

# Journal Pre-proof

Mutations in SARS-CoV-2; Consequences in structure, function, and pathogenicity of the virus

Behnaz Bakhshandeh, Zohreh Jahanafruz, Ardeshir Abbasi, Matin Babae Goli, Mahya Sadeghi, Mohammad Sadeq Mottaqi, Maryam Zamani



PII: S0882-4010(21)00103-0

DOI: <https://doi.org/10.1016/j.micpath.2021.104831>

Reference: YMPAT 104831

To appear in: *Microbial Pathogenesis*

Received Date: 6 December 2020

Revised Date: 27 February 2021

Accepted Date: 1 March 2021

Please cite this article as: Bakhshandeh B, Jahanafruz Z, Abbasi A, Goli MB, Sadeghi M, Mottaqi MS, Zamani M, Mutations in SARS-CoV-2; Consequences in structure, function, and pathogenicity of the virus, *Microbial Pathogenesis* (2021), doi: <https://doi.org/10.1016/j.micpath.2021.104831>.

This is a PDF file of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability, but it is not yet the definitive version of record. This version will undergo additional copyediting, typesetting and review before it is published in its final form, but we are providing this version to give early visibility of the article. Please note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

© 2021 Published by Elsevier Ltd.

## **Mutations in SARS-CoV-2; Consequences in Structure, Function, and Pathogenicity of the Virus**

Behnaz Bakhshandeh<sup>1\*</sup>, Zohreh Jahanafrooz<sup>2</sup>, Ardeshir Abbasi<sup>3</sup>, Matin Babae Goli<sup>1</sup>, Mahya Sadeghi<sup>4</sup>, Mohammad Sadeq Mottaqi<sup>5</sup>, Maryam Zamani<sup>5</sup>

- 1- Department of Biotechnology, College of Science, University of Tehran, Tehran, Iran
- 2- Department of Biology, Faculty of Sciences, University of Maragheh, Maragheh, Iran
- 3- Department of Immunology, Faculty of Medical Sciences, Tarbiat Modares University, Tehran, Iran
- 4- Department of Biomedical Engineering, Islamic Azad University, Central Tehran Branch, Tehran, Iran
- 5- School of Biology, College of Science, University of Tehran, Tehran, Iran

Running title:

SARS-CoV2 Mutations and Related Pathogenicity

---

\* Address correspondence: Behnaz Bakhshandeh, Department of Biotechnology, College of Science, University of Tehran, P.O. Box: 14155-6455, Tehran, Iran. Tel: +98(21)66491622. Fax: +98(21)66413835. E-mail: b.bakhshandeh@ut.ac.ir

**Abstract**

The third pandemic of coronavirus infection, called COVID-19 disease, began recently in China. The newly discovered coronavirus, entitled SARS-CoV-2, is the seventh member of the human coronaviruses. The main pathogenesis of SARS-CoV-2 infection is severe pneumonia, RNAemia, accompanied by glass turbidity, and acute cardiac injury. It possesses a single-stranded positive-sense RNA genome which is 60–140 nm in diameter, and has a size of 26 to 32 kbp. Viral pathogenesis is accomplished with spike glycoprotein through the employment of a membrane-bound aminopeptidase, called the ACE2, as its primary cell receptor. It has been confirmed that various factors such as different national rules for quarantine and various races or genetic backgrounds might influence the mortality and infection rate of COVID-19 in the geographic areas. In addition to various known and unknown factors and host genetic susceptibility, mutations and genetic variabilities of the virus itself have a critical impact on variable clinical features of COVID-19. Although the SARS-CoV-2 genome is more stable than SARS-CoV or MERS-CoV, it has a relatively high dynamic mutation rate with respect to other RNA viruses. It's noteworthy that, some mutations can be founder mutations and show specific geographic patterns. Undoubtedly, these mutations can drive viral genetic variability, and because of genotype-phenotype correlation, resulting in a virus with more/lower/no decrease in natural pathogenic fitness or on the other scenario, facilitating their rapid antigenic shifting to escape the host immunity and also inventing a drug resistance virus, so converting it to a more infectious or deadly virus. Overall, the detection of all mutations in SARS-CoV-2 and their relations with pathological changes is nearly impossible, mostly due to asymptomatic subjects. In this review paper, the reported mutations of the SARS-CoV-2 and related variations in virus structure and pathogenicity in different geographic areas and genotypes are widely investigated. Many studies need to be repeated in other regions/locations for other people to confirm the findings. Such studies could benefit patient-specific therapy, according to genotyping patterns of SARS-CoV-2 distribution.

**Keywords**

SARS-CoV-2; COVID-19; Mutation; Pathogenicity; Pandemic Disease.

## 1 COVID-19 Disease

The COVID-19 disease first started spreading in China and then in almost every region of the globe. As more cases were confirmed, many countries worldwide announced restrictions for transportation and implemented quarantine [1]. Findings revealed that not only people with the disease who had a range of symptoms could transmit the virus, but even those who did not develop symptoms were carriers of the virus [2]. In some instances, people infected with the virus while not experiencing any symptoms of COVID-19 or even showing mild symptoms of the disease. On the other hand, signs of the disease may change from mild and moderate to acute (critical) and become severe, necessitating special care for the patient with the development of the acute respiratory distress syndrome (ARDS) syndrome and accumulation of fluids in the lungs [2-7]. According to epidemiological researches, the clinical outcomes (ranging from asymptomatic to death) are heavily influenced by factors such as age, demographic characteristics, clinical care quality, and other potential unknown parameters [8-10]. Additionally, a healthy diet is recognized as another substantial factor in preventing the development of the severe form of the disease by boosting the immune system. [11]. The presence of some underlying diseases (e.g., diabetes) in a person is another factor affecting the increased probability of COVID-19 development and the incidence of a severe condition of the disease (Figure 1) [12]. At the same time, the virus can cause bacterial pneumonia infections [13]. To ensure the presence of the virus in the body, polymerase chain reaction (PCR) molecular testing is done. The other technique is computed tomography of the chest. Additionally, the number of white blood cells, creatine phosphokinase (CPK), erythrocyte sedimentation rate (ESR), c-reactive protein (CRP), and lactate dehydrogenase D (LDHD) is measured in the blood [2, 14].

## 2 Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2)

SARS-CoV-2 belongs to a large family of viruses known as coronaviruses. Coronaviruses are the largest group of viruses belonging to the Nidovirales containing Coronaviridae families, Roniviridae, and Arteriviridae. Coronaviridae is a family of positive-sense single-stranded enveloped RNA viruses, with a genome between 26 and 32 kilobases in length. This viral group consists of four different genera, namely *Alpha*, *Beta*, *Gamma*, and *Delta*, and encompasses various human and animal coronaviruses [15]. Human coronavirus was first identified in the mid-1960s. Seven human coronaviruses have so far been identified that belong to the *Alpha* and *Beta* genera; as depicted in Figure 2, genus *Alpha*, including (human coronavirus NL63 (HCoV-NL63) and human coronavirus 229E (HCoV-229E)) and genus *Beta*, including (human coronavirus OC43 (HCoV-OC43), human coronavirus HKU1

(HCoV-HKU1), severe acute respiratory syndrome coronavirus (SARS-CoV), Middle East respiratory syndrome coronavirus (MERS-CoV) and current SARS-CoV-2) [16]. They bring about a wide range of diseases, including the common cold and respiratory diseases with mild and acute symptoms.

To date, five coronaviruses belonging to beta-coronavirus lineage have been detected in association with acute pneumonia in humans (Figure 2) [17]. SARS-CoV-2 is the third coronavirus after MERS and SARS-CoV. [18, 19].

Bats have been reported to be an essential host for some coronaviruses. Examples include SARS-CoV, a virus with acute respiratory symptoms that spread through China, killing hundreds of people in 2003-2004. After that, MERS-CoV was reported in the middle east in 2012; both bat-originated viruses caused the infection and death of many people [17, 20]. In 2020, a new disease with a growing global prevalence was named COVID-19 by the World Health Organization. The COVID-19 pandemic has recently emerged and then rapidly spread in more than 200 countries worldwide so far. According to the World Health Organization, 39.8 M global cases of COVID-19 were confirmed on October 16<sup>th</sup> with a mortality rate of 3.4%.

### **3 The Biology of SARS-CoV-2**

Structural and non-structural proteins in the virus genome are encoded in the open reading frame (ORF) sequences (Figure 3). Almost 14 ORF have been sequenced in the SARS-CoV-2 genome so far [21]. Molecular analysis of the new coronavirus indicated that it possesses a single-stranded positive-sense RNA genome, which its size varies between 29.8 kb and 29.9 kb, being 60–140 nm in diameter. Its genome is almost 80% homologous with SARS-CoV, and nearly 96% of its genes are shared with the bat coronavirus (Bat-CoV) [22]. Each of these ORFs is responsible for coding 17 proteins, both structural and non-structural, that drive various processes during the entire life of the virus from its survival to virulence power [23, 24]. The genome starts with 5' UTR, which consists of a variable number (between 6 and 11) of ORFs, including ORF1ab, ORF3, ORF6, ORF 7a, ORF 8, and ORF10. SARS-COV 2 has eight accessory proteins (3a, 3b, p6, 7a, 7b, 8b, 9b, and orf14), encoded by ORF3a, ORF6, ORF7a, ORF7b, and ORF8 genes, respectively. These genes contain vital information for virus replication, that has been preserved through successive generations [15].

#### **3.1 Non-structural proteins**

During the virus transformation phase, the first event is the expression of orf1ab and orf1a gene segments to produce two large polyproteins named pp1ab and pp1a. These two polyproteins are then cleaved into 16 Nsps, including Nsp1 to Nsp10 and Nsp12 to Nsp16 by

the action of two protease enzymes as part of each of those polypeptides named papain-like proteases (PLpro) [25]. These Nsps participate in many viral processes; for example, the Nsp3 is the largest component among other constituents that forms the replication and transcription complex of coronaviruses, which its cleavage by the PL proteinase would result in the promotion of cytokine expression and suppression of the host's innate immune response. 3CLPro protease as another Nsp has an essential role in RNA replication. Many of the Nsps encoded in ORF1ab, including Nsp5, Nsp3, and the RNA-dependent RNA polymerase (RdRp, Nsp12), are responsible for the assembly of the replicas-transcriptase complex in double-membrane vesicles (DMVs) [16]. Other replicas consist of proteins that facilitate viral replication and block the intrinsic host immune system by controlling the cellular machinery. Nsp8 and Nsp12 are two monomeric RdRps that are involved in the process of SARS-CoV-2 replication. Nsp8 replicates the virus genome through interacting with ORF6 accessory protein and Nsp9 replicas protein, by binding to RNA. In contrast to Nsp12, Nsp8 has primase capacity that makes it needless of any primers to initiate the viral replication process. In addition to these polymerases, the SARS-CoV-2 has a unique multimeric RNA polymerase composed of Nsp7 and Nsp8, which is responsible for both the initiation and elongation of the newly synthesized fragment of the viral genome [16, 23].

### **3.2 Structural proteins**

The 3' end of the SARS-CoV-2 genome consists of four structural protein-encoding genes, which include spike (S), membrane (M), envelope (E), and nucleocapsid (N). From entering the host cells to virus particle formation, various viral processes are dependent on these proteins (Figure 4) [21, 25, 26].

#### **3.2.1 S protein**

The spike or S glycoprotein is a transmembrane protein that resides on the outer surface of the virus, with nearly 150 kDa molecular weight that determines the host range and pathogenicity of the virus by its contribution to viral entry through mediating the attachment of the virus particle to the plasma membrane of the host cell. The S protein of SARS-CoV-1 and SARS-CoV-2 forms homo-trimers protruding in the viral surface that facilitates viral entry into the host cells via interacting with angiotensin-converting enzyme 2 (ACE2) that is their main receptor expressed in lower respiratory tract cells. However, in some circumstances, they might use CD209L as an alternative receptor. In contrast, MERS-CoV uses dipeptidyl peptidase 4 (DPP4) as the primary receptor. These viruses spread between

different cell types with the help of this mechanism that suppresses virus-neutralizing antibody responses [16, 21, 25].

### **3.2.2 M protein**

Membrane or M protein determines the shape of the virus envelope. It can function as a small transmembrane protein through binding to all other structural proteins. The M protein also contributes to the packaging of the viral RNA genome into a helical ribonucleocapsid during virion formation [16].

### **3.2.3 N protein**

The Nucleocapsid or N protein is the structural constituent of SARS-CoV-2 that is bound to the RNA genome of the virus. This protein takes part in various processes related to the viral genome, including the replication cycle of viruses, viral genome signaling, and the host cells' response to viral infections [22, 25].

### **3.2.4 E protein**

The structural evidence shows that the N-amino terminal of the envelope or E protein is a short hydrophilic sequence consisting of 7–12 amino acids. It is followed by a large hydrophobic middle segment with 25 amino acids known as the transmembrane domain and possesses a long hydrophilic C-terminus which constitutes the most of the protein. E protein is involved in the production and maturation processes of the virus through interacting with the host cell membrane protein [22].

## **4 The Mechanism of SARS-CoV-2 Infection**

The presence of S protein in the process of pathogenesis of SARS-CoV-2 is critical. Spike protein uses a membrane-bound aminopeptidase called ACE2 as the primary cell receptor that binds to dipeptidyl peptidase-4 on the virus [27]. ACE2 is widely expressed in the various organs, including the heart, lungs, gastrointestinal tract, and kidneys [27]. Signal transmission through ACE2 is disrupted when severe myocardial damage and dysfunction in the lungs and other organs such as the kidneys and heart are induced when the virus attaches to the ACE2 receptor [28]. Besides, the determination of target cell specificity is influenced by the S glycoprotein interaction with cellular receptors. The location of the receptor binding domain (RBD) in S1 domain of S protein is determined by the structure of each virus, which can be variable. The RBD regions are located at N Terminal or C1 Terminal of S1. SARS-CoV-2 uses some peptidases as their cellular receptors, although their entry occurs even in the absence of secondary enzymes [25, 27].

After fastening to the receptor, the virus enters the host cell cytosol by transmembrane protease serine 2 (TMPRSS2) or other proteases such as furin, and fuses the viral and cell

membranes. SARS-CoV-2 spike glycoprotein contains a cleft-like site. The furin detection location is crucial for detection by pyrolysis and, thus, in viral infection of the virus. This glycoprotein is broken down by furin proteinase into a subunit S1 and S2 like a host cell. Infection begins with the interaction of S1, containing RBD, with the host receptor ACE2, into the cell membrane. The S2 domain is a viral fusion protein and is involved in a mixture of virion glands. Immediately after the virus entrance to the cell, the S glycoprotein is cleaved by the CTSL cathepsin to kill the S2 fusion peptide, and then forms endocytosis and low-pH endosomes, hence activating the fusion of membranes within the endosomes. At the junction between the virus membrane and the cell membrane, a bundle is formed that is likely to release the virus genome into the cytoplasm. The following step in the infection cycle is the translation of replica genes from the RNA virus. The two polymer proteins pp1a and pp1ab are encoded by two large proteins, ORF rep1a and rep1b. A sliding sequence (5'-UUUAAAC-3') and a pseudoknot RNA help express these proteins, leading to a deformation of the rep1a reading frame in the ORF rep1b. Usually, the ribosome opens the pseudoknot structure and continues the translation process until it encounters a pause or reach the end of 11 repetitions in the codon. But sometimes, the ribosome is stopped by the pseudoknot, which prevents stretching and staying in the slippery sequence, and stops the nucleotide reading frame with a backward shift that results in the translation of pp1ab. After the translation process, viral RNA is synthesized, and viral replication complexes are collected. The set of viral replications by the transcription process results in cytopathic effects such as abnormalities in the vascular complexes. In the downstream region of the structural proteins, subgenomic RNAs act as mRNAs for the structural genes of the virus [29].

The next step is the binding and translating the of structural glycoproteins of virus, including S, M and E to ER. These proteins are secreted into the ER-Golgi intermediate chamber (ERGIC). Here, the viral genomes that encode structural proteins are encapsulated by the N-glycoprotein in the ERGIC membrane and become mature viruses. Protein-protein interaction is mainly mediated by M glycoprotein, which is responsible for collecting CoVs. M protein, expressed with E protein, is involved in the formation of virus-like particles (VLPs) as well as in the formation of coatings for CoVs. While N glycoprotein causing the formation of VLPs, a combination of genomes encoded in ERGIC enhances virus growth. In the ERGIC chamber, S glycoprotein interacts with M glycoprotein due to being essential for its synthesis. Because M protein is abundant compared to E protein, M protein interaction can be a major source of motivation to mask puberty. Glycoprotein changes the host secretion pathway and releases the accumulation of the virus in the host. The M protein attaches to the nucleocapsid



located at the C-terminus of endo domain M and signals viral assembly completion. After assembly, the viruses are transmitted by vesicles to the host cell membrane and are released by exocytosis. By being transported to the cell surface, the S protein combines cell fusion between infected and healthy cells, resulting in a large multinucleated cell that spreads the virus in the host and provides the conditions for the detection of specific antibodies against the virus (Figure 5) [29].

## 5 The SARS-CoV-2 Mutations and Related Alterations in The Virus

The pace of mutation in SARS-CoV-2 as an RNA virus is possibly  $10^{-4}$  replacement of bp for every year. Besides, variations may appear in the course of each genome replication cycle [30]. Single-nucleotide polymorphism (SNPs) is applied for the comparison of DNA sequences. It can be used for evolutionary investigations, being useful in identifying mutations in the coronavirus genome, where many mutations may occur by cause of an RdRp in the process of genome replication [31, 32]. Table 1 and Figure 6 summarize the reported mutations in various parts of SARS-CoV-2 over the last months.

Table 1. Some reported mutations in various parts of SARS-CoV-2

VIRUS GENOME REGIONS	MUTATION SITE	AMINO ACID CHANGE	MUTATION TYPE	REF.	
NSP2 (NT181...818)	241C>T	NA*	NA	[16, 33, 34]	
	514T>C	NA	Synonymous	[33]	
	1059C>T	T→I	Missense	[33]	
	1397	V→I	Missense	[33, 35]	
	2662C>T	NA	NA	[24]	
	2891	NA	NA	[35]	
	3036	NA	Synonymous	[35]	
	3037C>T	NA	Synonymous	[16, 33, 34]	
	ORF1AB (NT266...21,555)	8782C>T/Y	NA	Synonymous	[24, 33, 35, 36]
		11083G>T	NA	NA	[24, 33, 35]
14408C>T		P→L	Missense	[16, 33-35]	
14805C>T		NA	Synonymous	[33]	
17373C>T		NA	NA	[24]	
17746		P→L	Missense	[33, 35]	
17857		C→Y	Missense	[35]	
18060C>T		NA	Synonymous	[24, 33,	

				35]
<b>S (NT21,563...25,384)</b>	21707C>T	NA	NA	[24]
	23403A>G	D→G	Missense	[16, 33, 35]
	24034C>T/Y	NA	NA	[24]
<b>ORF3A(NT25,393...26,220)</b>	26143	G→V	Missense	[33, 35]
	26144G>T	G→V	Missense	[24]
<b>M (NT26,523...27,191)</b>	26729T>C/Y	NA	NA	[24]
<b>ORF8 (NT27,894...28,259)</b>	27046C>T	T→M	Missense	[33]
	28077G>C/S	NA	NA	[24]
	28144T>C	L→S and G→R	Missense	[16, 24, 35, 36]
<b>N (NT28,274...29,533)</b>	28881G>A	NA	NA	[33, 35]
	28882G>A	NA	NA	[16, 33]
	28883G>C	NA	NA	[16, 33]
	29095C>T	NA	Synonymous	[33, 36]

In a study done by Yin, 558 COVID-19 strains were investigated to calculate the frequencies of mutations. The first regular SNP mutation in the virus genome was in nsp2 (241 C>T), which is a critical genomic location. It has been discovered that the nsp2 mutation (241 C>T) has emerged with three other critical mutations including, (3037 C>T), (14408 C>T), and (23403 A>G), that bring about mutations at the amino acid level of RNA primase, and S protein. The mutations as mentioned earlier (241 C>T, 14408 C>T, and 23403 A>G) are located in proteins which are essential for the replication of RNA (241C>T, 14408 C>T), and the spike protein (23403A >G) since it attaches to receptors. This paper recognized that these four mentioned-mutations are common in the SARS-CoV-2 European isolates genomes, where the severity of the infection is mostly more intense than in the other geographical regions. The second amplest SNP mutation in the isolates was (28144 T>C), which has appeared at nsp8, where the amino acid is changed (Leu to Ser), and the third one (26144 G>T) was in nsp3. The nsp3 collaborates with two other proteins called nonstructural protein-4 and 6, to bring about DMVs, which are complexes of the membrane that assist in replicating and assembling RNA. The most common mutation (23403 A>G) was positioned in the S protein sequence. This mutation in D614G (23403 A>G) was placed in the supposed S1–S2 junction part, which splits the spike protein. Analysis of the results demonstrated that nsp8, which is an RNA primase, had the greatest number of mutations (28881 G>A, 28144 T>C, 28882 G>A and 28883 G>C) in comparison to the remaining proteins. Mutations in RNA polymerase or RNA primase might increase fidelity and present obstruction to the

usage of nucleotide analogs that are mutagenic. In strains from the US, three mutations in the sequences of nsp8 (28881 G>A, 28882 G>A, and 28883 G>C) have been discovered [16].

ORF1ab gene, which has the most length in the ORFs in the virus genome and is split into viral non-structural proteins (nsp1 to nsp16), was reported as a gene with a high number of missense mutations [23, 37].

After analyzing around 48,000 SARS-CoV-2 genome sequences and comparing them with the reference sequence, researchers demonstrated that there reside at least three clades according to geographical and genomic relevance. The clade common in Europe obtains the p.614(Asp to Gly) mutation associated with the S protein along with (14408 C>T), (241 C>T), and (3037 C>T). The other significant mutation was in the N protein (p.RG203KR) [34].

Based on a study that analyzed 220 genome sequences from COVID-19 patients, the scientists appraised the disposition of the virus mutations within four geographic regions by standardizing the number of genomes containing a likely mutation in each region. The appearance of mutations in nucleotide locations 3036, 8782, 11083, 28144, and 26143 has been confirmed. Besides, mutations were found in positions 1397, 2891, 14408, 17746, 17857, 18060, 23403, 28881, existing in sequences of ORF-1ab (1397, 2891, 14408, 17746, 17857, and 18060), S (23403) and N (28881). three silent mutations (locations 3036, 8782, and 18060) were observed among the twelve most common ones, considering that one mutation (location 11083) was not in the sequence of ORF. Also, mutations in these locations cause amino acids to change in this order: 1397 (Val to Ile), 14408 (Pro to Leu), 17746 (Pro to Leu), 17857 (Cys to Tyr), 23403 (Asp to Gly), 26143 (Gly to Val), 28144 (L to Ser). One mutation (location 28881) causes the substitution of 2 amino acids, particularly 28881 (Arg to Lys) and (Gly to Arg) [38].

According to a paper that has analyzed the SARS-CoV-2 genomes, 10 out of the 48 sequences merely had lost their base pairs at the start and the end. However, 80 alternatives contained mutations including, 43 missenses, 21 synonymous, three deletions, and 13 noncoding deletion types. The most frequent alternatives (13 specimens) were in ORF-1ab (8782 C>T) and ORF-8 (28144 T>C), and after that, in N (29095 C>T) (5 specimens). Mutations in locations (8782 C>T) and (28144T >C) were observed at the same time. Also, (29095 C>T) is detected in both of them. The two mutations (8782 C>T) and (29095 C>T) were synonymous ones; nonetheless, (28144 T>C) has altered amino acid in the ORF8 region (p. L84S). Out of the 43 missense alternatives, 30 ones were observed in ORF-1ab. All three deletion mutations found in this study, are in-frame ones and were observed in nsp1, which is

in ORF1ab. All of the noncoding deletion mutations were detected in 3'-UTR or 5'-UTR positions, and researchers suggest that they do not significantly influence the functions of the virus. Through various alignments with other SARS-CoV-2 ORF8 regions, they have found out that L84, which is engaged in (28144 T>C) (p.L84S), is not conserved [36]. Khailany et al. have confirmed these results in their separate research [23].

Wang et al. also analyzed 95 SARS-CoV-2 genome sequences. They compared the isolates with the reference sequence and depicted that the total homology was 99.99%, whereas 15 isolates were identical with the reference one. Also, the homology of the ORF-1a in the genome was 99.99%, whereas the homology in connection with the bulk of other parts being 100% while no alteration in regions of E, ORF-6, and ORF-7b, was found. The results show that by comparing the amino acids of the sequences, the overall homology was 99.99%, while the homology between the bulk of the isolates in individual parts was 100%. Mutations which occurred in more than three variants were observed in particular nucleotide positions including, ORF-1ab (2662, 8782, 11083, 17373, 18060), S (21707, 24034), ORF3A (26144), M (26729), ORF8 (28077, 28144) and N (28854, 29095). Notably, the highest mutation rate was detected in ORF8 (28144) which was about 30% (29 out of 95), where C took the place of T. Likewise, the mutation rate of position ORF-1ab (8782) was about 29% (28 out of 95), where mostly T took the place of C. Mutations in amino acid sequences appearing in more than three strains were p.3606(Leu to Phe), p.49(His to Tyr), p.860(Val to Gln), p.251(Gly to Val), p.62(Val to Leu), p.84(Leu to Ser), p.194(Ser to Leu). Besides, six deletion mutations were observed in 5 of the sequences. These mutations led to 4 various shortened variations in the sequence of amino acids. Moreover, two more deletion-type mutations were observed in 3'-UTR or 5'-UTR of the genome, respectively [24].

Yu et al. stated that 120 substitution locations were observed in 8 coding regions of the viral genomes, encompassing 79 transitions (66%) and 41 transversions (34%). The 120 substitution locations were engaged in 119 codons, containing 79 nonsynonymous (66%) and 40 synonymous (34%) mutations. Forty nonsynonymous mutations (51%) altered the biochemical characteristics of the original amino acids, and thus are probably engaged in the adjustment of the virus [39].

Multiple sequence alignment of 86 patient-isolated strains of the SARS-CoV-2 disclosed 93 mutations including 42 missense mutations in viral proteins except for the E protein, 29 missense mutations in the ORF1ab protein, and 8 of them in the S protein [40].

In another paper, ten viral genomes derived out of the NCBI database were analyzed through sequence alignment. They did not see any divergence in the amino acid arrangement of N and

M proteins. Two alternatives in terms of amino acid level, in spike protein, were detected. According to the reference sequence, the authors of this study suggested two SNP subtypes named "L" and "S". Two likely "L" and "S" mutations within the sequences of ORF1ab and ORF8 were discovered. One mutation was found through evaluating the sequence alignment of E protein, in which the mutation "His" was detected from a South Korean genome in contrast with "Leu" from alternative nine ones in the alignment. In the course of studying the spike protein, one mutation of "Trp" was found in a South Korean genome in contrast with "Ser" from the alternative nine ones. Also, one mutation of "Arg" from an Australian sequence was observed considering "Ser" from alternative nine ones [22].

Holland et al. discovered that one of the viral genomes had an 81 nucleotide in-frame deletion in the ORF-7a sequence, leading to a loss of 27 amino acids [41]. Comparable deletion mutations in the virus genomes has appeared, especially in the ORF-8 sequence, which may likely lessen the fitness of the SARS-CoV-2 [42, 43].

By computational tools, [44] established a reference genome sequence of SARS-CoV-2, and revealed 80 mutated bases at 47 locations. Besides, the most significant SNP mutation in the S protein was found to be A>G.

## **6 The SARS-CoV-2 Mutations and Its Related Variabilities in Clinical Manifestations**

It is confirmed that the different mortality and infection rate of COVID-19 in geographic areas might be influenced by various factors such as different national rules for quarantine and movement restrictions, and also different races or the genetic backgrounds of people [38]; in particular, SNP array data analyzing revealed that a specific haplotype of TMPRSS2 gene in a European population which results in upregulation of TMPRSS2 gene expression (as a crucial enzyme for SARS-CoV-2 entry into the host cell) in an androgen-dependent manner, showed association with disease severity in Italian population [45]. Moreover, viral mutations over time are regarded as the contributing factors in transmission and severity of the COVID-19 infection [38]. Whereas the SARS-CoV-2 genome is more stable than SARS-CoV or MERS-CoV, it has a relatively high dynamic mutation rate comparable to other RNA viruses [46]. It's noteworthy that, some mutations can be as founder mutations and show specific geographic patterns. For example, the D614G mutation in S protein is a founder mutation in Europe (appearance from 16<sup>th</sup> February 2020) [47]. Undoubtedly, these mutations can drive viral genetic variability, and because of genotype-phenotype correlation, resulting in a virus with more/lower/no decrease in natural pathogenic fitness or on the other hand, facilitating their rapid antigenic shifting in order to escape the host immunity and also creating a drug resistance virus, hence converting it to a more infectious or deadly virus

(Table 2) [47-49]. Issa et al. discovered 51 distinct non-synonymous amino-acid substitutions in the ORF3a proteins by analyzing 2,782 SARS-CoV-2 sequenced isolates. The functional domains of the ORF3a protein are related to pathogenicity and virulence [50].

Table 2. Correlation of SARS-CoV-2 mutations with pathogenicity.

Location/Source	SARS-CoV-2 Genome Variation	Associated Phenotypic Changes	Ref.	
Various Sequences Database (Such As Gisaid Database) Analyzing	p.614(Asp to Gly) in S (Missense)	Higher fatality rate	[51]	
	p.367(Val to Phe) in RBD of S (Missense) p.436(Trp to Arg) in RBD of S (Missense) p.364(Asp to Tyr) in RBD of S (Missense)	Higher infectivity (higher affinity to human ACE2 receptors)	[52]	
	p.614(Asp to Gly) in S (Missense)	Higher infectivity	[52]	
	p.475(Ala to Val) in RBD of S (Missense) p.452(Leu to Arg) in RBD of S (Missense) p.483(Val to Ala) in RBD of S (Missense) p.490(Phe to Leu) in RBD of S (Missense)	Resistant to multiple neutralizing antibodies	[52]	
	p.408(Arg to Ile) in RBD of S (Missense) p.455(Leu to Tyr) in RBD of S (Missense) p.486(Phe to Leu) in RBD of S (Missense) p.493(Gln to Asn) in RBD of S (Missense) p.498(Gln to Tyr) in RBD of S (Missense) p.501(Asn to Thr) in RBD of S (Missense) p.930(Ala to Val) in heptad repeat 1 (HR1) domain of S (Missense) p.936(Asp to Tyr) in heptad repeat 1 (HR1) domain of S (Missense)	Reducing stability of spike proteins	[53]	
	p.3691 in NSP6 p.9659 in ORF-10	Reducing the stability of the proteins structures	[54]	
	8782C>T in ORF-1ab (Synonymous Mutation) 28144T>C in ORF-8 (Missense) 29095C>T in N (Synonymous Mutation)	Potential Effects On Transmission and Severity of COVID-19	[23]	
	<b>Singapore</b>	Elimination Of ORF8 (Deletion Of 382nt)	Attenuated Phenotype	[43]
	<b>During Passaging Virus in Vero-E6 Cells</b>	15-30nt Deletion in S1/S2 Junction Region	Attenuated Phenotype	[49]
	<b>China</b>	23403A>G in S (Missense) 23902A>C in S (Missense) 23607A>A in S (Missense)	Strengthen Phenotype	[47]

22303T>G and 22301A>C in S (Missense) 27775T>C, 27776T>G, And 27777G>A in S (Missense) 22205G>C in S (Missense)		
--	--	--

Eleven SARS-CoV-2 Chinese patient-isolated viral genome sequences were analyzed through ultra-deep sequencing regarding the mutations in the genome. Then researchers evaluated their infectivity in the Vero-E6 cells by Cycle threshold values (Ct) to quantify the viral load through *in-vitro* assays that resulted in identification of a total of 33 mutations. They recognized that these viral isolates demonstrate up to 270-fold differences in viral load and cytopathogenic effects (CPE) during spreading among Vero-E6 cells [55]. After researching 12,343 SARS-CoV-2 genome patient-isolated sequences worldwide, Toyoshima et al. found 131 mutations with frequencies of more than 10%. After multiple-regression analysis, they discovered that the frequency of p.614(Asp to Gly) in spike protein, and P4715L in ORF1ab had depicted positive correlations with the fatality rate of the virus ( $r = 0.45$ ,  $P = 0.070$ ;  $r = 0.49$ ,  $P = 0.047$ , respectively). They also found out that several epitopes which include locations of the two mentioned-mutations are likely to attach to HLA molecules [51].

To clarify the effect of a mutation on virus fitness, Yao et al. infected Vero-E6 cells (high similarity of its ACE2 or virus gate protein to human one) by employing 11 patient-derived SARS-CoV-2 isolates. They showed that mutations in RBD of S protein, as one of the most variable segments of SARS-CoV-2, has influenced the pathogenicity of SARS-CoV-2. According to the mentioned study, S-D614G has a significantly positive effect on viral replication [47]. Following Yao et al. study, another research showed that S protein is highly mutated, and some of its mutations such as D614G can potentially increase virus infection activity and also the severity of disease [48]. One study showed that a 15-30nt deletion in the S1/S2 junction region which happened during passaging virus in Vero-E6 cells, reduced the infection rate of the SARS-CoV-2. S1/S2 junction region is cleaved during virus entry to the cell by TMPRSS2 and therefore, deletion mutations in this region can attenuate the infection rate of the virus. So far, the pre-mentioned deletion in the S1/S2 junction region has not been detected in isolated sequences from COVID-19 patients [49].

Interestingly, it was suggested that mutations of the S1/S2 junction region could potentially increase vaccine development [49]. In another study, it was reported that 8782 C>T in ORF1-ab (synonymous mutation), 28144 T>C in ORF-8 (missense), 29095 C>T in the N gene

(synonymous mutation) are the most common mutations till 5<sup>th</sup> April 2020 and may have some effects on transmission and severity of COVID-19 in most area. Notably, genomic sequences analysis showed that the transition of C to T is considered the most common nucleotide variation, leading to significantly higher frequencies of synonymous mutation and missense mutations rather than deletion and insertion ones [23].

After bio-informatic analysis of 351 genome sequences of the SARS-CoV-2 patients, researchers detected two mutations in the recent isolates, involving nsp6 and adjacent regions of the ORF10 (at AA locations: 3691 and 9659). Analytics applying a Maximum Likelihood tree implies that both of the mentioned-mutations reduce the stability of the structures of the proteins [54].

Causative agents of the virus genome mutations are various, including fidelity of RdRp (as a main replication/transcription enzyme of the virus), host enzymes, effects of physical and chemical mutagens, and genetic recombination [38, 56]. By analyzing genomic sequences from the GISAID database derived from COVID-19 patients, it was reported that viruses with a missense (P to L) mutation in position 14408 of RdRp have a higher number of mutations compared to viruses without the mutation in RdRp [38]. Also, after RdRp 14408nt mutation occurrence (from 16<sup>th</sup> February 2020), the pattern of other variations has been changed, and new versions have been introduced to the virus genome in Europe. However, similar mutation pattern in North America was not observed in the SARS-CoV-2 genome sequences isolated from infected people, which had RdRp 14408nt mutation. As though, the simultaneous occurrence of RdRp 14408nt mutation with nsp3 3036nt, S protein 23403nt, and nucleocapsid phosphoprotein 28881nt in Europe and with nsp13 (helicase) 17746nt and 17857nt, and nsp14 (ExoN) 18060nt in North America had happened [38].

Therefore, the co-appearance of the other mutations with RdRp 14408nt mutation is also geographically distinct. Despite the possible effect of RdRp 14408nt mutation in the occurrence of the aforementioned mutations, other unknown factors have also influenced them. The frequency of RdRp 14408nt in Europe is significantly higher than in North American and Asia. Detected virus genome variation from European and North American patients (mainly Europe) showed a higher number than Asia during the viral spread out of China, and as a result, the infection has a higher mortality rate out of the China [38].

Despite the reported effect of some virus genome mutations on the severity of COVID-19, so far, the researchers found less evidence of the imposed modification by the identified mutations (such as D614G in virus S protein) on the virus's transmissibility [56].



Through *in-silico* methods, researchers identified the most damaging mutations in two of the functional domains of the viral S protein, including RBD (p.408 (Arg to Ile), p.455 (Leu to Tyr), p.486 (Phe to Leu), p.493 (Gln to Asn), p.498 (Gln to Tyr), and p.501 (Asn to Thr)) and heptad repeat 1 (HR1) domain (p.930 (Ala to Val), p.936 (Asp to Tyr)) [53]. Another study investigated the S protein mutations and their biological importance by analyzing 80 SARS-CoV-2 variants and 26 mutant ones. They discovered that variants with p.614 (Asp to Gly) mutation were quite more infectious and most of the variants with a mutation in the RBD domain were less infectious. Sequences containing p.475 (Ala to Val), p.452 (Leu to Arg), p.483 (Val to Ala), and p.490 (Phe to Leu) showed resistance to several neutralizing antibodies [52].

## 7 Conclusion

In addition to various known and unknown factors and host genetic susceptibility, mutations and genetic variabilities of the virus itself have a critical impact on variable clinical features of COVID-19 infection. Overall, at this time, detection of all the appeared mutations in SARS-CoV-2 and their relation with pathological changes is nearly impossible, mainly because of asymptomatic subjects. Notably, the asymptomatic situation in some affected persons with the mentioned disease could be due to attenuated occurred mutations in SARS-CoV-2, which needs to be clarified in future studies. Furthermore, to confirm the results of the findings, many studies need to be repeated in other regions/locations for other people. Such studies could favor patient-specific therapy, by revealing genotyping patterns of SARS-CoV-2 distribution.

## Conflict of Interest

The authors declare no conflicts of interest.

## Funding

The authors received no specific funding for this work.

## References

- [1] Pourghasemi HR, Pouyan S, Heidari B, Farajzadeh Z, Shamsi SRF, Babaei S, et al. Spatial modelling, risk mapping, change detection, and outbreak trend analysis of coronavirus (COVID-19) in Iran (days between 19 February to 14 June 2020). *International Journal of Infectious Diseases*. 2020.
- [2] Singhal T. A review of coronavirus disease-2019 (COVID-19). *The Indian Journal of Pediatrics*. 2020:1-6.

- [3] Devaux CA, Rolain JM, Colson P, Raoult D. New insights on the antiviral effects of chloroquine against coronavirus: what to expect for COVID-19? *International journal of antimicrobial agents*. 2020;105938.
- [4] Hoffmann M, Kleine-Weber H, Schroeder S, Krüger N, Herrler T, Erichsen S, et al. SARS-CoV-2 cell entry depends on ACE2 and TMPRSS2 and is blocked by a clinically proven protease inhibitor. *Cell*. 2020.
- [5] Zhang X, Cai H, Hu J, Lian J, Gu J, Zhang S, et al. Epidemiological, clinical characteristics of cases of SARS-CoV-2 infection with abnormal imaging findings. *International Journal of Infectious Diseases*. 2020.
- [6] Tetro JA. Is COVID-19 receiving ADE from other coronaviruses? . *Microbes and infection*. 2020;22:72-3.
- [7] Wu D, Wu T, Liu Q, Yang Z. The SARS-CoV-2 outbreak: what we know. *International Journal of Infectious Diseases* 2020.
- [8] Qiu H, Wu J, Hong L, Luo Y, Song Q, Chen D. Clinical and epidemiological features of 36 children with coronavirus disease 2019 (COVID-19) in Zhejiang, China: an observational cohort study. *Lancet Infect Dis*. 2020.
- [9] Liu K, Chen Y, Lin R, Han K. Clinical features of COVID-19 in elderly patients: A comparison with young and middle-aged patients. *J Infect*. 2020.
- [10] Bakhshandeh B, Sorboni SG, Javanmard AR, Mottaghi SS, Mehrabi MR, Sorouri F, et al. Variants in ACE2; potential influences on virus infection and COVID-19 severity. *Infect Genet Evol*. 2021:104773.
- [11] Grant WB, Lahore H, McDonnell SL, Baggerly CA, French CB, Aliano JL, et al. Evidence That Vitamin D Supplementation Could Reduce Risk of Influenza and COVID-19 Infections and Deaths. *Nutrients*. 2020;12:988. .
- [12] Guo W, Li M, Dong Y, Zhou H, Zhang Z, Tian C, et al. Diabetes is a risk factor for the progression and prognosis of COVID -19. *Diabetes Metab Res Rev*. 2020:e3319.
- [13] Lai CC, Shih TP, Ko WC, Tang HJ, Hsueh PR. Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) and corona virus disease-2019 (COVID-19): the epidemic and the challenges. *International journal of antimicrobial agents*. 2020:105924.
- [14] Li X, Geng M, Peng Y, Meng L, Lu S. Molecular immune pathogenesis and diagnosis of COVID-19. *Journal of Pharmaceutical Analysis*. 2020.
- [15] Ghandi NA, Allameh SF, Saffarpour R. All about COVID-19 in brief. *New Microbes New Infect*. 2020;35:100678. .
- [16] Yin C. Genotyping coronavirus SARS-CoV-2: methods and implications. *Genomics*. 2020.
- [17] Walls AC, Park YJ, Tortorici MA, Wall A, McGuire AT, Velesler D. Structure, function, and antigenicity of the SARS-CoV-2 spike glycoprotein. *Cell* 2020.
- [18] Machhi J, Herskovitz J, Senan AM, Dutta D, Nath B, Oleynikov MD, et al. The natural history, pathobiology, and clinical manifestations of SARS-CoV-2 infections. *Journal of Neuroimmune Pharmacology*. 2020:1-28.
- [19] Wille M, Holmes EC. Wild birds as reservoirs for diverse and abundant gamma-and deltacoronaviruses. *FEMS microbiology reviews*. 2020;44:631-44.
- [20] Dashraath P, Jeslyn WJL, Karen LMX, Min LL, Sarah L, Biswas A, et al. Coronavirus disease 2019 (COVID-19) pandemic and pregnancy. *American journal of obstetrics and gynecology* 2020.
- [21] Wu A, Peng Y, Huang B, Ding X, Wang X, Niu P, et al. Genome Composition and Divergence of the Novel Coronavirus (2019-nCoV) Originating in China,. *Cell Host & Microbe Commentary* 2020;27:325-8.
- [22] Chang TJ, Yang DM, Wang ML, Liang KH, Tsai PH, Chiou SH, et al. Genomic analysis and comparative multiple sequences of SARS-CoV2. *Journal of the Chinese Medical Association*. 2020;83:537-43.
- [23] Khailany RA, Safdar M, Ozaslan M. Genomic Characterization of a Novel SARS-CoV-2. *Gene Rep*. 2020;19:100682.

- [24] Wang L, Wang Y, Ye D, Liu Q. Review of the 2019 novel coronavirus (SARS-CoV-2) based on current evidence. *Int J Antimicrob Agents*. 2020;55:105948.
- [25] Astuti I, Srafil Y. Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2): An overview of viral structure and host response. *Diabetes Metab Syndr*. 2020;14:407–12. .
- [26] Kaul D. An overview of coronaviruses including the SARS-2 coronavirus e Molecular biology, epidemiology and clinical implications,. *Current Medicine Research and Practice*. 2020;10: 54e64.
- [27] Ratre YK, Kahar N, Bhaskar LVKS, Bhattacharya A, Verma HK. Molecular mechanism, diagnosis, and potential treatment for novel coronavirus (COVID-19): a current literature review and perspective. *3 Biotech* 11. 2021;94.
- [28] Bansal M. Cardiovascular disease and COVID-19. *Diabetes & Metabolic Syndrome: Clinical Research & Reviews*. 2020.
- [29] Raj S, Chandel V, Rathi B, Kumar D. Understanding the Molecular Mechanism(s) of SARS-CoV2 Infection and Propagation in Human to Discover Potential Preventive and Therapeutic Approach. DOI: 1020944/preprints2020040285v1 2020.
- [30] Su S, Wong G, Shi W, Liu J, Lai ACK, Zhou J, et al. Epidemiology, Genetic Recombination, and Pathogenesis of Coronaviruses. *Trends Microbiol*. 2016;24:490-502.
- [31] Domingo E, Holland JJ. RNA virus mutations and fitness for survival. *Annu Rev Microbiol*. 1997;51:151-78.
- [32] May RM, McLean AR, Pattison J, Weiss RA, Holmes EC, Rambaut A. Viral evolution and the emergence of SARS coronavirus. *Philosophical Transactions of the Royal Society of London Series B: Biological Sciences*. 2004;359:1059-65.
- [33] Mishra A, Pandey AK, Gupta P, Pradhan P, Dhamija S, Gomes J, et al. Mutation landscape of SARS-CoV-2 reveals three mutually exclusive clusters of leading and trailing single nucleotide substitutions. *bioRxiv*. 2020:2020.05.07.082768.
- [34] Mercatelli D, Giorgi FM. Geographic and Genomic Distribution of SARS-CoV-2 Mutations. *Frontiers in Microbiology*. 2020;11.
- [35] Pachetti M, Marini B, Benedetti F, Giudici F, Mauro E, Storici P, et al. Emerging SARS-CoV-2 mutation hot spots include a novel RNA-dependent-RNA polymerase variant. *J Transl Med*. 2020;18:179.
- [36] Koyama T, Platt D, Parida L. Variant analysis of COVID-19 genomes. 2020.
- [37] Phan T. Genetic diversity and evolution of SARS-CoV-2. *Infection, Genetics and Evolution*. 2020;81:104260.
- [38] Pachetti M, Marini B, Benedetti F, Giudici F, Mauro E, Storici P, et al. Emerging SARS-CoV-2 mutation hot spots include a novel RNA-dependent-RNA polymerase variant. *J Transl Med*. 2020;18:179-.
- [39] Yu WB, Tang GD, Zhang L, Corlett RT. Decoding the evolution and transmissions of the novel pneumonia coronavirus (SARS-CoV-2 / HCoV-19) using whole genomic data. *Zool Res*. 2020;41:247-57.
- [40] Phan T. Genetic diversity and evolution of SARS-CoV-2. *Infect Genet Evol*. 2020;81:104260.
- [41] Holland LA, Kaelin EA, Maqsood R, Estifanos B, Wu LI, Varsani A, et al. An 81 nucleotide deletion in SARS-1 CoV-2 ORF7a identified from sentinel surveillance in 2 Arizona (Jan-Mar 2020). *J Virol*. 2020.
- [42] Muth D, Corman VM, Roth H, Binger T, Dijkman R, Gottula LT, et al. Attenuation of replication by a 29 nucleotide deletion in SARS-coronavirus acquired during the early stages of human-to-human transmission. *Sci Rep*. 2018;8:15177.
- [43] Su YC, Anderson DE, Young BE, Zhu F, Linster M, Kalimuddin S, et al. Discovery of a 382-nt deletion during the early evolution of SARS-CoV-2. *bioRxiv*. 2020.
- [44] Chatterjee S. An Overview of Mutations Occurring within the Coronavirus-2 Genome: Mutations Data Reporting on SARS-CoV-2. *SSRN Electronic Journal*. 2020.
- [45] Asselta R, Paraboschi EM, Mantovani A, Duga S. ACE2 and TMPRSS2 variants and expression as candidates to sex and country differences in COVID-19 severity in Italy. *medRxiv*. 2020.

- [46] Lu R, Zhao X, Li J, Niu P, Yang B, Wu H, et al. Genomic Characterisation and Epidemiology of 2019 Novel Coronavirus: Implications for Virus Origins and Receptor Binding. *The Lancet*. 2020;395:565-74.
- [47] Yao H, Lu X, Chen Q, Xu K, Chen Y, Cheng L, et al. Patient-derived mutations impact pathogenicity of SARS-CoV-2. *medRxiv*. 2020.
- [48] 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. 2020.
- [49] Lau SY, Wang P, Mok BWY, Zhang AJ, Chu H, Lee ACY, et al. Attenuated SARS-CoV-2 variants with deletions at the S1/S2 junction. *Emerging Microbes & Infections*. 2020;9:837-42.
- [50] Issa E, Merhi G, Panossian B, Salloum T, Tokajian S. SARS-CoV-2 and ORF3a: Nonsynonymous Mutations, Functional Domains, and Viral Pathogenesis. *mSystems*. 2020;5.
- [51] Toyoshima Y, Nemoto K, Matsumoto S, Nakamura Y, Kiyotani K. SARS-CoV-2 genomic variations associated with mortality rate of COVID-19. *Journal of Human Genetics*. 2020.
- [52] Li Q, Wu J, Nie J, Zhang L, Hao H, Liu S, et al. The Impact of Mutations in SARS-CoV-2 Spike on Viral Infectivity and Antigenicity. *Cell*. 2020;182:1284-94.e9.
- [53] Ahamad S, Kanipakam H, Gupta D. Insights into the structural and dynamical changes of spike glycoprotein mutations associated with SARS-CoV-2 host receptor binding. *J Biomol Struct Dyn*. 2020:1-13.
- [54] Benvenuto D, Angeletti S, Giovanetti M, Bianchi M, Pascarella S, Cauda R, et al. Evolutionary analysis of SARS-CoV-2: how mutation of Non-Structural Protein 6 (NSP6) could affect viral autophagy. *Journal of Infection*. 2020;81:e24-e7.
- [55] Yao H, Lu X, Chen Q, Xu K, Chen Y, Cheng L, et al. Patient-derived mutations impact pathogenicity of SARS-CoV-2. *medRxiv*. 2020:2020.04.14.20060160.
- [56] van Drop L, Richard D, Shaw LP, Acman M, Balloux F. No evidence for increased transmissibility from recurrent mutations in SARS-CoV-2,. *bioRxiv*. 2020.

**Figure Captions**

**Figure 1:** Possible influencing factors on variable symptoms and severity of COVID-19.

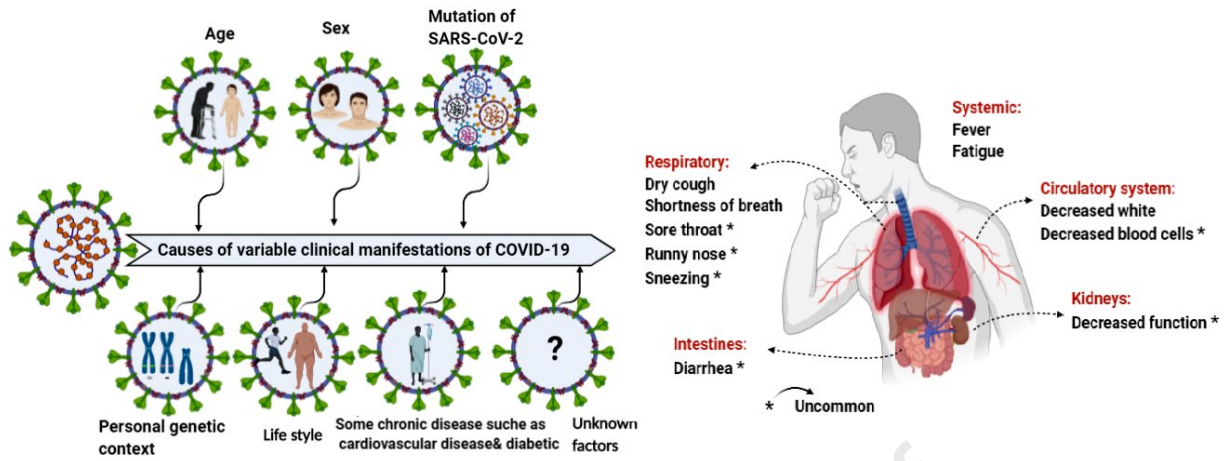
**Figure 2:** Classification and outbreaks of coronaviruses in human society.

**Figure 3:** The genomic organization of SARS-CoV-2. ORF1ab, as the largest gene, encodes the pp1ab protein that contains 15 Nsps (nsp1-nsp10 and nsp12-nsp16). It is located at the 5' end of the genome. The 3' terminus of the SARS-CoV-2 genome contains four structural genes, which encode four essential structural proteins: envelope (E), spike protein (S), nucleocapsid protein (N), membrane (M) and a gene for RNA-dependent RNA polymerase.

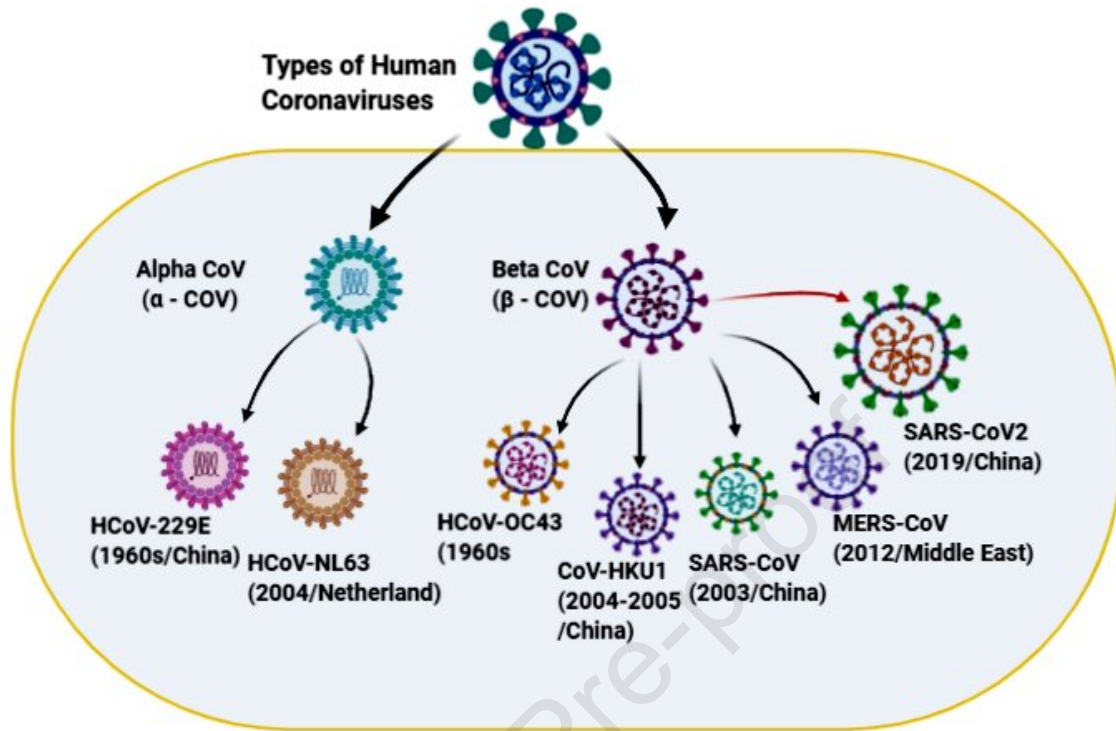
**Figure 4:** Schematic structure of SARS-CoV-2 with the minimal set of structural proteins.

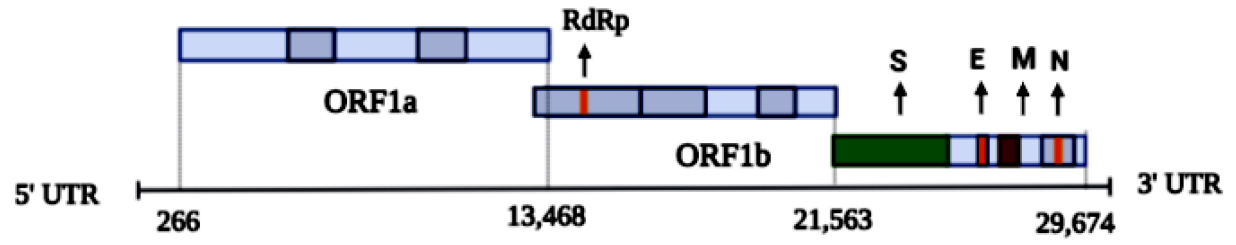
**Figure 5:** The mechanism of coronavirus infection and spread. Coronaviruses bind to ACE-2 receptors on the surface of the target cell via their spike proteins. Next, the ACE2 receptor will be cleaved by the TMPRSS2 that results in the activation of the spike protein through which the virus enters the cell.

**Figure 6:** Linear positions of the reported mutations in various regions of the SARS-CoV-2 genome.



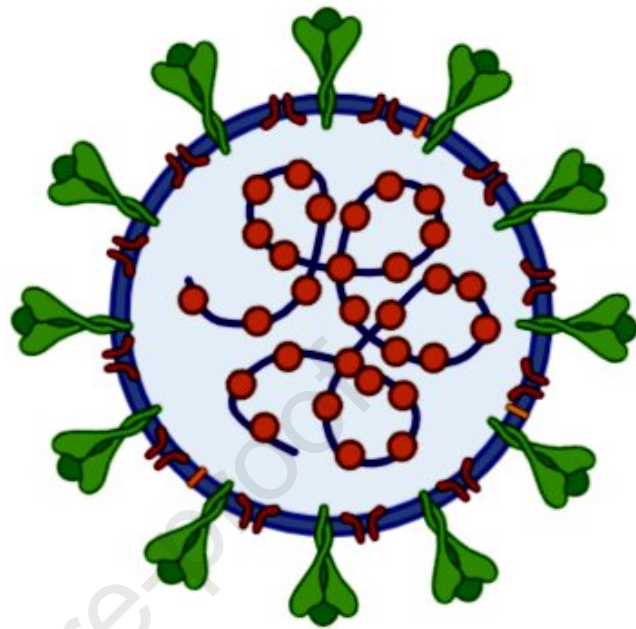
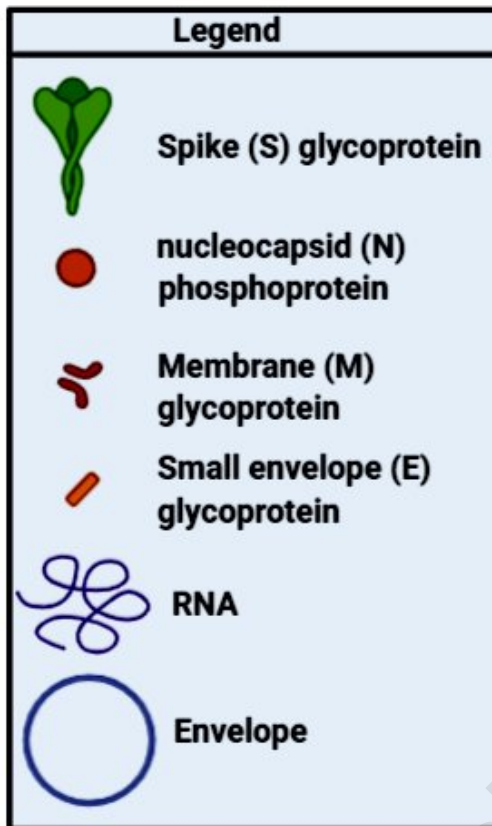
Journal Pre-proof

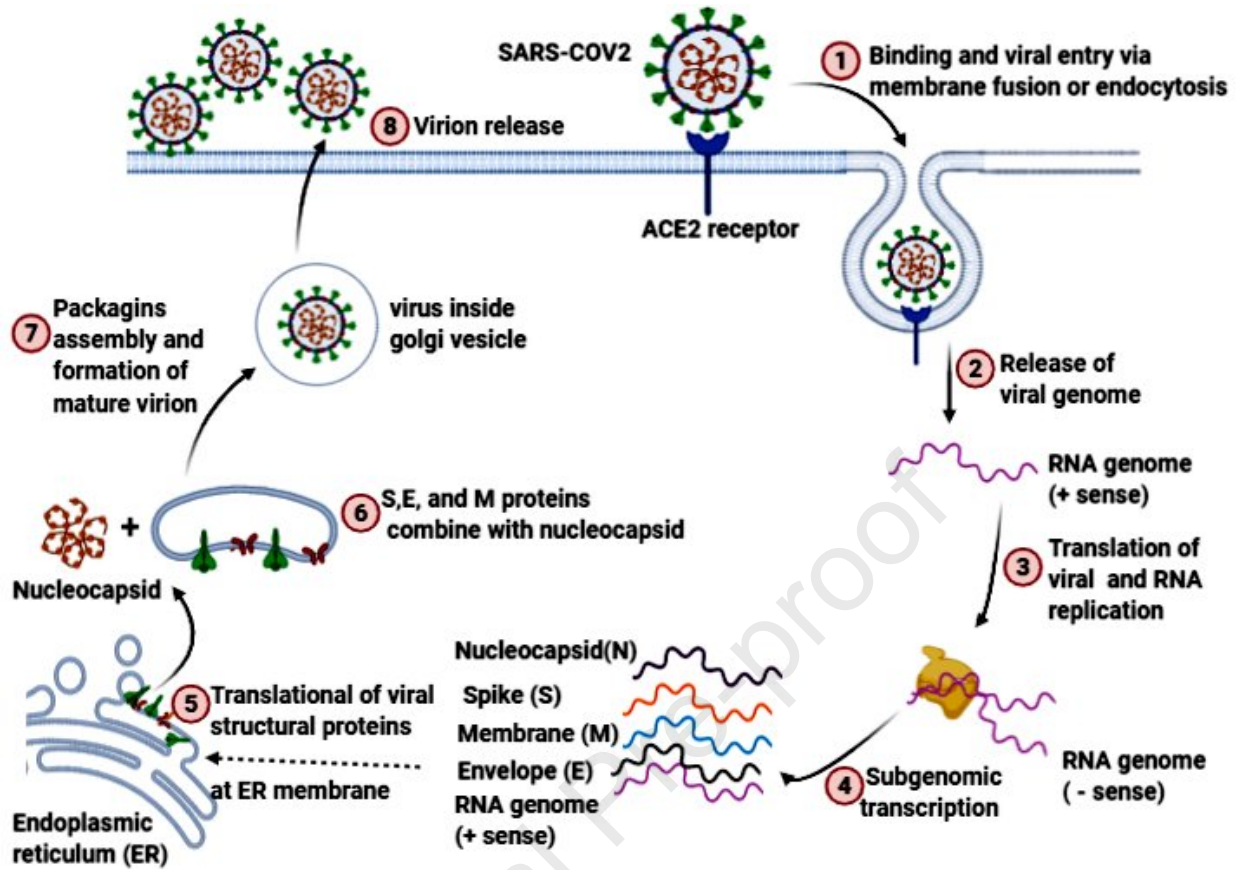


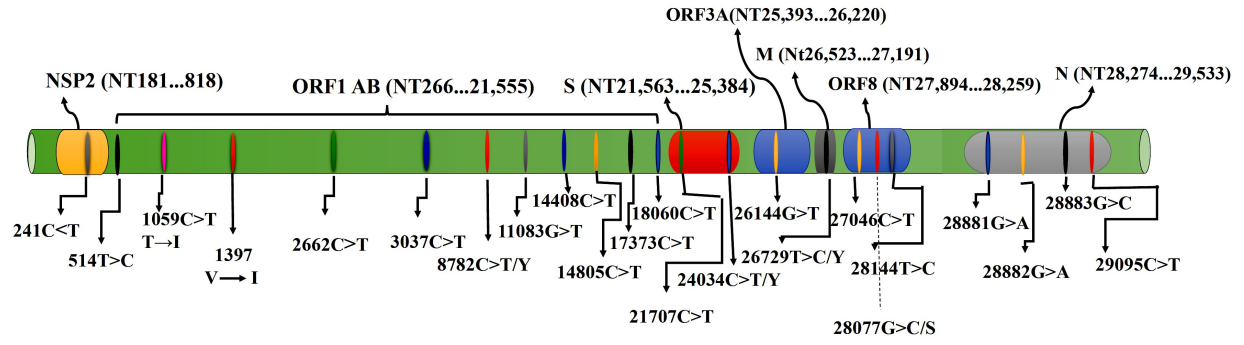


Journal Pre-proof









**Highlights:**

1. The severity rate of COVID-19 is influenced by virus mutations and genetic variabilities.
2. SARS-CoV-2 genome presents high dynamic mutation rate in comparison to other RNA viruses.
3. Because of genotype-phenotype correlation, virus mutations result in various pathogenicity and infectivity.
4. Effective treatment of COVID-19 disease is based on the detection of the SARS-CoV-2 mutation patterns.

**Declaration of interests**

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

Journal Pre-proof