Arch Biol Sci. 2020;72(4):575-585 https://doi.org/10.2298/ABS201024051H

Infection with different strains of SARS-CoV-2 in patients with COVID-19

Hayder O. Hashim[1], Mudher K. Mohammed[2], Mazin J. Mousa[1], Hadeer H. Abdulameer[3], Alaa T.S. Alhassnawi[4], Safa A. Hassan[5] and Mohammed Baqur S. Al-Shuhaib[6,*]

[1]*Department of Clinical Laboratory Sciences, College of Pharmacy, University of Babylon*, Babil 51001, Iraq
[2]*Department of Pharmacy, Al-Manara College of Medical Science*, Iraq
[3]*University of Kufa, Faculty of Education for Girls*, Iraq
[4]*Department of Biology, College of Science, University of Babylon*, Babil 51001, Iraq
[5]*Alfadhel Training and Development Company / Babylon Branch*, Iraq
[6]*Department of Animal Production, College of Agriculture, Al-Qasim Green University*, Al-Qasim, Babil 51001, Iraq

***Corresponding author**: mohammed79@agre.uoqasim.edu.iq; baquralhilly_79@yahoo.com

The manuscript is available as a preprint at the following web server address: https://www.preprints.org/manuscript/202009.0375/v1, which received the following DOI: https://doi.org/10.20944/preprints202009.0375.v1.

Received: October 24, 2020; **Revised**: November 10, 2020; **Accepted**: November 13, 2020; **Published online**: November 13, 2020

Abstract: The biological diversity of SARS-CoV-2 was assessed by investigating the genetic variations of the spike glycoprotein of patients with COVID-19 in Iraq. Sequencing identified fifteen novel nucleic acid variations with a variety of distributions within the investigated samples. The electropherograms of all identified variations showed obvious co-infections with two different viral strains per sample. Most samples exhibited three nonsense single nucleotide polymorphism (SNPs), p.301Cdel, p.380Ydel and p.436del, which yielded three truncated spike glycoproteins, respectively. Network and phylogenetic analyses indicated that all viral infections were derived from multiple viral origins. Results inferred from the specific clade-based tree showed that some viral strains were derived from European G-clade sequences. Our data demonstrated the absence of single-strain infection among all investigated samples in the studied area, which entails a higher risk of SARS-CoV-2 in this country. The identified high frequency of truncated spike proteins suggests that defective SARS-CoV-2 depend on helper strains possessing intact spikes during infection. Alternatively, another putative ACE2-independent route of viral infection is suggested. To the best of our knowledge, this is the first report to describe co-infection with multiple strains of SARS-CoV-2 in patients with COVID-19.

Keywords: co-infection; SARS-CoV-2; COVID-19; spike glycoprotein; stop mutations

INTRODUCTION

Coronavirus disease 2019 (COVID-19) is a pandemic of a severe acute respiratory syndrome (SARS) caused by the novel coronavirus SARS-CoV-2 that was first reported in Wuhan, China in late 2019. These viral particles are attributed to the beta coronaviridæ family, which is responsible for the highest infectivity and pathogenesis in contemporary history [1,2]. Since its emergence as a newly identified pathogenic virus, SARS-CoV-2 has exhibited a rapid spread in a relatively short time scale with a catastrophic global pandemic sequel on healthcare systems and economies, with multifarious social aspects.

Despite preventive measures imposed by many governments around the world, a discrepancy in the cumulative mortality rates between states in America and even between western European countries by mid-June 2020 has been observed. This limitation in viral confrontation can be attributed to many causes, such as failure to adhere to measures of social distancing, the presence of concomitant chronic diseases strongly correlated to mortality (comorbidities), the preparedness of health systems to cope with the pandemic, and other related factors [3]. Despite the rapid progress of phylogenetic analysis and genetic mapping of SARS-CoV-2, the incidence of concomitant mixed infections

with different strains has not been reported [4]. SARS-CoV-2 is similar to SARS-CoV, which resulted in an epidemic in China in 2002. Nucleotide substitutions have been suggested as the major route of natural evolution through which SARS-CoV has developed to SARS-CoV-2. However, both viral entities use their spike (S) glycoprotein to recognize the angiotensin type 2 (ACE2)-converting enzyme on cell surfaces for subsequent internalization into the host cytoplasm. The high prevalence of SARS-CoV-2 infection and the acceleration of infection spread raises many questions concerning the emergence of new strains by mutations, however, it has not been stated whether these developed viral strains are more virulent than the original SARS-CoV strain [5]. Early results obtained from analysis of the genetic sequence showed the presence of several mutations in SARS-CoV-2 compared to the genetic sequence of the original SARS-CoV viral strain from Wuhan [6]. Furthermore, new strains have been registered with higher virulence than the original Chinese SARS-CoV-2 isolates. These newly diagnosed strains emerged as the result of a point mutation of D614G that causes changes in the viral spike glycoprotein by rendering it more resistant to cleavage during viral replication [7]. More than 30 distinguishable mutations in different sites of the coding region of SARS-CoV-2 genomic sequences have been reported, with viral strains possessing these mutations and exhibiting a distinct distribution in different locations around the world [8].

It should be noted that there are several alternative routes of viral entry into human cells other than through hooking to ACE2, for example through sialic acid [9], CD147-spike protein [10], CD 209L glycoprotein [11] and by endocytosis [12]. These mechanisms may enhance virulence by acting as substitutes for spike-ACE2 interaction and viral internalization. Though mutations involving the receptor-binding domain can generate inactive spike proteins with decreased viral virulence, the coexistence of multiple strains with the potential to simultaneously infect the same individuals could resolve this issue. Recent genetic analyses of the genomic sequences of SARS-CoV-2 reported that these viruses evolved into two types of infections that are defined by two different genetic variations across the viral strains sequenced to date [13]. Variable severity of selective pressure may be applied to some of these strains. Thus, major, minor, or both types of viral infections can emerge [14]. These findings strongly support the urgent need for studies that combine both genomic data and epidemiological data of patients with COVID-19 worldwide. It is well known that this pattern of infection is called concurrent infection or co-infection [15]. Co-infections can be due to either horizontal transmission of multiple strains or repeated extensive exposure before recovery [16]. The interaction of multiple strains with each other during co-infection can lead to potential differences in epidemiological behavior [17]. This alteration can include the incidence of a cyclical pattern and strain replacement, which can enhance viral co-infectivity through several postulated mechanisms [18-20]. Given the rapid spread of COVID-19, investigations are underway to better understand the genetic variation underlying the emergence of the original outbreak, starting from Wuhan city by these viral particles. Accordingly, the investigation of mutation rates of SARS-CoV-2 in places far removed from Wuhan is worthwhile. This study was designed to characterize the genetic variations of SARS-CoV-2 transmitted in Iraq, a country with an ongoing acceleration of COVID-19 cases since the beginning of 2020.

MATERIALS AND METHODS

Viral samples

Nineteen nasopharyngeal swabs from patients with COVID-19 were kindly provided by the medical staff of the Iraqi Health Laboratory, Ministry of Health. Detection of the presence of SARS-CoV-2 was conducted by the medical staff of the Babylon Central Health Laboratory using the COVID-19 SARS-CoV-2 Real-TM kit (Sacace, Italy). Swabs were collected in universal viral transport medium (Capricorn Scientific, Germany) following the World Health Organization (WHO) guidelines. No personal data of patients were provided, and the specimens were labeled with random numbers so that test results could not be linked to the involved individuals. Ethics approval was not sought because the collected viral swaps were part of laboratory validation for these patients. The secondary use of the anonymous pathological specimens was certified by the Ministry of Health Review Board (1318/24-08-2020).

RNA extraction and reverse transcription

For RNA extraction from nasopharyngeal swabs, a nucleic acid purification kit based on the magnetic bead method was used following the manufacturer's instructions (Cat No. DA0630, DAAN Gene, China). The purity and quantity of the viral RNA were determined using NanoDrop (Biodrop, μLITE, UK). For downstream studies, samples with a 260 nm:280 nm ratio higher than 1.7 were used in subsequent analysis. The first-strand cDNA was synthesized using AccuPower® RocketScriptTM RT PreMix (Bioneer, South Korea). The RT-PCR experiment was started using 2 μL of total RNA, 100 pmols of random hexamer, in a final reaction volume of 20 μL. Thermocycling was conducted using the Biometra TRIO thermal cycler (Analytik Jena, Germany). Primer annealing was at 15°C for 10 min, cDNA synthesis was at 42C° for 30 min, and heat inactivation was at 95°C for 5 min.

PCR

One PCR fragment of 795 bp was designed in this study using NCBI Primer-BLAST software [21]. This fragment partially covered 264 amino acids, starting from 217 to 480 amino acid residues, of the viral spike glycoprotein of SARS-CoV-2. The sequences of the forward and reverse primers were 5'-CCCTCAGGGTTTTTCGGCTT-3', and 5'-TTACAAGGTGTGCTACCGGC-3', respectively. Lyophilized oligonucleotides were purchased from Bioneer (Daejeon, South Korea). The PCR amplification reactions were conducted using the Bioneer PCR reaction mixture. The amplification conditions for the designed amplicon were empirically optimized using a PCR thermocycler (Nexus, Eppendorf, Hamburg, Germany). The amplification reaction was started by one cycle of denaturation at 95°C for 5 min, followed by 35 cycles of denaturation at 95°C for 30 s, annealing at 59°C for 30 s, and elongation at 73°C for 60 s, and finalized by polymerization at 73°C for 5 min. The PCR products were verified by electrophoresis on a 1.5% agarose gel.

DNA sequencing

All amplified fragments were subsequently exposed to sequencing reactions following the instructions of Macrogen Laboratories (Macrogen, Geumchen, South Korea). The database of the S gene nucleic acids was retrieved from the NCBI server (GenBank accession number MT847228.1). The retrieved sequences were then visualized and annotated by BioEdit ver, 7.1. (DNASTAR, Madison, USA). The correct reading frames of the viral variants in the analyzed spike glycoprotein were determined using the ExPASy translate tool [22]. The electropherograms of all identified nucleic acid variations were manually validated using SnapGene Viewer ver. 4.0.4 (http://www.snapgene.com). Only clear electropherograms were considered for variations of the SARS-CoV-2 sequences. The appropriate reading frame of the observed variants was aligned with its corresponding reference sequences within the viral glycoprotein using the UniProtKB server [23].

In silico prediction

The 3D structure of the entire amino acid residue of the spike glycoprotein 6VXX was retrieved from the protein data bank (PDB) webserver (https://www.rcsb.org/). The effects of the identified missense variants on spike structure, function and biological stability were predicted using state-of-the-art *in silico* tools, SIFT [24], PROVEAN [25], PolyPhen-2 [26], SNAP2 [27], ConSurf [28], and mCSM [29]. The cumulative outcomes of the utilized tools were assessed for each identified missense SNP.

Data analysis

Based on the nucleic acid haplotypes generated by the DNA sequence polymorphism (DnaSP) software ver. 6.12.01 [30], the exact distances between haplotypes were matched by a tight-span walker network using population analysis with reticulate tree (PopART) software v. 4.8.4 [31]. The phylogenetic relationships among the identified viral variations were generated by constructing a neighbor-joining-based unrooted phylogenetic tree. All the reference sequences incorporated within this tree were retrieved from the NCBI database (https://www.ncbi.nlm.nih.gov/). Subsequently, another unrooted neighbor-joining tree was constructed to assess the exact clade of the investigated viral strains. In this specific tree, the five main known (G, L, S, O and V) viral clades were incorporated alongside our investigated sequences. The incorporated clades were retrieved from the GISAID database to represent several

Asian, European, African, and American countries [32]. Both trees were annotated by the iTOL server [33]. Linkage disequilibrium (LD) analysis between the detected SNPs of the SARS-CoV2 sequences was conducted using Haploview software, ver. 4.2 [34]. LD values were estimated using squared allele frequency correlations (r^2) between pairs of the analyzed loci.

RESULTS

Genetic polymorphism and prediction

All investigated samples showed co-infection with more than one strain, as all identified variations in the electropherograms were found to exhibit two peaks per individual variant (Fig. 1). Co-infection for all identified variants was confirmed in all cases. Viral strains with major electropherogram reads were designated as type-A strains, while those exhibiting minor reads were designated as type-B strains. A total of fifteen nucleic acid substitutions in the coding region of the viral spike protein were identified in both type-A and type-B strains. Most of the identified mutations were represented by eight missense SNPs (p.274T>S, p.291C>G, p.346R>K, p.357R>I, p.357R>K, p.391C>G, p.408R>K, and p.452L>R), and four silent SNPs (p.316S=, p.366S=, p.394N=, and p.399S=) (Table 1). Interestingly, the remaining three SNPs (p.301Cdel, p.380Ydel, and p.436del) caused premature termination at three separate positions in the entire structure of the mature spike glycoprotein (protein ID QMU94980.1). The observed variations were deposited in the NCBI database under GenBank accession numbers MT936988-MT937006 for type-A viral strains, and MT937007-MT937025 for type-B viral strains.

Cumulative *in silico* prediction was conducted for all identified missense SNPs in the viral spike glycoprotein using both sequence- and structure-based tools. All prediction tools revealed highly deleterious effects of the two missense SNPs on the structure, function and stability of the viral spike protein, namely p.291C>G, and p.391C>G, while the other missense SNPs pointed to less deleterious effects on the analyzed viral protein (Table 2).

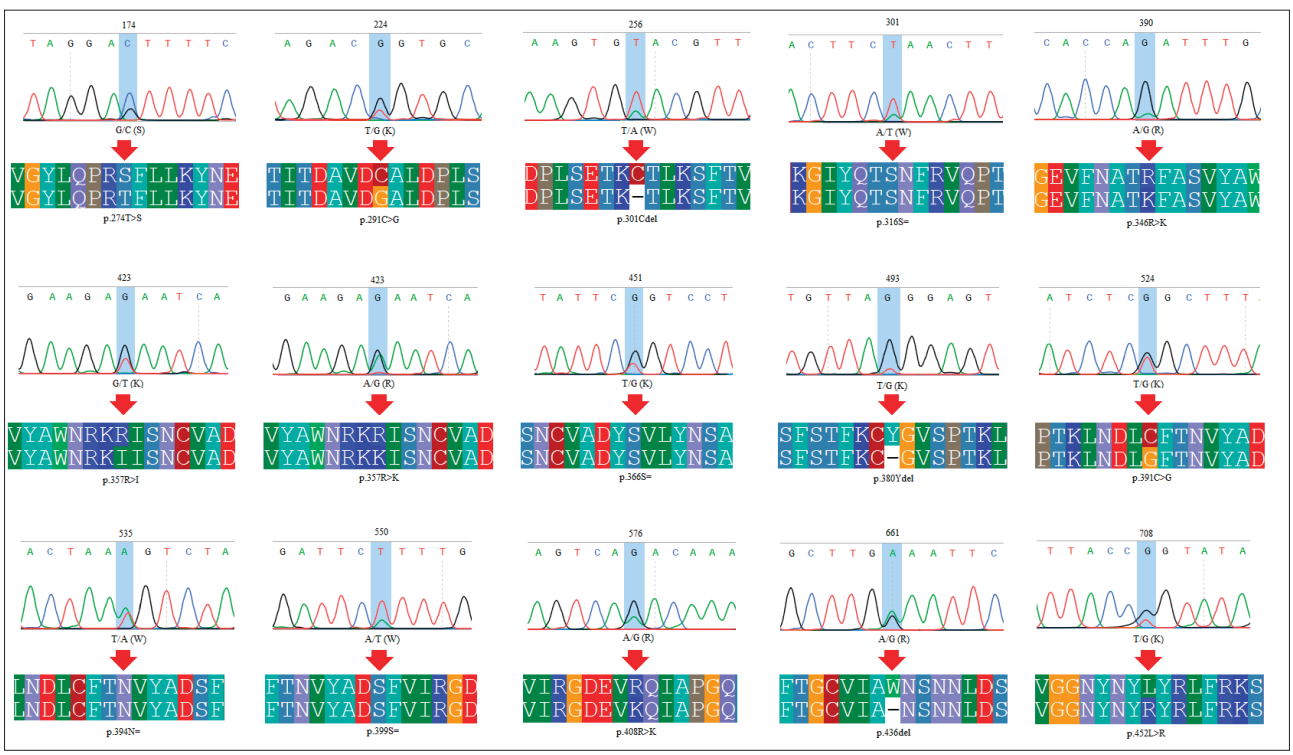

Fig. 1. Electropherograms of the detected nucleic acid substitutions in 795 bp amplicons showed clear co-infections with all observed variations in the S gene contained in different SARS-CoV-2 strains. The final effect (silent, missense, nonsense) of each identified SNP is indicated by red arrows.

Table 1. Nucleic acids and the amino acid variations of SARS-CoV-2 samples within the entire structure of the spike glycoprotein.

Nucleotide position	Nucleotide variation in the reference sequence	Sample No.	Mixed infection type	Amino acid position	Amino acid change	Type of SNP	SNP summary
174	G/C	S2, S7, S14, S19	Type-B	T274	T>S	Missense	c.174C>G (p.274T>S)
224	T/G	S1, S5, S9, S15, S17-S19	Type-A	C291	C>G	Missense	c.224T>G (p.291C>G)
256	T/A	S2-S5, S7-S9, S12, S14, S16, S18	Type-B	C301	C>-	Nonsense	c.256T>A (p.301Cdel)
301	A/T	S2, S4, S5, S7, S9, S11-S19	Type-B	S316	S=	Silent	c.301T>A (p.316S=)
390	A/G	S4, S6-S9, S11, S12, S15-S18	Type-B	R346	R>K	Missense	c.390G>A (p.346R>K)
423	G/T	S4, S8, S10, S11, S15, S16, S18, S19	Type-B	R357	R>1	Missense	c.357G>T (p.357R>I)
423	G/A	S1-S3, S5-S7, S9, S13, S14, S17	Type-B	R357	R>K	Missense	c.423G>A (p.357R>K)
451	T/G	S7, S9, S11, S13, S18, S19	Type-A	S366	S=	Silent	c.451T>G (p.366S=)
493	T/G	S1-S6, S8-S17	Type-A	Y380	Y>-	Nonsense	c.493T>G (p.380Ydel)
524	T/G	S3, S6, S7, S9, S11, S12, S17, S19	Type-A	C391	C>G	Missense	c.524T>G (p.391C>G)
535	T/A	S3, S8, S17-S19	Type-A	N394	N=	Silent	c.535T>A (p.394N=)
550	A/T	S11, S14, S15, S19	Type-A	S399	S=	Silent	c.399A>T (p.399S=)
576	A/G	S1-S3, S6, S9, S14-S17	Type-B	R408	R>K	Missense	c.576G>A (p.408R>K)
661	G/A	S1, S2, S6-S10, S13, S15-S18	Type-A	W436	W>-	Nonsense	c.661G>A (p.436del)
708	T/G	S1, S2, S4, S7-S10, S17-S19	Type-A	L452	L>R	Missense	c.708T>G (p.452L>R)

Table 2. *In silico* analysis of the deleterious effects of the observed missense SNPs in the S gene contained in SARS-CoV-2 using several different bioinformatics tools.

In silico tool	T274S	C291G	R346K	R357I	R357K	C391G	R408K	L452R
SIFT	Neut	Del	Neut	Neut	Neut	Del	Neut	Del
PROVEAN	Neut	Del	Neut	Neut	Neut	Del	Neut	Neut
PolyPhen-2	Del	Del	Neut	Del	(Neut)	Del	Del	Neut
I-Mutant2	Del	Del	Del	Del	Del	Del	Del	Del
PhD SNP	Neut	Del	Neut	Neut	Neut	Del	Neut	Del
SNAP2	Neut	Del	Neut	Neut	Neut	Del	Del	Neut
ConSurf	Del	Del	Neut	Neut	Neut	Del	Del	Neut
mCSM	Del	Del	Del	Neut	Del	Del	Del	Del

Del –deleterious effect of the missense SNP; Neut –neutral effect of the missense SNP.
Prediction outcomes of PolyPhen-2 were used to classify the missense SNP as benign, possibly damaging, or probably damaging, according to the score ranging from 0 to 1, respectively. Values >0.5 are predicted as deleterious.

Network analysis

A specific tight span walker network was generated from haplotype data of the observed variants of both types of identified viral strains. This network was developed to describe the interrelationship between the identified variant haplotypes from one side, and the reference viral sequences from the other side. A high level of diversity of type-A (Hap 2 to Hap 20) and type-B (Hap 21 to Hap 34) viral infections was inferred from the generated network. The results of this network indicated the presence of multiple viral sources for all identified haplotypes within the co-infections. These multiple sources or Hap 1 sequences were not restricted to a particular country, as both type-A and type-B strains were derived from viral strains of many countries on different continents (Fig. 2).

Phylogenetic analysis

Further clarifications of the observed network-based data were made by constructing a neighbor-joining phylogenetic tree. Another layer of confirmation of the multiple sources of the investigated viral strains was demonstrated. The precise phylogenetic positions of the investigated viral strains with related SARS-CoV-2 sequences were identified (Fig. 3A).

Due to the high number of identified SNPs of the investigated viral SARS-CoV-2 strains, both type-A and type-B sequences occupied unique positions within the known SARS-CoV-2 sequences deposited in the NCBI database.

In addition to type-A and type-B viral sequences, the number of incorporated organisms within the tree was 151 and all were attributed to SARS-CoV-2. Type-A viruses were clustered together in one clade, and type-B viruses also exhibited the same pattern

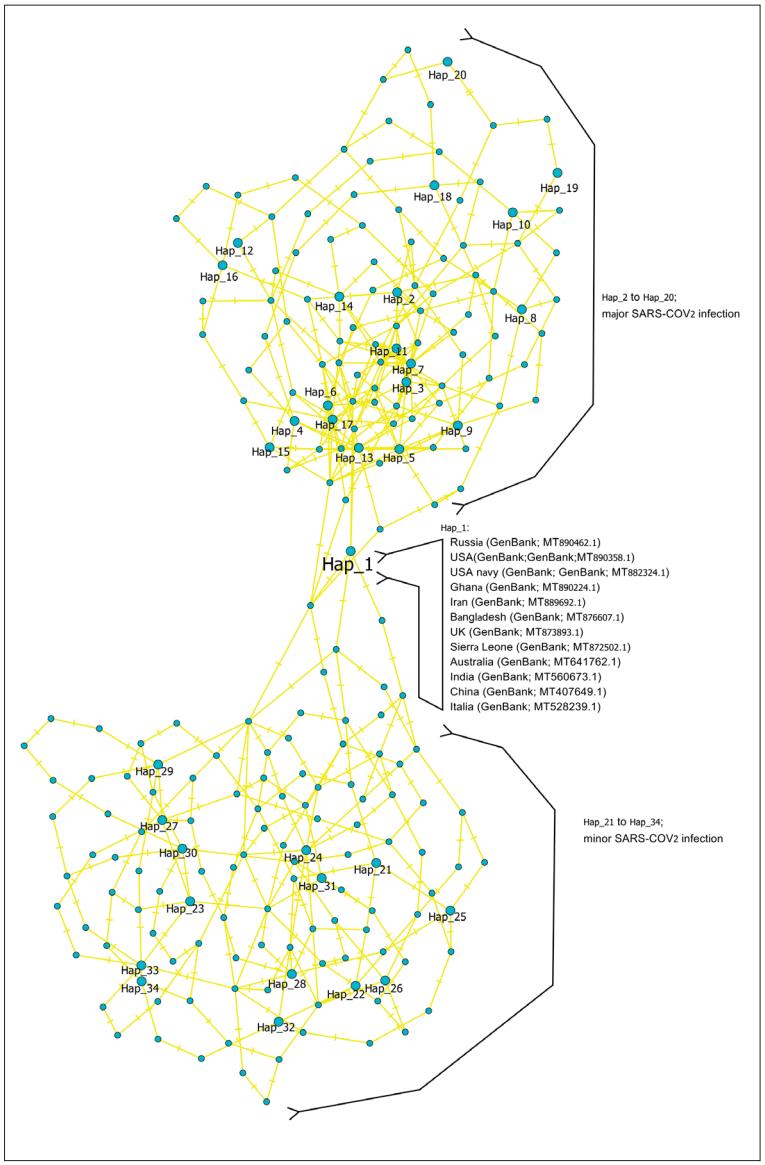

Fig. 2. Haplotype network analysis of SARS-CoV-2 viruses. Hap 1 represents the deposited reference sequence from different regions around the world. Hap 2 to Hap 20 represent the major, type-A infection; Hap 21 to Hap 34 represent the minor, type-B infection.

of positioning in an adjacent clade. Therefore, our neighbor-joining phylogenetic data confirmed the separation of the identified type-A and type-B viruses from each other. The tree-based results were validated by network-based data, which also proved the close positioning of both types of viral infections in the tree. This tree showed that both investigated viral types were derived from mixed evolutionary viral sources which were equally shared among many viral sequences without being attributed to a particular country, continent or race. However, the data obtained from the developed phylogenetic tree was not sufficient to identify the clade from which our viral samples were descended. Accordingly, representative samples for the five main known viral clades (G, L, O, V and S) were constructed from the GISAID database and aligned side by side with our samples. Based on these clades, several pieces of evidence were obtained from this tree. Most importantly, three strains of type-A, S1A, S3A and S6A originated from the viral representative of the G clade (Fig. 3B). Other clear-cut data were not recognized and all other strains were represented by a unique clade outside the five commonly known clades of SARS-CoV-2. Both trees suggested that while the two observed viral types were closely related to each other, they were distantly related to the other strains within the tree.

Linkage disequilibrium analysis

More details concerning the potential linkage across viral strains were obtained using data retrieved from major type-A and minor type-B viruses identified in this study. To assess the extent of homologous recombination among the investigated viral strains, LD plots of both A and B viral types were developed. In the type-A variants, population genetic data showed only one close linkage between 145 and 493 loci in the amplified 795 bp fragment, while the other loci of the type-A variants exhibited low R-squared values (Fig. 4A). Since no close linkage was observed among the identified SNPs, high recombination rates of these SNPs were confirmed. The same pattern of distribution of SNPs was also demonstrated for type-B variants in which no linkage among SNPs was shown (Fig. 4B). Since most variants were at very high frequencies, it is not surprising that many pairs had very low r^2 values. Accordingly, high recombination among almost all investigated SARS-CoV-2 strains was confirmed.

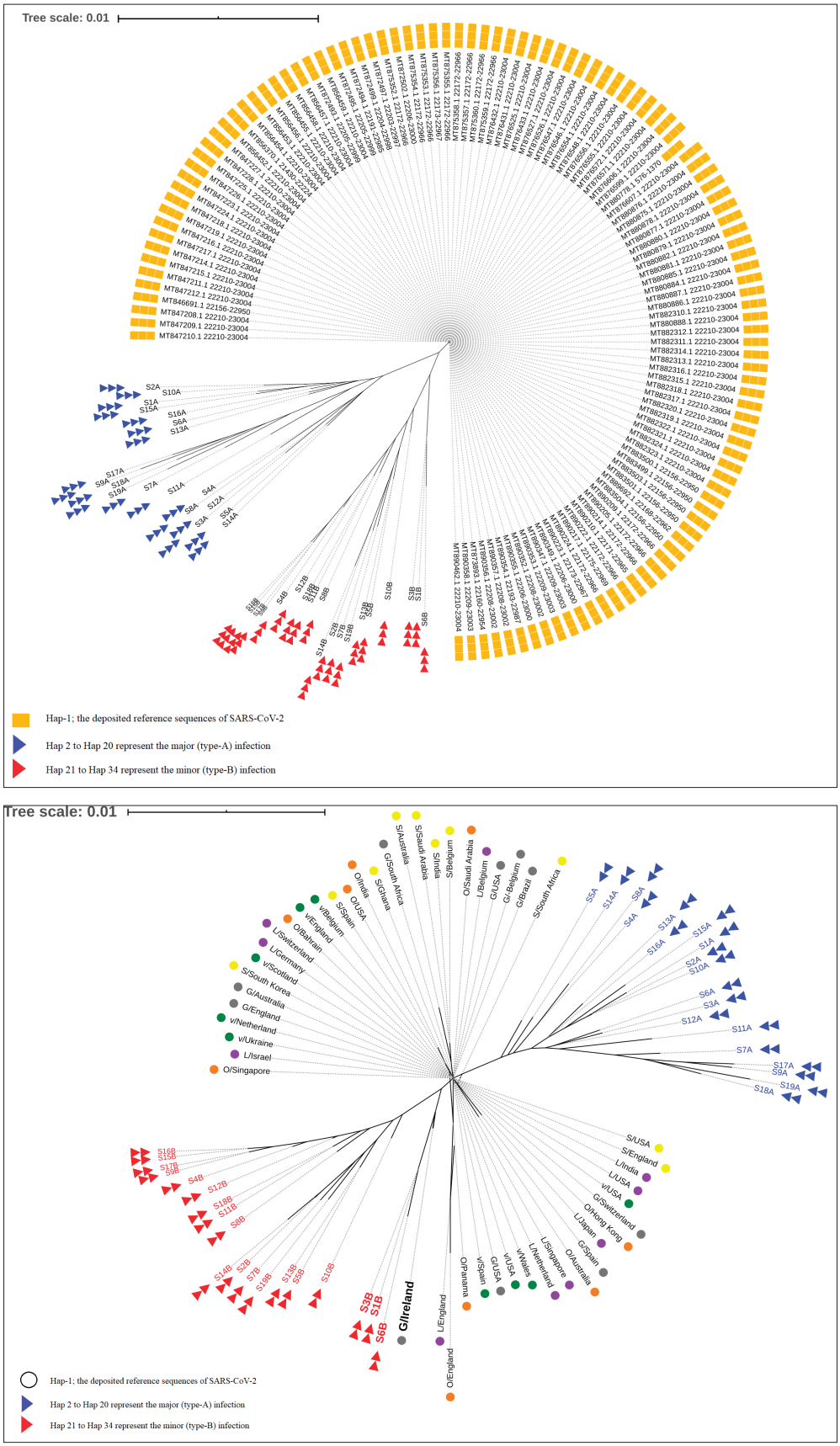

Fig. 3. NCBI-based and GISAID-based phylogenetic trees of genetic variants of the S gene for SARS-CoV-2 Iraqi strains. **A** – in the NCBI-based tree, the nucleic acid variations were identified among the investigated viral strains. The blue and red color triangles refer to type-A (S1A-S19A) and type-B (S1B-S19B) variants, respectively. The dark yellow colored boxes refer to the deposited NCBI GenBank accession numbers of reference sequences of SARS-Cov2. **B** – in the GISAID-based tree, the main clades of the viral strains were identified. The blue and red color triangles refer to type-A (S1A-S19A) and type-B (S1B-S19B) variants, respectively. Other colored circles refer to the deposited GISAID GenBank accession numbers of the main clades of the SARS-Cov2. In both trees, the number "0.01" at the top refers to the degree of the evolutionary scale range.

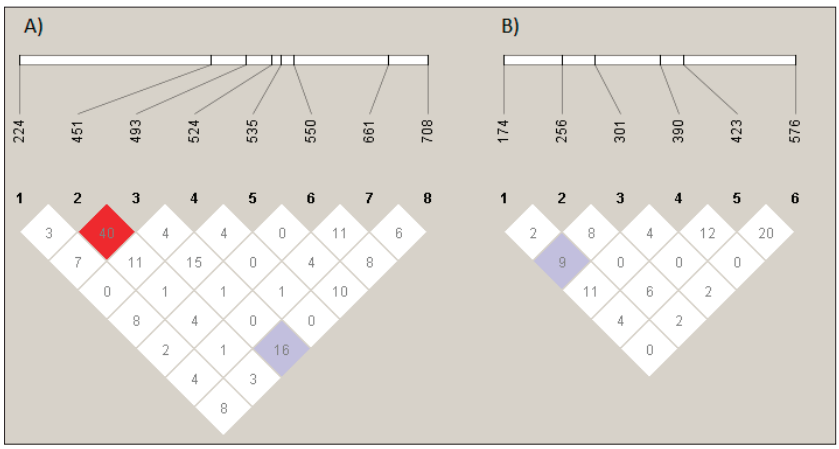

Fig. 4. Graphical representation of linkage disequilibrium (LD) among all possible pairs of fourteen variants observed in the amplified 795 bp fragment of the SARS-Cov-2 S gene using a standard D'/LOD Color Scheme. LD values are expressed as differences in the shade of each color. **A** – LD plot for 8 sites that have major alleles (type-A). **B** – LD plot for 7 sites that have minor alleles (type-B). The numbers at the top of the figure indicate the coordinate of sites in the 795 bp amplicons.

DISCUSSION

The patterns of molecular variations of the receptor-binding domain of SARS-CoV-2 spike protein were investigated in 19 COVID-19 cases. Our results identified an unprecedented rate of genetic variations in SARS-CoV-2, represented by fifteen novel SNPs in the coding regions of the spike glycoprotein. Within these SNPs, three unexpected nonsense SNPs were identified since electropherograms of all variations inferred clear co-infections in the entire viral samples. Although this study was conducted in the Babylon province within a geographic area spanning no more than 150 km, the number of identified SNPs in the samples isolated from this region exhibited a higher frequency than strains from nearby regions or other places around the world [35,36].

Although it is difficult to explain how SARS-CoV-2 strains attained transmissibility with such mixed strains, the limited hygienic conditions in Iraq could not be excluded from the explanation. Nevertheless, the high number of identified SNPs may be associated with the increased number of variations upon transmission from the country from which the virus originally spread. This is consistent with the variations of SARS-CoV-2 detected in samples collected from recently infected individuals as compared to earlier reports during the pandemic. Considering that the most likely point of origin of SARS-CoV-2, one would expect cases from China to exhibit a lower number of mutations [37]; the high number of identified SNPs observed herein are not consistent with the reported relatively low level of SARS-CoV-2 genomic mutations [38]. Instead, the identified SNPs are a real manifestation of the rapidly evolved transmission of this virus [39]. Although the coding region is the most frequent locus for mutational events, it is unusual to detect such a high rate of mutations in this region of the spike glycoprotein. However, the more variable the environment an RNA virus faces, the higher the mutation rate as a result of this challenge [40].

Results inferred from the phylogenetic tree obtained in this study indicated that there were multiple sources for the analyzed SARS-CoV-2 strains, suggesting that the Iraqi SARS-CoV-2 sequences are derived from many international sources and not from any particular country. Meanwhile, specific clade-based tree results showed that some of the samples originated from the European G clade of the SARS-CoV-2 strains. Generally, the G and GR clades prevail in Europe, clades S and GH have been mostly observed in America, and the reference L clade is mostly represented by sequences isolated in Asia where the virus likely originated [41]. These data suggest that it might be possible for some Iraqi strains to be derived from some viral European strains of the G clade. However, the high mutation rates observed in other Iraqi strains have made it difficult to assign these strains to any particular clade. Furthermore, the high recombination rates inferred from both local viral types can be explained by the ability of these viral sequences to communicate with each other [42].

It is well established that the spike glycoprotein enables SARS-CoV-2 to attach to ACE2 receptors on target cells. The nonsense mutations detected herein and which involve the receptor-binding site of the spike glycoprotein suggest that the truncated viral spike protein is incapable of attachment and that downstream entry of any particle of SARS-CoV-2 could not be

accomplished. It would be expected that the truncated spike glycoproteins reduce or even abolish the ability of strains with these nonsense mutations to bind to ACE2 receptors; however, this is not the case in our study as co-infection with different viral strains was observed. When a viral strain possessed a truncated spike, an intact spike glycoprotein of another viral strain presumably compensated for the deficiency of this truncation in such a way that both viral strains could effectively invade the host; according to the co-infection strategy, strains with truncated spike glycoproteins relied on the other co-infecting strain with intact spike glycoproteins for its infectivity. This finding requires the presence of recessive and dominant strains to interact with each other in infecting host cells. However, the presence of three highly frequent nonsense SNPs in the receptor-binding domain of spike glycoprotein may also be explained by the presence of alternative pathways of host cellular invasion. This is supported by a recent finding of the reliance of some SARS-CoV-2 strains on ACE2-independent routes of infection [43]. However, further information regarding this matter is lacking. Accordingly, serious concern was raised by the discovery of co-infection with different viral strains, which may magnify the severity of infections and the mortality rates in investigated areas [44].

In addition to truncated proteins resulting from SNPs, it was found that the observed missense SNPs were also responsible for defects in the structure, biological activity and stability of spike glycoproteins. Two missense SNPs (p.291C>G and p.391C>G) had deleterious effects on spike glycoprotein of different viral strains. Of note is the positioning of the p.291C>G SNP upstream of all three detected nonsense mutations (p.301Cdel, p.380Ydel, and p.436del), which suggests a more damaging role for this amino acid substitution on the spike glycoprotein. As in the case of nonsense SNPs, the presence of co-infecting viral strains could compensate for this damaging effect. However, the high frequency of this and other deleterious SNPs could not be ignored, adding another question regarding the binding between SARS-CoV-2 and ACE2 [12]. Unfortunately, the actual clinical status of the investigated patients was not provided, and we could not assess the clinical performance and severity of the infection in patients with these SNPs.

The presence of co-infection in all investigated viral strains contests airborne-based transmission of infection with SARS-CoV-2 particles. Because of the inability to be replicated outside an animal host, non-human routes could transmit one strain without association with another viral strain. Thus, co-infection may be triggered by direct human-to-human contact whereby an infected person transmits viral particles that appear to favor viral co-infection [36]. This pattern of infection could also be cumulative, which may explain the high severity and mortality among medical staff exposed to further risks of viral infection [45].

Co-infection with different SARS-CoV-2 strains can present an additional risk for everyone infected in an area. Moreover, the danger of co-infection can reach beyond this region to include other parts of Iraq and potentially other parts of the Middle East. If this high rate of mutations is ongoing in Iraq, other concerns about false-negative PCR results could also emerge. Therefore, more data on these mutations and co-infections are required to establish their importance in the severity and mortality of individuals infected with SARS-CoV-2.

CONCLUSIONS

This study revealed the presence of viral co-infection with all identified SNPs in the coding region of the SARS-CoV-2 spike protein. The identified SNPs in two types of infection with A- and B-type virus strains have not been detected in SARS-CoV-2 sequences. Three truncated spike glycoproteins were produced from three identified nonsense mutations, p.301Cdel, p.380Ydel, and p.436del. This study suggests that strains possessing a truncated spike glycoprotein are aided by helper strains with an intact 3D structure of the same protein. Interaction between both viral types was suggested by the presence of co-infection in all analyzed samples. Due to the high frequency of the identified nonsense mutations, the absence of intact spike glycoproteins raises questions about the effectiveness of ACE2 receptors in host cell invasion. Hence, questions should also be considered about the effectiveness of vaccines designed to bind with the intact 3D structure of the spike glycoprotein. Co-infection and the high-frequency of nonsense mutations identified in all investigated viral samples

point to an urgent need for further investigations to examine associations of viral SNPs with clinical records of patients with SARS-CoV-2.

Funding: This manuscript was partially supported by the Training and Development Company / Babylon Branch, Iraq (No. 020, 1).

Acknowledgments: The authors would like to thank the medical staff of the Iraqi Health Laboratory, Ministry of Health, for providing the coded samples of patients with COVID-19.

Author contributions: HOM developed the work, helped in the design and experimental work. MKM contributed to the study design and carried out the experimental work. MJM was the experimental advisor. HHA, ATSA, and SAH contributed to the experimental work. MBSA performed the analyses and wrote the manuscript.

Conflict of interest disclosure: The authors declare that they have no potential conflicts of interest with respect to the research, funding, authorship, or publication of this article.

REFERENCES

1. Fauver JR, Petrone ME, Hodcroft EB, Shioda K, Ehrlich HY, Watts AG, Vogels CB, Brito AF, Alpert T, Muyombwe A, Razeq J. Coast-to-coast spread of SARS-CoV-2 during the early epidemic in the United States. Cell. 2020;181(5):990-6. e5.
2. Wu F, Zhao S, Yu B, Chen YM, Wang W, Song ZG, Hu Y, Tao ZW, Tian JH, Pei YY, Yuan ML. A new coronavirus associated with human respiratory disease in China. Nature. 2020;579(7798):265-9.
3. Cohen J. "Were behind the curve": US hospitals confront the challenges of large-scale coronavirus testing. Science. 2020;3:4.
4. Hu Y, Riley LW. Dissemination and co-circulation of SARS-CoV2 subclades exhibiting enhanced transmission associated with increased mortality in Western Europe and the United States. medRxiv 20152959 [Preprint]. 2020 [cited 2020 Jul 13]. Available from: https://doi.org/10.1101/2020.07.13.20152959
5. Phan T. Genetic diversity and evolution of SARS-CoV-2. Infect Genet Evol. 2020;81:104260.
6. Korber B, Fischer W, Gnanakaran SG, Yoon H, Theiler J, Abfalterer W, Foley B, Giorgi EE, Bhattacharya T, Parker MD, Partridge DG. Spike mutation pipeline reveals the emergence of a more transmissible form of SARS-CoV-2. bioRxiv 069054 [Preprint]. 2020 [cited 2020 Apr 29]. Available from: https://doi.org/10.1101/2020.04.29.069054
7. Bhattacharyya C, Das C, Ghosh A, Singh AK, Mukherjee S, Majumder PP, Basu A, Biswas NK. 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 075911 [Preprint]. 2020 [cited 2020 May 4]. Available from: https://doi.org/10.1101/2020.05.04.075911
8. Zheng S, Fan J, Yu F, Feng B, Lou B, Zou Q, Xie G, Lin S, Wang R, Yang X, Chen W. Viral load dynamics and disease severity in patients infected with SARS-CoV-2 in Zhejiang province, China, January-March 2020: retrospective cohort study. BMJ-Brit Med J. 2020;369:m1443.
9. Robson B. Bioinformatics studies on a function of the SARS-CoV-2 spike glycoprotein as the binding of host sialic acid glycans. Comput. Biol Med. 2020;122:103849.
10. Wang K, Chen W, Zhou YS, Lian JQ, Zhang Z, Du P, Gong L, Zhang Y, Cui HY, Geng JJ, Wang B. SARS-CoV-2 invades host cells via a novel route: CD147-spike protein. BioRxiv 988345 [Preprint]. 2020 [cited 2020 Mar 14]. Available from: https://doi.org/10.1101/2020.03.14.988345
11. Jeffers SA, Tusell SM, Gillim-Ross L, Hemmila EM, Achenbach JE, Babcock GJ, Thomas WD, Thackray LB, Young MD, Mason RJ, Ambrosino DM. CD209L (L-SIGN) is a receptor for severe acute respiratory syndrome coronavirus. Proc Natl Acad Sci USA. 2004;101(44):15748-53.
12. Wang H, Yang P, Liu K, Guo F, Zhang Y, Zhang G, Jiang C. SARS coronavirus entry into host cells through a novel clathrin-and caveolae-independent endocytic pathway. Cell Res. 2008;18(2):290-301.
13. Tang X, Wu C, Li X, Song Y, Yao X, Wu X, Duan Y, Zhang H, Wang Y, Qian Z, Cui J. On the origin and continuing evolution of SARS-CoV-2. Natl Sci Rev. 2020;7(6):1012-23.
14. Essa S, Owayed A, Altawalah H, Khadadah M, Behbehani N, Al-Nakib W. Mixed viral infections circulating in hospitalized patients with respiratory tract infections in Kuwait. Adv Virol. 2015;2015:714062.
15. Li J, Kou Y, Yu X, Sun Y, Zhou Y, Pu X, Jin T, Pan J, Gao GF. Human co-infection with avian influenza and seasonal influenza viruses, China. Emerg Infect Dis. 2014;20:1953-5.
16. Liu Y, Ning Z, Chen Y, Guo M, Liu Y, Gali NK, Sun L, Duan Y, Cai J, Westerdahl D, Liu X. Aerodynamic analysis of SARS-CoV-2 in two Wuhan hospitals. Nature. 2020;582(7813):557-60.
17. Zhang XS. Strain interactions as a mechanism for dominant strain alternation and incidence oscillation in infectious diseases: seasonal influenza as a case study. PLoS ONE. 2015;10(11):e0142170.
18. Brundage JF. Interactions between influenza and bacterial respiratory pathogens: implications for pandemic preparedness. Lancet Infect Dis. 2006;6:303-12.
19. Goldstein E, Cobey S, Takahashi S, Miller JC, Lipsitch M. Predicting the epidemic sizes of Influenza A/H1N1, A/H3N2, and B: A Statistical Method. PLoS Med. 2011;8(7):e1001051.
20. Tsuchihashi Y, Sunagawa T, Yahata Y, Takahashi H, Toyokawa T, Odaira F, Ohyama T, Taniguchi K, Okabe N. Association between seasonal influenza vaccination in 2008-2009 and pandemic influenza A(H1N1) 2009 infection among school students from Kobe, Japan, April-June 2009. Clin Infect Dis. 2012;54:381-3.
21. Ye J, Coulouris G, Zaretskaya I, Cutcutache I, Rozen S, Madden T. Primer-BLAST: A tool to design target-specific primers for polymerase chain reaction. BMC Bioinformatics. 2012;13:134.
22. Gasteiger E, Gattiker A, Hoogland C, Ivanyi I, Appel RD, Bairoch A. ExPASy: the proteomics server for in-

depth protein knowledge and analysis. Nucleic Acids Res. 2003;31(13):3784-8.
23. Morgat A, Lombardot T, Coudert E, Axelsen K, Neto TB, Gehant S, Bansal P, Bolleman J, Gasteiger E, De Castro E, Baratin D. Enzyme annotation in UniProtKB using Rhea. Bioinformatics. 2020;36(6):1896-901.
24. Ng PC, Henikoff S. Predicting the effects of amino acid substitutions on protein function. Annu Rev Genom Hum Genet. 2006;22(7):61-80.
25. Choi Y, Sims GE, Murphy S, Miller JR, Chan AP. Predicting the functional effect of amino acid substitutions and indels. PLoS ONE. 2012;7:e46688.
26. Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, Kondrashov AS, Sunyaev SR. A method and server for predicting damaging missense mutations. Nat Methods. 2010;7(4):248-9.
27. Smigielski EM, Sirotkin K, Ward M, Sherry ST. dbSNP: a database of single nucleotide polymorphisms. Nucleic Acids Res. 2000;28:352-5.
28. Ashkenazy H, Erez E, Martz E, Pupko T, Ben-Tal N. ConSurf 2010: calculating evolutionary conservation in sequence and structure of proteins and nucleic acids. Nucl Acids Res. 2010;38:W529-33.
29. Pires DEV, Ascher DB, Blundell TL. mCSM: predicting the effects of mutations in proteins using graph-based signatures. Bioinformatics. 2012;30:335-42.
30. Librado P, Rozas J. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics. 2009;25:1451-2.
31. French N, Yu S, Biggs P, Holland B, Fearnhead P, Binney B, Fox A, Grove-White D, Leigh JW, Miller W, Muellner P. Evolution of Campylobacter species in New Zealand. In: Sheppard SK, Méric G, editors. Campylobacter ecology and evolution. Caister Academic Press; 2014. p. 221-40.
32. Elbe S, Buckland-Merrett G. Data, disease and diplomacy: GISAID's innovative contribution to global health. Glob Chall. 2017;1:33-46.
33. Letunic I, Bork P. Interactive tree of life (iTOL) v4: recent updates and new developments. Nucl Acids Res. 2019;47(W1):W256-9.
34. Barrett JC, Fry B, Maller JDM J, Daly MJ. Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics. 2005;21(2):263-265.
35. Memish ZA, Aljerian N, Ebrahim SH. Tale of three seeding patterns of SARS-CoV-2 in Saudi Arabia. Lancet Infect Dis. 2020;https://doi.org/10.1016/S1473-3099(20)30425-4.
36. Ghinai I, McPherson TD, Hunter JC, Kirking HL, Christiansen D, Joshi K, Rubin R, Morales-Estrada S, Black, SR, Pacilli M, Fricchione MJ. First known person-to-person transmission of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) in the USA. Lancet Infect Dis. 2020;395(10230):1137-44.
37. Mercatelli D, Giorgi FM. Geographic and Genomic Distribution of SARS-CoV-2 Mutations. Front Microbiol. 2020;22;11:1800.
38. Ceraolo C, Giorgi FM. Genomic variance of the 2019-nCoV coronavirus. J Med Virol. 2020;92(5):522-8.
39. Sanche S, Lin Y, Xu C, Romero-Severson E, Hengartner N, Ke R. High Contagiousness and Rapid Spread of Severe Acute Respiratory Syndrome Coronavirus 2. Emerg Infect Dis. 2020;26(7):1470-7.
40. Duffy S. Why are RNA virus mutation rates so damn high? PLoS Biol. 2018;16(8):e3000003.
41. Andersen KG, Rambaut A Lipkin, WI, Holmes EC, Garry RF. The proximal origin of SARS-CoV-2. Nat Med. 2020;26:450-2.
42. Brufsky, A. Distinct viral clades of SARS-CoV-2: implications for modeling of viral spread. J Med Virol. 2020;92(9):1386-90.
43. Morniroli D, Giannì ML, Consales A, Pietrasanta C, Mosca F. Human Sialome and Coronavirus Disease-2019 (COVID-19) Pandemic: An Understated Correlation? Front Immunol. 2020;11:1480.
44. Cox MJ, Loman N, Bogaert D, O'grady J. Co-infections: potentially lethal and unexplored in COVID-19. Lancet Microbe. 2020;1(1):e11.
45. Liu W, Li Z-D, Tang F, Wei M-T, Tong Y-G, Zhang L, Xin ZT, Ma MJ, Zhang XA, Liu LJ, Zhan L. Mixed infections of pandemic H1N1 and seasonal H3N2 viruses in 1 outbreak. Clin Infect Dis. 2010;50:1359-65.