

Article

Unexpected Co-infection with Different Strains of SARS-CoV-2 in Patients with COVID-19

Hayder O.Hashim ¹, Mudher K. Mohammed ², Mazin J. Mousa ³, Hadeer H. Abdulameer ⁴, Alaa T.S. Alhassnawi ⁵, Safa A. Hassan ⁶ and Mohammed Baqur S. Al-Shuhaib ^{7,*}

1. Department of Clinical Laboratory Sciences, College of Pharmacy, University of Babylon, Babil 51001, Iraq; phar.hayder.obayes@uobabylon.edu.iq; orcid: <https://orcid.org/0000-0001-8933-0994>

2. Department of Pharmacy, Al-Manara College of Medical Science, Iraq; ph.mudher79@gmail.com

3. Department of Clinical Laboratory Sciences, College of Pharmacy, University of Babylon, Babil 51001, Iraq; phar.mazin.jaafar@uobabylon.edu.iq or mazin.mousa@gmail.com; orcid: <https://orcid.org/0000-0003-2209-010X>

4. Faculty of Education for Girls, University of Kufa; hedear1.h@gmail.com

5. Department of Biology, College of Science, University of Babylon, Babil 51001, Iraq; alaataark79@yahoo.com

6. Babylon branch, Alfadhel for Training and Development Company; al7507242@gmail.com

7. Department of Animal Production, College of Agriculture, Al-Qasim Green University, Al-Qasim, Babil 51001, Iraq;

* Correspondence: mohammed79@agre.uoqasim.edu.iq or baquralhilly_79@yahoo.com; orcid: <https://orcid.org/0000-0002-6458-2068>

Running title: variations of SARS-CoV-2 spike glycoprotein

Abstract: There is a rising global concern for the ongoing outbreak of SARS-CoV-2 due to its high transmission rate and unavailability of treatment. Through the binding of its spike glycoprotein with angiotensin type 2 (ACE2), SARS-CoV-2 can efficiently get in the cells of patients and start its pandemic cycle. Herein, the biological diversity of SARS-CoV-2 infection was assessed in Babylon province of Iraq by investigating the possible genetic variations of the spike glycoprotein. A specific coding region of 795 bp within the viral spike (S) gene was amplified from 19 patients who suffered from obvious symptoms of SARS-CoV-2 infection. Sequencing results identified fifteen novel nucleic acid variations with a variety of distributions within the investigated samples. The electropherograms of all the identified variations showed obvious co-infections with at least two different viral strains per sample. Within these co-infections, the majority of samples exhibited three nonsense single nucleotide polymorphism (SNP)s, p.301Cdel, p.380Ydel, and p.436del, which yielded three truncated SARS-CoV-2 spike glycoproteins of 301, 380, and 436 amino acids length, respectively. The network and phylogenetic analyses indicated that for all viral infections were derived from multi-ancestral origins. Results inferred from the specific clade-based tree entailed that some viral strains were derived from European G-clade sequences. In conclusion, our data demonstrated the absence of any single strain infection among all investigated viral samples in the studied area, which may entail a higher risk of SARS-CoV-2 in this country. Through the identified high frequency of truncated spike proteins, we suggest that defective SARS-CoV-2 may depend on helper strains having intact spikes in its infection. Alternatively, another putative ACE2-independent route of viral infection way also suggested. To the best of our knowledge, this is the first report to describe the co-infection of multiple strains of SARS-CoV-2 in patients with COVID-19.

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

1. Introduction

Coronavirus disease 2019 (COVID-19) is a pandemic of a severe acute respiratory syndrome (SARS) caused by novel coronavirus SARS-CoV-2, which has first been reported in Wuhan, China in late 2019. It has been classified as one of the three beta coronaviridae that cause the highest infectivity and pathogenesis in contemporary history after SARS-CoV-1, and MERS-CoV [1, 2]. Since its emergence as a newly identified pathogenic virus, SARS-CoV-2 has a widely known rapid spread in a relatively short time scale with a catastrophic pandemic sequel all around the globe on healthcare systems, economy, and social aspects.

Despite the preventive measures imposed by many governments around the world, the discrepancy in the cumulative mortality rates between the states in America and even between the western European countries by mid-June, 2020 has been observed. This limitation in the viral confrontation can be attributed to many causes, such as failure to adhere to measures of social distancing, presence of concomitant chronic diseases that are strongly correlated to mortality (co-morbidities), the preparedness of health systems capabilities to cope with the pandemic, and other related factors [3]. Despite the rapid progress of the phylogeny and the genetic map of the SARS-CoV-2, the incidence of concomitant mixed infections with different strains has not been reported [4]. SARS-CoV-2 is highly similar to SARS-CoV that has 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 been developed to SARS-CoV-2. However, both viral entities use their spike (S) glycoprotein to recognize angiotensin type 2 (ACE2) on cell surfaces for the subsequent internalization into the cytoplasm of the host. The high rate of prevalence of SARS-CoV-2 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]. The early results obtained from the 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 a result of a point mutation of D614G that causes changes in viral spike glycoprotein by rendering it more resistant to cleavage during viral replication [7]. Furthermore, recent reports illustrated that there are more than 30 distinguishable mutations in different sites of the coding region of SARS-CoV-2 genomic sequences. These viral strains having these mutations have been known to exhibit effective distribution in different locations around the world [8].

It deserves to note that there are several alternative routes of viral entry into human cells other than through hooking to ACE2, for example through sialic acid [9], CD 147 [10], CD 209L [11] and by endocytosis [12]. These mechanisms may enhance virulence and act as substitutes for spike-ACE2 interaction and viral internalization. Any mutation involving the receptor-binding domain may cause the generation of inactive spike proteins that have been expected to decrease viral virulence. However, this issue can be solved by the coexistence of multiple strains with the simultaneous potential to infected the same individual. It is well known that this pattern of infection is called concurrent infection or co-infection [13]. The co-infections can be originated by direct transmission of multiple strains or in a sequential pattern when an infected person is re-infected with another strain before recovery [14]. The interaction of multiple strains with each other's during co-infection can lead to an alteration of epidemiological behavior [15]. This alteration can include the incidence of cyclical pattern and strain replacement which may enhance the viral co-infectivity by several postulated mechanisms [16, 17, 18]. Given the rapid spread of COVID-19 globally, investigations are underway to better understand the genetic variation and emergence of this outbreak starting from the original Wuhan city to the last region these viral particles have been attended. It may also be rational to explore the rate of mutations of SARS-CoV-2 in remote places from Wuhan, such as Iraq, which is a country with noticeably increased cases of COVID-19 since the beginning of 2020. Taken all of these data in our consideration, this study was designed to reveal the genetic variations of SARS-CoV-2 in Babylon province, which is

one of the Iraqi cities suffering from an ongoing acceleration in the reported pandemic with COVID-19.

2. Materials and Methods

2.1. Viral samples

Nineteen nasopharyngeal swabs from patients with COVID-19 were kindly provided from the medical staff of Iraqi Health Laboratory, Ministry of Health. The detection of the presence of this 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 a universal viral transport medium (Capricorn Scientific, Germany) following the World Health Organization (WHO) guidelines. All the involved swabs were from six different areas in Babylon province (Supplementary File). No other personal data of patient identifiers were provided, and specimens were labeled with random numbers so that test results cannot be linked to the involved individuals. Ethics approval was not sought because the collected viral swabs were part of laboratory validation for those patients. The secondary use of these anonymous pathological specimens was certified by the Ministry of Health Review Board (1318/24-08-2020).

2.2. RNA Extraction and reverse transcription

For RNA extraction from Nasopharyngeal swabs, a nucleic acid purification kit based on a 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, μ LITE, UK). For downstream studies, samples with the 260 nm/280 nm ratio higher than 1.7 were used in subsequent analysis. The first-strand cDNA was synthesized using AccuPower® RocketScript™ RT PreMix (Bioneer, Korea). The RT-PCR experiment was started using 2 μ l of total RNA, random hexamer 100 pmoles, in a final reaction volume of 20 μ l. Thermocycling was conducted on Biometra TRIO thermal cycler (Analytik Jena, Germany). Primer annealing condition was adjusted at 15 C° for 10 min, and cDNA synthesis was started at 42 C° for 30 min and heat inactivation was adjusted to 95 C° for 5 min.

2.3. PCR

One PCR fragment of 795 bp was designed in this study using NCBI Primer-BLAST software [19]. The main designing approach for this fragment was made to partially cover 264 amino acids, starting from 217 to 480 amino acid residues, of the viral spike glycoprotein of the SARS-CoV-2. The sequences of the forward and reverse primers were 5'-CCCTCAGGGTTTTTCGGCTT-3', and 5'-TTACAAGGTGTGCTACCGGC-3', respectively. The 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 sec, annealing at 59°C for 30 sec, and elongation at 73°C for 60 sec, and finalized with a polymerization at 73°C for 5 min. PCR products were verified by electrophoresis on 1.5% agarose gel.

2.4. DNA sequencing

All amplified fragments were subsequently exposed to sequencing reactions following the instructions of Macrogen laboratories (Macrogen, Geumchen, Seoul, Korea). The database of the S gene nucleic acids was retrieved from NCBI server (GenBank accession number MT847228.1). The retrieved sequences were then visualized and annotated by BioEdit ver, 7.1. (DNASTAR, Madison). The correct reading frames of the viral variants in the analyzed spike glycoprotein were determined

using the ExPASy translate server [20]. 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 in the observed 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 surface glycoprotein using the UniProtKB server [21].

2.5. *In silico* prediction

The 3D structure of the entire amino acid residues of the spike glycoprotein, 6vxx, was retrieved from the protein data bank (PDB) webserver (<https://www.rcsb.org/>). Subsequently, the effects of the identified missense variants on spike structure, function, and biological stability were predicted using state-of-the-art *in silico* tools, namely SIFT [22], PROVEAN [23], PolyPhen [24], SNAP2 [25], ConSurf [26], and mCSM [27]. The cumulative outcomes of utilized tools were assessed for each identified missense SNP.

2.6. Data analyses

Based on the nucleic acid haplotypes generated by the DNA sequence polymorphism (DnaSP) software ver. 6.12.01 [28], the exact distances between haplotypes were matched by a tight-span walker network using the population analysis with reticulate tree (PopART) software v. 4.8.4 [29]. The phylogenetic relationships among the identified viral variations were generated by constructing a neighbor-joining-based unrooted comprehensive tree. All the reference sequences incorporated within this tree were retrieved from the NCBI database (<https://www.ncbi.nlm.nih.gov/>). Subsequently, another unrooted neighbour-joining tree was constructed to assess the exact clade of the investigated viral strains. In this specific tree, the main known five (G, L, S, O, and V) viral clades were incorporated alongside with our investigated sequences. The incorporated clades were retrieved from the GISAID database to represent several Asian, European, African, and American countries around the world [30]. Both trees were annotated by the iTOL server [31]. Linkage disequilibrium (LD) analysis between the detected SNPs of the SARS-CoV2 sequences was conducted using haploview software, ver. 4.2 [32]. LD values were estimated using squared allele frequency correlations (r^2) between the pairs of the analyzed loci.

3. Results

3.1. Genetic polymorphism and prediction

All the investigated samples showed co-infection with more than one strain as all identified variations within electropherograms were found to exhibit double peaks per individual variant (Fig 1).

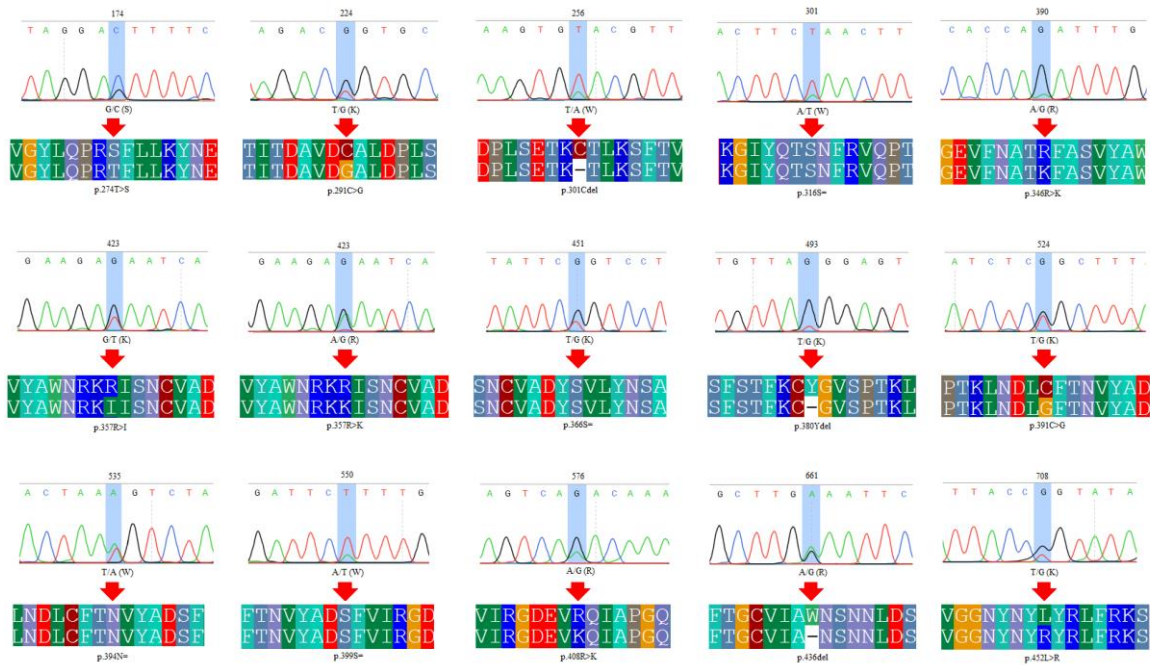


Figure 1. Electropherograms of the detected nucleic acid substitutions in the 795 bp amplicons showed clear co-infections in all variations observed in the *S* gene within SARS-CoV-2 strains. The final effect (silent, missense, or nonsense) of each identified SNP was indicated by red arrows.

The co-infection patterns for all identified variants were also confirmed in all analyzed electropherograms. Since all investigated viral samples exhibited co-infection, viral strains with the major electropherograms reads assigned type-A strains, while those exhibited minor reads were assigned type-B strains. A total of fifteen nucleic acid substitutions were identified in the coding region of the viral spike protein in both type-A or type-B strains. The majority of these 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 rest three SNPs (p.301Cdel, p.380Ydel, and p.436del) were predicted to cause premature termination in the nascent viral spike polypeptide in three separate positions in the entire structure of the mature spike glycoprotein (protein ID QMU94980.1).

Table 1. Nucleic acids and their 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>I	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

451	T/G	S7, S9, S11, S13, S18, S19	Type-A	S366	S=	Silent	(p.357R>K) 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)

The observed variations were deposited in NCBI database under the GenBank accession numbers MT936988 - MT937006 for type-A viral strains and MT937007 - MT937025 for type-B viral strains.

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

Table 2. The *in silico* analysis of the deleterious effects of the observed missense SNPs in the S gene on the SARS-CoV-2, using several different bioinformatics tools. the symbol “Del” refers to the deleterious effect of the missense SNP, while “Neut” refers to the neutral effect of the missense SNP.

<i>In silico</i> 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
PolyPhen2	Del	Del	Neut	Del	0.003	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

3.2. Network analyses

A specific tight span walker network was generated from haplotypes data of the observed variants of both types of the identified viral strains. This network was developed to describe the accurate genetic distances between the identified variants haplotypes from one side and the most relative 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 infection was inferred from the currently generated network. However, the results of this network indicated the presence of multi-common ancestor for all identified haplotypes within the co-infection conditions. These multi-ancestor sequences, or Hap 1, were not restricted by a particular country as both type-A and type-B strains were equally derived from viral strains of many countries in different continents worldwide (Fig 2).

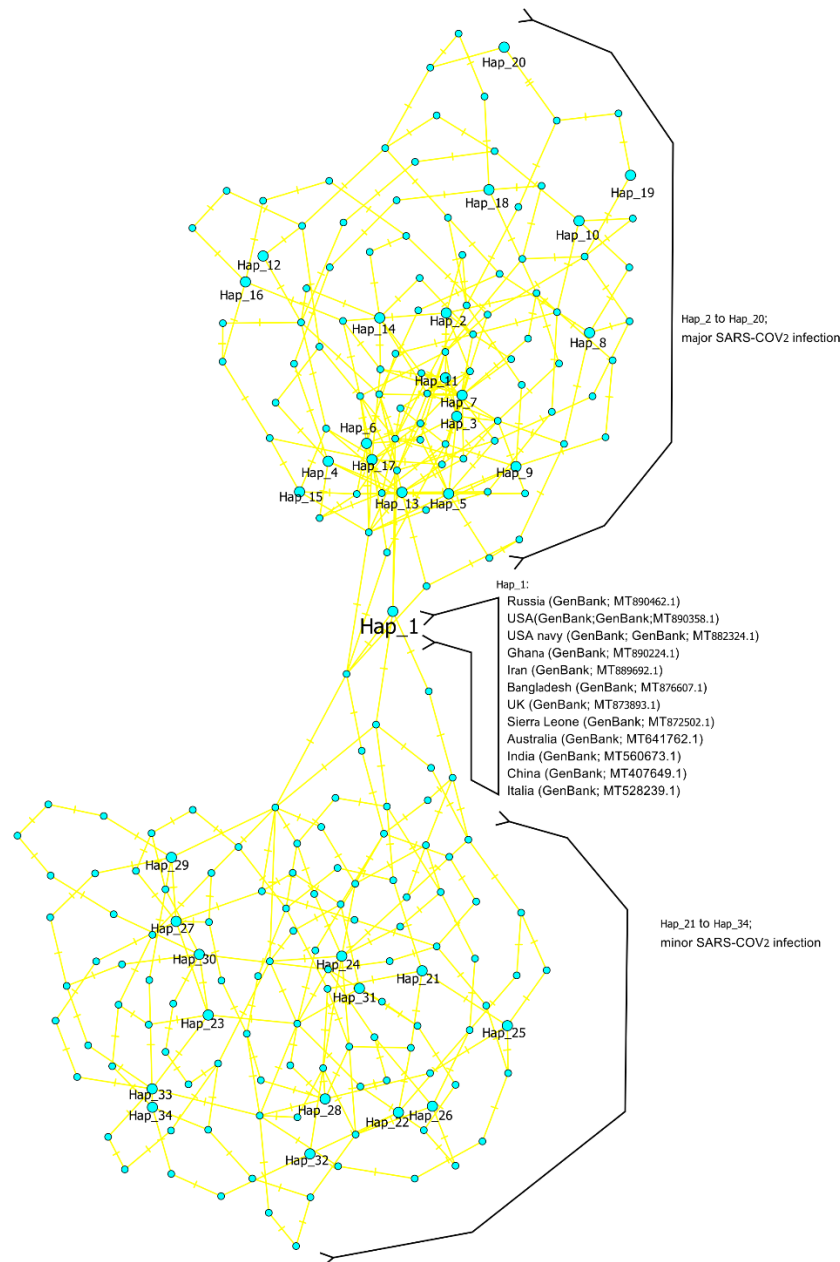


Figure 2. Haplotype network analysis of SARS-CoV-2 viruses. Hap 1 represents the deposited reference sequence from different regions around the world. Hap 2 to Hap 20 represent the major (type-A) infection. Hap 21 to Hap 34 represent the minor (type-B) infection.

3.3. Comprehensive Phylogenetic analysis

Further clarifications of the observed network-based data were made by constructing a neighbour-joining comprehensive tree. Another layer of confirmation about the multi-common ancestor of the identified viral strains was demonstrated with further details. The precise phylogenetic positioning of the investigated viral strains with the related SARS-CoV-2 sequences was identified (Fig 3).

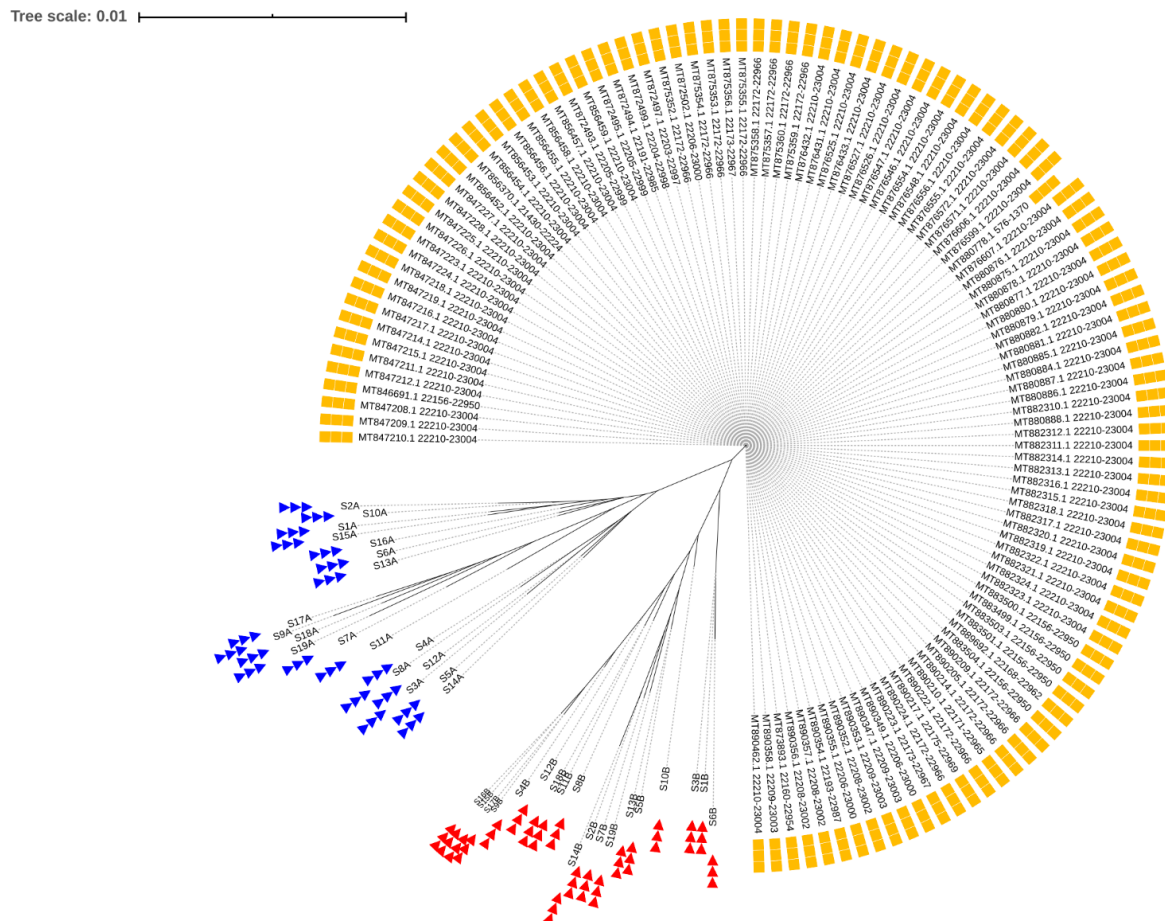


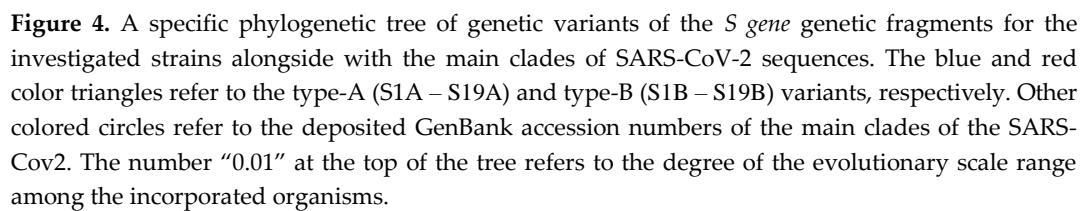
Figure 3. The comprehensive phylogenetic tree of genetic variants of the *S gene* genetic fragments for the SARS-CoV-2 Iraqi strains. The blue and red color triangles refer to the type-A (S1A – S19A) and type-B (S1B – S19B) variants, respectively. The dark yellow colored boxes refer to the deposited GenBank accession numbers of reference sequences of the SARS-Cov2. The number “0.01” at the top of the tree refers to the degree of the evolutionary scale range among the incorporated organisms.

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

In addition to type-A and type-B viral sequences, the total incorporated organisms within the tree were 151, all of them were attributed to the same organism, SARS-CoV-2. Viruses of the type-A were clustered together in one clade, and likewise, viruses of the type-B were also exhibited the same pattern of positioning in an adjacent clade. Therefore, our neighbour-joining comprehensive data confirmed the separation of the identified viral type-A and type-B from each other. The tree-based results were validated the network-based data due to the close positioning of both types of viral infections altogether in the tree. Furthermore, this tree showed that both investigated viral types were derived from one common multi-ancestors' origin, which was also equally shared among many viral sequences around the world without being attributed to a particular country, race, or continent.

However, data obtained from the currently developed comprehensive tree was not sufficient to identify the clade to which our viral samples were descended from. Accordingly, representative samples for the main known five 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 specific tree. Most importantly, three strains of type-A, S1A, S3A, and S6A, were originated from the viral representative of G clade (Fig 4). Meanwhile, no other clear

Tree scale: 0.01



More details concerning the potential linkage across viral strains were conducted using data retrieved from the major A viral type and the minor B viral type identified in this study. To assess the range of recombination among the investigated viral strains, LD plot of both A and B viral types were developed. In the studied type-A variants, haploplot view data showed the presence of only one close linkage between 145 and 493 loci in the amplified 795 bp fragments. Meanwhile, the other loci of the type-A variants exhibited low r^2 values (Fig 5a). Since no close linkage was observed among the majority of the identified SNPs, high recombination rates of these SNPs were confirmed. The same case was also demonstrated in type-B variants, in which all SNPs were not found to exhibit any association among each other (Fig 5b). 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 the investigated SARS-CoV2 strains was confirmed.

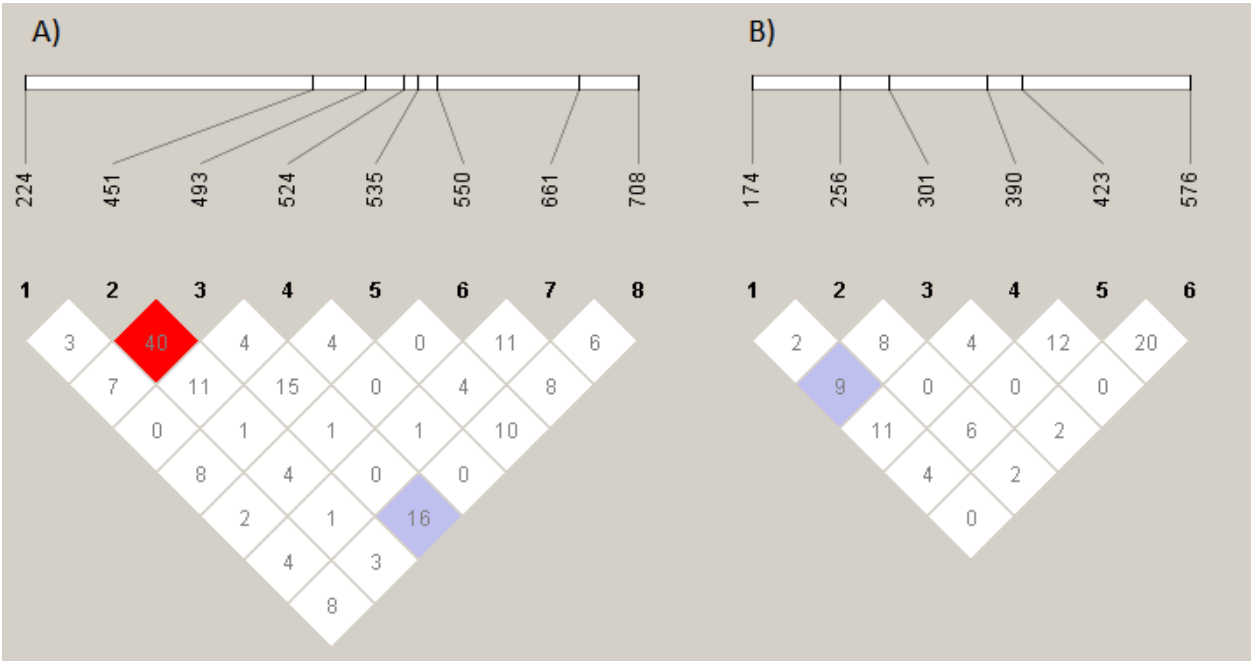


Figure 5. Linkage disequilibrium (LD) graphical representation among all possible pairs of fourteen variants observed in the amplified 795 bp fragment of the SARS-Cov-2 S gene using standard D'/LOD Color Scheme. LD values are expressed as differences in the shade of each color. A) LD plot for 8 sites that have major alleles (type-A). B) LD plot for 7 sites that have minor alleles (type-B). The numbers given at the top of the figure indicate the coordinate of sites in the 795 bp amplicons.

4. Discussion

The patterns of molecular variations of the receptor-binding domain of SARS-CoV-2 spike protein were investigated in this study among 19 infections with COVID19. Our results identified an unprecedented rate of genetic variations in the 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. As well, electropherograms of all variations inferred clear co-infections in the entire viral samples. Though this study was conducted in Babylon province, within a geographic area spanning no more than 150 km, the number of identified SNPs in samples isolated from this region exhibited very high frequency than the other strains from the nearby regions or even from other places around the world [33, 34].

Though it is not easy to explain the way through which SARS-CoV-2 strains attained transmissibility with the presence of such mixed strains, the humble hygienic situations in Iraq could not be excluded from explanation. Nevertheless, the reason for the high number of the currently identified SNPs may be associated with the increased number of variations upon transmission from the original country from which this virus has first spread. This suggestion is consistent with the increasingly reported variations of SARS-CoV-2 detected in increasing frequency in samples collected from people infected recently compared to early reports in the pandemic. Taking into consideration that the most likely point of origin of SARS-CoV-2, one would expect China to have a lower number of mutations [35]. However, the high numbers of the identified SNPs in this study are not consistent with the reported slow changes or relatively low variations of SARS-CoV-2 genomic mutations [36]. Instead, the currently identified SNPs represent a real manifestation of rapidly evolved transmission of this virus [37]. Although the coding region is the most prevalent 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 environments an RNA virus faces, the more an increased mutation rate would be preferred as a result of these challenges [38].

Results inferred from the comprehensive tree of this study found that there was a multi-common ancestor for the analyzed SARS-CoV-2 strains, which entailed that Iraqi SARS-CoV-2 sequences have been derived from many international sources around the world without being originated from any particular country. Meanwhile, specific clade-based tree results showed that some of our samples were originated from the European G clade of SARS-CoV-2 strains. Generally, the G and GR clades prevail in Europe, while the clade S and GH have been mostly observed in America. Whereas the reference clade, L clade, is mostly represented by sequences isolated from Asia, where the virus has likely originated [39]. These data suggested that it might be possible for some Iraqi strains to be derived from some viral European strains having G clade. However, the high mutation rates observed in the 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 in this study may be explained by the ability of these viral sequences to be communicated with each other away from the common multi-ancestor origin. However, it is still unclear if the speed of the viral transmission observed in Iraq may be the consequence of the clade's differences [40].

It is well established that the spike glycoprotein enables SARS-CoV-2 to attach onto target cell ACE2 receptors. Nonsense mutations detected in this study involving the receptor-binding site of the spike glycoprotein suggest that the truncated viral spike protein may not be able to undergo its scheduled task of attachment. Thus, the downstream entry of any particle of SARS-CoV-2 could not be accomplished in this way. In the currently identified results, the truncated spike glycoproteins may reduce or even abolish the direct ability of strains having these nonsense mutations to bind with ACE2 receptors. However, this is not the case in our study due to the co-infection with different viral strains. One when the viral strains exerted a truncated spike, another intact spike glycoprotein of another viral strain would presumably compensate the deficiency of this truncation in such a way both viral strains could invade the host efficiently. According to the putative co-infection strategy, strains having truncated spike glycoproteins would rely upon the other co-infecting strains having intact spike glycoproteins in its infection. This finding requires the presence of recessive dominant strains to interact with each other in infecting host cells. However, the presence of three high frequent nonsense SNPs in the receptor-binding domain of spike glycoprotein in this study may also be explained by the presence of alternative pathways of host cellular invasion. This possible suggestion has been supported by a recent finding that described the possible reliance of some SARS-CoV-2 strains on ACE2-independent routes in infection [41]. However, further information in this regard is still missing, and many investigations have not yet released sufficient data around this issue. Accordingly, there is a serious concern raised by the discovery of co-infection with different viral strains [42], which may magnify the severity of infections and mortality rates in the investigated areas.

In addition to the truncated proteins caused by nonsense SNPs, the effect of missense SNPs was also involved in damaging the structure, biological activity, and stability of spike glycoproteins. These data were applied to two missense SNPs (p.291C>G and p.391C>G) that exhibited obvious deleterious effects on spike glycoprotein on these viral strains. Noteworthy, p.291C>G SNP was positioned upstream of all three detected nonsense mutations (p.301Cdel, p.380Ydel, and p.436del), which suggested a more damaging role for this amino acid substitution on the resulting spike glycoprotein. As in the case of nonsense SNPs, the presence of co-infecting viral strains may also compensate for this damaging effect. However, the high frequency of this deleterious SNP, as well as other relatively deleterious SNPs, could not be ignored, which also adds another question regarding the genuine binding between SARS-CoV-2 and ACE2 [12]. Unfortunately, the actual clinical status of all investigated patients was not provided. For this reason, we were not able to assess the clinical performance and severity of infection of patients having these specific SNPs revealed by this study.

The presence of co-infection in all investigated viral strains may refute the airborne-based transmission of infection of SARS-CoV-2 particles. Due to the inability to replicate outside the

animal host, non-human routes may transmit one strain without being associated with another viral strain. Thus, the co-infection may be triggered by the direct human to human contact, in which the infected person transmits the viral particles for recipients, which seems more favoured for viral co-infection [34]. As a reason for a human to human contact, the pattern of infection could also be cumulative, which may explain the high severity and mortality among medical staff in hospitals that are exposed to multiple strains sequentially [43].

The obvious co-infection with different SARS-CoV-2 strains in all analyzed samples may bring about additional risks for everyone infected in the investigated area. Moreover, it is consequent to say that the danger of co-infection for all investigated patients may exceed this region to include other portions in Iraq or potentially other parts of the Middle East. However, if this reported high rate of mutations is being ongoing in Iraq, other concerns about PCR false negative-results may also emerge. Therefore, more data on these mutations and co-infections are mandatory to establish their importance in the severity and mortality of SARS-CoV-2.

5. Conclusions

This study showed the presence of clear viral co-infection in all identified SNPs in the coding region of the SARS-CoV-2 spike protein. All the identified SNPs exerted two types of infection (A and B type), which have not been detected so far 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 having a truncated spike glycoprotein are aided by helper strains with an intact 3D structure of the same protein. The 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, this study raises questions about the effectiveness of ACE2 receptors of the host cellular invasion in the absence of intact spike glycoproteins. Hence, questions should also be considered about the effectiveness of vaccines designed to bind with the intact 3D structure of the spike glycoprotein. Therefore, the co-infection and high-frequency nonsense mutations identified in all investigated viral samples imply the urgent need for further immediate investigations to associate the currently observed viral SNPs with clinical records of the patients with SARS-CoV-2.

Acknowledgments: Authors like to thank the medical staff of Iraqi Health Laboratory, Ministry of Health for providing coded samples of patients with COVID-19. This manuscript was partially supported by for training and development company / Babylon branch, Iraq (No. 020, 1).

Conflicts of Interest: The authors declare no conflict of interest.

Supplementary Materials: The following are available online at www.mdpi.com/xxx/s1, Figure S1: Places of origin of SARS-CoV-2 samples taken from Babylon province, Iraq.

References

1. Fauver JR, Petrone ME, Hodcroft EB, Shioda K, Ehrlich HY, Watts AG, et al. Coast-to-coast spread of SARS-CoV-2 during the early epidemic in the United States. *Cell*. 2020; 181(5): 990-996.e5. <https://doi.org/10.1016/j.cell.2020.04.021> PMID: 32386545
2. 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): 265-9. <https://doi.org/10.1038/s41586-020-2012-7> PMID: 32015508

3. Cohen J. “Were behind the curve”: US hospitals confront the challenges of large-scale coronavirus testing. *Science* 2020; 3: 4. <https://doi.org/10.1126/science.abb6856> PMID: 32399515
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* 2020.07.13.20152959; <https://doi.org/10.1101/2020.07.13.20152959>
5. Phan T. Genetic diversity and evolution of SARS-CoV-2. *Infect Genet Evol.* 2020; 1;81: 104260. <https://doi.org/10.1016/j.meegid.2020.104260> PMID: 32092483
6. Korber B, Fischer W, Gnanakaran SG, Yoon H, Theiler J, Abfalterer W, et al. Spike mutation pipeline reveals the emergence of a more transmissible form of SARS-CoV-2. *bioRxiv* 2020.04.29.069054. <https://doi.org/10.1101/2020.04.29.069054>
7. Bhattacharyya C, Das C, Ghosh A, Singh AK, Mukherjee S, Majumder PP, 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* 2020.05.04.075911. <https://doi.org/10.1101/2020.05.04.075911>
8. Zheng S, Fan J, Yu F, Feng B, Lou B, Zou Q, et al. Viral load dynamics and disease severity in patients infected with SARS-CoV-2 in Zhejiang province, China, January-March 2020: retrospective cohort study. *BMJ.* 2020; 369: m1443. <https://doi.org/10.1136/bmj.m1443> PMID: 32317267
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. <https://doi.org/10.1016/j.compbiomed.2020.103849> PMID: 32658736
10. Wang K, Chen W, Zhou YS, Lian JQ, Zhang Z, Du P, et al. SARS-CoV-2 invades host cells via a novel route: CD147-spike protein. *BioRxiv* 2020.03.14.988345. <https://doi.org/10.1101/2020.03.14.988345>
11. Jeffers SA, Tusell SM, Gillim-Ross L, Hemmila EM, Achenbach JE, Babcock GJ, et al. CD209L (L-SIGN) is a receptor for severe acute respiratory syndrome coronavirus. *Proc Natl Acad Sci USA.* 2004; 101(44): 15748–15753. <https://doi.org/10.1073/pnas.0403812101> PMID: 15496474
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. <https://doi.org/10.1038/cr.2008.15> PMID: 18227861

13. Li J, Kou Y, Yu X, Sun Y, Zhou Y, Pu X, et al. Human co-infection with avian influenza and seasonal influenza viruses, China. *Emerg Infect Dis*. 2014; 20:1953–1955. <https://doi.org/10.3201/eid2011.140897> PMID:25340661
14. Liu Y, Ning Z, Chen Y, Guo M, Liu Y, Gali NK, et al. Aerodynamic analysis of SARS-CoV-2 in two Wuhan hospitals. *Nature*. 2020; 582(7813): 557–60. <https://doi.org/10.1038/s41586-020-2271-3> PMID: 32340022
15. 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. <https://doi.org/10.1371/journal.pone.0142170> PMID: 26562668
16. Brundage JF. Interactions between influenza and bacterial respiratory pathogens: implications for pandemic preparedness. *Lancet Infect Dis*. 2006; 6: 303–312. [https://doi.org/10.1016/S1473-3099\(06\)70466-2](https://doi.org/10.1016/S1473-3099(06)70466-2) PMID: 16631551
17. 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. <https://doi.org/10.1371/journal.pmed.1001051> PMID: 21750666
18. Tsuchihashi Y, Sunagawa T, Yahata Y, Takahashi H, Toyokawa T, Odaira F, et al. 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: 381–383. <https://doi.org/10.1093/cid/cir787> PMID: 22100572
19. 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. <https://doi.org/10.1186/1471-2105-13-134> PMID: 22708584
20. 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): 3784–3788. <https://doi.org/10.1093/nar/gkg563> PMID: 12824418
21. Morgat A, Lombardot T, Coudert E, Axelsen K, Neto TB, Gehant S, et al. Enzyme annotation in UniProtKB using Rhea. *Bioinformatics*. 2020; 36(6): 1896–1901. <https://doi.org/10.1093/nar/gkg563> PMID: 12824418

22. Ng PC, Henikoff S. 2006. Predicting the effects of amino acid substitutions on protein function. *Annu Rev Genom Hum Genet.* 2006; 22(7): 61–80. <https://doi.org/10.1146/annurev.genom.7.080505.115630> PMID: 16824020
23. 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. <https://doi.org/10.1371/journal.pone.0046688> PMID: 23056405
24. Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, et al. A method and server for predicting damaging missense mutations. *Nat Methods.* 2010; 7(4): 248–249. <https://doi.org/10.1038/nmeth0410-248> PMID: 20354512
25. Smigielski EM, Sirotkin K, Ward M, Sherry ST. dbSNP: a database of single nucleotide polymorphisms. *Nucleic Acids Res.* 2000; 28: 52–355. <https://doi.org/10.1093/nar/28.1.352> PMID: 10592272
26. 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–W533. <https://doi.org/10.1093/nar/gkq399> PMID: 20478830
27. Pires DEV, Ascher DB, Blundell TL. mCSM: predicting the effects of mutations in proteins using graph-based signatures. *Bioinformatics* 2012; 30: 335–342. <https://doi.org/10.1093/bioinformatics/btt691> PMID: 24281696
28. Librado P, Rozas J. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* 2009; 25: 1451–1452. <https://doi.org/10.1093/bioinformatics/btp187> PMID: 19346325
29. French N, Yu S, Biggs P, Holland B, Fearnhead P, Binney B, et al. Evolution of *Campylobacter* species in New Zealand. In: Sheppard SK, Méric G (eds) *Campylobacter ecology and evolution*. Caister Academic Press, Norfolk, 2014; 221 – 240. url: <https://www.caister.com/campy2>
30. Elbe S, Buckland-Merrett G. Data, disease and diplomacy: GISAID's innovative contribution to global health. *Glob. Chall.* 2017; 1: 33-46. <https://doi.org/10.1002/gch2.1018> PMID: 31565258
31. Letunic I, Bork P. Interactive tree of life (iTOL) v4: recent updates and new developments. *Nucleic Acids Res.* 2019; 47(W1): W256–W259. <https://doi.org/10.1093/nar/gkz239> PMID: 30931475
32. Barrett JC, Fry B, Maller JDM J, Daly MJ. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics.* 2005; 21(2): 263-265. <https://doi.org/10.1093/bioinformatics/bth457> PMID: 15297300

33. Memish ZA, Aljerian N, Ebrahim SH. Tale of three seeding patterns of SARS-CoV-2 in Saudi Arabia. *Lancet Infect Dis.* 2020; S1473-3099(20)30425–4. [https://doi.org/10.1016/S1473-3099\(20\)30425-4](https://doi.org/10.1016/S1473-3099(20)30425-4) PMID: 32442522
34. Ghinai I, McPherson TD, Hunter JC, Kirking HL, Christiansen D, Joshi K, et al. 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): 1137–1144. [https://doi.org/10.1016/S0140-6736\(20\)30607-3](https://doi.org/10.1016/S0140-6736(20)30607-3) PMID: 32178768
35. Mercatelli D, Giorgi FM. Geographic and Genomic Distribution of SARS-CoV-2 Mutations. *Front Microbiol.* 2020; 22;11: 1800. <https://doi.org/10.3389/fmicb.2020.01800> PMID: 32793182
36. Ceraolo C., Giorgi FM. Genomic variance of the 2019-nCoV coronavirus. *J Med Virol.* 2020; 92(5): 522–528. <https://doi.org/10.3389/fmicb.2020.01800> PMID: 32027036
37. 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): 1470–1477. <https://dx.doi.org/10.3201/eid2607.200282> PMID: 32255761
38. Duffy S. Why are RNA virus mutation rates so damn high? *PLoS Biol.* 2018; 16(8): e3000003. <https://doi.org/10.1371/journal.pbio.3000003> PMID: 30102691
39. Andersen KG, Rambaut A, Lipkin WI, Holmes EC, Garry RF. The proximal origin of SARS-CoV-2. *Nat Med.* 2020; 26: 450–452. <https://doi.org/10.1038/s41591-020-0820-9> PMID: 30102691
40. Brufsky, A. Distinct viral clades of SARS-CoV-2: implications for modeling of viral spread. *J. Med. Virol.* 2020; 10.1002/jmv.25902. <https://doi.org/10.1002/jmv.25902> PMID: 32311094
41. Morniroli D, Gianni ML, Consales A, Pietrasanta C, Mosca F. Human Sialome and Coronavirus Disease-2019 (COVID-19) Pandemic: An Understated Correlation? *Front Immunol.* 2020; 11. <https://doi.org/10.3389/fimmu.2020.01480> PMID: 32655580
42. Cox MJ, Loman N, Bogaert D, O'grady J. Co-infections: potentially lethal and unexplored in COVID-19. *The Lancet Microbe.* 2020; 1(1): e11. [https://doi.org/10.1016/S2666-5247\(20\)30009-4](https://doi.org/10.1016/S2666-5247(20)30009-4) PMID: 32835323
43. Liu W, Li Z-D, Tang F, Wei M-T, Tong Y-G, Zhang L, et al. Mixed infections of pandemic H1N1 and seasonal H3N2 viruses in 1 outbreak. *Clin Infect Dis.* 2010; 50: 1359–1365. <https://doi.org/10.1086/652143> PMID: 20380566



© 2020 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).