Al-Rafidain J Med Sci. 2024;6(2):149-155.
DOI: https://doi.org/10.54133/ajms.v6i2.777

A J M S
Online ISSN (2789-3219)

Research Article

Determine the Whole Genome Sequences of SARS-COV-2 Isolated from Iraqi Patients Using NGS Method

Israa Hashem Al-Zubaidy[1]*, Saife Al-Ahmer[1], Mohammed Issa Aldafaee[2]

[1]*Institute of Genetic Engineering and Biotechnology for Post Graduate Studies, University of Baghdad, Baghdad, Iraq;* [2]*Medical City, Baghdad, Iraq*

Received: 26 March 2024; Revised: 27 May 2024; Accepted: 31 May 2024

Abstract

Background: Next-generation sequencing (NGS) can monitor the transmission of COVID-19 and viral alterations. ***Objectives***: To provide information about testing techniques and infection control measures, as well as to direct the development of vaccines and treatments for the Corona virus. ***Methods***: Six Iraqi SARS-CoV-2 strains were investigated using whole-genome sequencing using the next-generation sequencing method. The sequencing was carried out with an Illumina MiSeq system, and phylogenetic analysis was carried out for all Iraqi sequences retrieved from GISAID. ***Results***: The analysis of the isolates from this study showed that all the sequences from the most recent wave, which happened in the summer of 2022, were primarily clustered in the 20A clades and the 21K, 21L (Omicron) clade, as determined by the GISAID and Nextclade systems. On the other hand, the PANGO system revealed that six sequences were of the BA.1 lineage in Iraq, while four were of the BA.2 lineage. We found that throughout the country's subsequent pandemic waves, SARS-CoV-2 clades and their lineages exhibited circulation patterns and dominance. ***Conclusions***: NGS continues to supply vital COVID-19 evidence to academics, vaccine and medication makers, and public health regulators.

Keywords: SARS corona virus-2, WGS, NGS.

تحديد تسلسل الجينوم الكامل ل SARS-COV-2 المعزول من المرضى العراقيين بطريقة معرفة تسلسل الجيل التالي

الخلاصة

الخلفية: يمكن لمعرفة تسلسل الجيل التالي مراقبة انتقال COVID-19 والتغيرات الفيروسية. **الأهداف**: توفير معلومات حول تقنيات الاختبار وتدابير مكافحة العدوى، وكذلك توجيه تطوير اللقاحات والعلاجات لفيروس كورونا. **الطرق**: تم فحص ست سلالات عراقية من SARS-CoV-2 باستخدام تسلسل الجينوم الكامل باستخدام طريقة تسلسل الجيل التالي. تم إجراء التسلسل باستخدام نظام Illumina MiSeq ، وتم إجراء تحليل النشوء والتطور لجميع التسلسلات العراقية التي تم استردادها من GISAID. **النتائج**: أظهر تحليل العزلات من هذه الدراسة أن جميع التسلسلات من الموجة الأخيرة ، والتي حدثت في صيف عام 2022 ، كانت متجمعة بشكل أساسي في مجموعات 20A و 21K، 21L (Omicron) ، على النحو الذي حدده نظامي GISAID و Nextclacdeمن ناحية أخرى ، كشف نظام PANGO أن ستة تسلسلات كانت من سلالة BA.1 في العراق ، بينما كانت أربعة من سلالة BA.2. وجدنا أنه خلال الموجات الوبائية اللاحقة في البلاد، أظهرت مجموعات SARS-CoV-2 وأنسابها أنماط دوران وهيمنة. **الاستنتاجات**: توفر طريقة معرفة تسلسل الجيل التالي أدلة COVID-19 الحيوية للأكاديميين وصانعي اللقاحات والأدوية ومنظمي الصحة العامة.

* ***Corresponding author***: Israa H. Al-Zubaidy, Institute of Genetic Engineering and Biotechnology for Post Graduate Studies, University of Baghdad, Baghdad, Iraq; Email: israa.hashem1100a@ige.uobaghdad.edu.iq

Article citation: Al-Zubaidy IH, Al-Ahmer S, Aldafaee MI. Determine the Whole Genome Sequences of SARS-COV-2 Isolated from Iraqi Patients by NGS Method. *Al-Rafidain J Med Sci.* 2024;6(2):149-155. doi: https://doi.org/10.54133/ajms.v6i2.777

INTRODUCTION

The members of the Coronaviridae are known to infect a wide variety of hosts, causing various diseases and symptoms, such as SARS, COVID-19, and MERS. CoV was discovered during the 1960s; it caused the COVID-19 pandemic and has caused widespread disease, fatalities, and disruptions to daily life [1,2]. A group of closely related viruses known as CoVs can infect people's respiratory systems and cause minor symptoms that can lead to death [3, 4]. A coronavirus is an encapsulated virus with a positive single-stranded RNA genome. Four genera, α, β, γ, and δ, can be distinguished based on the serotype and genomic features [5–7]. Since then, 216 countries and territories have contracted the disease. The WHO proclaimed COVID-19 to be a pandemic on January 30, 2020 [8,9]. With a single-stranded positive-sense RNA genome between 26 and 35 kb, SARS-CoV-2 codes for about 27 proteins, some of which share similarities with proteins with established activities. In contrast, others are uncertain, unidentified, or presumptive [10,11]. There are fourteen open reading frames in the genome of SARS-CoV-2. These frames

code for sixteen nonstructural proteins (nsp1–16), four structural proteins (spike (S), envelope (E), membrane (M), and nucleocapsid (N), and nine accessory proteins. The angiotensin-converting enzyme 2 (ACE2), a cellular receptor for protein S, has a role in all of these viral proteins and helps enhance the entry of SARS-CoV-2 into host cells. In addition, the transmembrane proteases serine 2 and cathepsin L can only let SARS-CoV-2 into cells at the plasma membrane surface and endosomal compartments [12]. Whole-genome sequencing (WGS) has been essential in providing insights into several newly discovered viral epidemics, including Zika, Ebola, Usutu, and yellow fever viruses [13]. The detection of viral presence and the comprehension of the SARS-CoV-2 virus's epidemiology are also frequently accomplished using next-generation gene sequencing (NGS) techniques. Nevertheless, despite the accuracy and dependability of NGS systems, their practical applicability is sometimes constrained by the additional expenses and skills required [14]. Following the initial COVID-19 epidemic, extensive worldwide next-generation sequencing (NGS)-based initiatives were launched to investigate the genetic insights of SARS-CoV2, due in great part to the efficacious function of WGS during many outbreaks. The first SARS-CoV2 whole-genome sequencing was released on January 5, 2020. Several nations contributed the full genome sequence of SARS-CoV2 to the Global Initiative on Sharing All Influenza Data (GISAID) database after the virus's initial full genome sequencing [15,16]. Because COVID-19-positive patients from Iraq are different from those from its neighboring countries, a full WGS-based analysis is needed to rule out the major differences linked to the country's SARS-CoV2 strains and find out if similar mutations are present in other countries with similar prevalence patterns. This study aims to evaluate the NGS technologies for diagnosing, surveilling, and surveying SARS-CoV-2. The classification of isolates into separate clades, variations, and lineages was accomplished with the assistance of three well-established global platforms: GISAID, Nextstrain, and PANGO. As a result, we sequenced and described SARS-CoV2's whole genome. Phylogenetic analysis was conducted in addition to WGS to compare the genomes of the current SARS-CoV2 strain with the available genomes. This study aims to provide information regarding diagnosis, infection control measures, and guidance for the development of vaccines and treatments for the Corona virus.

METHODS

Study design and samples collection

In this study, we collected 50 samples of nasopharynx (NP) swabs and oropharynx (OP) from patients suspected of having COVID-19, based on clinical symptoms reported by doctors. Patients' samples were collected from Imam Hassan Al-Mujtaba Teaching Hospital in Karbala, Iraq, and the Central Public Health Laboratory (CPHL) from June 2022 to August 2022. Specimens were then directly inserted into a viral transport medium (VTM) and samples were classified into main groups according to gender, as shown in Table 1.

Table 1: Distribution of patients according to infection and gender

Gender	Positive	Negative	*p*-value
Male	35.65±16.14	34.6±12.96	
Female	32.96±9.14	36.63±15.58	0.752
Total	34.51±13.58	35.39±13.88	
p-value		0.771	

Values were expressed as mean±SD.

Detection of COVID-19

RNA was isolated from samples obtained from nasopharyngeal swabs as well as oropharyngeal (OP) swabs using the AddPrep Viral Nucleic Acid Extraction Kit (Addbio, South Korea). The QubitTM RNA Broad Range (BR) assay kit (Catalogue Number Q10211) was utilized to perform quantification, which was the method of quality control (QC) for the extracted RNA.

PCR Amplification

RNA was extracted and analyzed by using a real-time reverse transcriptase polymerase chain reaction (rRT-PCR) assay to confirm infection [17,18]. Polymerase chain reaction (PCR) in real-time: The reaction consisted of 20 µL of RNA and 10 µL of 2 µL reaction buffer, which was supplied with the Coronavirus COVID-19 Nucleic Acid Detection Kit (Fluorescence RT-PCR Method, AeHealth, UK). The reverse transcription process involved thermal cycling at 50 °C for 10 minutes, followed by 95 °C for 3 minutes, and then 40 cycles of 95 °C for 10 seconds and 55 °C for 30 seconds. The laboratories that participated in the study utilized Bio-Rad, USA. The whole genome sequencing was conducted on samples from six diverse Iraqi locales using the NGS technique. These specimens had threshold (Ct) values ranging from 15 to 17 and were assessed using samples from six different regions in Iraq.

NGS library preparation and whole-genome sequencing

The PCR product was normalized to a concentration of 0.2 ng/L. The AmpliSeq Library PLUS kit for Illumina was utilized to prepare the paired-end sequencing library, which involved the implementation of the following primary procedures: The process involves several steps: 1) DNA quantification and dilution; 2) target amplification; 3) partial amplicon digestion; 4) index ligation; 5) library cleaning; and 6) Equilzer workflow. The pooled library was subsequently prepared for paired-end sequencing utilizing the MiSeq (Illumina) platform, employing a sequencing reagent cartridge known as the MiSeq Reagent Micro Kit v2, with a total of 300 cycles (Cat. No. MS-103-1002).

NGS bioinformatics analysis

Paired-end NGS data were analyzed using publicly available bioinformatics software. The FastQC program (version 0.11.8) assisted in examining the read quality of FASTQ files in the first step of the process [19]. By using the Trimmomatic tool, version 0.39 [20], it was possible to get rid of low-quality base calls (Q < 30) and index adapter sequences from both ends of sequenced reads. Using the default settings for the Burrows-Wheeler Aligner (BWA, version 0.7.17), the reads that had been filtered were aligned with the Wuhan reference genome (GenBank accession number: NC_045512) [21]. This was accomplished by using the default settings. For the viral genome alignment, the reference sequence of SARS-CoV2 (accession number: NC_045512) was utilized as a control [22]. The National Genomics Data Centre (NGDC) of China was utilized in order to carry out the processes of genome annotation and variant calling [23] (Figure 1).

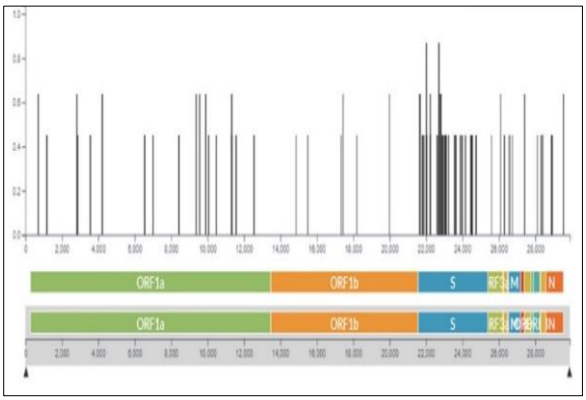

Figure 2: The map of gene annotation where the exact site of mutation appeared.

Erroneous data may be brought to the forefront during the mapping process due to the presence of duplicated reads. For this reason, Sambamba (version 0.68), which is utilized to remove duplicated reads,. By utilizing mapping information like the start location and CIGAR string, we can identify and differentiate duplicate reads. The search for genetic variation was carried out using the bulk sequence data that was produced. A copy of the reference genome is kept on the National Centre for Biotechnology Information (NCBI) website (https://www.ncbi.nlm.nih.gov/). Following the elimination of duplicates using Sambamba and the identification of variations using SAMTools, information regarding each variant was acquired, and it was categorized according to chromosomes or scaffolds.

Phylogenetic analysis

SARS-CoV2 genomes were acquired from available sources using FASTQ files retrieved from the GISAID and NCBI databases. The phylodynamic approach of Augur Nextstrain was employed to cluster a collective of 85 genomes of SARS-CoV2, which encompasses the present sequenced genome (accession number: MW242667) [24]. The alignment of the sequences was performed using MAFFT [25] using the Wuhan reference genome (accession number: NC_045512.2). We generated the phylogenetic tree using the IQ-TREE software [26]. We used the software FigTree v1.4.4 [27] to visualize the phylogenetic tree. The Codon Tree pipeline produces phylogenetic trees of bacteria. The proposed methodology involves utilizing the amino acid and nucleotide sequences obtained from a predetermined set of global protein families (PGFams). We randomly select these programs to construct an alignment. We then generate a tree structure by identifying the variations within the selected sequences. This tutorial's focus is on the Codon Trees pipeline. For each of the selected genes from the PGFams, both the protein (amino acid) and gene (nucleotide) sequences are utilized. MUSCLE is employed for the alignment of protein sequences, while the Codon_align function in BioPython is utilized for the alignment of nucleotide-coding gene sequences. A PHYLIP-formatted file was utilized to record a concatenated alignment of proteins and nucleotides. Subsequently, a partition file for RaxML was prepared, which included a description of the alignment of the proteins as well as the positions of the first, second, and third codons. The support values are derived by the utilization of 100 iterations of the "rapid" bootstrapping feature in RaxML. The generated Newick file can be accessed for viewing purposes. Additionally, we recommend that researchers download the file and utilize FigTree to build an image of publication quality.

Statistical analysis

We used the Statistics Package for Social Science (SPSS), version 25 for Windows software, and the Microsoft package (Excel and Word) (IBM, Armonk, NY, USA) for statistical analysis. The data are normally distributed and shown as the mean ± standard deviation (SD). The Pearson correlation coefficients (p-value) were used to find the chi-square between variables and to see how statistically significant the difference between the groups was. A p-value of less than 0.05 was considered significant. We analyzed the study's results to align them with its objectives and presented them under the following headings: We used the Statistics Package for Social Science (SPSS), version 25 for Windows software, and the Microsoft package (Excel and Word) (IBM, Armonk, NY, USA) for statistical analysis. The data are normally distributed and shown as the mean ± standard deviation (SD). The Pearson correlation coefficients (p-value) were used to find the chi-square between variables and to see how statistically significant the difference between the groups was. A *p*-value of less than 0.05 was considered significant.

RESULTS

The present study reveals six complete genomes of SARS-CoV-2. These genomes were collected from people who tested positive for COVID-19 using real-time polymerase chain reaction (RT-PCR). These

individuals were obtained from Al-Imam Hassan Al-Mujtaba Teaching Hospital in Karbala, Iraq, and the Central Public Health Laboratory (CPHL) in Baghdad, Iraq. samples were classified into main groups according to gender, as shown in Table 2.

Table 2: Distribution of six strains SARS-CoV-2 by their Next-Generation Sequencing (NGS) according to the Wuhan reference sequences (NC045512); shows the set of mutations if the type of mutation insertion or deletion

Samples name	Clade	Pango lineage	WHO name	Mutations	Non-ACGTN	Ns (missing)	Coverage (%)
K1	20A	B.1.36	-	17	1	9760	67.2
K2	21K	BA.1	Omicron	40	4	5208	82.5
K3	21L	BA.2	Omicron	24	10	17850	40.1
K4	21L	BA.2	Omicron	47	3	6130	79.5
K5	21L	BA.2	Omicron	49	2	4634	84.5
K6	21L	BA.2	Omicron	24	6	18282	38.8

The investigation of the whole genome sequence and the resulting variable amino-acid composition using the genome detective web server (https://www.genomedetective.com) was conducted on the Iraqi SARS-CoV-2 retrieved whole genome sequences aligned with the Wuhan reference sequences (NC045512); the obtained results are displayed in Table 2. It gives a full picture of the genetic sequences that were looked at, showing their evolutionary connections (clade inference) and different traits like mutations, ambiguities, gaps, insertions, frameshifts, and premature stop codons. It can be valuable for understanding the genetic variation and potential functional implications of the analyzed sequences. In the context of phylogenetic trees, divergence refers to the evolutionary distance or amount of time that has passed since two or more species shared a common ancestor (K4, K5, K6), which have a similar divergence of 73. Genetic differences, like the number of mutations or changes in sequences, unite them in a cluster, allowing us to infer the relationships and evolutionary history among species. K2 exhibited a divergence of 62 points. According to the phylogenetic tree in Figure 6, the SARS-CoV-2 strains of this study (K1–K6) have diverged from a common ancestor. Divergence in this context refers to the accumulation of genetic mutations in the viral genome over time. These mutations occur as the virus replicates and spreads through human populations. The more mutations that accumulate in a strain, the more genetically distinct it becomes from the original strain that caused the COVID-19 pandemic. We align the mutations with the regions that lack data. Figure 2 displays the missing data, highlighted in gray. Missing data could be quickly checked out, as it's distributed on the genome, clustering in one area or many small missing segments.

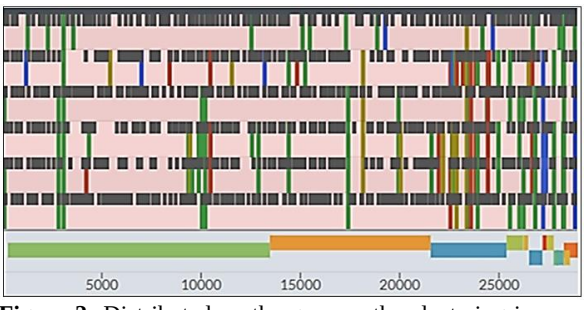

Figure 2: Distributed on the genome the clustering in one area.

Table 2 clearly demonstrates that all these clades carry the 21L BA2 Omicron clade, further confirming their genetic similarity. Figure 3 illustrates the clustering of clades 4, 5, and 6, and Table 2 suggests that these clades lack any private mutations, which, if present, would have separated them. Table 2 clearly demonstrates that all those clades carry the 21L 1L BA2 omicron clade, further confirming their Simi branch and butene tic consistency.

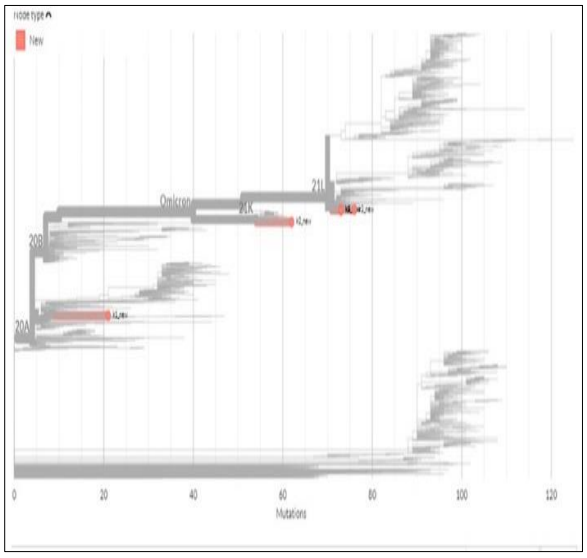

Figure 3: The clades 4, 5, and 6 clustered together.

The whole-genome mutations are illustrated in Figures 3, 4, 5, and 6. Strain K3 also carried the clade (21L) BA2 Omicron and was located on the same branch, but it separated from the rest of the phylogenetic tree due to the presence of a private mutation, as shown in Figure 4 B and C.

Table 2 shows that K2 belongs to a different branch than the previous ones, indicating a different genetic constituency and referring to it as clade 21K. K1 is referred to as 20A B.1.36. In the context of phylogenetic trees, divergence refers to the evolutionary distance or time elapsed since two or more species (K4, K5, K6) shared a common ancestor, with a similar divergence of 73. Genetic differences, such as the number of mutations or changes in sequences, unite them in a cluster, allowing us to infer the relationships and evolutionary history among species.

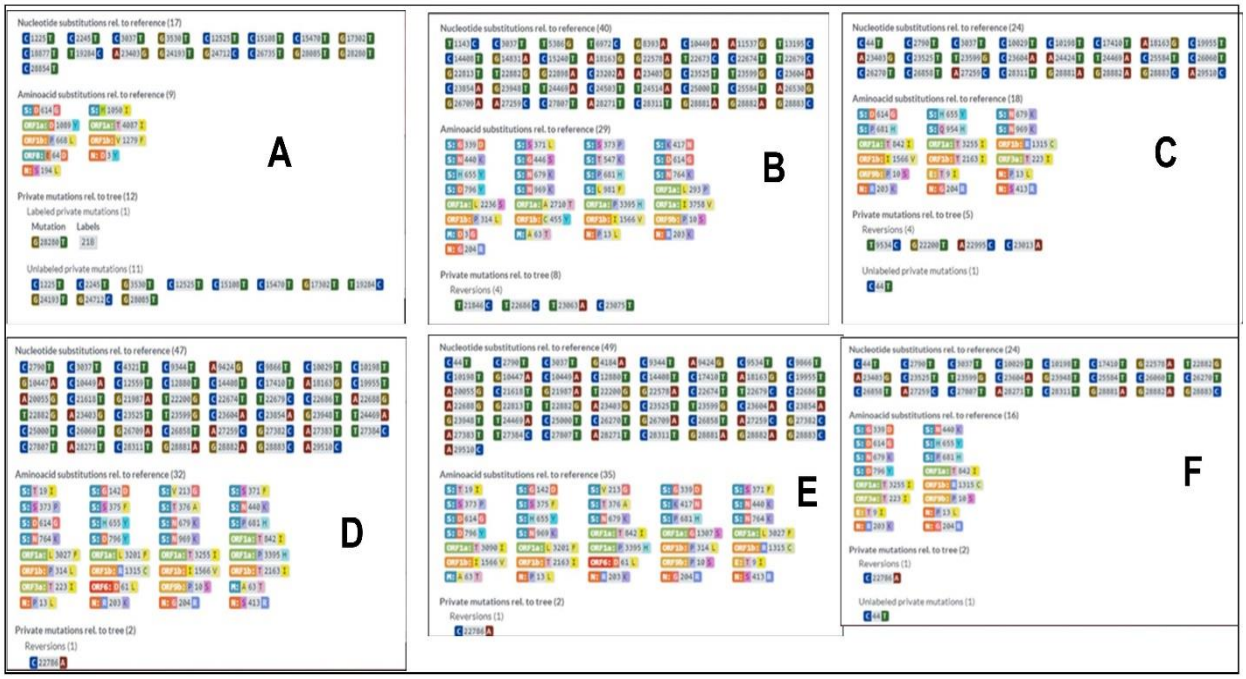

Figure 4: The 20A clade and the whole genome sequence mutations. The k1 sample shows 1 labeled mutation, indicative of sequence quality problems, potentially a contamination with 21B (A); the k2 sample refers to the 21K clade and the whole genome sequence mutations (B); the k3 sample refers to the 21L clade and the whole genome sequence mutations (C); the k4 sample refers to the 21L clade and the whole genome sequence mutations (D); the k5 sample refers to the 21L clade and the whole genome sequence mutations (E); the k6 sample refers to the 21L clade and the whole genome sequence mutations (F).

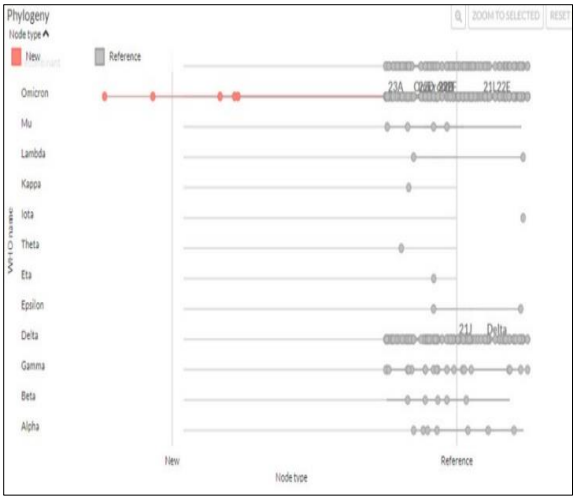

Figure 5: Nodes in a phylogenetic tree represent common ancestors of the species or groups and all the strains in figure 9 tree refer to the same ancestor, which is Omicron. The branching pattern of the tree represents the evolutionary relationships among these species, with each node representing a point in time when a common ancestor existed.

According to the phylogenetic tree (Figure 6), the SARS-CoV-2 strains of this study (K1–K6) have diverged from a common ancestor. Divergence in this context refers to the accumulation of genetic changes (mutations) in the viral genome over time. These mutations occur as the virus replicates and spreads through human populations. The more mutations that accumulate in a strain, the more genetically distinct it becomes from the original strain that caused the COVID-19 pandemic.

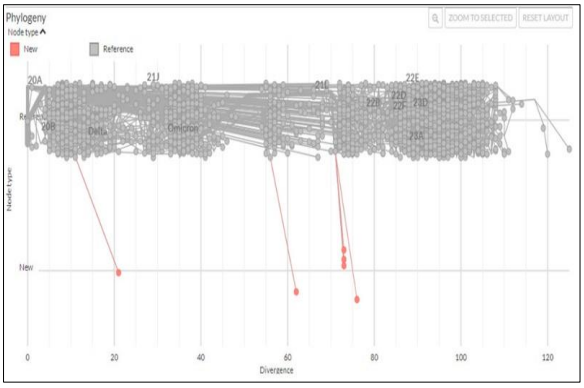

Figure 6: Shows correlation between divergence and node type.

Figure 7 shows that six study samples were chosen for phylogenetic analysis, and all six samples (K1, K2, K3, K4, K5, and K6) were included within the global clades funded at the time of the study. Researchers have highlighted the study samples in red.

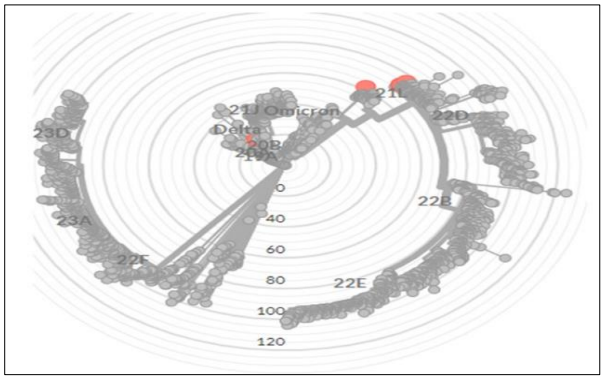

Figure 7: The phylogenetic tree that is based on the distribution of clades within the world.

When we compared the isolates with each other in terms of genetic variation, as shown in Figure 8, it became evident that there were differences. The presence of three distinct branches, which divided the isolates into three genetic groups divergent in terms of mutation types and placed the strains within three different clusters, clearly demonstrated the variations among them. The phylogenetic tree was constructed based on sequence data in GISAID data using the next strain server (Next Clade Tool). The gray color represents the GISAID clades, while the orange circle represents the Iraqi SARS-CoV-2 variants.

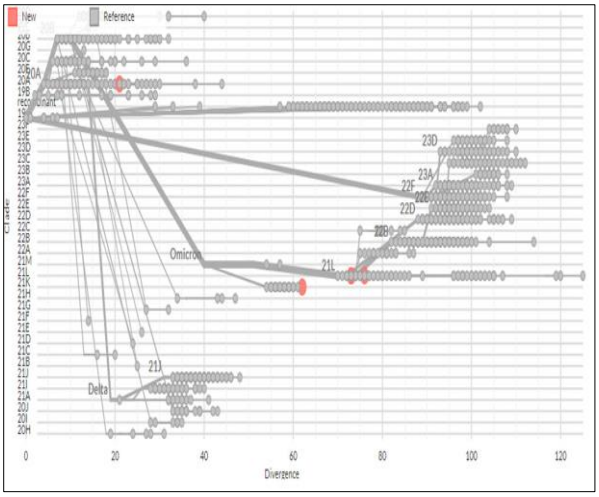

Figure 8: Phylogenetic SARS-CoV-2 variant analysis according to correlation between divergence of clades.

DISCUSSION

The World Health Organization's coronavirus dashboard indicates that the Omicron subvariant (BA.2) is associated with the most recent COVID-19 wave that occurred in Iraq in 2022. The current investigation found all three full genomes of SARS-CoV-2 were obtained from patients with RT-PCR-positive COVID-19, originating from Krbalaa, Al-Muthana, Misan, and Baghdad provinces in Iraq. In Table 2, the majority of samples belong to omicron BA.1 and BA.2. The number of mutations ranges from 17 to 49, and the sample coverage ranges from 38.8% to 84.5%. Therefore, the unusual mutations discovered in these samples could be crucial for creating antiviral vaccines or treatments that work. The evolutionary characteristics of this virus are significant because they add to our understanding of viral changes and help identify the best location to target with a vaccination or antiviral medication. To determine the clades and lineages of our findings, we uploaded our sequences to the GISAID database. A20 appeared in sample K1, and when the infected person was investigated, it was found that he was traveling to Australia [28]. Based on the current findings, it is thought that SARS-CoV-2 strains have been repeatedly brought into Iraq, particularly by foreign visitors. The most well-recognized worldwide lineage, B.1, has been further split into more than 70 sub-lineages. According to our findings, the regular arrival of foreign tourists in Iraq helped the SARS-CoV-2 spread, causing different clades to move around over time. Governmental actions, such as limiting meetings indoors and outdoors, may break up and disrupt community transmission chains. However, the fact that foreign tourists often bring diseases has also changed how diseases are spread. More epidemiological and genomic surveillance is required to obtain an accurate assessment of SARS-CoV-2 diversity. We compared the SARSCoV-2 sequences we had with the reference genome (NC 045512.2) to find genomic variations. The results of this investigation revealed several changes in the S-gene, followed by ORF1ab, N-gene, and M-gene in the Omicron VOC of SARS-CoV-2 sequences. The genes with the lowest number of mutations were ORF3a, E-gene, ORF6, ORF7b, and ORF9b. The S-gene with the most mutations codes for a structural protein that helps the virus attach to host cell receptors and is very important in deciding the types of cells that can be infected. The variant analysis revealed that the spike glycoprotein included a D614G mutation in each of our instances. The D614G variant of SARS-CoV-2 appears to be the most common virus circulating in Iraq. Up to August 2021, 97.83 percent of all spike sequencing samples across 198 nations will have the D614G mutation [29]. Certain studies propose that positive selection accounts for the worldwide domain of D614G based on sequence distribution and phylogenetic tree analysis, whereas a founder effect accounts for the European domain [30]. Research indicates that individuals with the D614G mutation have greater respiratory tract viral loads than SARS-CoV-2 patients without the mutation [31]. Individuals with the D614G mutation who were infected with SARS-CoV-2 often experienced moderate COVID-19 symptoms, but those without the mutation only experienced mild symptoms [32]. So, this change may make SARS-CoV-2 more contagious by improving the RBD's ability to bind to human ACE2 and decreasing the contacts between S1 and S2. [33]. There were 549 mutations and 53 distinct variations found in 47 isolates, according to one investigation. Like our findings, that study discovered that the most common mutation in spike glycoprotein is D614G [34].

Conclusion

The SARS-CoV-2 isolates obtained from Iraqi COVID-19 patients came from several geographic locations associated with the BA.1 and BA.2 pango lineages. These strains differ from the original hCoV19/Wuhan/NC 045512.2/2019 strain in several ways, including the presence of multiple mutations in the nonstructural proteins and spike protein. Patients from Iraq had a genome sequence that resembled the global SARS-CoV-2 genome. Nonetheless, we might presume that Australia, Europe, Iran, India, and the United States are the origins of SARS-CoV-2 in Iraq.

Conflict of interests

No conflict of interests was declared by the authors.

Funding source

The authors did not receive any source of fund.

Data sharing statement

Supplementary data can be shared with the corresponding author upon reasonable request.

REFERENCES

1. Sulayman N. Deep learning-based predictive model of mRNA vaccine deterioration: An analysis of the stanford COVID-19 mRNA vaccine dataset. *Baghdad Sci J*. 2023;20(4 (SI)):1451-1458. doi: 10.21123/bsj.2023.8504.
2. Zhu FC, Li YH, Guan XH, Hou LH, Wang WJ, Li JX, et al. Safety, tolerability, and immunogenicity of a recombinant adenovirus type-5 vectored COVID-19 vaccine: a dose-escalation, open-label, non-randomised, first-in-human trial. *Lancet*. 2020;395(10240):1845-1854. doi: 10.1016/S0140-6736(20)31208-3.
3. Hamid MK. Impact of COVID-19 vaccine on hearing status of young ages (Medical College students as a sample). *Baghdad Sci J*. 2023;20(4):1498-1506. doi: 10.21123/bsj.2023.8694.
4. Yousif MG, Hashim K, Rawaf S. Post COVID-19 effect on medical staff and doctors' productivity analysed by machine learning. *Baghdad Sci J*. 2023;20(4):1507-1519. doi: 10.21123/bsj.2023.8875.
5. Zhou P, Yang X-L, Wang X-G, Hu B, Zhang L, Zhang W, et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin. nature. 2020;579(7798):270-273. doi: 10.1038/s41586-020-2012-7.
6. Almashhadani AD, AL-Thwani AN. Determination of angiotensin-converting enzyme 2 (ACE2) receptor level in samples of Iraqi patients infected with COVID-19. *Iraqi J Biotechnol*. 2022;21(2)178-182.
7. Auda IG, Auda J, Salih RH. SARS-CoV-2 and other Coronaviruses: A matter of variations. *Al-Kindy Coll Med J*. 2023;19(1):5-10.doi: 10.47723/kcmj.v19i1.927.
8. Guo YR, Cao QD, Hong ZS, Tan YY, Chen SD, Jin HJ, et al. The origin, transmission and clinical therapies on coronavirus disease 2019 (COVID-19) outbreak–an update on the status. *Military Med Res*. 2020;7:1-10. doi: 10.1186/s40779-020-00240-0.
9. Abduallah ZS, AL-Thwani AN, Almashhadani AD, Mohsin MI. Estimation of CRP and some hematological parameters with COVID-19 patients using ANOVA as a statistical tool. *Nat Volatile Essent Oils*. 2021;8(5):8915-8919.
10. Zalzala HH. Diagnosis of COVID-19: facts and challenges. *New Microbes New Infect*. 2020;38:100761. doi: 10.1016/j.nmni.2020.100761.
11. Kradi AM, Al-Faisal AHM, Turki AM. Alteration of M and N genes of corona virus 2 (SARS-CoV-2). *Iraqi J Biotechnol*. 2022;21(2):39-45.
12. Hoffmann M, Kleine-Weber H, Schroeder S, Krüger N, Herrler T, Erichsen S, et al. SARS-CoV-2 cell entry depends on ACE2 and TMPRSS2 and is blocked by a clinically proven protease inhibitor. *Cell*. 2020;181(2):271-280. doi: 10.1016/j.cell.2020.02.052.
13. Oude Munnink BB, Nieuwenhuijse DF, Stein M, O'Toole Á, Haverkate M, Mollers M, et al. Rapid SARS-CoV-2 whole-genome sequencing and analysis for informed public health decision-making in the Netherlands. *Nature Med*. 2020;26(9):1405-1410. doi: 10.1038/s41591-020-0997-y.
14. Alpdagtas S, Ilhan E, Uysal E, Sengor M, Ustundag CB, Gunduz O. Evaluation of current diagnostic methods for COVID-19. *APL Bioengineer*. 2020;4(4):041506. doi: 10.1063/5.0021554.
15. Elbe S, Buckland-Merrett G. Data, disease, and diplomacy: GISAID's innovative contribution to global health. *Global Challenges*. 2017;1(1):33-46. doi: 10.1002/gch2.1018.
16. Shu Y, McCauley J. GISAID: Global initiative on sharing all influenza data–from vision to reality. *Eurosurveillance*. 2017;22(13):30494. doi: 10.2807/1560-7917.ES.2017.22.13.30494.
17. Ghazzi JJ, Fadhil HY, Aufi IM. Impact of SARS-COV-2 Variants on the Infection Severity among Iraqi Patients. *Iraqi J Sci*. 2023:3263-3272. doi: 10.24996/ijs.2023.64.7.7.
18. Al-Janabi MK, Al-Ahmer SD, Al-Ganabi AD, Aufi IM, editors. Rapid and accurate molecular detection of adenovirus associated with gastroenteritis children in Baghdad city. *AIP Conference Proceedings*. 2022; 2398(1). doi: 10.1063/5.0093770.
19. Andrews S. Fast QC: a quality control tool for high throughput sequence data. Cambridge, United Kingdom; 2010.
20. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*. 2014;30(15):2114-2120. doi: 10.1093/bioinformatics/btu170.
21. Li H, Durbin R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics*. 2009;25(14):1754-1760. doi: 10.1093/bioinformatics/btu170.
22. Wu F, Zhao S, Yu B, Chen YM, Wang W, Song ZG, et al. A new coronavirus associated with human respiratory disease in China. *Nature*. 2020;579(7798):265-269. doi: 10.1038/s41586-020-2008-3.
23. Database resources of the national genomics data center in 2020. *Nucleic Acids Res*. 2020;48(D1):D24-D33. doi: 10.1093/nar/gkz913.
24. 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. doi: 10.1093/bioinformatics/bty407.
25. Katoh K, Asimenos G, Toh H. Multiple alignment of DNA sequences with MAFFT. *Methods in Molecular Biology (Clifton, N.J.)*. 2009:537:39-64. doi: 10.1007/978-1-59745-251-9_3.
26. Nguyen LT, Schmidt HA, Von Haeseler A, Minh BQ. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evolu*. 2015;32(1):268-274. doi: 10.1093/molbev/msu300.
27. Rambaut A, Holmes EC, O'Toole Á, Hill V, McCrone JT, Ruis C, et al. A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology. *Nature Microbiol*. 2020;5(11):1403-1407. doi: 10.1038/s41564-020-0770-5.
28. GISAID. lineage nomenclature aids in genomic epidemiology studies of active hCoV-19 viruses. Clade and Lineage Nomenclature, March 2, 2021. Available at: https://gisaid.org/resources/statements-clarifications/clade-and-lineage-nomenclature-aids-in-genomic-epidemiology-of-active-hcov-19-viruses/
29. Alhussien TAA, Fadhil HY. Analysis of mutations in conserved and susceptible regions across the whole genome sequencing analysis for SARS-CoV-2 in Iraqi patients. *Iraqi J Sci*. 2023:56-64. doi: 10.24996/ijs.2023.64.1.6.
30. Dearlove B, Lewitus E, Bai H, Li Y, Reeves DB, Joyce MG, et al. A SARS-CoV-2 vaccine candidate would likely match all currently circulating variants. *Proc Natl Acad Sci USA*. 2020;117(38):23652-23662. doi: 10.1073/pnas.2008281117.
31. Korber B, Fischer WM, Gnanakaran S, Yoon H, Theiler J, Abfalterer W, 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. doi: 10.1016/j.cell.2020.06.043.
32. Cleemput S, Dumon W, Fonseca V, Abdool Karim W, Giovanetti M, Alcantara LC, et al. Genome detective coronavirus typing tool for rapid identification and characterization of novel coronavirus genomes. *Bioinformatics*. 2020;36(11):3552-3555. doi: 10.1093/bioinformatics/btaa145.
33. Michel CJ, Mayer C, Poch O, Thompson JD. Characterization of accessory genes in coronavirus genomes. *Virol J*. 2020;17:1-13. doi: 10.1186/s12985-020-01402-1.
34. Mahmood ZS, Fadhil HY, Hussein TAA, Ad'hiah AH. The severity of coronavirus disease 19: Profile of inflammatory markers and ACE (rs4646994) and ACE2 (rs2285666) gene polymorphisms in Iraqi patients. *Meta Gene*. 2022;31:101014. doi: 10.1016/j.mgene.2022.101014.