNV |
| | 6176 | GAT > AAT | D 1971 N | Aspartic acid>Asparagine | Non-synonymous SNV |
| | 6513 - 6515 | | del2083/2083 | del2083/2083 | Non-synonymous deletion |
| | 6516 | TTA > -TA | L 2084 I | Leucine>Isoleucine | Non-synonymous SNV |
| | 7036 | TTA > TTT | L 2257 F | Leucine>Phenylalanine | Non-synonymous SNV |
| | 7488 | ACT > ATT | T 2408 I | Threonine>Isoleucine | Non-synonymous SNV |
| | 8393 | GCT > ACT | A 2710 T | Alanine>Threonin | Non-synonymous SNV |
| | 9344 | CTT > TTT | L 3027 F | Leucine>Phenylalanine | Non-synonymous SNV |
| | 9474 | GCT > GTT | A 3070 V | Alanine>Valine | Non-synonymous SNV |
| | 9534 | ACT > ATT | T 3090 I | Threonine>Isoleucine | Non-synonymous SNV |
| | 9866 | CTT > TTT | L 32201 I | Leucine>Isoleucine | Non-synonymous SNV |
| | 10029 | ACC > ATC | T 3255 I | Threonin>Isoleucine | Non-synonymous SNV |
| | 10323 | AAG > AGG | K 3353 R | Lysine>Arginine | Non-synonymous SNV |
| | 10449 | CCC > CAC | P 3395 H | Proline>Histidine | Non-synonymous SNV |
| | 11405 | GTC > TTC | V 3714 F | Valine>Phenylalanine | Non-synonymous SNV |
| | 11285-11293 | | del3674/3676 | del3674/3676 | Non-frame shift deletion |
| | 11537 | ATT > GTT | I 3758 V | Isoleucine>Valine | Non-synonymous SNV |
| | 12534 | ACT > ATT | T 409 I | Threonine>Isoleucine | Non-synonymous SNV |
| **ORF1b (13468...21555)** | | | | | |
| | 13756 | ATA > GTA | I 97 V | Isoleucine>Valine | Non-synonymous SNV |
| | 14408 | CCT > CTT | P 314 L | Proline>Leucine | Non-synonymous SNV |
| | 14821 | CCA > TCA | P 452 S | Proline>Serine | Non-synonymous SNV |
| | 15641 | AAT > AGT | N 725 S | Asparagine>Serine | Non-synonymous SNV |
| | 15982 | GTA > ATA | V 839 I | Valine>Isoleucine | Non-synonymous SNV |
| | 16744 | GGT > AGT | G 1093 S | Glycine>Serine | Non-synonymous SNV |
| | 17410 | GGT > TGT | R 1315 C | Arginine>Cisteine | Non-synonymous SNV |

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 16 August 2023 doi:10.20944/preprints202308.1167.v1



| | | | | |
|---|---|---|---|---|
| 18163 | ATA > GTA | I 1566 V | Isoleucine>Valine | Non-synonymous SNV |
| 18433 | GAT > CAT | D 165 H | Aspartic acid>Histidine | Non-synonymous SNV |
| 19999 | GTT > TTT | V 2178 F | Valine>Phenylalanine | Non-synonymous SNV |
| 20003 | GAT > GGT | P 2179 G | Proline>Glycine | Non-synonymous SNV |

**S (21563...25384)**

| | | | | |
|---|---|---|---|---|
| 21765 - 21770 | TACATG > - - - | del69/70 | del69/70 | Non-synonymous deletion |
| 21789 | ACT > ATT | T 76 I | Threonine>Isoleucine | Non-synonymous SNV |
| 21846 | ACT > ATT | T95I | Threonine>Isoleucine | Non-frame shift deletion |
| 21987 | GGT > GAT | G142D | Glycine>Aspartic acid | Non-synonymous SNV |
| 21987 - 21995 | | del142/144 | del142/144 | Non-frame shift deletion |
| 21996 | TAC > -AC | Y 145 D | Tyrosine>Aspartic acid | Non-synonymous SNV |
| 22194 - 22196 | AAT > A-- | del211/211 | del211/211 | Non-synonymous deletion |
| 22197 | TTA > -TA | L 212 I | Leucine>Isoleucine | Non-synonymous SNV |
| 222000 | GTG > GGG | V 213 G | Valine>Glycine | Non-synonymous SNV |
| 22578 | GCT > GAT | G339D | Glycine>Aspartic acid | Non-synonymous SNV |
| 22599 | AGA > AAA | R346K | Arginine>Lysine | Non-synonymous SNV |
| 22673 | T > C | S371L | Serine>Leucine | Non-synonymous SNV |
| 22674 | C > T | S 373 P | Serine>Proline | Non-synonymous SNV |
| 22686 | TCC > TTC | S 375 F | Serine>Phenylalanine | Non-synonymous SNV |
| 22688 | ACT > GCT | T 376 A | Threonine>Isoleucine | Non-synonymous SNV |
| 22786 | AGA > AGC | R408S | Arginine>Serine | Non-synonymous SNV |
| 22813 | AAG > AAT | K 417 N | Lysine>Asparagine | Non-synonymous SNV |
| 22882 | AAT > AAG | N440K | Asparagine>Lysine | Non-synonymous SNV |
| 22898 | GGT > AGT | G446S | Glycine>Serine | Non-synonymous SNV |
| 23013 | GAA > GCA | E 484 A | Glutamic acid > isoleucine | Non-synonymous SNV |
| 22992 | AGC > AAC | S477N | Serine>Asparagine | Non-synonymous SNV |
| 22995 | ACA > AAA | T478K | Threonine>Lysine | Non-synonymous SNV |
| 23040 | CAA > CGA | Q493R | Glutamine>Arginine | Non-synonymous SNV |
| 23048 | G > A | G496S | Glycine>Serine | Non-synonymous SNV |
| 23055 | A > G | Q498R | Glutamine>Arginine | Non-synonymous SNV |
| 23063 | AAT > TAT | N501Y | Asparagine>Tyrosine | Non-synonymous SNV |
| 23075 | TAC > CAC | Y505H | Tyrosine>Histidine | Non-synonymous SNV |
| 23202 | ACA > AAA | T547K | Threonine>Lysine | Non-synonymous SNV |
| 23403 | GAT > GGT | D614G | Aspartic acid>Glycine | Non-synonymous SNV |
| 23525 | CAT > TAT | H655Y | Histidine>Tyrosine | Non-synonymous SNV |
| 23599 | T > G | N679K | Asparagine>Lysine | Non-synonymous SNV |
| 23604 | CCT > CAT | P681H | Proline>Histidine | Non-synonymous SNV |
| 23854 | AAC > AAA | N764K | Asparagine>Lysine | Non-synonymous SNV |
| 23948 | GAT > TAT | D796Y | Aspartic acid>Tyrosine | Non-synonymous SNV |
| 24130 | ACC > AAA | N856K | Asparagine>Lysine | Non-synonymous SNV |
| 24424 | CAA > CAT | Q954H | Glutamine>Histidine | Non-synonymous SNV |
| 24469 | AAT > AAA | N969K | Asparagine>Lysine | Non-synonymous SNV |
| 24503 | CCT > TTT | L981F | Leucine>Phenylalanine | Non-synonymous SNV |

 doi:10.20944/preprints202308.1167.v1



| | | | | | |
|---|---|---|---|---|---|
| **ORF3a (25393…26220)** | | | | | |
| | 25471 | GAT > TAT | D 27 Y | Aspartic acid>Tyrosine | Non-synonymous SNV |
| | 26060 | ACT > ATT | T 223 I | Threonine>Isoleucine | Non-synonymous SNV |
| **M (26523... 27191)** | | | | | |
| | 26530 | GAT > GGT | D 3 G | Aspartic acid>Glycine | Non-synonymous SNV |
| | 26577 | CAA > GAA | Q 19 E | Glutamine>Glutamic acid | Non-synonymous SNV |
| | 26709 | GCT > ACT | A 63 T | Alanine>Threonin | Non-synonymous SNV |
| **ORF6 (27202…27387)** | | | | | |
| | 27269 | AAA > -AA | K 23 * | K23* | Non-synonymous SNV |
| | 27266 - 27268 | TTA > - - - | del22/23 | del22/23 | Non-frame shift deletion |
| **ORF9b (28284…28577)** | | | | | |
| | 28311 | CCC > TCC | P 10 S | Proline>Serine | Non-synonymous SNV |
| **N (28274…29533)** | | | | | |
| | 28881 | AGG > AAA | R 203 K | Arginine>Lysine | Non-synonymous SNV |
| | 28882 | AGG > AAA | R203 K | Arginine>Lysine | Non-synonymous SNV |
| | 28883 | GGA > ACG | G 204 R | Glycine>Arginine | Non-synonymous SNV |
| | 28311 | CCC > CTC | P 13 L | Proline>Leucine | Non-synonymous SNV |
| | 28725 | CCT > CTT | P 151 L | Proline>Leucine | Non-synonymous SNV |
| | 29000 | GGC > AGC | G 243 S | Glycine>Serine | Non-synonymous SNV |
| | 29005 | CAA > CAC | Q 244 H | Glutamine>Histidine | Non-synonymous SNV |
| | 29510 | AGT > CGT | S 413 R | Serine > Arginine | Non-synonymous SNV |

## 4. Results and Discussions

In early January 2020, it was discovered that the novel coronavirus was the source of many pneumonia cases that were reported in China in late December 2019 but were of unknown origin (35,36). The virus was named severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), the causative agent of coronavirus disease 2019 **(**COVID-19). The virus has spread worldwide despite significant efforts to restrict the disease in China, and the WHO classified COVID-19 as a global pandemic in March 2020 (37). SARS-COV-2 is characterized by a single-stranded RNA making up their genomes with positive polarity, meaning that the RNA base sequences are oriented in a manner that is identical to the later messenger RNA (mRNA). The SARS-CoV-2 genome is the biggest RNA genome known to exist, measuring to 26.4–31.7 kilobases (38).

In April 2022, we uploaded our isolated virus strain genome sequences between waves four and five for genome analysis in the GISAID database. According to the WHO coronavirus dashboard of the reported cases in Iraq ( https://covid19.who.int/region/emro/country/iq) (Figure 2), the samples of the latest fifth wave belonged to the Omicron variant (BA.2). This finding is of some kind, as it has never been discovered in previous studies in Iraq. This unique discovery highlights the potential evolution and diversification of SARS-CoV-2 in Brazil.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 16 August 2023 doi:10.20944/preprints202308.1167.v1



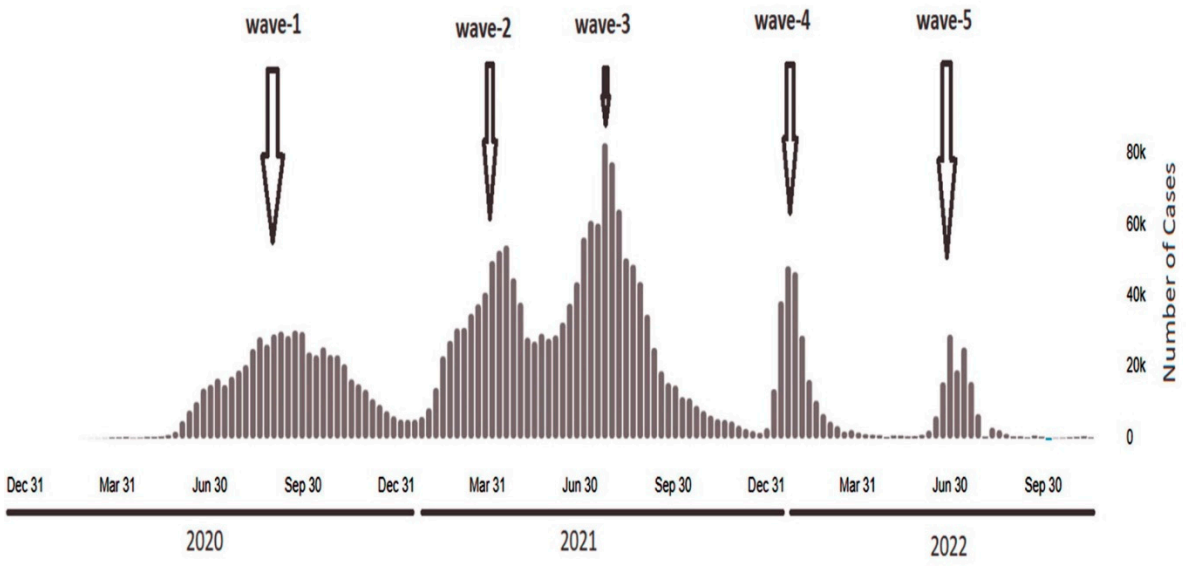


**Figure 2.** Daily confirmed cases of SARS-COV-2between 2021 and 2022 on the WHO Coronavirus dash-board in Iraq ( https://covid19.who.int/region/emro/country/iq).

Analysis of the Omicron VOC of SARS-CoV-2 sequences revealed that the S gene had the most changes, followed by the ORF1ab, N, M, ORF6, ORF3a, ORF9b, and E genes. Among these genes, ORF3a, ORF6, E-gene, and ORF9b had the least number of mutations. Because spike glycoprotein (S) is the primary target for both therapy and diagnosis, it is also an essential protein that defines viral host affinity and pathogenesis. The structural protein encoded by the S gene with the greatest mutations acts as a protein that binds to viruses for host cell receptors to recognize the host range (39). More than 90% of the neutralizing antibodies in COVID-19 convalescent plasma, which is subunit S1, are anti-RBD; the S1 viral protein is the most immunodominant (40). There are a total of 38 non-synonymous mutations in this gene, including T 76 I, T95I, G142D, Y 145 D, L 212 I, V 213 G, G339D, R346K, S371L, S373P, S375F, T376A, R408S, K417N, N440K, G446S, S477N, T478K, Q493R, G496S, Q498R, N501Y, Y505H, T547K, D614G, H655Y, N679K, P681H, N764K, D796Y, N856K, Q954H, N969Kand L981Fdel211/211 del3674/3676 del69/70 del142/144 del22/23 del68 /70, del142 /145, and del211 /212 deletions (Table 1). According to previous reports, the mutation potential of this gene is higher than that at other genomic locations (41). The global prevalence of the S protein D614G variant has progressively increased over time and is currently present in approximately 74% of all known variants, according to the GISAID SARS-CoV-2 database, from June 25, 2020. On January 24, 2020, one of the most common SARS-CoV-2 mutations, D614G, which changes the amino acid glycine (G) with a non-polar side chain from the amino acid aspartate (D) with a polar negatively charged side chain, was first identified in China. (42). It was discovered that D614G enhances the infectivity, transmission rate, and effectiveness of cellular entry of SARS-CoV-2 across a wide spectrum of human cell types as a result of positive natural selection (43,44,45). One of the earliest and most common mutations is D614G, in which glycine is substituted with aspartic acid at position 614 (G614). The RBM domain mutation D614G increases the density of S proteins on the viral surface, thereby increasing infectivity (46). In spike protein, the dominant variant was D614G. This variant has also been preserved for Omicron and found in all other VOCs, which is related to increased transmission and infection rates, as well as viral escape from reactive antibodies (47,48).

In this study, we discovered that some of the other frequently occurring RBD changes improved ACE2 binding, including N501Y, S477N, and E484K, and the same substitutions were seen in the majority of VOCs, which are associated with greater transmissibility. In the existing change of VOC, the T478K, Q498R and Q493K mutations have demonstrated to increase the electrostatic potential, boosting the RBD-ACE2 binding affinity (49). Immune escape has also been associated with the

 doi:10.20944/preprints202308.1167.v1



existing change in E484K (50). A variety of VOCs in a previous study, including B.1.617.2 (E484K/E484Q B.1.351 (E484K), P.1, and B.1.1.529, showed E484 replacements (E484A) (51).

Notably, during the evolution of SARS-CoV-2, the S1-NTD in the present study additionally organized many mutations, including deletions idel69/70 and del 142/144. Similarly, frequently deleted regions in the NTD were those positioned at 69–70, 141–144, 146, 210, and 243-244; reported in the previous study most NTD mutations were found to change antigenicity or eliminate epitopes, enabling immune escape (52).

Most mutations are likely to have an impact on viral entry because S2 remains stable across SARS-CoV-2 and has a low mutation rate. Additionally, it is less antigenic than S1, which may be due to substantial N-linked glycosylation, and is consequently not subjected to as much selective pressure (53). The BA.1.BA.1.1 and BA.1.18 (Omicron variant), in this study, unexpectedly display several S2 changes, including D796Y, N856K, Q954H, N969K, and L981F. Similarly, a previous study reported that the Omicron variant B.1.1.529 had the following S2 substitutions: D796Y, N856K, Q954H, N969K, and L981F. (54). The global sustainable dissemination of B.1.1.529 indicates that these mutations are beneficial. It is not yet known how these mutations affect the pathogenicity of the virus and the polyclonal mAb response.

ORF1a/b is a key target for SARS-CoV-2 nucleic acid assays. The non-structural proteins encoded by ORF1a and ORF1b are required for reproduction, maintenance, and viral DNA repair (55). RdRp (nsp12), 3-chymotrypsin-like protease (3CLpro), nsp5, (PLpro, nsp3) also known as major protease or Mpro, and papain-like proteinase protein are a few of these proteins that antiviral medications used to treat COVID-19 (56). ORF1ab produces an RNA-dependent RNA polymerase enzyme and a helicase protein that are both necessary for viral replication (57). The most frequently altered non-structural protein of ORF1ab was nsp3, which also had a deletion at the amino acid sites del2083/2083, A 2710 T, and K 856 R. Variants T 3255 I and I 3758 V with non-frameshift deletions (del3674/3676) were found on nsp4 and nsp6, respectively. Transmembrane proteins NSP3, NSP4, and NSP6 modulate host immunity and enhance the functionality of host cell organelles for viral reproduction (58). The largest NSP and a crucial part of viral transcription and replication are NSP3, which is present at this stage. The host cell membrane is where transcription and replication of the virus genome occur (59).

The nucleocapsid (N) is a viral protein or gene of significance in diagnostics (nucleic acid and antigen detection) and a unique vaccine formulation (60). It is essential for viral assembly, budding, and the recipient cell reaction to viral infection. Its function is to maintain the structure of the genome within the membrane (61). R203K and G204R mutations were the most frequent types of N-proteins in the present study (62). A previous study reported that SARS-CoV-2 variants have improved virulence and transmission owing to R203K and G204R mutations. Following these two mutations, P13L was discovered in the present study. These mutations have been documented in other geographical regions (63). An essential T-cell epitope containing P13. Consequently, a change in this location may affect the characteristics of the epitope, as well as the cellular immune system's reaction to the virus (64).

In addition to spike protein changes, nucleocapsid protein mutations are crucial for the transmission of the pandemic virus (65). The N gene in the omicron contains a significant number of deletions, which have been observed to affect diagnostics, primarily the primer binding of a few commercially available kits. However, it is unclear how these alterations affect viral pathogenicity. However, the accessory proteins ORF6 and ORF9b inhibit innate immunity, signaling pathways, and interferon (IFN) expression by concentrating on the MAVS adapter associated with mitochondria (66). Previous studies have reported that the existing ORF9b gene mutation was>85% prevalent across all Omicron sequences (n = 70) (67).

Finally, the genome sequences of the located sequenced SARS-COV-2 were compared to those from the USA, UK, Germany, Austria, Kenya, Poland, Denmark, Kyrgyzstan, Malaysia, Morocco, Singapore, and Lithuania (Table 2). The results show how closely related our isolates are to previously sequenced SARS-COV-2 genomes, such as those from Poland, USA, and the UK, which share the same lineage as BA.1, while the other mentioned countries share similar lineages to (BA.1.1).

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 16 August 2023 doi:10.20944/preprints202308.1167.v1



Interestingly, the subvariant (BA.2) was observed only in Duhok, which was isolated from USA (https://nextstrain.org/),(https://www.epicov.org/) (Figure 3). These Pango nomenclatures help researchers and public health organizations worldwide monitor the transmission and spread of SARS-CoV-2, including variants of concern (https://cov-lineages.org/).

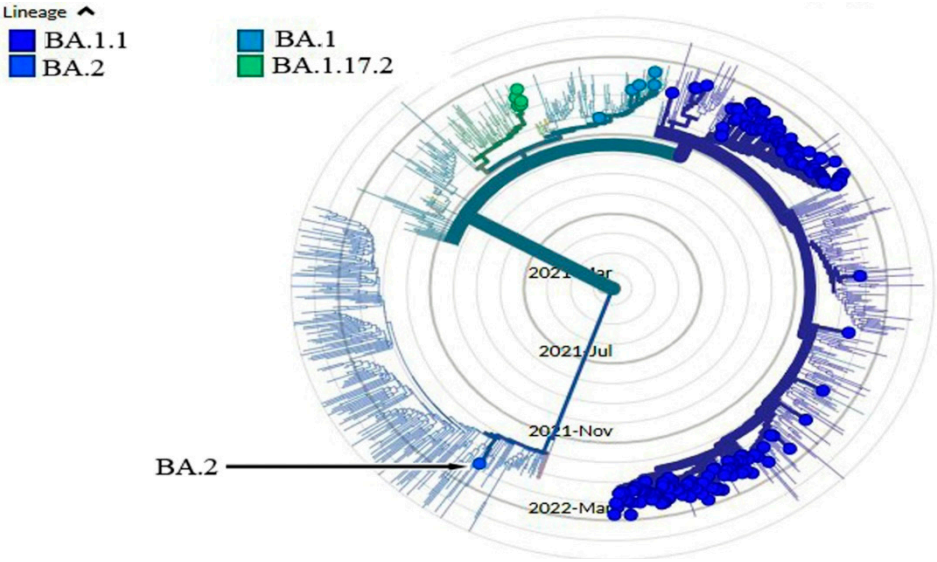


**Figure 3. T**he phylogenetic tree illustrates the distribution of the pango lineage in Iraq, BA.2 (Accession ID: EPI_ISL_12604503) was first observed in Duhok, Iraq (https://nextstrain.org/),(https://www.epicov.org/).

**Table 2.** The table shows the distribution of closely related genomes to uploaded genomes according to the Pango lineage in the present study. The analyzed data were downloaded from GISAID databases, and the uploaded data in this study were compared with other neighboring countries downloaded from GISAID databases and GenBank. For each query sequence, the table gives the sequence number within the input data and the Accession ID (identified from the name provided in the input data). The other columns provide information on the closest related genome, including the match distance, match quality, Accession ID, collection date, submission date, lineage, and country of origin of the matched genome (https://nextstrain.org/) (https://www.epicov.org/).

| Upload Acc. ID | Distance | Quality | Accession ID Of Close Related Genome | Collection date | Submission date | lineage | Country/State |
|---|---|---|---|---|---|---|---|
| EPI_ISL_12604438 | 1 | 0.959 | EPI_ISL_10185095 | 2/1/2022 | 2/22/2022 | BA.1.1 | United Kingdom / England |
| EPI_ISL_12604442 | 0 | 0.998 | EPI_ISL_12604518 | 2/10/2022 | 5/9/2022 | BA.1.1 | Iraq / Kurdistan / Duhok |
| EPI_ISL_12604444 | 0 | 0 | EPI_ISL_11163451 | 1/26/2022 | 3/18/2022 | BA.1.1 | USA / Tennessee |
| EPI_ISL_12604448 | 0 | 1 | EPI_ISL_11501531 | 2/18/2022 | 3/28/2022 | BA.1.1 | Canada / Saskatchewan |
| EPI_ISL_12604478 | 0 | 0.974 | EPI_ISL_9222066 | 1/16/2022 | 1/28/2022 | BA.1.1 | Germany / Hamburg |
| EPI_ISL_12604481 | 0 | 0.943 | EPI_ISL_9021360 | 1/16/2022 | 1/24/2022 | BA.1.1 | United Kingdom / England |
| EPI_ISL_12604482 | 0 | 1 | EPI_ISL_9041699 | 1/5/2022 | 1/24/2022 | BA.1.1 | USA / Maryland |
| EPI_ISL_12604483 | 0 | 0.999 | EPI_ISL_13528111 | 2/6/2022 | 6/28/2022 | BA.1.1 | Austria / Styria / Liezen |
| EPI_ISL_12604487 | 0 | 0.977 | EPI_ISL_9830331 | 1/20/2022 | 2/12/2022 | BA.1 | Poland /Swietokrzyskie Voivodeship |
| EPI_ISL_12604488 | 0 | 0.947 | EPI_ISL_8456340 | 12/28/2021 | 1/7/2022 | BA.1.1 | United Kingdom / Scotland |

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 16 August 2023 doi:10.20944/preprints202308.1167.v1



| | | | | | | | |
|---|---|---|---|---|---|---|---|
| EPI_ISL_12604489 | 1 | 1 | EPI_ISL_17134891 | 1/13/2022 | 2/28/2022 | BA.1 | USA / Arkansas |
| EPI_ISL_12604490 | 0 | 0.997 | EPI_ISL_10676026 | 2/11/2022 | 3/4/2022 | BA.1.1 | United Kingdom / Wales |
| EPI_ISL_12604495 | 1 | 0.94 | EPI_ISL_11955489 | 1/12/2022 | 4/11/2022 | BA.1.1 | United Kingdom / England |
| EPI_ISL_12604496 | 1 | 0.942 | EPI_ISL_10042180 | 2/2/2022 | 2/17/2022 | BA.1.1 | USA / Illinois |
| EPI_ISL_12604501 | 0 | 1 | EPI_ISL_12604497 | 2/10/2022 | 5/9/2022 | BA.1.1 | Iraq / Kurdistan / Duhok |
| EPI_ISL_12604502 | 0 | 0.994 | EPI_ISL_12604412 | 2/5/2022 | 5/9/2022 | BA.1.1 | Iraq / Kurdistan / Duhok |
| EPI_ISL_12604503 | 0 | 1 | EPI_ISL_11252814 | 3/3/2022 | 3/21/2022 | BA.2 | USA / Ohio |
| EPI_ISL_12604507 | 1 | 0.999 | EPI_ISL_11041477 | 3/7/2022 | 3/15/2022 | BA.1.1 | United Kingdom / Scotland |
| EPI_ISL_12604508 | 0 | 1 | EPI_ISL_10011583 | 12/28/2021 | 2/17/2022 | BA.1.1 | Kenya / Migori |
| EPI_ISL_12604509 | 1 | 0.954 | EPI_ISL_9517295 | 1/14/2022 | 2/4/2022 | BA.1.1 | Iraq / Baghdad |
| EPI_ISL_12604510 | 0 | 0.991 | EPI_ISL_7908939 | 12/10/2021 | 12/21/2021 | BA.1.1 | Denmark / Syddanmark |
| EPI_ISL_12604514 | 1 | 0.903 | EPI_ISL_9835746 | 1/11/2022 | 2/13/2022 | BA.1.1 | United Kingdom / England |
| EPI_ISL_12604516 | 0 | 1 | EPI_ISL_10181069 | 2/2/2022 | 2/22/2022 | BA.1.1 | United Kingdom / Scotland |
| EPI_ISL_12604517 | 0 | 0.999 | EPI_ISL_10497679 | 1/11/2022 | 3/1/2022 | BA.1.1 | USA / Michigan |
| EPI_ISL_12604521 | 0 | 1 | EPI_ISL_17128823 | 1/20/2022 | 3/21/2022 | BA.1.1 | USA / Arkansas |
| EPI_ISL_12604527 | 1 | 0.986 | EPI_ISL_9998212 | 2/10/2022 | 2/17/2022 | BA.1.1 | United Kingdom / Scotland |
| EPI_ISL_12604528 | 1 | 1 | EPI_ISL_13674544 | 1/13/2022 | 7/6/2022 | BA.1.1 | USA / Minnesota |
| EPI_ISL_12604532 | 2 | 0.989 | EPI_ISL_9233949 | 12/29/2021 | 1/28/2022 | BA.1.1 | Germany / Rhineland-Palatinate |
| EPI_ISL_12604845 | 1 | 0.946 | EPI_ISL_10240172 | 12/22/2021 | 2/23/2022 | BA.1.1 | Lithuania / Kauno apskritis |
| EPI_ISL_12604846 | 0 | 1 | EPI_ISL_12604443 | 1/24/2022 | 5/9/2022 | BA.1.1 | Iraq / Kurdistan / Duhok |
| EPI_ISL_12604847 | 0 | 1 | EPI_ISL_9041699 | 1/5/2022 | 1/24/2022 | BA.1.1 | USA / Maryland |
| EPI_ISL_12604848 | 0 | 1 | EPI_ISL_9985416 | 2/4/2022 | 2/16/2022 | BA.1.1 | Poland / Lodzkie / Lodz |
| EPI_ISL_12604451 | 0 | 0.954 | EPI_ISL_10249214 | 2022 | 2/23/2022 | BA.1 | United Kingdom / England |
| EPI_ISL_12604457 | 1 | 0.94 | EPI_ISL_10557396 | 2/9/2022 | 3/2/2022 | BA.1.1 | USA / Colorado |
| EPI_ISL_12604460 | 2 | 0.971 | EPI_ISL_13711474 | 1/4/2022 | 7/8/2022 | BA.1.1 | Kyrgyzstan / Chui |
| EPI_ISL_12604463 | 1 | 1 | EPI_ISL_12604423 | 1/19/2022 | 5/9/2022 | BA.1.1 | Iraq / Kurdistan / Duhok |
| EPI_ISL_12604471 | 0 | 0.932 | EPI_ISL_9224622 | 12/30/2021 | 1/28/2022 | BA.1 | Germany / North Rhine-Westphalia |
| EPI_ISL_12604476 2 0.954 | 2 | 0.954 | EPI_ISL_12058290 | 12/19/2021 | 4/14/2022 | BA.1.1 | USA / California / Placer County |
| EPI_ISL_12604526 | 0 | 0.907 | EPI_ISL_11405543 | 2/27/2022 | 3/25/2022 | BA.1.1 | South Korea |

The detection of the Omicron BA.2 subvariant in the fifth wave of COVID-19 cases in Iraq indicates a significant and unique discovery. This finding stresses the demand for continuing genomic surveillance efforts, worldwide data sharing, and reporting of emerging variants and

subvariants. These measures are crucial for motivating prompt public health responses and minimizing the impact of the pandemic.

Compared to other VOCs, the SARS-CoV-2 genome's high mutation rate, particularly on the spike protein, may promote viral transmission and immunological evasion. Additionally, the development of novel vaccines that incorporate the Omicron variety as a potential reference strain has become necessary because of the assortment of these large mutations in the immunogenic epitopes of the Spike protein. More research is needed to determine the infectivity and efficacy of current vaccines against Omicron.

## 5. Conclusion

In this study, 40 SARS-CoV-2 genomes from a clinical sample collected from Duhok, Iraqi-Kurdistan, were sequenced. The genome contains additional mutations and known mutations (non-synonymous mutations). Nucleotide sequences contain position-specific mutations such as single-nucleotide variants. Additional analyses were performed on the amino acid substitutions to determine the functional stability of the proteins. Our findings showed that the SARSCoV-2 Omicron variant proteins have 95 different variant locations in their coding regions. The spike protein has the most mutations, followed by six structural proteins (ORF1a, N, M, ORF6, ORF9, and E). We found fewer mutations in ORF6, ORF9, and E proteins than in S proteins and ORF1a. A non-synonymous substitution was observed in all protein variants in the whole-genome sequence. In addition, we observed an omicron subvariant (BA.2) in Duhok, which has not been previously recorded in other studies from Iraq. These new subvariant (BA.2) findings provide data on the mutation pattern in circulating variants within the country, which could help public health authorities issue and update control roles for SARS-CoV-2 emergence.

Taking everything into account, we propose that future studies should focus on the impacts and functions of these genomic variants on virus infectivity, pathogenesis, and severity. Nevertheless, developing new vaccines, particularly multivalent vaccines containing multiple VOCs, could boost the control of the latest infections.

**Author Contributions:** SMT , contributed to the study conception and design , wrote the manuscript , collected and analyzed data. JMA and MAM critically revised the final manuscript and confirm the authenticity of all the raw data. All authors read and approved the final manuscript.

**Funding:** No funding was received for this study.

**Institutional Review Board Statement** Not applicable.

**Informed Consent Statement** Not applicable

**Data Availability Statement:** The Global Initiative on Sharing All Personal Data (GISAID) has received SARS-CoV-2 sequences from this investigation.

**Conflicts of Interest** The authors declare that they have no competing interests.

## References

1. Kim D, Lee J-Y, Yang J-S, Kim JW, Kim VN, Chang H, et al. The architecture of SARS-CoV-2 transcriptome. Cell. 2020;181(4):914-21. e10.
2. Sidiqi KR, Sabir DK, Ali SM, Kodzius R, et al. Does early childhood vaccination protect against COVID-19?. Frontiers in molecular biosciences, 2020, 7: 120
3. Huremović D. Psychiatry of pandemics: a mental health response to infection outbreak: Springer; 2019.
4. Shaibu JO, Onwuamah CK, James AB, Okwuraiwe AP, Amoo OS, Salu OB, et al. Full length genomic sanger sequencing and phylogenetic analysis of Severe Acute Respiratory Syndrome
5. Coronavirus 2 (SARS-CoV-2) in Nigeria. PloS one. 2021;16(1):e0243271.
6. Aldiabat K, Kwekha Rashid A, Talafha H, Karajeh A,et al. The extent of smartphones users to adopt the use of cloud storage. J Comput Sci. 2018;14(12):1588-98.1
7. Alhayani B, Abbas ST, Mohammed HJ, Mahajan HB,et al. Intelligent secured two-way image transmission using corvus corone module over WSN. Wireless Personal Communications. 2021;120(1):665-700.

8. Hasan HS, Abdallah AA, Khan I, Alosman HS, Kolemen A, Alhayani B,et al. Novel unilateral dental expander appliance (udex): a compound innovative materials. Computers, Materials and Continua. 2021:3499-511.
9. Mostafaei S, Sayad B, Azar MEF, Doroudian M, Hadifar S, Behrouzi A, et al. The role of viral and bacterial infections in the pathogenesis of IPF: a systematic review and meta-analysis. Respiratory research. 2021;22(1):1-14.
10. Hui EK-W. Reasons for the increase in emerging and re-emerging viral infectious diseases. Microbes and infection. 2006;8(3):905-16
11. Domingo E. Viruses at the edge of adaptation. Virology. 2000;270(2):251-3.
12. Taheri M, Rad LM, Hussen BM, Nicknafs F, Sayad A, Ghafouri-Fard S,et al. Evaluation of expression of VDR-associated lncRNAs in COVID-19 patients. BMC Infectious Diseases. 2021;21(1):1-10.
13. Ibrahim FM, Alkaim A, Kadhom M, Sabir DK, Salih N, Yousif E,et al. Chemistry of selected drugs for SARS-CoV-2 inhibition: tested in vitro and approved by the FDA. Chem Int. 2021;7(3):212-6.
14. Khan S, Hussain A, Vahdani Y, Kooshki H, Hussen BM, Haghighat S, et al. Exploring the interaction of quercetin-3-O-sophoroside with SARS-CoV-2 main proteins by theoretical studies: A probable prelude to control some variants of coronavirus including Delta. Arabian Journal of Chemistry. 2021;14(10):103353.
15. Elbe S, Buckland-Merrett G. Data, disease and diplomacy: GISAID's innovative contribution to global health. Global challenges. 2017;1(1):33-46.
16. Zhu N, Zhang D, Wang W, Li X, Yang B, Song J, et al. others.(2020). A novel coronavirus from patients with pneumonia in China, 2019. New England Journal of Medicine
17. Chen, Z., Azman, A. S., Chen, X., Zou, J., Tian, Y., Sun, R., ... & Yu, H. Global landscape of SARS-CoV-2 genomic surveillance and data sharing. Nature genetics, 2022, 54.4: 499-507
18. Khare S, Gurry C, Freitas L, Schultz MB, Bach G, Diallo A. . GISAID's role in pandemic response. China CDC weekly, 2021, 3.49: 1049
19. Motayo BO, Oluwasemowo OO, Olusola BA, Akinduti PA, Arege OT, Obafemi YD. Evolution and genetic diversity of SARS-CoV-2 in Africa using whole genome sequences. International Journal of Infectious Diseases, 2021, 103: 282-287
20. World Health Organization. WHO Announces Simple, Easy-To-Say Labels for SARS-CoV-2 Variants of Interest and Concern.(2021). 2022.
21. Elbe S, Buckland-Merrett G. Data, disease and diplomacy: Data, disease and diplomacy: GISAID's innovative contribution to global health. *Global challenges*, 2017, 1.1: 33-46.
22. Hadfield J, Megill C, Bell SM, Huddleston J, Potter B, Callender C, et al. Nextstrain: real-time tracking of pathogen evolution. Bioinformatics, 2018, 34.23: 4121-4123
23. Rambaut A, Holmes EC, O'Toole A, Hill V, McCrone JT, Ruis C, et al. A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology. Nature microbiology, 2020, 5.11: 1403-1407
24. Choi, J.Y.; Smith, D.M. SARS-CoV-2 variants of concern. *Yonsei Med. J.* **2021**, *62*, 961–968.
25. Merza MA, Haleem Al Mezori AA, Mohammed HM, et al. COVID-19 outbreak in Iraqi Kurdistan: the first report characterizing epidemiological, clinical, laboratory, and radiological findings of the disease. Diabetes MetabSyndr. 2020;14(4):547-554. doi: 10.1016/j.dsx.2020.04.047
26. Merza MA, Aswad SM, Sulaiman HM, et al. Clinical and epidemiological characteristics and outcomes of Coronavirus disease-19 patients in a large longitudinal study. Int J Health Sci (Qassim). 2021;15(4):29-41
27. OCHA. Panel study IV: impact of COVID-19 on small- and medium-sized enterprises in Iraq - June 2020 to June 2021. https://reliefweb.int/report/iraq/panel-study-iv-impact-covid-19-small-and-medium sizedenterprisesiraq-june-2020-june. Accessed June 2022.
28. Al-Rashedi NAM, Alburkat H, Hadi AO, et al. High prevalence of an alpha variant lineage with a premature stop codon in ORF7a in Iraq, winter 2020-2021. PLoS One. 2022;17(5):e0267295. Published May 26, 2022.
29. Rudaw. Kurdistan region, Iraq record first cases of COVID-19 Omicron variant. Rudaw.net.https://www.rudaw.net/english/middleeast/iraq/060120222. Accessed January 6, 2022.
30. Worldometers. Worldometer's COVID-19 data. https://www.worldometers.info/coronavirus/country/iraq/. Accessed February 25, 2022.
31. Ong SWX, Chiew CJ, Ang LW, et al. Clinical and virological features of SARS-CoV-2 variants of concern: a retrospective cohort study comparing B.1.1.7 (Alpha), B.1.315 (Beta), and B.1.617.2 (Delta) [published online ahead of print, Aug 23, 2021]. Clin Infect Dis. 2021;ciab721
32. Finlay C, Brett A, Henry L, Yuka J, Franck K, Neale B. Increased transmissibility and global spread of SARS-CoV-2 variants of concern as at June 2021. Euro Surveill. 2021; 26(30)
33. Al-Rashedi NAM, Licastro D, Rajasekharan S, Monego D, Marcello A, Munahi MG, et al. Genome sequencing of a novel coronavirus SARS-CoV-2 isolate from Iraq. Microbiol Resour Announc. 2021; 10 (4):e01316–20. https://doi.org/10.1128/MRA.01316-20 PMID: 33509990
34. Elbe, S.; Buckland-Merrett, G. Data, disease and diplomacy: GISAID's innovative contribution to global health. *Glob. Challenges* **2017**, *1*, 33–46.

35. Aksamentov, I.; Roemer, C.; Hodcroft, E.B.; Neher, R.A. Nextclade: Clade assignment, mutation calling and quality control for viral genomes. *J. Open Source Softw.* **2021**, *6*, 3773.
36. Huang, C., Wang, Y., Li, X., Ren, L., Zhao, J., Hu, Y., ... & Cao, B., et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *The lancet*, 2020, 395.10223: 497-506.
37. Woo PC, Lau SK, Lam CS, Lai KK, Huang Y, Lee P, et al. Comparative analysis of complete genome sequences of three avian coronaviruses reveals a novel group 3c coronavirus. Journal of virology. 2009;83(2):908-17.
38. Xia X. Domains and functions of spike protein in Sars-Cov-2 in the context of vaccine design. Viruses. 2021;13(1):109.
39. Gupta D, Sharma P, Singh M, Kumar M, Ethayathulla A, Kaur P,et al. Structural and functional insights into the spike protein mutations of emerging SARS-CoV-2 variants. Cellular and Molecular Life Sciences. 2021;78(24):7967-89.
40. Mahajan S, Kode V, Bhojak K, Karunakaran C, Lee K, Manoharan M, et al. Immunodominant T-cell epitopes from the SARS-CoV-2 spike antigen reveal robust pre-existing T-cell immunity in unexposed individuals. Sci Rep. (2021) 11:13164. doi: 10.1038/s41598-021-92521-4
41. Wang C, Liu Z, Chen Z, Huang X, Xu M, He T, et al. The establishment of reference sequence for SARS-CoV-2 and variation analysis. Journal of medical virology. 2020;92(6):667-74.
42. Zhang L, Jackson CB, Mou H, Ojha A, Peng H, Quinlan BD, et al. SARS-CoV-2 spike-protein D614G mutation increases virion spike density and infectivity. Nature communications. 2020;11(1):19.
43. 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.
44. Zhang J, Cai Y, Xiao T, Lu J, Peng H, Sterling SM, et al. Structural impact on SARS-CoV-2 spike protein by D614G substitution. Science. 2021;372(6541):525-30.
45. Ahmad L. Implication of SARS-CoV-2 immune escape spike variants on secondary and vaccine breakthrough infections. Frontiers in immunology. 2021:4563.
46. Korber, B., Fischer, W. M., Gnanakaran, S., Yoon, H., Theiler, J., Abfalterer, W., ... & Montefiori, D. C., et al. Tracking changes in SARS-CoV-2 spike: evidence that D614G increases infectivity of the COVID-19 virus. *Cell*, 2020, 182.4: 812-827. e19.
47. Li Q, Wu J, Nie J, Zhang L, Hao H, Liu S, et al. The impact of mutations in SARS-CoV-2 spike on viral infectivity and antigenicity. Cell. 2020;182(5):1284-94. e9.
48. Shah M, Woo HG. Omicron: a heavily mutated SARS-CoV-2 variant exhibits stronger binding to ACE2 and potently escapes approved COVID-19 therapeutic antibodies. Frontiers in immunology. 2022;12:6031.
49. Barton MI, MacGowan SA, Kutuzov MA, Dushek O, Barton GJ, Van Der Merwe PA,et al. Effects of common mutations in the SARS-CoV-2 Spike RBD and its ligand, the human ACE2 receptor on binding affinity and kinetics. Elife. 2021;10:e70658.
50. Greaney, A. J., Loes, A. N., Crawford, K. H., Starr, T. N., Malone, Greaney, A. J., Loes, A. N., Crawford, K. H., Starr, T. N., Malone, et al. Comprehensive mapping of mutations in the SARS-CoV-2 receptor-binding domain that affect recognition by polyclonal human plasma antibodies. *Cell host & microbe*, 2021, 29.3: 463-476. e6.
51. Khateeb J, Li Y, Zhang H. Emerging SARS-CoV-2 variants of concern and potential intervention approaches. Critical Care. 2021;25(1):1-8.
52. Barrett CT, Neal HE, Edmonds K, Moncman CL, Thompson R, Branttie JM, et al. Effect of mutations in the SARS-CoV-2 spike protein on protein stability, cleavage, and cell-cell fusion function. bioRxiv. 2021.
53. Harvey, W. T., Carabelli, A. M., Jackson, B., Gupta, R. K., Thomson, E. C., Harrison, E. M., ... & Robertson, D. L.., et al. SARS-CoV-2 variants, spike mutations and immune escape. *Nature Reviews Microbiology*, 2021, 19.7: 409-424.
54. Mishra SK, Tripathi T. One year update on the COVID-19 pandemic: Where are we now? Acta tropica. 2021;214:105778.
55. Guruprasad K. Geographical distribution of amino acid mutations in human SARS-CoV-2 orf1ab poly-proteins compared to the equivalent reference proteins from China. ChemRxiv. (2021). 10.33774/chemrxiv-2021-lf2zd-v2
56. Graham RL, Sparks JS, Eckerle LD, Sims AC, Denison MR,et al. SARS coronavirus replicase proteins in pathogenesis. Virus research. 2008;133(1):88-100.
57. Thomas S. Mapping the nonstructural transmembrane proteins of severe acute respiratory syndrome coronavirus 2. Journal of Computational Biology. 2021;28(9):909-21.
58. Wolff G, Melia CE, Snijder EJ, Bárcena M,et al. Double-membrane vesicles as platforms for viral replication. Trends in microbiology. 2020;28(12):1022-33.
59. Diao B, Wen K, Zhang J, Chen J, Han C, Chen Y, et al. Accuracy of a nucleocapsid protein antigen rapid test in the diagnosis of SARS-CoV-2 infection. Clinical Microbiology and Infection. 2021;27(2):289. e1-. e4.
60. Gao T, Gao Y, Liu X, Nie Z, Sun H, Lin K, et al. Identification and functional analysis of the SARS-COV-2 nucleocapsid protein. BMC microbiology. 2021;21(1):1-10.

61. Ramesh S, Govindarajulu M, Parise RS, Neel L, Shankar T, Patel S, et al. Emerging SARS-CoV-2 variants: a review of its mutations, its implications and vaccine efficacy. Vaccines. 2021;9(10):1195.
62. Washington NL, Gangavarapu K, Zeller M, Bolze A, Cirulli ET, Barrett KMS, et al. Emergence and rapid transmission of SARS-CoV-2 B. 1.1. 7 in the United States. Cell. 2021;184(10):2587-94. e7.
63. Jung, C.; Kmiec, D.; Koepke, L.; Zech, F.; Jacob, T.; Sparrer, K.M.J.; Kirchhoff, F. Omicron: What Makes the Latest SARS-CoV-2 Variant of Concern So Concerning? *J. Virol.* **2022**, *96*, e02077-21.
64. de Silva, T.I.; Liu, G.; Lindsey, B.B.; Dong, D.; Moore, S.C.; Hsu, N.S.; Shah, D.; Wellington, D.; Mentzer, A.J.; Angyal, A.; et al. The Impact of Viral Mutations on Recognition by SARS-CoV-2 Specific T Cells. *iScience* **2021**, *24*, 103353
65. Washington N.L., Gangavarapu K., Zeller M., Bolze A., Cirulli E.T., Schiabor Barrett K.M., Larsen B.B., Anderson C., White S., Cassens T., et al. Emergence and rapid transmission of SARS-CoV-2 B.1.1.7 in the United States. *Cell.* 2021;184:2587–2594.e7.
66. Gordon D.E., Hiatt J., Bouhaddou M., Rezelj V.V., Ulferts S., Braberg H., Jureka A.S., Obernier K., Guo J.Z., Batra J. Comparative host-coronavirus protein interaction networks reveal pan-viral disease mechanisms. *Science.* 2020;370:eabe9403
67. Kannan, S. R., Spratt, A. N., Sharma, K., Chand, H. S., Byrareddy, S. N., & Singh, K..Omicron SARS-CoV-2 variant: Unique features and their impact on pre-existing antibodies. J Autoimmun. 2022 Jan; 126: 102779.