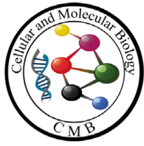
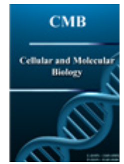
Cell. Mol. Biol. (ISSN: 1165-158X)



Journal Homepage: www.cellmolbiol.org

Cellular and Molecular Biology



Original Article

Dominance of SARS-CoV-2 Delta AY.33 sublineage and Omicron BA.1.1 sublineage in Erbil city/Kurdistan region of Iraq

Sazan Moffaq Abdulaziz[1*], Asmaa Ameen Ghareeb[2], Mohammed Omar Rahman[3], Sayran Hamad Haji[4]

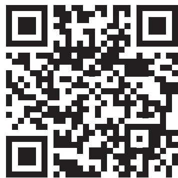
[1] Department of Medical Laboratory Technology, Erbil Technical Health and Medical College, Erbil Polytechnic University, Erbil, Kurdistan-Region, Iraq
[2] Anesthesia Technology Department, College of Science, Cihan University-Erbil, Kurdistan Region, Iraq
[3] Medical Microbiology Department, College of Science, Cihan University-Erbil, Kurdistan Region, Iraq
[4] Department of Clinical Analysis, College of Pharmacy, Hawler Medical University, Erbil, Iraq

Article Info

Article history:

Received: February 11, 2024
Accepted: June 17, 2024
Published: October 31, 2024

Use your device to scan and read the article online



Abstract

This study aimed to analyze the genetic characteristics of a sample of SARS-CoV-2 strains circulated in Erbil City from the 15th of October 2021 and the 5th of January 2022 focusing on their evolutionary feature including lineages, sublineages and clades. Following confirmation of the SARS-CoV-2 positivity of throat and nasopharyngeal swab specimens using qRT-PCR, 20 RNA extracts were subjected to NGS of the S gene and analysis in which only 12 matched the criteria of good sequences. Later, alignment was done with WIV04 reference sequence from Wuhan applying a number of bioinformatics tools. Then, based on sequences recorded in EpiCoV database/GISAID, related genomes to our sequences were identified. The PANGO system revealed that out of the 12 sequences, 10 were Delta (B.1.617.2) variants and two were Omicron (B.1.1.529). Seven out of 10 Delta sequences belonged to AY.33 sublineage and 2 were AY.4. Both Omicron sequences belonged to BA.1.1 sublineage. All Delta sequences belonged to the 21J Nextstrain subclade, meanwhile, both Omicron sequences were from 21K. Spike protein mutations in Delta variant varied, some were sublineage-specific, and others were unique, however, mutations generally were found in the N-terminal domain. Omicron variant appeared with 33 mutations, most of which were in the receptor-binding domain. On the whole, related sequences to our sequences were from Germany, the USA, Denmark, the UK, Iraq, Turkey and several other countries. These findings could provide insights into SARS-CoV-2 evolution nature and significant impact of amino acid changes in the spike protein on disease pathogenicity and emphasize the demand for continuous genomic surveillance globally.

Keywords: SARS-CoV-2, Spike protein, COVID-19, Phylogeny, Sublineages, Clades, Kurdistan.

1. Introduction

Since 2019, a new illness named coronavirus disease 19 (COVID-19) has been established to be caused by an enveloped RNA virus that belongs to coronaviridae family; the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) [1]. The disease has been characterized by different ranges of respiratory symptoms from mild to severe. COVID-19 pandemic with its huge number of morbidities and mortalities resulted in a worldwide health problem within a very short period. Similar to other RNA viruses, SARS-CoV-2 RNA genome underwent continuous mutation that impacted disease pathogenicity. The genome codes for a number of structural and non-structural proteins [2,3]. One of the genes is the S gene; a highly mutated gene that codes for SARS-CoV-2 Spike protein with a total length of 1237 amino acids (aa). It consists of a signal peptide located at the N-terminus (1-13 aa), the S1 subunit (14-685 aa) responsible for binding to angiotensin-converting enzyme 2 receptor (ACE2), and the S2 subunit (686-1273 aa) responsible for membrane fusion. The S1 subunit is composed of two domains; the N-terminal (NTD) (14-305 aa) and the receptor-binding (RBD) (319-541). Regions that build up the S2 subunit include the fusion peptide (FP) (788-806), heptapeptide repeat sequence 1 and 2 (HR1 and HR2) (912-984aa and 1163-1213aa length, respectively), transmembrane domain (TM) (1213-1237aa) and cytoplasmic domain (CM) (1237-1273 aa) [4,5]. Clinically, altered S protein has been reported to greatly affect virus pathogenicity, host immunity, antiviral therapy and vaccine efficacy [6,7]. In 2020, the World Health Organization (WHO) established a classification for SARS-CoV-2 into variants of interests (VOIs) and variants of concerns (VOCs) [8]. Two of the VOCs relatively received much attention; the Delta (B.1.617.2 lineage) and the Omicron (B.1.1.529 lineage). The Delta variant arose during the third wave in India in October 2020. This variant showed increased resistance to neutralizing antibodies and relatively increased expression of the S glycoprotein and improved binding with the ACE2 receptor [9]. On the other hand, the Omicron variant originated from South Africa

* Corresponding author.
E-mail address: sazan.abdulaziz@epu.edu.iq (Sazan Moffaq Abdulaziz).
Doi: http://dx.doi.org/10.14715/cmb/2024.70.10.12

in late 2021. It rapidly replaced the delta variant during the fourth wave of SARS-CoV-2 infection [10, 11]. It has been characterized by excellent binding to ACE2 receptors and better transmission rates, but relatively lower rates of hospitalization and death compared to Delta variant [12,13]. Considering the genetic evolution and classification, three subclades of the Delta variant were identified (21A, 21I and 21J) and multiple Omicron subclades (21K, 21L, 21M, 22A-F and 23A-F) [14]. As the Delta lineage B.1.617.2 diversified, a dynamic classification into sublineages using AY.X was established generating 129 sublineages (AY.1-AY.133) and multiple Omicron B.1.1.529 sublineages designated as BA.X and recently XBB according to PANGO (Phylogenetic Assignment of Named Global Outbreak**)** dynamic nomenclature system [15]. Nearly 16.1 million SARS-CoV-2 genome sequences are readily available in EpiCoV database from GISAID (Global Initiative on Sharing All Influenza Data) [16]. Actually, it is significant to address the molecular and epidemiological features of this virus because of its highly mutated feature especially the S gene, and its clinical consequences. This study was conducted to characterize the S gene of a number of SARS-CoV-2 isolates through NGS sequencing and analysis in order to highlight mutation profile and identify possible clades, lineages and sublineages distributed in Erbil City/Kurdistan Region of Iraq during the pandemic. In addition, the possible genetic relatedness of each identified sequence to local and international sequences was investigated.

2. Materials and methods

2.1. Sample collection

Nasopharyngeal and throat swab specimens were obtained from hospitalized and non-hospitalized patients having signs and symptoms of COVID-19 from Erbil Public Health Laboratory or admitted to the three COVID-19-specific hospitals; Al Emarati, West Erbil Emergency and Lalav Hospitals in Erbil/Kurdistan Region of Iraq between the 15th of October 2021 and the 5th of January 2022. Specimens were collected in vials containing special viral transport medium (VTM) and transported to the lab. according to special guidelines distributed by the WHO.

2.2. RNA extraction and qRT-PCR

In the laboratory, SARS-CoV-2 detection was carried out using SARS-CoV-2 nucleic acid detection and extraction kit (Zybio/China); both RNA extraction and qRT-PCR-based virus detection system were integrated into a single kit. Following confirmation of the SARS-CoV-2 positivity, 20 RNA extracts (Ct value lower than 25) were selected for next-generation sequencing (NGS) because of the cost-effectiveness of this procedure as the study has not been funded by any company or organization.

2.3. Next-generation sequencing

As NGS was not available in Iraq at the time of the study, samples were sent to the Intergen Genetic Diagnosis and Research Center in Ankara/Turkey on dry ice. There, all the 20 RNA extracts were re-checked for RNA integrity. Using a standard protocol of Ipsogen RT kit (Qiagen/Germany), RNA extracts were reverse-transcribed to cDNA and stored at -20°C until processing. Primer Designer V.2.0 (Scientific& Educational Software) was used to create specific primers to amplify SARS-CoV-2 S gene. Then, the cDNAs ran in a 2% agarose gel and the obtained DNA bands were purified using NucleoFast® 96 PCR clean-up kit (Macherey-Nagel GmbH/Germany) and quantified using Nanodrop 1000 micro-volume spectrophotometer (Thermo Inc./USA). Later, NGS and analysis were carried out in which samples were first prepared for NGS using Nextera XT sample preparation kit (Illumina Inc./USA), then the sequencing process was carried out using MiSeq (Illumina Inc/ USA) according to protocols fixed in the kit.

2.4. Bioinformatics analysis

The sequencing reads were aligned and assembled using the WIV04 reference sequence from Wuhan and the BBMap alignment algorithm. Meanwhile, the process of variant calling and mutation detection against the reference genome were executed utilizing the Genome Analysis Toolkit (GATK). Furthermore, the annotation of the assembled sequences conducted using SnpEff. Prior to approval, quality control measures were applied to all sequence readings. For further analysis, we chose sequences that had a genome sequencing coverage of over 99% and a gap length of less than 40 base pairs. Finally, only 12 S gene sequences successfully passed quality control measures and were submitted later to the GISAID database. The GISAID database provided sequence accession numbers to all the 12 genome sequences; EPI_ISL_18139226, PI_ISL_18139227, EPI_ISL_18142862, EPI_ISL_18142863, EPI_ISL_18142864, EPI_ISL_18142865, EPI_ISL_18142866, EPI_ISL_18142867, EPI_ISL_18142868, EPI_ISL_18142869, EPI_ISL_18142870, EPI_ISL_18142871

In the present study, the AudacityInstant (v5.1.0) program was utilized, a specialized tool designed for doing comprehensive searches across the GISAID EpiCoV database. The use of such tool served a crucial role by effectively finding closely associated sequences in collected in nearly similar dates offering essential metadata for each recorded sequence. The metadata given encompassed essential details, including clade, lineage, location, variant, and collection date.

2.5. Lineage and phylogenetic analysis

The developmental origins of the 12 SARS-CoV-2 S gene sequences were studied. Lineage identification was performed using Pangolin system (v3.1.14). Moreover, clades were determined using the Nextclade sequence analysis and the GISAID database tools. In addition, an attempt to find out the most closely related sequence to each sequence under study was done including sequences from the neighboring countries (Turkey, Saudi Arabia, Jordan, Iran, Kuwait and Syria) and a phylogenetic tree was generated using the Neighbor-Joining method as implemented in Molecular Evolutionary Genetics Analysis (MEGA) version 11.

3. Results

Twelve nasopharyngeal and throat swab samples collected from COVID-19 patients in Erbil City/Kurdistan Region of Iraq between the 15th of October 2021 and the 5th of January 2022 were subjected to primary identification of SARS-CoV-2 using qRT-PCR followed by the S gene sequencing using NGS method. Following bioinformatics analysis with the aid of EpiCoV database/GISAID and alignment of the obtained sequences with a reference

sequence from Wuhan (WIV04), it was observed that ten of the sequences were Delta (B.1.617.2) variant and two were Omicron (B.1.1.529).

3.1. PANGO lineage distribution of the sequences and related genomes

All the Delta variant S gene sequences in this study were from the B.1.617.2 PANGO lineage which was further diversified into three sublineages; sublineage AY.33 was the most predominant lineage (7 out of 10), sublineage AY.4 (2 out of 10) and a single AY.36. Both Omicron variants were found to be from B.1.1.529 lineage and BA.1.1 sublineage.

On the whole, among the 16.1 million SARS-CoV-2 genomes in the EpiCoV database, 868 unique genomes were related or closely related to the Delta and 124 to the Omicron sequences with distances ranging from 0 to 3 and a match quality of 0.95 or higher. Generally, the related sequences to the Delta variants were mostly collected in Germany, USA, Denmark, UK, Iraq, Turkey and several other countries between the 21st of July 2021 and the 15th of December 2021, whereas the related sequences to the Omicron variants were mostly collected in UK, USA, Germany, Thailand, Japan, UAE and several other countries between the 21st of December 2021 and the 15th of March 2022 (Figure 1).

When the geographical distribution of the related genomes to each Delta sublineage was considered separately, new facts were observed (Figure 2). The most frequent related genomes for both AY.4 sublineages were mostly collected from the USA and the UK with proportions ranging from 46.6 to 52.5%. As mentioned earlier, AY.33 was the most frequent sublineage among the sequenced samples, but there was a great variation in the countries where the related genomes were identified. For instance, the most related genomes to AY.33 from sequence 2 (S2) were mostly identified in USA (46.6%) followed by the UK and Sweden. Exceptionally, the most related genomes to AY.33 from S3 were mostly isolated from Iraq (65.3%), specifically, the Kurdistan region of Iraq (as mentioned by EpiCoV database) followed by Estonia and Turkey. Moreover, the source of the related genomes to AY.33 from samples 5-8 were nearly similar as they were mostly identified in Denmark, Germany, Netherland and Turkey. Sweden was the main source of the related genomes to AY.33 from S10 with a proportion of 76%! Finally, the related genomes to the only AY.36 sublineage (S4) in this study were identified from many other countries at nearly similar rates such as the UK, Sweden, the USA, Germany, Denmark and Brazil.

3.2. Mutations in Delta sublineages

Figure 3A represents all mutations identified in Delta sublineages in this study following alignment with Wuhan (WIV04) wild-type SARS-CoV-2. Generally, different mutation classes have been observed in SARS-CoV-2 Delta sublineages including nonsynonymous single nucleotide variations (SNV), synonymous SNVs and a non-frameshift deletion. Surprisingly, the same sublineages showed variation in mutation type and number. For example, despite that samples 1 and 9 (S1 and S9) that belong to sublineage AY.4 shared 9 mutations namely; T19R, T95I, G142D, 156-158del, L452R, T478K, D614G, P681R and D950N, but N657N was found only in S9, whereas D80Y, L229L and F543F were specific to S1. The same observation was correct for sublineage AY.33 (S2, S3, S5-8 and S10) in which ten mutations shared among the sequences

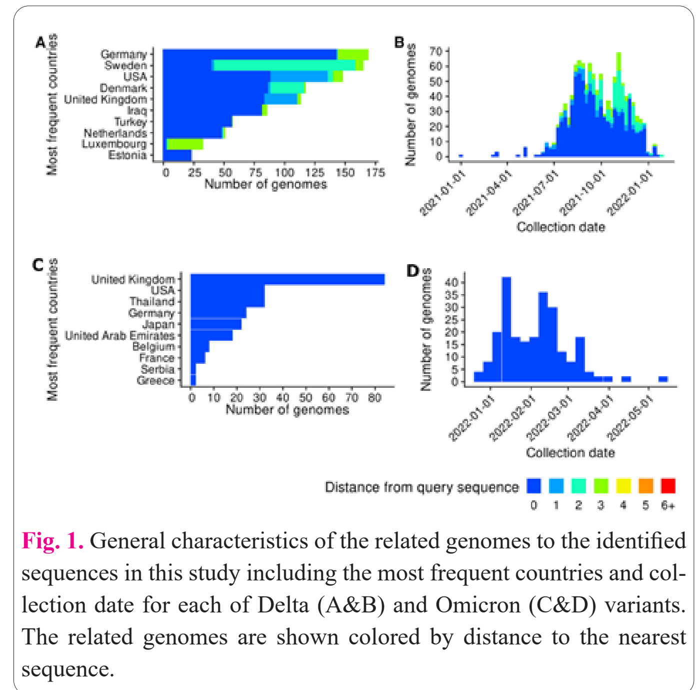


Fig. 1. General characteristics of the related genomes to the identified sequences in this study including the most frequent countries and collection date for each of Delta (A&B) and Omicron (C&D) variants. The related genomes are shown colored by distance to the nearest sequence.

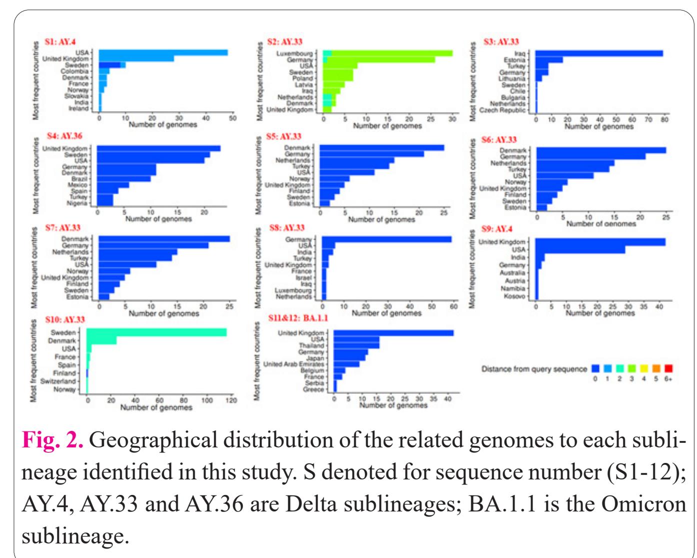


Fig. 2. Geographical distribution of the related genomes to each sublineage identified in this study. S denoted for sequence number (S1-12); AY.4, AY.33 and AY.36 are Delta sublineages; BA.1.1 is the Omicron sublineage.

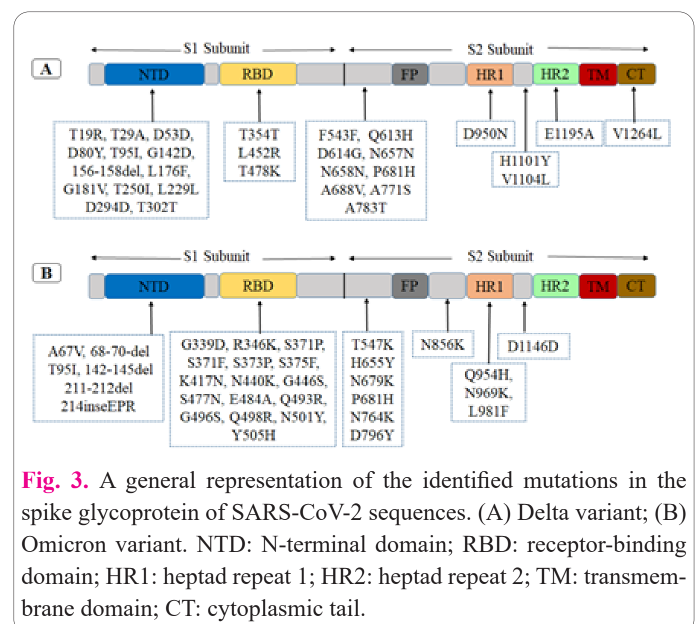


Fig. 3. A general representation of the identified mutations in the spike glycoprotein of SARS-CoV-2 sequences. (A) Delta variant; (B) Omicron variant. NTD: N-terminal domain; RBD: receptor-binding domain; HR1: heptad repeat 1; HR2: heptad repeat 2; TM: transmembrane domain; CT: cytoplasmic tail.

namely; T19R, T29A, G142D, 156-158del, T250I, L452R, T478K, D614G, P681R and D950N. However, S gene sequences from S2 and S10 express three additional but different SNVs (D294D, T345T and E1195A in S2 and L176F, Q613H and N658H in S10), meanwhile, sequences from S3, S5 and 6 had an extra SNV (G181V, T302T and A783T, respectively). The mutation set of the single AY.36 sublineage (S4) composed of T19R, T95I, G142D, 156-158del, L452R, T478K, D614G, P681R, D950N and V1104L.

3.3. Mutations in Omicron sublineage

SARS-CoV-2 S gene sequences 11 and 12 were identified as Omicron BA.1.1 sublineage and were identical. Actually, the Omicron variant was known with its large and unusual number of mutations. Thirty-three mutations were identified in both sequences following alignment with Wuhan (WIV04) wild type SARS-CoV-2 namely; A67V, 68-70-del, T95I, 142-145del and 211-212del, 214inseEPR, G339D, R346K, S371P, S371F, S373P, S375F, K417N, N440K, G446S, S477N, E484A, Q493R, G496S, Q498R, N501Y, Y505H, T547K, H655Y, N679K, P681H, N764K, D796Y, N856K, Q954H, N969K, L981F and D1146D (Figure 3B). On the whole, mutations were mostly localized in the RBD (Figure 3B).

3.4. Phylogenetic analysis of SARS-CoV-2 Sequences

Phylogenetic analysis of the studied sequences showed that all the Delta (B.1.617.2) variant sequences belonged to the 21J Nextstrain Delta subclade and both sequences of Omicron variant (B.1.1.529) were from 21K Nextstrain Omicron subclade (Figure 4A). Yet, phylogeny of the genome sequences to the nearest sequence was also performed in which the branching indicates the evolutionary differences (distances) among the related sequences (Figure 4B). Generally, Delta sequences were variable in their origin. Only one sequence (S2) was shown to be closely related to a sequence from Iraq whereas all the other sequences were found to group with sequences from other countries such as Sweden, Finland and the USA. Among the neighboring countries, sequences from Turkey and Saudi Arabia showed close proximity. The same phylogeny showed a great evolutionary distance between the isolated sequences and the Wuhan (WIV04) SARS-CoV-2 wild type. On the other hand, both Omicron sequences were almost grouped with sequences from the USA.

4. Discussion

In this study, molecular characterization of 12 SARS-CoV-2 sequences through S gene sequencing and analysis using a number of bioinformatics tools and EpiCoV/GISAID database. Nasopharyngeal and throat swab specimens were collected in Erbil City/Kurdistan Region of Iraq from the 15th of October 2021 and the 5th of January 2022. The results showed that among the studied sequences, 10 identified as Delta (B.1.617.2 lineage) and two as Omicron (B.1.1.529 lineage) variants; it seems that it was the period of transition from Delta to Omicron variant in Erbil City during the study period. Infection with Delta and Omicron variants and their sublineages expanded rapidly in large number of countries across the globe via international travelling in spite of the restricted instructions forced by most of the countries [17]. As it was evident in the results (Figures 1 and 2), 11 out of the 12 studied sequences were closely related to sequences from different European countries and the USA indicating numerous external viral introducing sources during the study period. A single sequence (sublineage AY.33 from Sequence 3) could be regarded as the domestic sequence as 65.3% of the related genomes were recorded in Iraq and the Kurdistan Region.

Dynamic classification of Delta B.1.617.2 lineage into sublineages was established within the PANGO classification system using "AY.X" [15]. Sublineages were characterized by occurrence of specific mutations and distinct geographical distribution and infection properties [18]. Among Delta sequences, 7 out of 10 belonged to AY.33 sublineage. This sublineage appeared in June 2021 and increased in prevalence worldwide, comprising the main sublineage being recorded in the Kurdistan Region of Iraq, Denmark, Germany, Sweden, the Netherland and Turkey until the end of December 2021 [19]. Another sublineage; AY.4, was detected among Delta sequences. This sublineage was highly recorded in July and August 2021 (nearly 40% of all Delta sequences in EpiCoV database), however, a relatively lower rate (10.9%) was reported by Yadouleton et al. (2021) [20], highlighting the fact that SARS-CoV-2 infection is challenging to fit within a

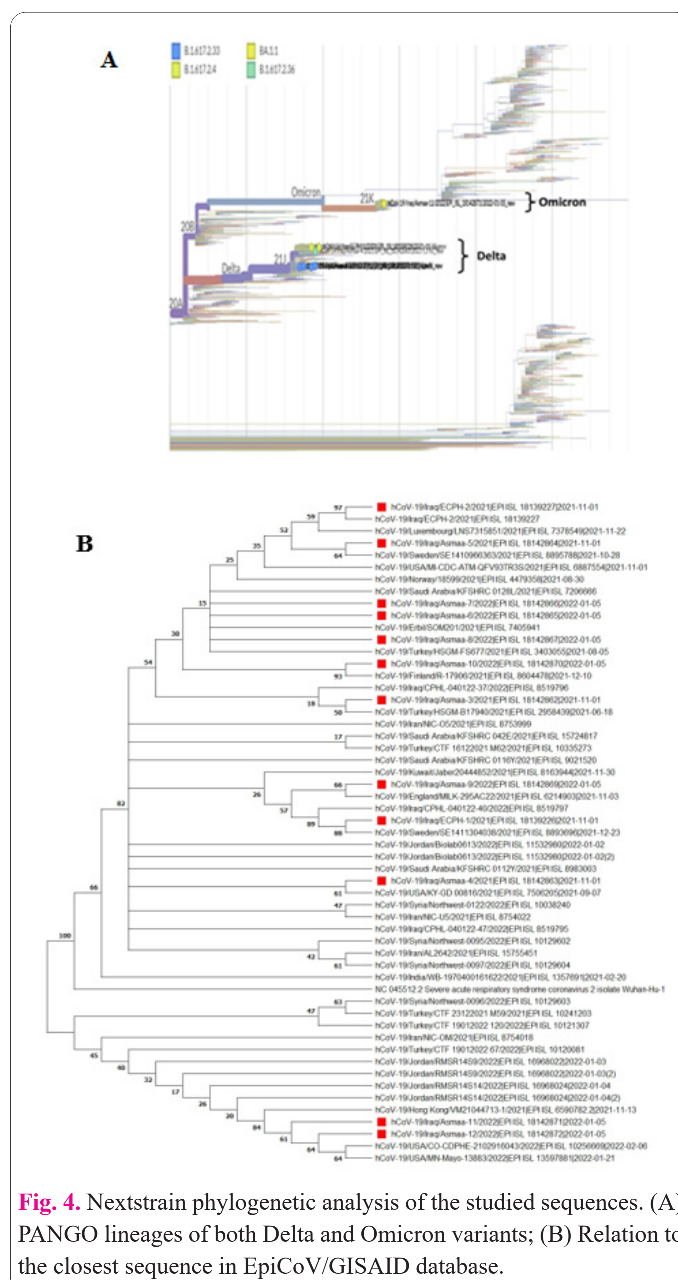


Fig. 4. Nextstrain phylogenetic analysis of the studied sequences. (A) PANGO lineages of both Delta and Omicron variants; (B) Relation to the closest sequence in EpiCoV/GISAID database.

specific regulation. The prevalence rate of AY.4 obviously declined (2.5%) by the end of the year. Yet, a single Delta AY.36 sublineage was identified in our study in which the related genomes were reported in the UK, Sweden, the USA, Germany, Denmark and Brazil. Globally, AY.36 was a rare sublineage identified only in 11 countries in very low rates (less than 5% of all Delta sequences in GISAID). In a noteworthy, Ozer et al. (2022) reported 85% prevalence rate of this sublineage in their study in Oyo state/ Nigeria [21].

Numerous sublineages of Omicron B.1.1.529 lineage evolved and were readily recorded in EpiCoV database. Both identified sequences of the Omicron variants in the present study belonged to BA.1.1 sublineage. The BA.1 and BA.1.1 were the first and the most prevalent sublineages globally at the beginning of the fifth wave of COVID-19, but were quickly replaced by the BA.2 sublineage in March 2022 [22,23].

S gene mutations were identified following alignment with the reference sequence. On the whole, Delta sublineages were variable in their mutation profiles., but carried the unique set of mutations; L452R, T478K, and P681R reported by the WHO as signature mutations of Delta variant [8]. L452R and T478K are located in the RBD and found to enhance the interaction with the ACE2 receptor, virus replication, transmission and immune escape [24,25]. P681R generated as proline replaced by arginine at position 681 near S1/S2 cleavage site. This nonsynonymous mutation was found to greatly enhance virus replication and transmissibility [26] and partially decrease neutralizing antibody binding affinity as well [27]. Yet, all Delta sublineages shared some additional nonsynonymous mutations including T19R, T95I, G142D, D614G and D950N, in addition to 156-158 deletion. The same mutation set was also reported by Ghareeb and Abdulaziz (2023) [28]. NTD is a target site for neutralizing Antibodies (Abs) such as 4A8. It was observed that T19R and G142D mutations interfere with Ab binding, whereas 156-158 deletion contributes to the vaccine-escape abilities of the virus by changing the NTD structure [29,30]. T95I originated in Eta (B.1.525) and Iota (B.1.526) VOIs and structural analysis revealed that this mutation might alter the sidechain conformation to weaken the interaction with antibodies [31]. Interestingly, the asparagine-to-glycine substitution at amino acid position 614 (D614G) was found in all SARS-CoV-2 sequences in this study, including Omicron variant. This mutation was one of the most critical mutations that appeared in Alpha variant and conserved in other VOCs and VOIs, but the detection rate was variable. Globally, around 80% of SARS-CoV-2 sequences were found to possess this mutation. Studies revealed that this mutation dramatically boosts replication potential of the virus thereby increasing the transmission rate [32,33] without any link, however, with severe clinical outcomes [34]. Within the same group mentioned above is D950N. it is located near S1/S2 cleavage site and enhanced virus replication and transmissibility as P681R, however, P681R in particular, inflicted more advantages to the virus as it enhanced S1/S2 cleavage more efficiently. D950N was also linked to decreased binding potential of neutralizing antibodies [35]. Considering every sublineage separately, new facts are disclosed. For instance, all the AY.33 genomes carried two additional mutations; T29A and T250I, in the NTD, both mutations were specified to sublineage AY.33 by Mazouri et al. (2022). The substitution of threonine to alanine at position 29 and to isoleucine at positions 250 in the NTD resulted in a change in the secondary structure of the spike protein which might have functional consequences [36]. Moreover, sequence-specific mutations were also observed within this sublineage. One of the significant mutations in this group was G181V in the NTD which was specifically found in Sequence 3; the sequence was dominant in Iraq including Kurdistan Region at that time and was detected in less than 50% of the related genomes globally. G181V significantly affected ACE2 binding thereby increasing the infectivity [37]. In the RBD, little is about its impact of a rare mutation Q613H, however, given its proximity to D614G, it was suggested to enhance virus transmissibility along with the latter mutation [38]. Conversely, mutations like A783T and E1195A were only observed in sequences 6 and 2, respectively, but nothing is known about their role in virus pathogenesis. Interestingly, both AY.4 sequences showed nearly similar mutation pictures, except for D80Y mutation which was found only in one sequence and it was recorded in fewer than 50% of the related genomes in EpiCoV database. Scarce studies mentioned the role of this mutation, but Li et al. (2023) stated its role in increasing the infectivity of SARS-CoV-2 [39]. The single genome identified as AY.36 Delta sublineage appeared with single-specific mutation in the C-terminus of the S2 subunit; V1104L. the same mutation was fixed in AY.20, AY.22 and AY.31 Delta sublineages as well, but the functional role of this mutation has not been established yet [21]. Gathering all the previous facts together, one can conclude that variation in these nonsynonymous mutations in Delta sublineages is the basis of first, variation in the clinical picture of COVID-19 from mild to severe, and second, the high morbidity and mortality rates associated with infection with Delta variant.

Yet, both sequences of Omicron BA.1.1 sublineage carried the same number (33 mutations) and set of mutations in the S gene, in which 16 mutations located in the RBD namely; G339D, R346K, S371P, S371F, S373P, S375F, K417N, N440K, G446S, S477N, T478K, E484A, Q493R, G496S, N501Y, and Y505H. Remarkably, prior studies documented each of the aforementioned alterations and were found to increase spike protein-ACE2 binding potential thereby enhancing virus infectivity and transmissibility. However, they differ in their binding affinity, for instance, Q493R and N501Y had the highest binding affinity to ACE2 whereas N440K and E484A conferred the lowest binding affinity [25,26,40]. K417N was documented previously as the signature mutation of BA.1.1 sublineage [23]. The triple mutation set; K417N, E484A and N501Y, was found to be significantly related to immunological escapes and evasion from neutralizing antibodies [41]. In addition, mutations in the NTD included A67V, 68-70-del, T95I, 142-145del, 211-212del, 214inseEPR were found to be less advantageous for the virus as reported by Kumar et al. (2022) in their extensive analysis of the spike protein of different Omicron sublineages and found that in contrast to the RBD, mutations in the NTD resulted in a negative electrostatic potential indicating that Omicron variant bound less efficiently to ACE2; a possible interpretation for Omicron's decreased severity and hospitalization [40]. It's still unclear how the Omicron sublineages appeared with this unusual number of mutations in the spike pro-

tein. Since this variation was originally identified in immunocompromised patients in South Africa, there was less selection pressure on the virus, and the extended time of infection may have contributed to such a significant evolution [42,43].

In this study, the phylogenetic analysis indicated that all the Delta variant sequences belonged to 21J Nextstrain Delta subclade and both sequences of Omicron variant were from 21K Nextstrain Omicron subclade (Figure 4A). On the whole, Delta variant comprised 3 subclades; 21A, 21I and 21J, in which subclade 21J was the most prevalent clade which dominated until the end of the COVID-19 third wave [44,45]. By the end of 2021, Omicron variant emerged and the first clade; 21K, was named. This clade comprised BA.1 and its sublineages; the most prevalent Omicron sublineage [46]. The phylogenetic analysis, in addition, revealed that the closest sequence to our sequences (except for sequence 2 which matched sequences from Iraq) were from Sweden, Finland, the USA, Saudi Arabia and Turkey, indicating that these countries might be the basic sources of the SARS-CoV-2 found in Erbil during the study period.

In conclusion, it seems that the study period was the period of transition from the third to the fourth wave of COVID-19 in Erbil City because both Delta and Omicron variants were identified among the sequences. AY.33 was the prevalent Delta sublineage that included a domestic sequence. BA.1.1 sublineage of Omicron was the first sublineage identified in the study area. Variations in the identified sublineages and their related genomes from EpiCoV/GISAID database were greatly attributed to the main role of public transport in virus transmission. The rapidly mutated genome of the virus, especially the S gene, interprets variable mutation patterns that greatly impact virus transmission, infectivity, and immune and vaccine escape abilities. Comprehending the temporal patterns of mutation and molecular characterization of SARS-CoV-2 would be beneficial for clinical purposes, public health planning, and control strategy development. Similar studies are continuously needed to understand the molecular characteristics of the virus and surveillance of the circulating strains locally and globally.

Conflict of interests

The authors have no conflicts with any step of the article preparation.

Ethics approval and consent to participate

All procedures performed in this study were in accordance with the institutional and/or national research committee's ethical standards and with the 1964 Helsinki Declaration and its later amendments or comparable ethical standards.

Informed consent

Informed consent was obtained from all individual participants included in the study.

Availability of data and material

Spike gene sequences of SARS-CoV-2 of this study were deposited in the Global Initiative on Sharing all Individual Data (GISAID).

Acknowledgments

We greatly acknowledge the directorate of Erbil Public Health Laboratory, Al Emarati, West Erbil Emergency and Lalav Hospitals in Erbil/Kurdistan Region of Iraq for their incredible assistance.

Authors' contributions

Asmaa Ameen Ghareeb: Sample collection and laboratory procedures and manuscript writing, Sazan Moffaq Abdulaziz: Research design, supervision, manuscript writing and editing, Mohammed Omar Rahman, bioinformatics analysis, Sayran Hamad Haji, manuscript writing and editing.

Funding

None.

References

1. Challen R, Dyson L, Overton CE, Guzman-Rincon LM, Hill EM, Stage HB, et al (2021) Early epidemiological signatures of novel SARS-CoV-2 variants: establishment of B.1.617.2 in England. medRxiv. doi: 10.1101/2021.06.05.21258365.
2. COVID-19 Omicron Delta study group (2023) Clinical progression, disease severity, and mortality among adults hospitalized with COVID-19 caused by the Omicron and Delta SARS-CoV-2 variants: A population-based, matched cohort study. PloS One;18(4):e0282806. doi: 10.1371/journal.pone.0282806.3.
3. Ghareeb AA, Sazan Moffaq Abdulaziz (2024) Genetic analysis of SARS-CoV-2 spike gene using Next Generation Sequencing from COVID-19 patients in Erbil/Iraq. Cell Mol Biol 70(6):7-13. doi: 10.14715/cmb/2024.70.6.2
4. Ghareeb AA, Abdulaziz SM (2023) Clinical & laboratory markers as predictors for severity and mortality in COVID-19. Polytech J 13(1). https://polytechnic-journal.epu.edu.iq/home/vol13/iss1/20.
5. Huang Y, Yang C, Xu X Feng, Xu W, Liu S (2020) Structural and functional properties of SARS-CoV-2 spike protein: potential antivirus drug development for COVID-19. Acta Pharmacol Sin 41(9):1141–1149. doi: 10.1038/s41401-020-0485-4.
6. Xia S, Zhu Y, Liu M, Lan Q, Xu W, Wu Y, et al (2020) Fusion mechanism of 2019-nCoV and fusion inhibitors targeting HR1 domain in spike protein. Cell Mol Immunol 17(7):765-767. doi: 10.1038/s41423-020-0374-2.
7. Wrapp D, Wang N, Corbett KS, Goldsmith JA, Hsieh CL, Abiona O, et al (2020) Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. Science 13 367(6483):1260–1263. doi: 10.1101/2020.02.11.944462.
8. Wu A, Peng Y, Huang B, Ding X, Wang X, Niu P, et al (2020) Genome Composition and Divergence of the Novel Coronavirus (2019-nCoV) Originating in China. Cell Host Microbe 27(3):325–328. doi: 10.1001/jama.2020.2648.
9. https://www.who.int/activities/tracking-SARS-CoV-2-variants. Tracking SARS-CoV-2 variants. https://www.who.int/activities/tracking-SARS-CoV-2-variants
10. Starr TN, Greaney AJ, Hilton SK, Ellis D, Crawford KHD, Dingens AS, et al (2020) Deep Mutational Scanning of SARS-CoV-2 Receptor Binding Domain Reveals Constraints on Folding and ACE2 Binding. Cell 182(5):1295-1310. doi: 10.1016/j.cell.2020.08.012.
11. Sahoo J, Samal K (2021) World on Alert: WHO Designated South African New COVID Strain (Omicron/B.1.1.529) as a Variant of Concern. Biotica Research Today 3(11): 1086-1088. doi: 10.13140/RG.2.2.25160.03847.
12. Lewnard JA, Hong VX, Patel MM, Kahn R, Lipsitch M, Tartof SY (2022) Clinical outcomes associated with SARS-CoV-2 Omicron (B.1.1.529) variant and BA.1/BA.1.1 or BA.2 subvariant infection in Southern California. Nat Med 28(9):1933–1943. doi:

10.1038/s41591-022-01887-z.

13. Van Goethem N, Chung PYJ, Meurisse M, Vandromme M, De Mot L, Brondeel R, et al (2022) Clinical Severity of SARS-CoV-2 Omicron Variant Compared with Delta among Hospitalized COVID-19 Patients in Belgium during Autumn and Winter Season 2021–2022. Viruses 14(6):1297. doi: 10.3390/v14061297.
14. Wolf JM, Wolf LM, Bello GL, Maccari JG, Nasi LA (2023) Molecular evolution of SARS-CoV-2 from December 2019 to August 2022. J Med Virol 95(1):e28366. doi: 10.1002/jmv.28366
15. PANGO lineages: latest epidemiological lineages. 2023. https://cov-lineages.org/
16. Khare S, Gurry C, Freitas L, Schultz MB, Bach G, Diallo A, et al (2021) GISAID's Role in Pandemic Response. China CDC Wkly 3(49):1049–1051. doi: 10.1080/10903127.2020.1838676.
17. Campbell F, Archer B, Laurenson-Schafer H, Jinnai Y, Konings F, Batra N, et al (2021) Increased transmissibility and global spread of SARS-CoV-2 variants of concern as at June 2021. Eurosurveillance 26(24). doi: 10.2807/1560-7917.ES.2021.26.24.2100509.
18. Rambaut A, Holmes EC, O'Toole Á, Hill V, McCrone JT, Ruis C, et al (2020) A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology. Nat Microbiol 5(11):1403–1407. doi: 10.1038/s41564-020-0770-5.
19. Ntagereka PB, Oyola SO, Baenyi SP, Rono GK, Birindwa AB, Shukuru DW, et al (2022). Whole-genome sequencing of SARS-CoV-2 reveals diverse mutations in circulating Alpha and Delta variants during the first, second, and third waves of COVID-19 in South Kivu, east of the Democratic Republic of the Congo. Int J Infect Dis 122:136–143. doi: 10.1016/j.ijid.2022.05.041.
20. Yadouleton A, Sander AL, Adewumi P, De Oliveira Filho EF, Tchibozo C, Hounkanrin G, et al (2022) Emergence of SARS-CoV-2 Delta Variant, Benin, May–July 2021. Emerg Infect Dis 28(1):205–209. doi: 10.3201/eid2801.211909.
21. Ozer EA, Simons LM, Adewumi OM, Fowotade AA, Omoruyi EC, Adeniji JA, et al (2022) Multiple expansions of globally uncommon SARS-CoV-2 lineages in Nigeria. Nat Commun 13(1):688. doi: 10.1038/s41467-022-28317-5.
22. Fan Y, Li X, Zhang L, Wan S, Zhang L, Zhou F (2022) SARS-CoV-2 Omicron variant: recent progress and future perspectives. Signal Transduct Target Ther 7(1):1–11. doi: 10.1038/s41392-022-00997-x.
23. Chatterjee S, Bhattacharya M, Nag S, Dhama K, Chakraborty C (2023) A Detailed Overview of SARS-CoV-2 Omicron: Its Sub-Variants, Mutations and Pathophysiology, Clinical Characteristics, Immunological Landscape, Immune Escape, and Therapies. Viruses 15(1):167. doi: 10.3390/v15010167.
24. Motozono C, Toyoda M, Zahradnik J, Saito A, Nasser H, Tan TS, et al (2021) SARS-CoV-2 spike L452R variant evades cellular immunity and increases infectivity. Cell Host Microbe 29(7):1124-1136. doi: 10.1016/j.chom.2021.06.006.
25. Pascarella S, Ciccozzi M, Zella D, Bianchi M, Benedetti F, Benvenuto D, et al (2021) SARS-CoV-2 B.1.617 Indian variants: Are electrostatic potential changes responsible for a higher transmission rate? J Med Virol 93(12):6551–6556. doi: 10.1002/jmv.27210.
26. Shiehzadegan S, Alaghemand N, Fox M, Venketaraman V (2021) Analysis of the Delta Variant B.1.617.2 COVID-19. Clin Pract 11(4):778–784. doi: 10.3390/clinpract11040093.
27. Cascella M, Rajnik M, Aleem A, Dulebohn SC, Di Napoli R (2023) Features, Evaluation, and Treatment of Coronavirus (COVID-19). In: StatPearls. Treasure Island (FL): StatPearls Publishing. http://www.ncbi.nlm.nih.gov/books/NBK554776/
28. Ghareeb AA, Abdulaziz SM (2023) Next Generation Sequencing of SARS-CoV-2 Spike Gene from COVID-19 Patients with Characterization of Some Laboratory Parameters [M Sc]. [Erbil]: Erbil Polytechnic University. https://scholar.google.com/citations?view_op=view_citation&hl=en&user=HzicfSMAAAAJ&citation_for_view=HzicfSMAAAAJ:YsMSGLbcyi4C
29. Mishra T, Dalavi R, Joshi G, Kumar A, Pandey P, Shukla S, et al (2022) SARS-CoV-2 spike E156G/Δ157-158 mutations contribute to increased infectivity and immune escape. Life Sci Alliance 5(7):e202201415. doi: 10.26508/lsa.202201415.
30. Shen L, Triche TJ, Bard JD, Biegel JA, Judkins AR, Gai X (2021) Spike Protein NTD mutation G142D in SARS-CoV-2 Delta VOC lineages is associated with frequent back mutations, increased viral loads, and immune evasion. Infectious Diseases (except HIV/AIDS). doi: 10.1101/2021.09.12.21263475
31. Aleem A, Akbar Samad AB, Vaqar S (2023) Emerging Variants of SARS-CoV-2 and Novel Therapeutics Against Coronavirus (COVID-19). In: StatPearls. Treasure Island (FL): StatPearls Publishing. Available from: http://www.ncbi.nlm.nih.gov/books/NBK570580/
32. Korber B, Fischer WM, Gnanakaran S, Yoon H, Theiler J, Abfalterer W, et al (2020) Tracking Changes in SARS-CoV-2 Spike: Evidence that D614G Increases Infectivity of the COVID-19 Virus. Cell 182(4):812-827. doi: 10.1016/j.cell.2020.06.043.
33. Omotuyi IO, Nash O, Ajiboye OB, Iwegbulam CG, Oyinloye EB, Oyedeji OA, et al (2020) Atomistic simulation reveals structural mechanisms underlying D614G spike glycoprotein-enhanced fitness in SARS-COV-2. J Comput Chem 41(24):2158–2161. doi: 10.1002/jcc.26383.
34. Volz E, Hill V, McCrone JT, Price A, Jorgensen D, O'Toole Á, et al (2021) Evaluating the Effects of SARS-CoV-2 Spike Mutation D614G on Transmissibility and Pathogenicity. Cell 184(1):64-75. doi: 10.1016/j.cell.2020.11.020.
35. Lemmin T, Kalbermatter D, Harder D, Plattet P, Fotiadis D (2020) Structures and dynamics of the novel S1/S2 protease cleavage site loop of the SARS-CoV-2 spike glycoprotein. J Struct Biol X 4:100038. doi: 10.1016/j.yjsbx.2020.100038.
36. El Mazouri S, Bendani H, Boumajdi N, Aanniz T, Kandoussi I, Amzazi S, et al (2022) Phylogeography and genomic analysis of SARS-CoV-2 delta variant in Morocco. J Infect Dev Ctries 16(8):1258–1268. doi: 10.3855/jidc.16513.
37. Hanifa M, Salman M, Fatima M, Mukhtar N, Almajhdi FN, Zaman N, et al (2023) Mutational analysis of the spike protein of SARS-COV-2 isolates revealed atomistic features responsible for higher binding and infectivity. Front Cell Dev Biol 10. doi: 10.3389/fcell.2022.940863.
38. Peacock TP, Penrice-Randal R, Hiscox JA, Barclay WS (2021) SARS-CoV-2 one year on: evidence for ongoing viral adaptation. J Gen Virol 102(4).001584. doi: 10.1099/jgv.0.001584.
39. Li M, Du J, Liu W, Li Z, Lv F, Hu C, et al (2023) Comparative susceptibility of SARS-CoV-2, SARS-CoV, and MERS-CoV across mammals. ISME J 17(4):549–60. doi:10.1038/s41396-023-01368-2.
40. Kumar S, Karuppanan K, Subramaniam G (2022) 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 94(10):4780–4791. doi:10.1002/jmv.27927.
41. Wang L, Cheng G (2022) Sequence analysis of the emerging SARS-CoV-2 variant Omicron in South Africa. J Med Virol 94(4):1728–1733. doi:10.1002/jmv.27516.
42. Dhawan M, Saied AA, Mitra S, Alhumaydhi FA, Emran TB, Wilairatana P (2022) Omicron variant (B.1.1.529) and its sublineages: What do we know so far amid the emergence of recombinant variants of SARS-CoV-2? Biomed Pharmacother Biomedecine Pharmacother 154:113522. doi:10.1016/j.biopha.2022.113522.
43. Khandia R, Singhal S, Alqahtani T, Kamal MA, El-Shall NA,

Nainu F, et al (2022) 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 209:112816. doi:10.1016/j.envres.2022.112816.

44. Goletic T, Konjhodzic R, Fejzic N, Goletic S, Eterovic T, Softic A, et al (2021) Phylogenetic pattern of SARS-CoV-2 from COVID-19 patients from Bosnia and Herzegovina: lessons learned to optimize future molecular and epidemiological approaches. Biomol Biomed 21(4):484–7. doi: https://doi.org/10.17305/bjbms.2020.5381. doi:10.17305/bjbms.2020.5381.

45. Wolf JM, Wolf LM, Bello GL, Maccari JG, Nasi LA (2023) Molecular evolution of SARS-CoV-2 from December 2019 to August 2022. J Med Virol 95(1):e28366. doi:10.1002/jmv.28366.

46. Gopalan V, Chandran A, Arumugam K, Sundaram M, Velladurai S, Govindan K, et al (2023) Distribution and Functional Analyses of Mutations in Spike Protein and Phylogenic Diversity of SARS-CoV-2 Variants Emerged during the Year 2021 in India. J Glob Infect Dis 15(2):43–51. doi:10.4103/jgid.jgid_178_22.