**ISSN: 0067-2904**

# Analysis of Mutations in Conserved and Susceptible Regions Across the Whole Genome Sequencing Analysis for SARS-CoV-2 in Iraqi Patients

**Thaer A. Abed Alhussien, Hula Y. Fadhil ***
*Biology department, College of Science, University of Baghdad, Baghdad, Iraq*

Received: 27/1/2022 Accepted: 29/4/2022 Published: 30/1/2023

**Abstract**

This study aimed to get a better understanding of molecular epidemiology and genetic variation in the spike glycoprotein as a key viral component involved in viral entrance into host cells and as a potential vaccination target. Three Iraqi SARS-CoV-2 strains were investigated using whole-genome sequencing, with two of them clustering into the 20A (GH) clade, and the remaining strain is clustered in 20E (GV) clade, belonging to the B.1.36.1 and B.1.177.80 lineage, respectively. Whole-genome sequencing of the viral RNA samples revealed nine sporadic nonsynonymous uncommon mutations with frequency ranged from 0.00 to 0.19%. The ORF1ab, ORF1a, ORF3a, S, N, intergenic, ORF7 and ORF8 areas have seen the most changes. Furthermore, in all of our sequences, we discovered a D614G (aspartic acid to glycine) mutation in spike protein that co-occurred with an NSP12 P323L (viral RNA-dependent RNA polymerase) mutation. The findings point to several viral introductions in Iraq and provide new genetic information on SARS-CoV-2 at the worldwide level. Pathogenesis, diagnostics and vaccine development require information such as SNPs and mutations.

**Keywords**: Coronavirus disease 2019; Whole genome sequence; SNP; Nonsynonymous mutation; RdRp mutation; Illumina NovaSeq 6000.

# تحليل الطفرات في المناطق الثابتة والمعرضة للتأثر من خلال تسلسل الجينوم الكامل لـ فايروس السارس-2 في المرضى العراقيين

**ثائر علي عبد الحسين , حلا يونس فاضل***
قسم علوم الحياة, كلية العلوم, جامعة بغداد , بغداد, العراق

**الخلاصة**

هدفت هذه الدراسة إلى الفهم الواسع للوبائيات الجزيئية والتنوع الجيني في البروتين السكري الشوكي كعامل فيروسي مهم مرتبط بدخوله لخلايا المضيف وكهدف للقاح. اعتماداً على تسلسل الجينوم الكامل، حللت ثلاثة عزلات عراقية من فايروس SARS-CoV-2 والتي اثنين منها ارتبطت في كليد 20 A (GH)، والسلالة المتبقية ارتبطت في سلسلة 20 E (GV)، والتي تنتمي إلى B.1.36.1 و B.1.177.80. على التوالي. كشف تسلسل الجينوم الكامل لعينات الحمض النووي الريبي الفيروسي عن تسع طفرات متفرقة غير معروفة وغير شائعة بتردد يتراوح من 0.00 إلى 0.19٪. حدثت معظم الاختلافات في مناطق ORF1ab و ORF1aو ORF3a و S و N و intergenic و ORF7 و ORF8. علاوة على ذلك ، وجدت طفرة

*Email: hulayounis@yahoo.com

D614G(حمض الأسبارتيك إلى الجلايسين) في بروتين الشوكي وحدثت مع طفرةNSP12 P323L RNA الفيروسي المعتمد على (RNA polymerase) في جميع التسلسلات الجينومية. تشير النتائج إلى طفرات متعددة للفيروس في العراق وإضافة بيانات جينومية جديدة عن SARS−CoV−2 على المستوى الدولي. مثل هذه المعلومات المتعلقة بطفرات SNP من الممكن ان تساهم بصورة فاعلة في فهم الإمراضية و تشخيص المرض ، فضلا عن تطوير اللقاح.

## 1. Introduction

SARS coronavirus 2 (SARS-CoV-2) first appeared in Wuhan and then spread across the continents, resulting in the COVID-19 pandemic [1]. Despite considerable changes between the current and previously known SARS-CoV genomes, the cause of the epidemic remains unknown. In the past, genetic research has been successfully applied to the sequence map of genome-level variation, its characteristics, functions and structure highly depend on the genes that exist in the organism's genome [2]. However, it is still necessary to understand the molecular basis of the pathogenicity of SARS-CoV-2. For this reason, comparative genomics can help to gain a deeper understanding of the pathogenesis of COVID-19 [3]. Globally, the sequencing of genomes has been released and placed in public databases such the Global Initiative for Sharing all Influenza Database (GISAID) [4]. With these genomic datasets, it is conceivable and crucial to expose the evolutionary processes of SARS-CoV-2 in order to comprehend the circulating genomes and which portions of the genome alter are based on these kinds. In the genome pool, there exist distinct evolutionary groupings. SNP-based approaches have been utilized in several studies to detect emerging virus types and show genomic areas involved in transmission and evolution [5]. A competitive subtype was identified in another investigation based on the D614G mutation in the spike protein, which promotes binding to the ACE2 receptor on the host cell surface [6]. Iraq is the country with the highest number of cases of coronavirus disease (COVID-19). The first confirmed SARS-CoV-2 case was reported in February 2020 from an Iranian student who had traveled from Iran. Since then, the number of cases, whether imported or locally, has increased. However, data on the molecular epidemiology of the SARS-CoV-2 strain circulating in Iraq remains scarce. The full-length genome sequences of three SARS-CoV-2 isolates from various parts of Iraq (Al-Najaf-Al-Ashraf, Salahuddin, and Baghdad provinces) were used in this work to better understand the molecular epidemiology and genetic heterogeneity and to identify the single nucleotide polymorphism (SNP), including the changes in amino acids related to the spike glycoprotein which is an essential viral component linked to host cell entrance and a vaccine target.

## 2. Materials and methods

### 2.1. Specimens collection:

COVID-19 virus specimens were gathered from hospitalized individuals between December 2020 and February 2021 at Baghdad medical city hospitals. The study practical component took place at Baghdad's Central Public Health Laboratory (CPHL) and Molecular biology laboratory at the College of Science, Department of Biology, University of Baghdad. Specimens were collected from nasopharyngeal (NP) and oropharyngeal (OP) swabs, which were subsequently placed straight into viral transport medium (VTM). Nasopharyngeal (NP) or nasopharyngeal (OP) specimen (200µl) was used for viral RNA extraction via the Exiprep TM Plus Viral DNA/RNA Kit (Bioneer, Korea). SARS-CoV-2 infection was detected using a particular real-time reverse transcription-polymerase chain reaction (RT-PCR) on nasopharyngeal swab specimens according to the AccuPower® SARS-CoV-2 Multiplex Real-Time RT-PCR Kit, and recognizing particular E, N, and RdRp gene sequences [7]. Three SARS-CoV-2 highly positive specimens with threshold (Ct) values ranging between (15-17)

and the NGS technique, was used to evaluate samples from three distinct Iraqi areas for the whole genome sequencing.

### 2.2. Whole-genome sequencing

The extracted RNA was converted to cDNA using the GOScript™ Reverse Transcriptase System First Strand cDNA synthesis kit (Promega, USA) by using an oligo primer as following, a 4 µl of extracted RNA and 1 µl of ligo primer were incubated at 70 °C then chilled in ice for 5 min. For prepared reverse transcriptase mixture, a 20 µl reaction with 0.5 µl of GOScriptTM reverse transcriptase enzyme was set up, 1 µl of dNTPs, 2 µl of $MgCl_2$, 0.5 µl of ribonuclease inhibitor, 4 µl of isolated RNA, 1 µl of Oligo 7 µl of nuclease free water, 4 µl of GOScript™ 5X reaction buffer. Thermal cycling was carried out at 25 °C for 5 min for reverse transcription, 42 °C for 60 min for cDNA and then 70 °C for 15 min for reverse transcriptase inactivation. The nanodrop was used to determine the cDNA concentration considering optical density (O.D) at 260 nm and 280 nm wave lengths. dsDNA purity = $A_{260}$ / $A_{280.}$ A ratio between 1.8 and 2.0 indicated that the cDNA sample is pure. Complementary DNA was subjected to whole genome sequencing on the Illumina NovaSeq 6000 utilizing next generation sequencing (NGS) (Illumina, San Diego, CA, USA).

### 2.3. Nucleotide sequence Analysis and Data Quality

Raw sequencing data was obtained in FASTQ format and processed for variant calling and consensus sequences generation. First, the FASTQ sequences' quality, length and assembly were evaluated by the FastQC program, and then adapter sequences were trimmed from reads using Cutadapt software v.1.12. Processed reads were aligned to hCoV-19/Wuhan/Hu-1/2019 (GenBank accession number: NC 045512.2) is a reference genome from Wuhan, China using bowtie v.1.1.2. Variant calling and generation of consensus sequences were performed by using SAMTools v.1.9. The number of high confidence base calls (consensus sequence variants of the assembly) that disagree with the reference bases for the genomic site of interest, was used to identify single nucleotide polymorphisms. The nucleotide sequences of whole genome were analyzed using Empowering the Development of Genomics Expertise (EDGE) bioinformatics tools (https://edge-covid19.edgebioinformatics.org/). All three whole genome sequences of SARS-CoV-2 were uploaded to the Global Initiative for Sharing All Influenza Database GISAID and EPI ISL 3402236, EPI ISL 3402237, and EPI ISL 3402238 were the accession numbers [9].

### 2.4. Nucleotide Sequences and phylogenetic analysis

The three SARS-CoV-2 Iraqi strains whole genome sequences were submitted to the NCBI database under the accession codes MZ327801.1, MZ324812.1, and MZ330784.1. The Multiple Sequence Comparison by Log-Expectation (MUSCLE) software, which is part of the Molecular Evolutionary Genetics Analysis package version 10, was used to align multiple sequences (MEGA X) [10]. Phylogenetic analysis were performed using the Maximum Likelihood (ML) method for molecular evolution and phylogenetic analysis. The Kimura 2 parameter model was used for the calculation of genetic distance by nucleotide dataset alignment with 1000 bootstrap replicates. Since the purpose of our phylogenetic study was to discover the evolutionary links between our viral samples and the other SARS-CoV-2 viruses, the tree was rooted at the oldest virus, hCoV19/Wuhan/Hu-1/2019.

## 3. Results

The viral isolate discovered from patient-1 (hCoV-19/Iraq/Hula-02/2020, EPI ISL 3402237) was found to be hCoV-19/Iraq/Hula-02/2020, according to whole-genome sequencing on 14 December 2020, from Iraq /Al-Najaf-Al-Ashraf, belonged to the GV clade.

Whereas those from patient-2 (hCoV19/Iraq/Hula-01/2021, EPI ISL 3402236) on 7 February 2021 from Iraq/Salahuddin and patient-3 (hCoV19/Iraq/Hula-03/2021, EPI_ISL_3402238) on 26 February 2021 from Baghdad displayed the GH clade. The virus from patient-1 has 10 amino acid alterations in eight proteins, according to WGS, including Spike (A222V**,** D614G, L18F), N (A220V), NS3 (S117G), NS7a (A105S), NSP12 (P323L**)**, NSP13(S166L**)**, NSP14 (T516I) and NSP16 (R216N). Nine amino acid alterations in five proteins were discovered in patient-2 including (Spike A520S, D614G, N679K), N (S194L), NS3 (Q57H, R122I), NSP8 (R51C) and NSP12 (A185V, P323L). Also nine amino acid mutations in sex proteins were discovered in patient-3 including Spike (D614G, P681H), N (S194L), NS3 (G172R, Q57H, S180P), NS7a (Q94stop), NSP8 (R51C) and NSP12 (P323L). We aligned our genomic sequence with the SARS-CoV-2 NCBI Reference Sequence Wuhan (NCBI GenBank, NC_045512) to investigate variants throughout the genome. Not all single nucleotide variant (SNVs) causes amino acid changes from the sample.

### 3.1. **Uncommon and Unique mutation**

This study revealed that each protein has several mutations at the amino acid level, and there are very few uncommon mutations (Table 1). For instance, in the genome sequence of hCoV-19/Iraq/Hula-02/2020 EPI_ISL_3402237, a unique mutation (NS3 S117G 0.00% of all samples with NS3 sequence) was observed in nonstructural protein 3 at position 117. The hCoV-19/Ecuador/HGSQ-USFQ-010/2020 was the first isolate with this AA alteration reported in March 2020. This AA alteration was discovered most recently in the isolate of hCoV-19/Spain/CL-COV05691/2021, collected in June 2021. The particular isolate (hCoV-19/Iraq/Hula-02/2020 EPI_ISL_3402237) that had this mutation was collected in Iraq in December 2020. Similarly, another uncommon mutation (NS7a Q94stop) in ORF7a was seen in position 94 of sequence from isolate hCoV-19/Iraq/Hula-03/2021 EPI_ISL_3402238 2021-02-26 already occurred in 0.05% of all samples with nonstructural protein 7a (NS7a) sequence in 51 countries. The hCoV-19/Greece/51/2020 was the first isolate with this AA alteration, acquired in March 2020. The most recent AA modification was found in isolate hCoV-19/Denmark/DCGC-147502/2021, collected in August 2021. Interestingly, a mutation in Spike-N679K, a spike protein mutation, was seen in sequence from isolate hCoV-19/Iraq/Hula-01/2021EPI_ISL_3402236 2021-02-07, already occurred in 0.12% of all isolates containing the Spike sequence in 62 countries. The hCoV-19/USA/NY-NYUMC79/2020 was the first isolate with this AA alteration, acquired in March 2020. The most recent AA change was found in isolate hCoV-19/USA/NY-PRL-2021 0809 00F14/2021, collected in August 2021. This would mean that the virus strain in sample hCoV-19/Iraq/Hula 01/2021EPI_ISL_3402236 2021-02-07, with a mutation similar to that isolated in USA, was already circulating in Iraq after March 2020. Two patients with nonstructural protein 8 R51C mutations (NSP8 R51C) have the same variants. The current study found that these two patients were in the same lineage, B.1.36.1. The current study propose that this viral genotype similarity is due to the lineage spread.

**Table 1:** Detected uncommon and unique mutations

| Mutation | Gene | Type of mutation | Worldwide frequency | Number of cases in the world | Cases in this study |
|---|---|---|---|---|---|
| **T516I** | NSP14 | non synonymous | 0.08% | 60 | Patient-1 |
| **S117G** | ORF3a | non synonymous | 0.00% | 8 | Patient-1 |
| **Q94stop** | ORF7a | non synonymous | 0.05% | 51 | patient-3 |
| **R51C** | ORF8 | non synonymous | 0.07% | 58 | Patient -2, patient-3 |
| **A105S** | ORF7a | non synonymous | 0.04% | 49 | Patient-1 |
| **N679K** | Spike | non synonymous | 0.12% | 62 | Patient -2 |
| **A520S** | Spike | non synonymous | 0.2 % | 69 | Patient -2 |
| **S166L** | NSP13 | non synonymous | 0.14 % | 63 | Patient-1 |
| **S180P** | ORF3a | non synonymous | 0.19% | 33 | patient-3 |

### 3.2. Common Mutation

#### 3.2.1. Spike_D614G

The D614G variant of the SARS-CoV-2 Spik (aspartic acid (SD614) and glycine (SG614) at residue 614) showed improved replication and transmissibility. According to the mutation analysis, all our samples contained the D614G mutation in the spike glycoprotein, which is a common mutation in the Iraqi samples.

#### 3.2.2. Non-structural protein 12 (NSP12 P323L)

Effect on the non-structural protein 12 (NSP12), the viral RNA-dependent RNA polymerase. We observed that (P323L) mutation co-occurs with spike D614G mutation in all our samples. P323L has been found in 197 countries 2581232 times (95.53% of all samples with the NSP12 sequence). The first strain with this AA alteration was hCoV-19/Japan/20200409-129/2020, which was collected in January 2020. The most recent AA alteration was found in the strain hCoV-19/USA/KS-KHEL-4589/2021, which was collected in August 2021 (GISAID, 2021)

#### 3.2.3. N_S194L, N -A220V

This mutation is related to higher virulence, evolutionary ability and characteristics that are beneficial to the virus. However, whether this type of mutation will affect the function of the protein or even the infectivity of the virus needs to be further clarified.

### 3.3. Analysis of Phylogenetic and Evolutionary relationship

At the time of writing, full genome sequences for SARS-CoV-2 were obtained from the (http://www.gisaid.org/) and NCBI (http://www.ncbi.nlm.nih.gov/genbank) databases. All sequences that were duplicated as well as those that had missing areas were eliminated. The 33 sequences were chosen from 28 different nations. Iraqi SARS-CoV-2 genomes were dispersed separately in several groups in the phylogenetic tree in Figure 1, with strains mostly from Asia and Europe. GH clade viruses from Asia, the Middle Eastnd Europe grouped two out of three Iraqi genomes (EPI ISL 3402236 and EPI ISL 3402238). The viral sample (EPI ISL 3402237) was assigned to the clade GV viruses, which are mostly found in Europe (Germany, Estonia, England, Switzerland, Iceland, Italy and France).

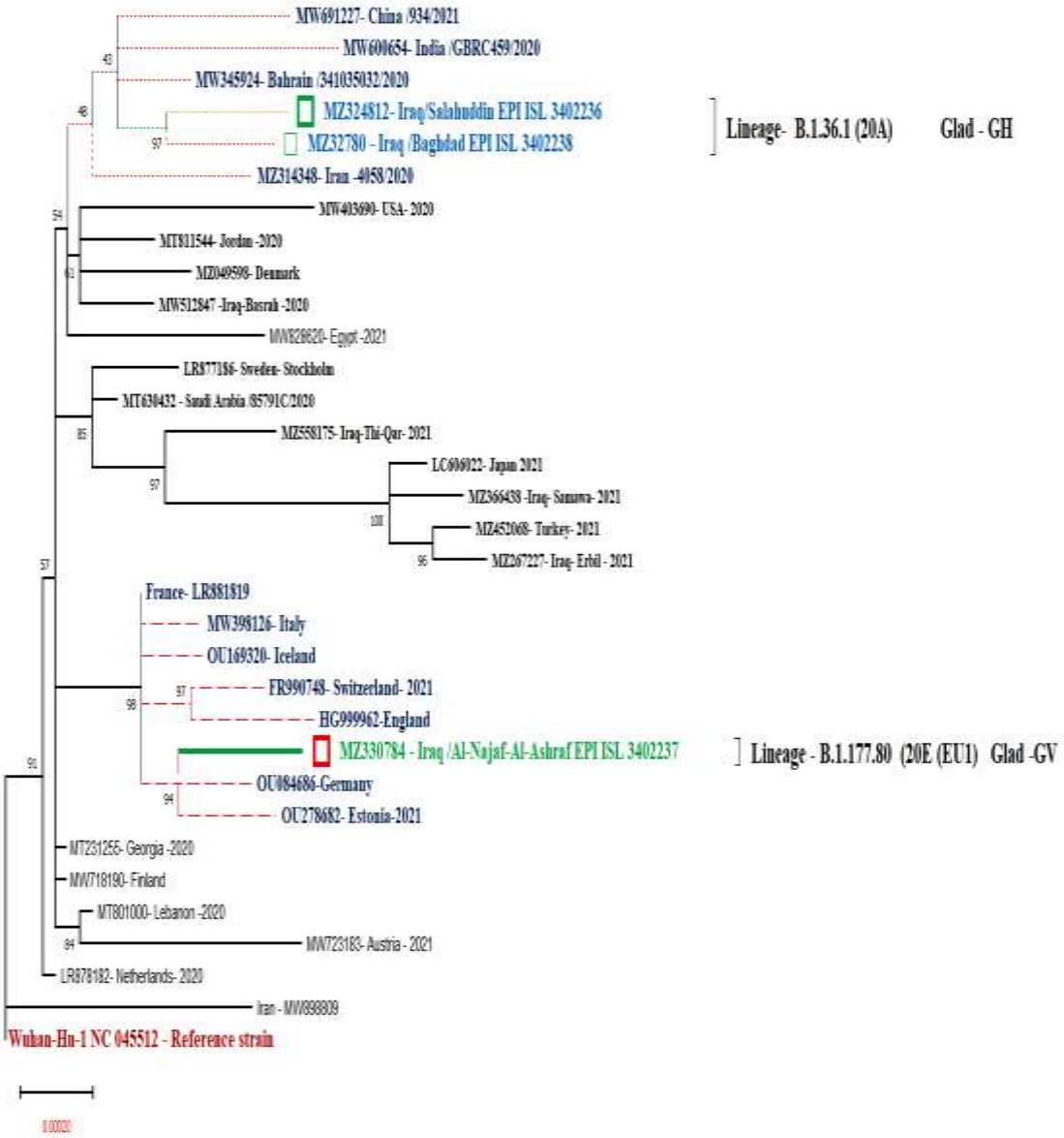


**Figure 1:** Phylogenetic tree comprises three SARS-CoV-2 genome sequences from this investigation (n = 28 complete genome sequences), as well as sequences from other nations. MEGA X software was used to rebuild the phylogenetic tree using the Maximum Likelihood technique bootstrap 1000 repetitions (Kumar, *et al*., 2018). SARS-CoV-2 genomes were subsampled globally and supplemented with additional complete genomes from Iraqneighboring countries. Iraqi strains in this study are denoted by blue and green arrows. The Wuhan reference genome (GenBank accession number: NC 045512.2) has a bigger typeface.

## 4. Discussion

The current study reports three complete genomes of SARS-CoV-2, taken from RT-PCR positive COVID-19 patients, originated from Al-Najaf-Al-Ashraf, Salahuddin and Baghdad provinces in Iraq. These sequences were uploaded to the Global Initiative for Sharing All Influenza Database (GISAID) to determine the global frequency of our genetic variations. We discovered that nine sporadic non synonymous mutations were based on the database results. The frequency of these uncommon mutations ranged from 0.00 to 0.19 percent. The mutations were not observed before in the samples isolated in Iraq. As a result, the uncommon

mutations found in these samples may be important for developing effective antiviral vaccines or therapeutics. The phylogenetic properties of this virus are important to contribute to the knowledge of viral mutations to specify the optimal region for use as a vaccine target or antiviral agent. Therefore, as a result, we uploaded our sequences to the GISAID database to determine the clades and lineages of our results. For this reason, all 33 complete and high coverage sequences isolated in different countries were selected to create a phylogenetic tree. Based on specific combination of genetic variations, the (GISAID) divides phylogenetic clusters into seven clades (S, L, V G, GH, GR, GV). Depending on the unique variant profile, two out of three of our samples (EPI_ISL_3402236 and EPI_ISL_3402238) collected in February 2021 belonged to lineage B.1.36.1 were clustered in GH (A20) clade of SARS-CoV-2 viruses from countries in Asia, the Middle East and Europe. However, the EPI_ISL_3402236 and EPI_ISL_3402238 genomes were in diverse roots from other strains because of their uncommon mutations. Whereas the other sample (EPI_ISL_3402237), collected in December 2020, belonged to lineage B.177.80 and clustered in the GV(E20) clade, that was closely linked to the other SARS-CoV-2 viruses in evolution, including those from Europe collected between August and December 2020, during the pandemic early months of disease transmission [10]. Through the current results it is believed that SARS-CoV-2 strains have been introduced into Iraq many times, especially through international travelers. Lineage B.1 is the most commonly known global lineage, and it has been subdivided into over 70 sub-lineages. According to our data, SARS-CoV-2 transmission in Iraq was impacted by frequent and recurrent introductions by international visitors, resulting in the circulation of different clades throughout time. Although governmental initiatives such as limits on indoor and outdoor gatherings may serve to limit and combat community transmission chains, the disease transmission dynamics have also been altered by regular introductions by overseas visitors. To get an exact estimate of SARS-CoV-2 diversity, greater epidemiological and genomic surveillance is needed. To identify genomic variants of SARS-CoV-2 sequences, we compared our sequences to the reference genome (NC 045512.2). According to the variant analysis, all of our cases had a D614G mutation in the spike glycoprotein. The predominant virus circulating in Iraq appears to be SARS-CoV-2 with the D614G mutation. Until August 2021, the D614G mutation will account for about 97.83 percent of all spike sequencing samples in 198 countries [8]. Based on sequence distribution and phylogenetic tree analysis, some studies suggest that the global domain of D614G is because of positive selection [11], whereas the domain of European D614G results from a founder effect [12]. According to the research, patients that have the D614G mutation are linked to higher virus loads in their respiratory tracts than SARS-CoV-2 patients without the mutation [11, 13]. Patients infected with SARS-CoV-2 and had the D614G mutation, frequently got moderate COVID-19 symptoms, whereas those who did not had mild symptoms [14-16]. As a result, by diminishing interactions between S1 and S2, this mutation may boost the viral infectivity of SARS-CoV-2 by improving the RBD attachment potential to human ACE2 [17]. Gained mutations can cause changes in host miRNA targeting, which can alter host immune responses during viral infection. Many variables determine the severity of COVID-19, including the patients' age, gender, comorbidities and immunological responses.

The first common SNP mutation in the SARS-CoV-2 genome is found in the leader sequence (241C>T), which is a critical genomic position for discontinuous sub-genomic replication [18]. We observed in our samples that (241C>T), 3037C>T, 14408C>T, and 23403A>G mutations co-occur in all samples, resulting in non-synonymous mutation in nsp3, RNA primerase (P323L) and amino acid mutations in spike glycoprotein (protein S, D614G). In viral isolates from Europe, where SARS-CoV-2 infections are more severe than in other

regions of the world, these four co-mutations are widespread. These four co-mutations, when combined, are likely to make the virus more transmissible [18]. In one study, it is reported that 549 mutations and 53 unique variants in 47 isolates. that study found that D614G is the most frequent mutation found in spike glycoprotein, which is like our findings [19]. Despite the fact that the D614G mutation is growing increasingly frequent in strains throughout the world, we discovered uncommon mutations in spike glycoprotein (Spike A222V, L18F, A520S, N679K, P681H) in our strains. Monitoring the circulation of these mutant strains in Iraq requires additional molecular epidemiological studies. This might also explain why SARS-CoV-2 genetic variations have such an influence on illness severity and clinical management. In addition, detecting alterations in the spike glycoprotein of SARS-CoV-2 is critical not only for inducing neutralizing antibodies but also for their function in receptor binding and penetration into host cells. After the detection of the G614 mutant and its worldwide spread, there was a major concern about the effect of the D614G mutation on vaccine efficacy as most vaccines were designed using the D614 virus [20]. However, it is vital to monitor SARS-CoV-2 spike protein genetic variants to discover possible escape mutations and future vaccine research programs. The P323L (C14408T) RdRp (viral RNA-dependent RNA polymerase) mutation, affecting NSP12, was found in all our sequenced isolates, and it is often seen in isolates from Europe and North America [21]. When compared to SARS-CoV variations, the D614G in Spike protein and the P323L in RNA polymerase discovered in novel variants, may boost SARS-CoV-2 infectivity [22]. Multiple points imported from Europe (UK, Italy, France) are also included in this data . Since the encoded polymerase is a significant target for current therapeutic polymerase inhibitors, mutations in RdRp are intriguing [22], and our study reveals a new mutation in nsp12.

## 5. Conclusion

The SARS-CoV-2 isolates collected from three Iraqi COVD-19 patients originated from different regions that belonged to GH and GV clades. These strains also have several mutations as compared to the listed mutations in the nonstructural proteins and spike protein, which can affect their characteristics as compared to the first hCoV19/Wuhan/NC 045512.2 / 2019 strain. The SARS-CoV-2 genome sequence of Iraqi patients is like that of the world. However, we can assume that the sources of SARS-CoV-2 in Iraq are from India, Iran, Europe and the United States. Due to the lack of information related to the patient's travel/contact history; it is impossible to make further inferences and obvious conclusions about the origin and source.

## Funding

This study received no particular support from governmental, private, or non-profit funding bodies.

## Conflict of interest

There were no conflicts of interest declared by the authors.

## Acknowledgments

The authors sincerely appreciated the kind medical personnel's collaboration at the National Central Public Health Laboratory in Baghdad.

## References

**[1]** G. Li *et al*.**,** "Coronavirus infections and immune responses", J *Med Virol*, vol. 92, no. 4, pp. 424-432, 2020. Available: doi: 10.1002/jmv.25685.

**[2]** M. Iyer *et al*.**,** "COVID-19: An update on diagnostic and therapeutic approaches" *BMB Rep*, vol. 53, no. 4, pp. 191-205, 2020. Available: doi: 10.5483/BMBRep.2020.53.4.080.

**[3]** P.Asrani *et al*., " Molecular Basis of Pathogenesis of Coronaviruses: A Comparative Genomics

Approach to Planetary Health to Prevent Zoonotic Outbreaks in the 21st Century", *OMICS: A Journal of Integrative Biology*, vol. 24, no. 11, pp, 2020. 634-644. Available: doi.org/ 10. 1089/omi.2020.0131.

**[4]** Y. Shu *et al*.**,** "GISAID: Global Initiative on Sharing All Influenza Data—From Vision to Reality", *Euro Surveill*, vol. 22, no. 13, pp. 30494, 2017. Available: 10.2807/1560-7917.ES.2017.22.13.30494.

**[5]** X.Tang *et al*.**,** " On the origin and continuing evolution of SARS-CoV-2", *National Science Review*, vol. 7, no. 6, pp. 1012–1023, 2020. Available:10.1093/nsr/nwaa036.

**[6]** C.Bhattacharyya *et al*., " Global spread of SARS-CoV-2 subtype with spike protein mutation D614G is shaped by human genomic variations that regulate expression of *TMPRSS2* and *MX1* genes", BioRxiv, 2020. Available: 10.1101/2020.05.04.075911.

**[7]** Z. Mahmood *et al*., "Estimation of Hematological Parameters of Disease Severity in Iraqi Patients with COVID-19" *Iraqi Journal of Science*, vol. 62, no. 10, pp: 3487-3496, 2021. Available: 10.24996/ijs.2021.62.10.8.

**[8]** GISAID, Pandemic coronavirus causing COVID-19 [Online], 2021.Available at https://platform. gisaid.org/epi3/cfrontend#8dc5e.

**[9]** S. Kumar *et al*., "MEGA X: Molecular Evolutionary Genetics Analysis across computing platforms", *Mol Biol Evol*, vol. 35, no. 6, pp. 1547-1549, 2018. Available: 10.1093/molbev /msy096.

**[10]** Gisaid-Clade, "GISAID - Clade and Lineage Nomenclature Aids in Genomic Epidemiology of Active hCoV-19 Viruses" (2020) [Online]. Available: https://www.gisaid.org/references /statements-clarifications/clade-and-lineage-nomenclature-aids-in-genomic-epidemiology-of-active-hcov-19-viruses.

**[11]** B. Korber *et al*., "Tracking changes in SARS-CoV-2 spike: evidence that D614G increases infectivity of the COVID-19 virus", *Cell*, vol. 182, no. 4, pp. 812–827, 2020. Available: 10.1016/j.cell.2020.06.043.

**[12]** B. Dearlove *et al*., "A SARS-CoV-2 vaccine candidate would likely match all currently circulating variants", *Proceedings of the National Academy of Sciences of the United States of America*, vol. 117, no. 38, pp. 23652–23662, 2020. Available:10.1073/pnas.2008281117.

**[13]** L. Zhang *et al*., "The D614G mutation in the SARS-CoV-2 spike protein reduces S1 shedding and increases infectivity, *BioRxiv*, 2020. Preprints 20200612. doi:10.1101/2020.06.12.148726.

**[14]** J. Hu *et al*., "The D614G mutation of SARS-CoV-2 spike protein enhances viral infectivity", *BioRxiv*, 2020. Available: https://doi.org/10.1101/2020.06.20.161323.

**[15]** S. Cleemput *et al*., "Genome Detective Coronavirus Typing Tool for rapid identification and characterization of novel coronavirus genomes", *Bioinformatics*, vol. 36, pp. 3552-3555, 2020. Available: 10.1101/2020.01.31.928796.

**[16]** R. Lorenzo-Redondo *et al*., "A unique clade of SARS-CoV-2 viruses is associated with lower viral loads in patient upper airways", *medRxiv*, Preprint. 2020. Available: 10.1101/2020.05 .19.20107144.

**[17]** Z. Mahmood *et al*, "Severity of coronavirus disease 19: Profile of inflammatory markers and ACE (rs4646994) and ACE2 (rs2285666) gene polymorphisms in Iraqi patients" *Meta Gene*, vol. 31, pp. 1-6, 2022. Available: 10.1016/j.mgene.2022.101014.

**[18]** C. Yin, "Genotyping coronavirus SARS-CoV-2: methods and implications", *Genomics*, vol. 112, no. 5, pp. 3588-3596, 2020. Available: 10.1016/j.ygeno.2020.04.016.

**[19]** M. Karamese *et al*., "Molecular Characterization, Phylogenetic and Variation Analyzes of SARS-CoV-2 strains in Turkey", *Future Microbiol*. 2021. Preprints 20200911. Available:10.2217/fmb-2021-0118.

**[20]** N. Lurie *et al*., "Developing Covid-19 Vaccines at Pandemic Speed", *N. Engl. J. Med.*, vol. 382, pp. 1969–1973, 2020. Available: 10.1056/NEJMp2005630.

**[21]** M. Pachetti *et al*., "Emerging SARS-CoV-2 mutation hot spots include a novel RNA-dependent-RNA polymerase variant", *Journal of Translational Medicine*, vol. 18, no. 1, pp.179, 2020. Available: 10.1186/s12967-020-02344-6.

**[22]** MA**.** Sirwan *et al*.**,** "Rapid, inexpensive methods for exploring SARS CoV-2 D614G mutation", *Meta Gene*, vol. 30, no. 4, 2021. Available: 10.1016/j.mgene.2021.100950.