Analysis of Whole-Genome Sequencing of SARS-CoV-2 Reveals Recurrent Mutations among Iranian Patients

Mohammad Hadi Abbasian [1, +], Bahar Mahdavi[2, +], Mohammadamin Mahmanzar[3, +], Samaneh Tokhanbigli [4], Mahsa Mollapour Sisakht[5], Bahman Moradi[6], Karim Rahimian [7, *]

1. Department of Medical Genetics, National Institute for Genetic Engineering and Biotechnology, Tehran, Iran.
2. Department of Molecular Biotechnology, Cell Science Research Center, Royan Institute for Biotechnology, ACECR, Isfahan, Iran
3. Department of Bioinformatics, Kish International Campus University of Tehran, Kish, Iran
4. Department of Molecular and Cellular Sciences, Faculty of Advanced Sciences and Technology pharmaceutical Sciences Branch, Islamic Azad University, Tehran, Iran (IAUPS)
5. Department of Biochemistry, Erasmus University Medical Center, P.O. Box 2040, 3000 CA Rotterdam, The Netherlands
6. Department of Biology, faculty of sciences, Shahid Bahonar University of Kerman, Iran
7. Bioinformatics and Computational Omics Lab (BioCOOL), Department of Biophysics, Faculty of Biological Sciences, Tarbiat Modares University, Tehran, Iran.

+ These authors have the same contribution

***Corresponding Author:**

Karim Rahimian, Bioinformatics and Computational Omics Lab (BioCOOL), Department of Biophysics, Faculty of Biological Sciences, Tarbiat Modares University, Tehran, Iran. karim.rahimian@modares.ac.ir

Running title: Recurrent SARS-CoV-2 mutations among Iranian patients

Abstract

Background

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is a new emerging coronavirus that causes coronavirus disease 2019 (COVID-19). Whole-genome tracking of the SARS-CoV-2 enhanced our understanding of the mechanism of disease, control, and prevent COVID-19 infections.

Materials and Methods

In the current study, we investigated 1221 SARS-CoV-2 protein sequences of Iranian SARS-CoV-2 in the public database of the GISAID from January 2019 to April 2022. Prepare a list of suitable samples and preprocess performed by python programming language. To compare and identify mutation patterns SARS-CoV-2 genome was aligned to the Wuhan-Hu-1 as a reference sequence.

Results

Our investigation revealed that spike-P323L, ORF9c-G50N, NSP14-I42V, spike-D614G, NSP4-T492I, nucleocapsid-R203K, nucleocapsid-G204R, membrane-A63T, membrane-Q19E, NSP5-P132H, envelope-T9I, NSP3-G489S, ORF3a-T24I, membrane-D3G, spike-S477N, Spike-D478K, nucleocapsid-S235F, spike-N501Y, nucleocapsid-D3L, and spike-P861H as the most frequent mutations among the Iranian SARS-COV-2 sequences. Furthermore, it was observed that more than 95 % of the SARS-CoV-2 genome, including NSP7, NSP8, NSP9, NSP10, NSP11, and ORF8, had no mutation when compared to the Wuhan-Hu-1. Finally, our data indicated the ORF3a-T24I, NSP3-G489S, NSP5-P132H, NSP14-I42V, envelope-T9I, nucleocapsid-D3L, membrane-Q19E, and membrane-A63T mutations might be one of the responsible factors for the surge in the SARS-CoV-2 omicron variant wave in Iran.

Discussion

Our results highlight the value of real-time genomic surveillance that help to identify novel SARS-CoV-2 variants and could be applied to update SARS-CoV-2 diagnostic tools, vaccine design, and understanding of the mechanisms of adaptation to a new host environment.

Introduction

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is an RNA virus which was first identified in Wuhan, China [1]. SARS-CoV-2 disseminates rapidly in the world, and The World Health Organization (WHO) officially announced coronavirus disease 2019 (COVID-19) a global pandemic in March 2020 [2]. This had spread worldwide and with 527,799,879 confirmed cases, including 6,300,437 deaths, as of 23 May 2022. The first COVID-19 case in Iran was confirmed in Qom on 19 February 2020 [3].

Previous studies have shown that epidemiological data combined with whole-genome sequence analysis is pivotal for public health decision-making [4]. The first sequence of SARS-CoV-2 was deposited in January 2020 [5]. The global initiative on sharing all influenza data (GSAID) (https://www.gisaid.org) is a global science initiative that facilitates analysis sequences of SARS-CoV-2 and monitors emerging new variants [6-8]. The SARS-CoV-2 carries one of the largest RNA genomes that varies from 29.8 kb to 29.9 kb, packed into small nucleocapsid (N) protein. (Figure1) The SARS-CoV-2 genome is arranged from 5′ to 3′ untranslated regions (UTRs) as non-structural genes (ORF1a/ORF1b replicase gene) and structural genes such as spike (S), envelope (E), membrane (M), and nucleocapsid, and accessory genes (ORF3a, ORF6, ORF7a, ORF7b, ORF8, ORF9b, and ORF9c) [9]. The S glycoprotein helps in the attachment of SARS-CoV-2 to the ACE-2 receptors of the host cells [10]. M protein helps to enclose mature virus particles in a membrane and the virion particles are assembled by the E protein. The SARS-CoV-2 genome is composed of 16 non-structural proteins (NSPs). The NSPs comprises various viral cysteine proteases such as main proteinase (NSP5), putative transmembrane domain (NSP6), RNA-dependent RNA polymerase (NSP12), helicase (NSP13), and 2'-O-methyltransferase (NSP16) that play a critical role in viral RNA transcription and replication [11].

It was evident that the pandemic could only be controlled with efficacious vaccines [12]. Vaccine development is the first and most favorable response to control the devastating impact of the COVID-19 pandemic. Thus, several new vaccines with different platforms such as mRNA, vector-based, inactivated, and nanoparticle-based vaccines are developed and approved in different countries to generate neutralizing antibodies upon immunization [13-15]. However, decreasing immunity and vaccine effectiveness against emerging variants, mainly variants of

concern (VOCs), is a significant challenge for the scientific world and the governments [16]. Vaccines are designed to induce immune responses against SARS-CoV-2 S glycoprotein [17]. Accumulating evidence shows that VOCs contain mutations in the SARS-CoV2 genome. For instance, the D614G, L452R, P681R, and E484K in spike mutations in VOC are potentially affected on transmissibility and virulence of SARS-CoV2 and reduce vaccine efficacy [18-20]. New studies have been found that delta and omicron variants reduce protection against COVID-19 [21, 22]. Therefore, continuous monitoring of SARS-CoV2 mutations is critical for controlling a better COVID-19 pandemic.

There are several SARS-CoV-2 genome sequence studies in Iran that efforts to elaborate on origins and transmission dynamics, as well as the effect on evolution pattern and disease spread. Eslami et al. analyzed 176 S glycoprotein sequences of Iranian SARS-COV-2 from April 2020 to May 2021. They found unique mutation E1202Q in HR2 subdomain that facilitates the virus membrane fusion process [23]. Fattahi et al. studied 50 whole-genome sequences of SARS-COV-2 from Iran from March to July 2020. The analysis of transmission dynamics proposed two introductions of the SARS-COV-2 into Iran. This study also showed that S mutation D614G increased infectivity and transmission of SARS-COV-2 [24]. In the study of Moradi et al., twenty-four SARS-COV-2 sequences between March to September 2020 were analyzed, and ORF1ab, S, N, intergenic, and ORF7 regions were the most common mutations of this virus [25]. The present study provides the most comprehensive description of the SARS-CoV-2 epidemic in Iran based on a genomic epidemiology approach. This study analyzed more than 1221 SARS-CoV-2 protein sequences of Iranian SARS-CoV-2 from January 2019 to April 2022 in Iran.

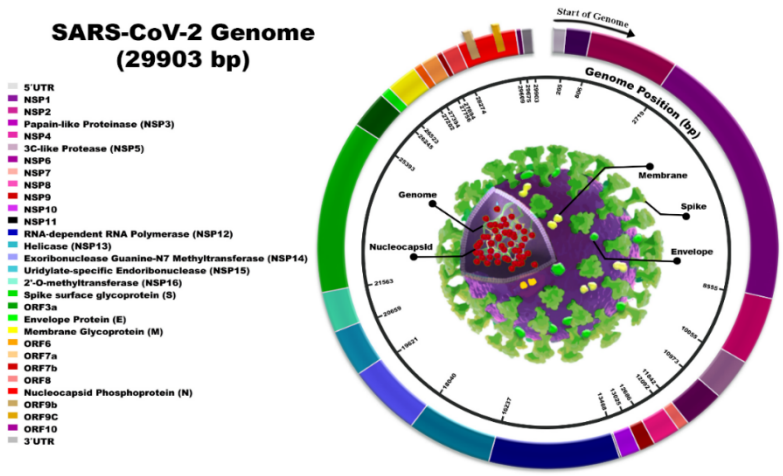

Figure 1. Schematic representation of 29,903 nucleotide base pair of SARS-CoV-2 single stranded RNA genome and its encoded proteins.

Materials and Methods

Sequence source

This study evaluated whole data regarding whole genome sequence of SARS-CoV-2 amino-acid (AA) sequences (AASs). All AASs were compared to the Wuhan-2019 virus 'EPI_ISL_402124' as the reference sequence. AASs of Genome sequences of SARS-CoV-2 samples were retrieved from the GISAID database (https://www.gisaid.org/) for different geographic locations in Iran [6, 8]. We have access to this database through Erasmus Medical Center's permission. The study design and flowchart of methods are summarized in Figure 2.

Sequence analyses and exclusion criteria

Python 3.8.0 software was utilized to preprocess FASTA files. Each difference between genome sequences of SARS-CoV-2 samples and reference was considered a mutation, and the location and substituted AA were reported. Non-human samples and those with less or more than length of SARS-CoV-2 genes and samples containing non-specified AAs (reported as X) were omitted.

The whole process was optimized by applying 'Numpy' and 'Pandas' libraries, as previously described [26, 27]. Briefly, for detection of mutations in reference and samples sequence, we used 'Refseq, and 'seq', respectively. For refitem, seqitem in zip (refseq, seq) If (refitem! =seqitem) report a new mutant. After extracting genome sequences of SARS-CoV-2, each sample's continent name and geographical coordinates were obtained and reported using pycountry-convert 0.5.8 software and 'Titlecase' library in Python to draw global prevalence maps of mutations. Figures were drawn using VMD 1.9.3 versions and GraphPad Prism, version 8.0.2.

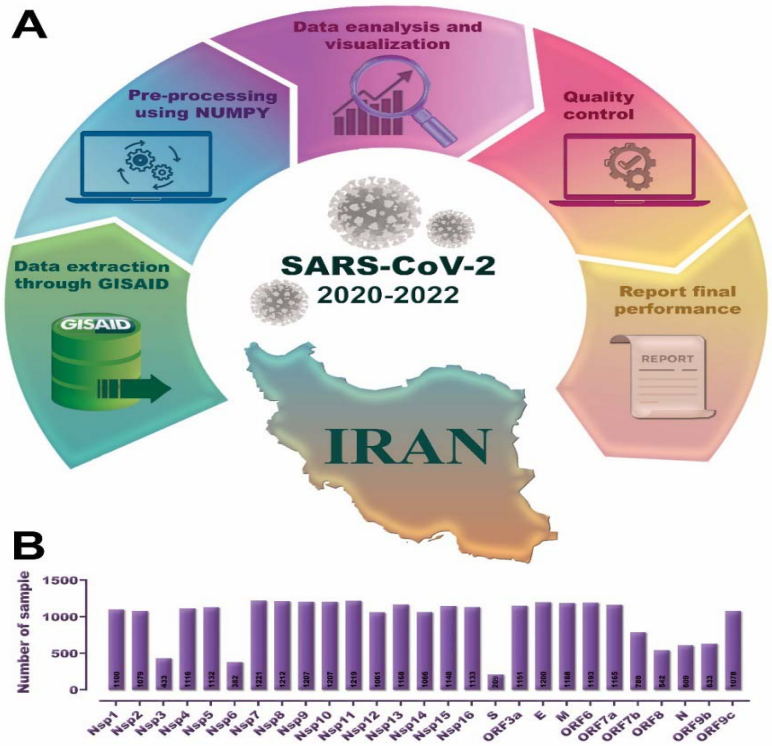

Figure 2. Study design. **(A)**. Flowchart of methods involved in this study. **(B)** Number of samples in each region

Results and Discussion

In the present study, we have compared the 1221 SARS-CoV-2 protein sequences from Iranian samples and approximately 10 million SARS-CoV-2 virus genomes circulating worldwide from December 2019 to April 2022. Data exported from the GISAID database and compared with the reference sequence Wuhan-Hu-1 (Accession NC_045512).

We reported mutation frequency in different regions of SARS-CoV-2 genome (Figure 3A). Among the common variants, we observed that 20 mutations represent >50% of the sequences including S-P323L (91.5%), ORF9c-G50N (82.1%), NSP14-I42V (61.2%), S-D614G (61.2%), NSP4-T492I (60.0%), N-R203K (57.3%), N-G204R (57.1%), M-A63T (56.9%), M-Q19E (54.9%), NSP5-P132H (54.3%), E-T9I (53.6%), NSP3-G489S (39.2%), ORF3a-T24I (39.2%), membrane-D3G (37.5%), S-S477N (28.2%), S-D478K (24.8%), N-S235F (22.4%), S-N501Y (21.5%), N-D3L (21.1%), and S-P861H (21.0%) (Figure 3B).

Previous genomics analysis in Iran reported six mutations that have a frequency of more than 50% of the sequences, including D614G, P323L, R203K, G204R, F105F, and C241T [28]. In

the United States, top mutations were reported as F106F, S76S, L7L, T85I, P323L, D614G, Q57H, L84S, Y541C, P504L, S24L, R203K, and G204R [29].

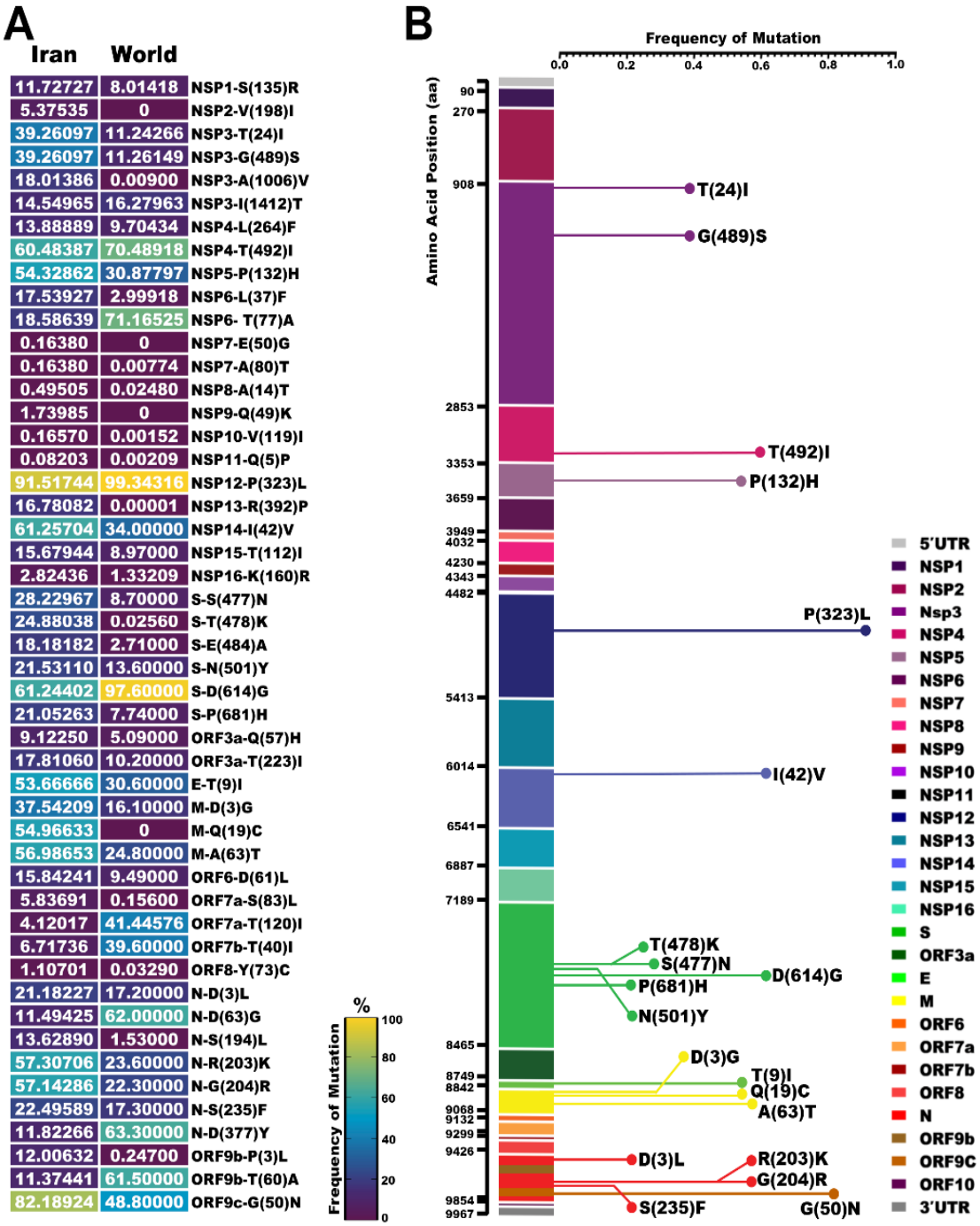

Figure 3. Graphical representation of SARS-CoV-2 mutation frequency. (**A**). Comparison of the highest percentage of mutation frequency in Iran and the world (**B**). Top 20 mutations with a frequency >20% in Iran in different regions of SARS-CoV-2 genome.

ORF1ab polyprotein

The ORF1a polyprotein of SARS-CoV-2 encompasses 11 NSP proteins, NSP1, NSP2, NSP3, NSP4, 3CL-like protease, NSP6, NSP7, NSP8, NSP9, NSP10, and NSP11.

SARS-CoV-2 NSP1, a virulence factor limiting protein synthesis by attaching directly to the human ribosome, suppresses the translation of host messenger RNAs (mRNAs). In our study, analyses suggest that the SARS-CoV-2 is highly conserved in the NSP1 and common mutations in NSP1 were S135R (11.7%) and R24C (0.5%).. The R24C (1.3%) was the most common mutation in the NSP1 of SARS-CoV-2 in the US and worldwide which has been suggested that destabilizing effect on the protein [30].

NSP2 is involved in coronavirus genome replication, SARS-CoV-2 infection and propose as a drug target for anti-SARS-CoV-2 drug development [31]. The structure and function of NSP2 in the SARS-CoV-2 have remained unknown [32]. Our study showed that the top mutations detected in NSP2 were V198I (5.3%), Y49C (2.3%), and R27C (1.6%). In the study of Cornillez-Ty et al. on the SARS-CoV, the deletion of NSP2 results in a modest reduction in viral titers [33]. Previous studies have been reported that SARS-CoV NSP2 interacts with prohibitin 1 (PHB1) and PHB2, which play a role in several cellular functions, including cell cycle, mitochondrial functions, and apoptosis [33, 34]. In addition, R27C was detected in 837 Indian SARS-CoV-2 strains[35]. Wang et al. reported that T85I may destabilize the structure of NSP2 but does not change the flexibility of NSP2 too much [29]. V198I was detected in India and was present in America, Europe, and Asia [35]. Gupta et al. studied a structure of SARS-CoV-2 NSP2 with cryo-EM experimental data and AlphaFold2 prediction. They suggested NSP2 interactions with endosomes, modulators of translation, and ribosomal RNA [36].

NSP3, also known as papain-like protease, the largest NSPs, plays a critical role in viral replication and function as a protease. In our study, just 11.3% of NSP3 was conserved, and the top mutations were G489S (3.9%), T24I (3.9%), A1006I (1.8%), and I1412 (1.4%). Vilar et al. analyzed 290,000 worldwide SARS-CoV-2 sequences from December 2019 to 30 December 2020. In this study, the mutation rate for T183, I1412, and A890 was 0.075 [37]. I1412T and

T183I were detected in 420 SARS-CoV-2 variants from global areas in January–March 2021[38, 39]. Furthermore, S126L and T33I were observed in SARS-CoV-2 variants from global areas and in Serbia, respectively [38, 40].

NSP3 with NSP4 rearrange host-derived membranes and plays a role in the replication of SARS-CoV [41]. A total of 32.8% of NSP4 was conserved and presented 40 mutations, including T492I (6%), L264F (1.3%), and T327I (1.2%) that were reported in previous studies [35, 42, 43].

NSP5 encodes chymotrypsin-like main protease and is essential for viral assembly and maturation of SARS-CoV-2. P132H (5.4%), K90R (1.1%), and G283S (0.7%) are more frequently mutations that found in our samples. K90R and L205V mutations in NSP5 were associated with positive selection in Brazil [44]. K90R was a frequent mutation in 675 sequences in India from March 2020 to April 2021 and the top mutation in the world [45]. G15S and K90R are also found in European, Chinese, and Icelandic strains [46]. The study of Fung et al. revealed that G15S and K90R suppressed IFN-β production [47]. SARS-CoV-2 NSP5 can induce the expression of cytokines IL-1β, IL-6, TNF-α, and IL-2 in Calu-3 and THP1 cells. Furthermore, they showed that NSP5 could activate the NF-κB signaling pathway and enhances cytokine expression [48].

NSP6 is a membrane protein of approximately 34 kDa and can be found in many coronaviruses. NSP6, NSP3 and NSP4 form a protein complex in the endoplasmic reticulum to create double-membrane vesicles, which are required to develop the virus replication/transcription complex [49]. In our study, T77A (1.8%), L37F (1.7%), and Y80C (0.4%) were common mutations in NSP6. The epidemiological analysis reported that the L37F mutation was associated with asymptomatic SARS-CoV-2 infection[50]. The transcriptome analysis of Sun et al. demonstrated that the L37F mutant failed to trigger pyroptosis and the downstream inflammasome pathway [51]. In our recent study, we showed that the T77A mutation is one of the top mutations globally, with a frequency of 69.8% [43].

The coronavirus RdRP machinery, which mediates viral replication, comprises of three proteins: NSP7, NSP8, and NSP12. The most common mutations in the NSP7 were A80T (0.16%) and E50G (0.16%).

NSP8 presents different point mutations, including A14T (1.3%) and R51C (0.66%). A14 to S and V were reported in 9692 SARS-CoV-2 [52]. R51C was also reported in Brazil [44].

In the RNA-synthesizing machinery, NSP9 is an essential RNA binding subunit. This protein form part of the replication and transcription complex (RTC) mediates all virus replication, overall pathogenicity, and viral genomic RNA reproduction. Q49K (1.7%) mutation was the most common mutation in our samples.

NSP10 has 139 residues, two zinc fingers, by interaction with NSP14 and NSP16 activates numerous replication enzymes [53, 54]. The result of our study showed that the most common mutation in this region was V119I (0.16%). Rogstam et al. showed that the crystal structure of the unbound form of SARS-CoV-2 NSP10 is similar to SARS-CoV NSP10 [55]. Furthermore, a recent study suggests that NSP10 has a role in SARS-CoV-2 replication machinery and that the nsp10–nsp14 complex has a relatively weak affinity [56].

The NSP11 comprises 13–23 residues in different CoVs species. NSP11 was the most conserved region in Iranian SARS-CoV-2 (99.9 %) and Q5H (0.08%) was the only mutation in this region. Consistent with our results, NSP11 presents no mutations compared with the Wuhan-2019 virus 'EPI_ISL_402124' in Brazil [44].

The NSP12, also known as the RNA-dependent RNA polymerase (RdRp), which is directly responsible for RNA synthesis, is a conserved region of SARS-CoV-2. RdRp is one of the pivotal druggable targets for coronaviruses [57]. In our study, the high-frequency mutations detected on the NSP12 were P323L (91.5%), G137C (4.5%), and G671S (7.6%). The P323L mutation for the first time was reported in Spain on 25 January 2020 The P323L mutation was one of the dominant mutations in the United States [29]. In this mutation, polar and has uncharged proline changes to leucine, which may not effect on NSP12 function. However, Wang et al. suggested that this mutation might enhance the transmission capacity of SARS-CoV-2 [29]. P323L was also a common mutation in 420 SARS-CoV-2 variants from global areas in January–March 2021 [38]

NSP13 of SARS-CoV-2 is a superfamily 1 helicase that unwinds a double-stranded RNA or DNA, and with NSP1 block interferon activation [58]. NSP13 inhibits type I interferons (IFN-I) production and enables SARS-CoV-2 to evade the host innate immune response [59]. Our

analysis showed that the most commonly occurring NSP13 mutations were R392 (1.6%) and P77L (0.47%). P77L was detected in the Delta variant and is one of the most frequently observed mutations in the world. This mutation increases the function of the NSP13 protein by binding to TBK1 and might suppress IFNß production, thereby making the SARS-CoV-2 more pathogenic [60]. Furthermore, P77L increases the function of the NSP13 protein [60].

NSP14 plays a critical role in viral RNA 5′ capping and has proofreading capability [61, 62]. Recent studies report that overexpression of NSP14 induces a near-complete shutdown in cellular protein synthesis [63]. In our study, NSP14 was 30 % conserved, and the top mutation in this region was I42V (61.2%) and A394V (4.8%). A394V was a positive selection site in the SARS-CoV-2 B.1.617 and is a common mutation globally [64].

NSP15 is conserved throughout the coronaviridae family, and its endonuclease function is essential for viral survival. According to recent findings, NSP15 nuclease activity is critical for evading host immune responses and could be a suitable target for an antiviral small molecule inhibitor [65, 66]. NSP15 was 77.3 % conserved and T112I (15.6%), K64R (0.6%), and H234Y (0.6 %) mutations were observed in our samples. H234Y was also observed in SARS-CoV-2 circulating during the third wave of the pandemic in Pakistan [67].

The transfer of a methyl group from S-adenosyl methionine (SAM) to RNA substrates is catalyzed by NSP16 [68]. Previous studies indicate that the NSP10-NSP16 complex is critical for CoVs' viral replication [54, 69]. In our study, NSP16 was 93.2% conserved, and the top mutations were K160R (2.8 %) and G77R (0.7 %).

Spike glycoprotein

The spike glycoprotein of SARS-CoV-2 is a homotrimeric class I fusion protein responsible for viral entry and binds to ACE2 on host cells. This glycoprotein is a critical target for both antibody-based therapeutics and vaccines development. Therefore, B.1.1.7, B.1.351, P.1, and three B.1.617 sublineages have exhibited mutations inside and near the ACE2-interacting surface of S protein. Our analysis identified that just 30.1 percent of the SARS-CoV-2 genome had not any mutant. More importantly, 43 percent of the SARS-CoV-2 genome had four or more mutations. (Figure 2 and Supplementary 1). Common mutations were observed including D614G

(61.2.3%), S477N (28.2%), T478K (24.8%), N501Y (21.5%), P681H (21.0%) and E484 (18.1%).

Within the SARS-CoV-2 spike protein, the mutation with glycine at the residue 614 (G614) is a highly variable site and previous studies indicated that D614G increases the infectivity of the COVID-19 virus. In the D614G mutation, the polar negative charged side amino acid aspartate (D) changes to amino acid glycine (G), a non-polar side chain. D614G and D614N were detected in 49 and 0.6 percent of our samples, respectively. The dates that were D614G first seen in China and Germany was on 24 January 2020 and was rare before March 2020 but increased steadily over time and now become one of the most prevalent mutations among all the continents [70, 71]. The first case with the D614G mutation in Iran was reported on 1 May 2020 and become major spike mutation after 1 September 2020 [28]. D614G is the second-highest frequency in the United States and has been associated with increased viral replication in primary human upper airway tissues [72]. It was well established that D614G mutations increases in fitness, infectivity, and fatality [73, 74]. Within the RBD, the S477N is the frequent mutation that has a pivotal role in the binding of the SARS-COV-2 spike with the hACE2 receptor [75]. D138Y, N501Y, and E484K were observed 14, 7, and 3 percent of mutations in our samples, respectively. These mutations were reported in the lineage P.1 and the B.1.1.28 variant [76]. N501Y and E484K occur at the receptor-binding motif (RBM) and increase binding affinity to hACE2 [77]. N501Y has one of the most frequent mutations in RDB and can influence the efficiency of vaccines and drugs target. Also, N501Y resistance to neutralization strengthens binding with hACE2, so increases virus infectivity [78, 79]. The E484K mutation occurs in different variants such as Delta sublineage B.1.617.2 and B.1.351 and has been suggested that reduces antibody neutralization [80]. P863H (6.9%) and P681H (3.7%) also were detected in our study. Previous reports have been suggested that these mutations decreased protein stability and improved viral fitness [81, 82].

ORF3a protein

The ORF3a protein composes of 275 amino acids that cause cytokine production and tissue inflammation. Our analysis revealed top mutations T223I (17.8%), Q57H (9.1%), and S26L (7.2%), on the ORF3a. The Q57H mutation was reported in Brazil and the United States. The amino acid glutamine (Q) changes with an amino acid histidine (H) in this mutation. Wang et al.

suggested that the Q57H mutation makes the SARS-CoV-2 more infectious in the United States [83]. S26L is one of the frequent mutations globally and is detected in 15,928 worldwide SARS-CoV-2 sequences [84].

Envelope protein

The E protein is a tiny 76-109 amino acid protein. Moreover, it works as an ion-channeling viroporin, facilitating viral release by damaging host membranes. This protein participates in viral assembly and budding by interacting with other CoV and host proteins. The E protein is largely conserved across coronaviruses. In our analysis, 45.5% of the SARS-CoV-2 genome had no mutant. Common mutations in the E protein were T9I (53.6%), P71L(0.25%), and S68F (0.16%).(Figure 2 and supplementary 2).

The frequency of P71L was 0.25% in our sample and previously was identified in the samples collected in the USA and Brazil populations. P71L mutations are located in the CT domain and a hydrophobic proline amino acid to another hydrophobic amino acid leucine that may not effect on protein structure [44]. T9I was one of the top mutations worldwide. In this mutation, hydrophilic amino acid changes to hydrophobic, thereby could positively modify the ability of membrane attachment and ER targeting by the E protein [44] and destabilize E protein structure [85]. S68F was the most commonly circulating variant in E in 295,000 SARS-CoV-2 genomes from December 2019 to December 2020. In the study of Mukhtar et al. T9I, P71L, and S68F shows a stabilizing effect [86, 87].

Membrane protein

The M protein is a viral assembly protein with 221 amino acids in three transmembrane helices and a cytoplasmic C-terminal domain. According to our results Q19E (54.9%), D3G (37.5%), I73M (15.5%) and A63T (6.9%), were the most common mutations in our sample. (Figure 2 and supplementary 3).The I73M is the characteristic mutation of B.1.1.413 lineage [28] and the I82T highly prevalent mutation in 176 countries/territories, and one of the top mutations worldwide [88].

ORF6 protein

ORF6 is an ER/Golgi membrane protein that prevents the nuclear import complex from forming. ORF6 is a small protein with 61 residues and a molecular weight of about 7 kDa. In a dose-

dependent way, Selinexor, a specific inhibitor of nuclear export, reduced SARS-CoV-2 ORF6-induced cellular toxicity [89]. In our sample D61L (15.8%) mutations were common mutations in this region.

ORF7a protein

The accessory ORF7a is composed of 121 amino acids and facilitates infection and pathogenesis. The rennet study indicated that ORF7 plays a role as an immunomodulating factor for human CD14+ monocytes [90]. While another study reported that ORF7a of SARS-CoV inhibits cellular translation [91]. According to our results, 22 mutations, including S83L (5.8%), T120I (4.1%), V82A (3.6%), and T39I (1.9%) were observed in Iran samples. S83L was also detected in North America and Oceania [92]. S83L also was observed SARS-CoV-2 mutations in clinical samples [93]. T39I is a frequently occurring mutation in the world [94]. Two common mutations V82A and T120I in ORF7a were also found in Pakistan and the USA, respectively [67, 95].

ORF7b protein

The function of SARS-CoV-2 ORF7b is not well investigated in SARS-CoV-2. ORF7b is located in the membrane of the endoplasmic reticulum ER of SARS-CoV-2, contains a transmembrane region, and functions in innate and adaptive immunity [96, 97]. Recent studies have shown that ORF7b may accelerate TNF-induced apoptosis in HEK293T cells and Vero E6 cells [98]. In our study, the most common mutation in ORF7b was T40I (6.7%) which was also observed in SARS-CoV-2 sequences from 45 countries [99].

ORF8 protein

ORF8 is a 121 amino-acid protein that functions in viral pathogenicity and replication in the host via the interferon pathway. Top mutations in ORF8 were Y73C (1.1%) and I121L (0.7%). In addition to the United States, V62L has been identified in 11 other nations. In the United States, two high-frequency mutations, L84S and S24L, were discovered [83]. I121L was observed in asymptomatic Indian individuals [100].

Nucleocapsid protein

The N protein of SARS-CoV-2 has 419 amino acids, involves different cellular processes, and plays a key role in the viral life cycle and host infection. N protein is involved in the packing of

RNA, the release of virus particles, and the formation of the ribonucleoprotein core. Among the nucleocapsid mutations R203K (57.3%), G204R (57.1%), S235F (22.4%), D3L (22.1 %), and S194L (13.6%) were the most common in our samples (Figure 2 and supplementary 4). Among these, S197L, R203K, and G204R were observed worldwide. In the mutation R203K, both Arginine (R) and Lysine (K) are positively charged, so this mutation may not effect on N protein structure and function. However, Glycine (G) is a nonpolar residue and its change to R positively charged Arginine (R) may destabilize the N protein structure [101]. G204R/R203K are co-occurring mutations detected in lineages B.1.1.7 (Alpha) and P.1 (Gamma) [102, 103]. R203K is a top mutation in the US and the world that increase the infectivity, fitness, and virulence of SARS-CoV-2. Through experimental investigations, Wu et al. realized that 203K/204R variants increase the infectivity and fitness of SARS-CoV-2 [104].

The S194L mutation has a role in the binding affinity of intraviral protein interactions and is associated with symptomatic patients [105, 106]. The S235F mutation is localized in the linker regions and highly stabilizing mutation [107]. D3L is localized in the unstructured N-terminal domain (NTD) [108]. Jian et al. sequenced five SARS-CoV-2 B.1.1.7 variant genomes and D3L, R203K, G204R, and S235F mutations observed in all five cases [109]. Zheng et al. studied the protein interaction of the SARS-CoV-2 N protein via affinity purification and mass spectrometry. They found that MOV10, EIF2AK2, TRIM25, G3BP1, ZC3HAV1, and ZCCHC3 proteins interact with N protein [110]. N of SARS-CoV-2 binds to La-related protein 1 (LARP1), UPF1, MOV10, and the Ras-GTPase-activating protein (GAP)-binding protein 1 (G3BP1) and G3BP2.

ORF9b protein

ORF9b is located within the N gene, which codes for a 97 amino acid long protein localized in the mitochondrial membrane. P3L (12.0%), T60A (11.3%), and S50L (1.5%) were common mutations in our study. T60A was observed in Delta and Delta Plus SARS-CoV-2 genomes and is common in worldwide SARS-CoV-2. ORF9b target Tom70, a mitochondrial import receptor that plays an important role in innate immune signaling [111]. Furthermore, ORF9, by mediating ATG5 induced autophagy in host cells [112]. The antibody responses were identified for ORF9b in COVID-19 convalescent patients [113].

ORF9c protein

ORF9c formerly also called ORF14 is a conserved region that presents in previous SARS-CoV [114]. In our study, ORF9c was 9.6% conserved and G50N (82.1%), G50W (5.3%), V49L (1.6%), L67F (0.6%), and P33S (0.6%) were the most common mutations in this region. ORF9c targeted Nod-like receptor NLRX1, proteinase-activated receptor 2 (F2RL1), and Nedd4 Family Interacting Protein 2 (NDFIP2) genes in the NF-κB pathway [111]. ORF9c interacts with membrane-related proteins, proteins in the mitochondria and endoplasmic reticulum (ER), and golgi that play a role in the protein biosynthesis and transport systems, affecting immune evasion, virulence, and pathogenesis [114]. G50N observed SARS-CoV-2 variants from the Philippines [115].

Mutational hotspots and conserved domains of the SARS-CoV-2 genome

Due to the viral genomic change, to overcome false-negative tests, drug resistance, and immune evasion, it has been suggested stable genes in the SARS-CoV-2 genome, which have conserved regions, are suitable targets for drug and vaccine designing [116, 117]. Furthermore, conserved regions are highly specific and sensitive markers for the detection of SARS-CoV-2 infection. The analysis of the SARS-CoV-2 genome conservation revealed that N, S, NSP12, and ORF9c are the lowest conservation among SARS-CoV-2 regions.

N gene is one of the most non-conserved genes in the SARS-CoV-2 [118]. The SARS-CoV-2 N protein has five domains: an N-terminal domain (NTD, amino acid residues 44–176), an RNA-binding domain (RBD), a serine-arginine (SR) rich-linker region (LKR), a dimerization domain, and a C-terminal domain (CTD, amino acid residues 248–369) [119]. The NTD binds to the 3′-end of the viral RNA, and is referred to as the RNA-binding domain, and plays a role in cell signaling [120]. In our study, 6.7% of the N protein has no mutations and 32.5% has four or more than four mutations (Figure 3C). The most common mutations in the N protein G204R, R203K, S235F, and S194L identified in this study are located within the SR rich-linker (182–247 amino acid). Accumulation of N gene mutations in the linker and the unstructured regions were also detected in Russian samples [118]. SR rich-linker has a different role in SARS-CoV, including oligomerization, phospho-regulation, and RNA and protein binding [121, 122].

The S is composed of 1273 amino acids and divided into the S1 and S2 subunits. The S1 subunit contains an N-terminal domain (NTD, 14–305 residues) and a receptor-binding domain (RBD, 319–541 residues). S2 subunit is composed of the fusion peptide (FP, 788–806 residues), heptapeptide repeat sequence 1 (HR1) (912–984 residues), HR2 (1163–1213 residues), TM domain (1213–1237 residues), and cytoplasm domain (1237–1273 residues) [123]. In the S protein sequence, 30.1 % have no mutations; however, 38.2% have four or more than four mutations. Common mutations were observed including D614G, D138Y, S477N, A262T, and N501Y. Our result revealed that two residues (508-635 and 127-254) are hot spot regions in the spike. Several studies have reported that the SARS-CoV-2 spike protein affects neutralization. The majority of SARS-CoV-2 neutralizing antibodies target either the RBD or the NTD of the viral spike [124].

Among common spike mutations in our study, E484K, S477N, and N501Y are in RBD. Several studies have been demonstrated that E484 with amino acid changes to K, Q, or P reduces neutralization by both monoclonal antibodies and human sera or plasma [80, 125]. Mutations at position 477 of the spike protein (S477G, S477N, and S477R) are also attenuate neutralizing immune responses [126].

The S protein-NTD could play a role in the prefusion-to-postfusion transition of the S protein and may have a high risk of immune evasion [127, 128]. In the NTD, two epitopes consist of residues 140–156 (N3 loop) and 246–260 (N5 loop) are important regions [128]. D138Y is located in S protein-NTD. Previous study has been identified that R246, Y144, K147, Y248, L249, and P251 are involved in the direct interaction of the NTD with many mAbs [127]. Another study reported the NTD mutations N148S, K150R, K150E, K150T, K150Q, and S151P as escape mutations [80]. Interestingly, the VOC Omicron, as a highly transmissible variant of concern, is characterized by 37 mutations in the sequence of the S protein, which includes 10 mutations in the N-terminal domain (NTD) and 15 mutations in the receptor binding domain (RBD) [129].

NSP 12 is the most conserved protein in coronaviruses, which its function associated with viral replication/transcription. The NSP12 of SARS-CoV-2 exhibits 96% sequence identity with SARS-CoV and 71% with MERS-CoV [111]. In our study, 5.9% of NSP12 had not any mutations, hot spot regions was 279_372, and conserved regions were 186-279 and 837-930.

CoV NSP12 also contains a nidovirus-unique N-terminal extension (amino acid 1-397) and a polymerase domain (amino acid 398-919) [130]. Common NSP12 mutations in our sample, P323L, G137C, and G137S were located in the polymerase domain.

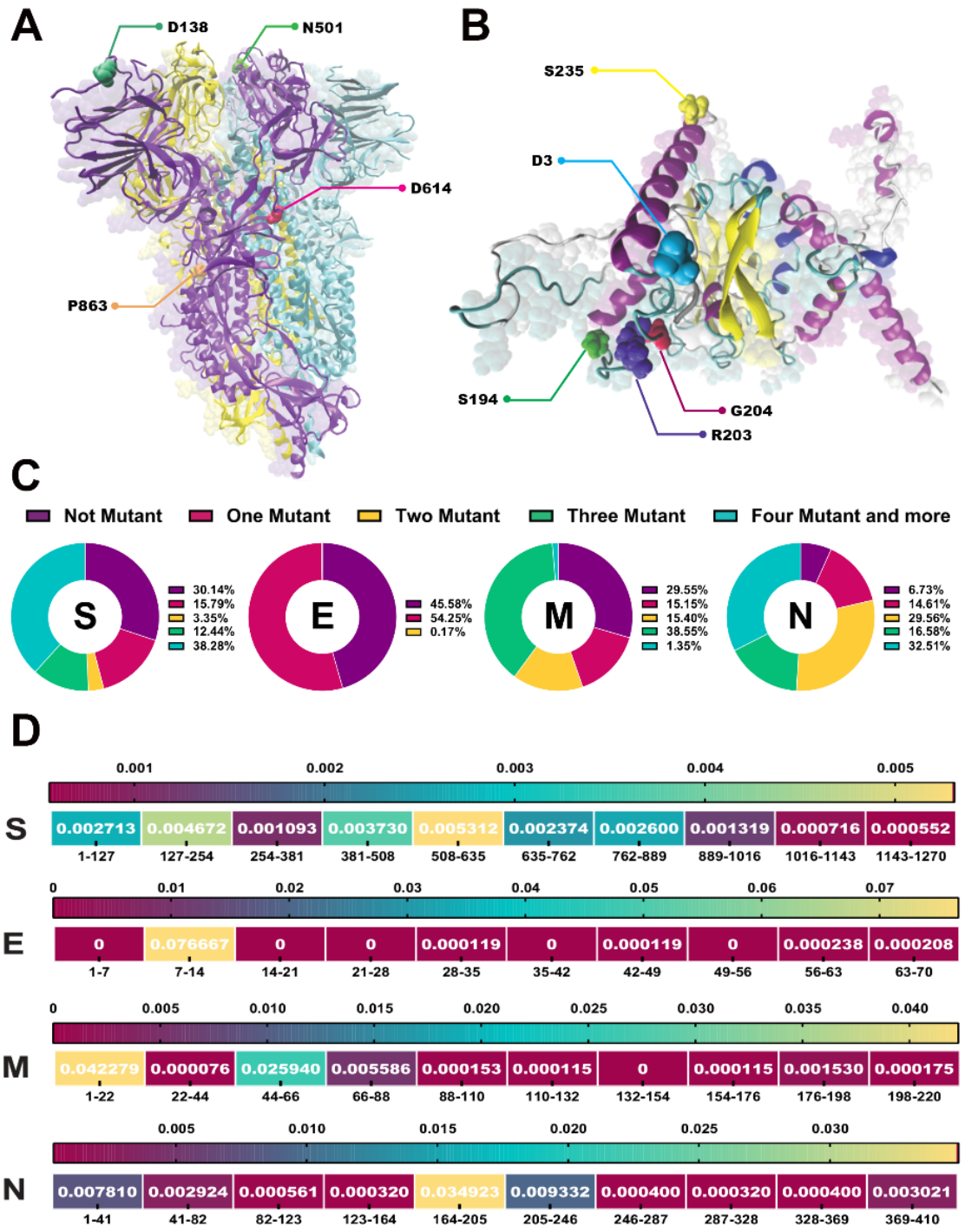

Figure 4. Conserved and hot spot region in structural proteins of SARS-CoV-2. (**A**) SARS-CoV-2 spike protein (PDB entry 6VYB) 3D model; Chain A, B, and C reported in violet, yellow and blue. (**B**) Structure of N protein predicted by I-Tasser and Protein backbones colored according to the secondary structure. (**C**) Pie charts show the proportion of the mutant in S, E, M, and N of SARS-CoV-2. (**D**) Heat map of genome conservation data showing the region that were differentially mutated in S, E, M, and N of SARS-CoV-2.

SARS-CoV-2 M gene is highly conserved between SARS-CoV (identity: 90.5%; similarity: 98.2%) and Bat and Pangolin isolates [131]. In our study, 73.36% of the M gene had no mutation, and I73M and I82T were the most common mutations in our sample.

The mutational analysis of our study revealed that only 45.5% of the SARS-CoV-2 E had no mutation. Rahman et al. analyzed 81,818 sequences of SARS-CoV-2 belonging to 159 countries or territories until 20 August 2020. They found that 1.2% (982/81818) strains possessed amino acid substitutions in 63 sites of the E protein. Previous studies showed that 98.8% of the E protein of globally circulating SARS-CoV-2 strains were conserved [85, 132].

In our analysis, no mutations were observed in more than 95 % of the SARS-CoV-2 genome including NSP7, NSP8, NSP9, NSP10, NSP11, and ORF8 when compared to the Wuhan-Hu-1 (Figure 4A). Previous Sequence analysis also revealed that NSP6, NSP9, NSP10, and NSP11 present no mutations in Brazilian SARS-CoV-2 samples (Figure 4A). In COVID-19 Egyptian patients, the frequency of mutations was high in ORF1ab, ORF3a S, and N proteins. The lowest mutation rate was observed in M, E, ORF7b, ORF7b, and ORF10 [133]. Kaushal et al. analyzed the USA SARS-CoV-2 genome between January to April 2020 and found that NSP7, NSP9, NSP10, NSP11, E, ORF6, and ORF7b proteins did not accumulate any mutation [134].

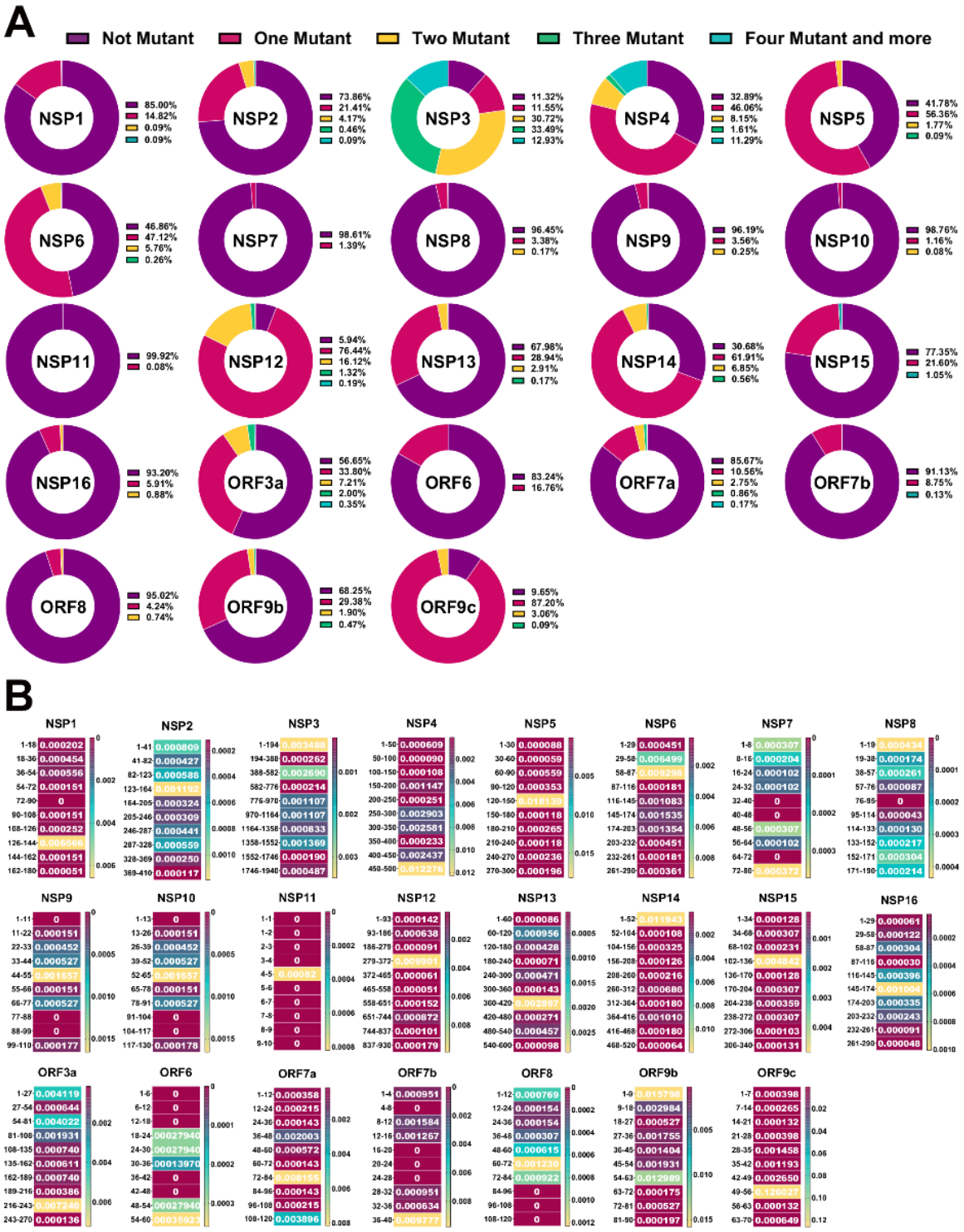

Figure 5. Conserved and hot spot regions in non-structural proteins of SARS-CoV-2. (**A**) Pie charts show the proportion of the mutants in SARS-CoV-2. (**B**) Heat map of genome conservation data showing the region that were differentially mutated in SARS-CoV-2.

Chronological trend of common SARS-CoV 2 mutations

Detecting and identifying circulating SARS-CoV-2 new variants and their consequences is pivotal in managing and controlling VOCs' spread. To elucidate how the amino acid changes in the SARS-CoV 2 genome are responsible for different outbreak waves, we explore the frequency of the top 20 mutations in Iran from January 2020 to April 2022. (Figure 6) From 19 February 2020, the first confirmed case of a COVID-19 report, Iran has experienced sixth waves until June 2022. (Figure 7A) The first wave (19 February 2020 - 15 May 2020), the second wave; summer 2020 (1 June 2020 - 31 August 2020), the third wave; autumn 2020 (1 September 2020 - 1 January 2021), the fourth wave; spring 2021 (1 April 2021 - June 2021), the fifth wave (15 July 2021 - 1 October), and the sixth wave of the pandemic (the Omicron variant) (10 January 2022 - 22 March 2022) (Figure 7B). There were five VOCs associated with increased SARS-CoV-2 transmissibility, several cases, and deaths in different countries worldwide. Alpha (lineage B.1.1.7 or UK variant), Beta (lineage B.1.351 or South Africa variant), Gamma (lineage P.1 or Brazil variant), Delta (lineage B.1.617.2 variants), and Omicron (lineage B.1.1.529).

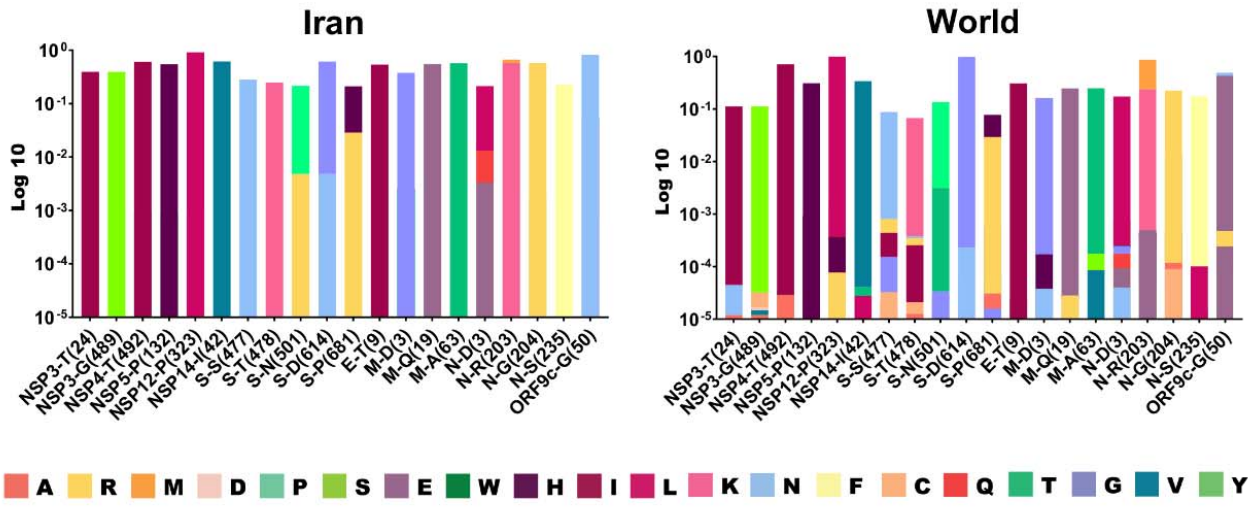

Figure 6. The amino acid substitutions of common SARS-CoV 2 mutations

More recently, it has been shown that approximately 96.5% of amino acid residues in the SARS-CoV-2 spike have undergone mutations since the outbreak of the COVID-19 [135]. Among them, it is well-known that D614G/N mutation enhanced the replication of the SARS-CoV-2 and

is the main characteristic of VOC. The D614G mutation was first identified in January 2020, and by the end of March 2020, it had increased in frequency worldwide and had become dominant worldwide until now. On 03/31/2020, the first spike mutations in D614 were detected in Iran and Figure 7C). Previous study reported that the D614G mutation was first detected in Iran in May 2020 [28]. S477N and S477G strengthen the binding of the SARS-COV-2 spike with the hACE2 receptor [75]. S477N is one of the Omicron RBD mutations S477N enhanced the binding of ACE2 by the Omicron [136]. Our study shows that S477N was first detected on 10/31/2020 in Iran and increased from October 2020 to February 2021. A previous study showed that three spike mutations, D138Y, S477N, and D614G, co-occur in B.1.1.* lineages [28]. First amino acid substitution in T478 was detected on 01/31/2020 worldwide, and then in August 2021, the mutation in T478 dramatically increased in the following months. T478K was first detected in Iran in August 2021. T478K is one of the Omicron RBD that has positive electrostatic potential at the surface for ACE2 binding [136]. N501Y/R was first detected in Iran in March 2021 and increased in December 2021. The N501Y is found in three variants of Concern, alpha, beta, and Gamma, and determines a tight interaction of the SARS-CoV-2 RBD [137], and seems to contribute to the immune escape [138]. P681R/H was first detected in Iran in April 2021 and increased in October 2021 and following months until March 2021. As previously observed, P681R was one of the mutations in the Delta variant that enhances SARS-CoV-2 fitness and is associated with enhanced pathogenicity [139, 140].

N protein from SARS-CoV-2 is a multivalent RNA-binding protein promoting genome packaging, interference in host translation, and RNA chaperoning [141]. N mutations D3L/E/Q, R203, and G204 were observed between May and August 2020 in Iran and then increased in September and December 2020 and became stable for several months (Figure 7C).

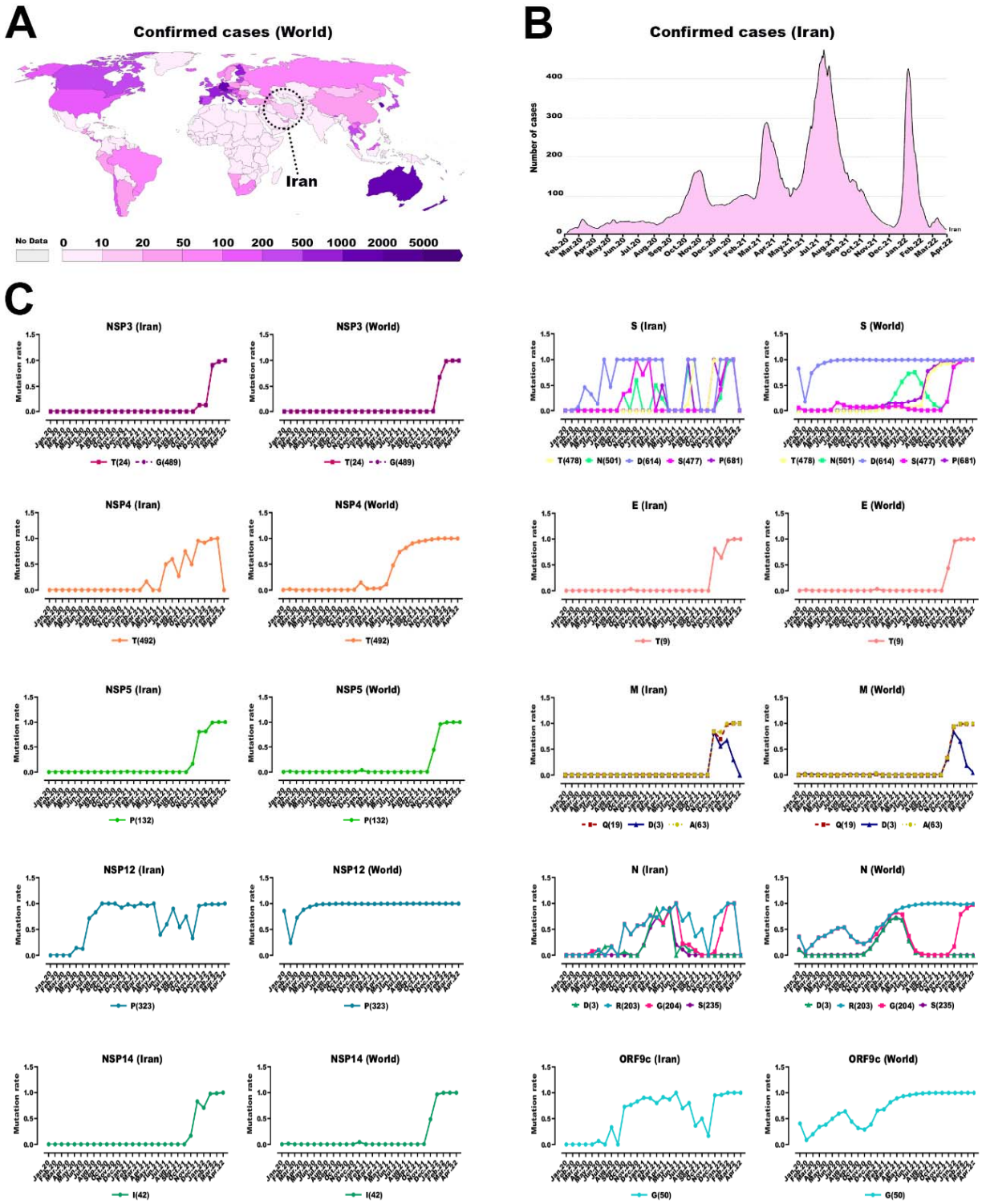

Figure 7. Timeline of common mutations in circulating SARS-CoV-2. (A) Confirmed COVID-19 cases in the world and (B) Iran. (C) Trends in common mutations and infections in Iran and the world from January 2020 to April 2022.

The NSP4 T492I mutation was detected in April 2021 in Iran and in May 2021 worldwide and then increased in the following months in Iran and worldwide (Figure 7C). T492I was found in the omicron variant and one of the SARS-CoV-2 non-Spike mutations co-occurring with S T478K (93.61%) [142, 143]. T492I mutation NSP4 protein is highly conserved and increased over time in different continents, especially in recent months [144]. T492I was also found in Pakistani SARS-CoV-2 Variants from 15 to 20 September 2021 [145].

NSP12 P323L appears in May 2020 and gradually increases and becomes stable in the following months (Figure 7C). Numerous studies have demonstrated that the P323L coevolved with D614G near 100% coexistence of P323L with D614G found in US SARS-CoV-2 in January 2020 - June 2020 [146]. Co-occurrence of P323L and D614G mutations has also been found in sequences obtained from Turkey [147]. Also, the P323L mutation was found in severely affected patients [148].

The first Omicron variant was identified in southern Africa in November 2021 and December 2021 in Iran. The Omicron variant is characterized by more than 32 mutations compared to the original virus, NSP3 K38R, V1069I, Δ1265, L1266I, A1892T, NSP4 T492I, NSP5 P132H, NSP6 Δ105-107, A189V, NSP12 P323L, and NSP14 I42V, E T9I, M D3G, Q19E, and A63T, N P13L, Δ31-33, R203K, G204R [142]. Interestingly, our results showed several common mutations, including NSP3 T24I and G489S, NSP5 P132H, NSP14 I42V, E T9I, M D3L Q19E, and A63T emerged in late November and December in Iran and worldwide and then increased dramatically until March 2022. (Figure 7C). These mutations can be considered a major driver of the COVID-19 surge in the sixth COVID-19 wave in Iran in January 2022.

Conclusion

This study analyzed SARS-CoV-2 genomes from Iranian samples from January 2020 to April 2022. SARS-CoV-2 mutations affect a different aspect of the COVID-19 pandemic. First, mutations change the sequence of primers and probes applied in PCR-based tests, so false-negative test results. Second, mutations in proteins such as S and N could alter vaccine targets, and vaccine efficacy for novel mutations is reduced. Therefore, according to the conserved and hotspot region in SARS-CoV-2 genomes, a novel multi-peptide subunit-based epitope vaccine candidate could localized design against COVID-19. Finally, monitoring of mutations in the SARS-CoV-2 genome could anticipate future viral drug resistance. In addition,

structure-based drug discovery is a promising therapeutic solution for the treatment of virus infections in the future that are based on specific molecular targets.

Declaration of competing interest

The authors declare that they have no conflicts of interest that might be relevant to the contents of this manuscript and the research was carried out regardless of commercial or financial relationships that may cause any conflict of interests.

Acknowledgments

The authors thank all of the researchers who have shared genome data openly via the Global Initiative on Sharing All Influenza Data (GISAID). We would like to thank Dr. Hossein Najmabadi for the review the manuscript and his insightful comments and suggestions.

Data availability

The raw data supporting the conclusions of this article is available in supplementary files.

Authors' contributions

K.R., M.M., M.M.S., S.T., M.H.A., and B.Mo. contributed to data collection; H.N., M.H.A., M.M., and B.Ma. contributed to study design; K.R., and M.M. designed workflow, code and data analysis; B.Ma., K.R., and M.M contributed to data visualization; M.H.A wrote the manuscript; B.Ma. S.T., and M.M. corrected the manuscript and provided useful comments; M.M., K.R., B.Ma. monitored the accuracy of additional data; B.Ma. designed graphical contents; K.R. have supervised the work.

References

1. Lu, H., C.W. Stratton, and Y.W. Tang, *Outbreak of pneumonia of unknown etiology in Wuhan, China: The mystery and the miracle.* Journal of medical virology, 2020. **92**(4): p. 401.
2. Organization, W.H., *WHO Director-General's opening remarks at the media briefing on COVID-19-11 March 2020*. 2020, Geneva, Switzerland.
3. Ghadir, M.R., et al., *The COVID-19 outbreak in Iran; the first patient with a definite diagnosis.* Archives of Iranian medicine, 2020. **23**(7): p. 503-504.
4. Munnink, B.B.O., et al., *Rapid SARS-CoV-2 whole-genome sequencing and analysis for informed public health decision-making in the Netherlands.* Nature medicine, 2020. **26**(9): p. 1405-1410.
5. Zhou, P., et al., *A pneumonia outbreak associated with a new coronavirus of probable bat origin.* nature, 2020. **579**(7798): p. 270-273.
6. Khare, S., et al., *GISAID's Role in Pandemic Response.* China CDC Weekly, 2021. **3**(49): p. 1049.

7. Elbe, S. and G. Buckland-Merrett, *Data, disease and diplomacy: GISAID's innovative contribution to global health.* Global challenges, 2017. **1**(1): p. 33-46.
8. Shu, Y. and J. McCauley, *GISAID: Global initiative on sharing all influenza data–from vision to reality.* Eurosurveillance, 2017. **22**(13): p. 30494.
9. Kim, D., et al., *The architecture of SARS-CoV-2 transcriptome.* Cell, 2020. **181**(4): p. 914-921. e10.
10. Lan, J., et al., *Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor.* Nature, 2020. **581**(7807): p. 215-220.
11. Yadav, R., et al., *Role of Structural and Non-Structural Proteins and Therapeutic Targets of SARS-CoV-2 for COVID-19.* Cells, 2021. **10**(4): p. 821.
12. Viana, J., et al., *Controlling the pandemic during the SARS-CoV-2 vaccination rollout.* Nature Communications, 2021. **12**(1): p. 1-15.
13. Polack, F.P., et al., *Safety and efficacy of the BNT162b2 mRNA Covid-19 vaccine.* New England Journal of Medicine, 2020.
14. Voysey, M., et al., *Safety and efficacy of the ChAdOx1 nCoV-19 vaccine (AZD1222) against SARS-CoV-2: an interim analysis of four randomised controlled trials in Brazil, South Africa, and the UK.* The Lancet, 2021. **397**(10269): p. 99-111.
15. Wang, H., et al., *Development of an inactivated vaccine candidate, BBIBP-CorV, with potent protection against SARS-CoV-2.* Cell, 2020. **182**(3): p. 713-721. e9.
16. Levin, E.G., et al., *Waning immune humoral response to BNT162b2 Covid-19 vaccine over 6 months.* New England Journal of Medicine, 2021. **385**(24): p. e84.
17. Martínez-Flores, D., et al., *SARS-CoV-2 vaccines based on the spike glycoprotein and implications of new viral variants.* Frontiers in Immunology, 2021. **12**.
18. Zhang, L., et al., *The D614G mutation in the SARS-CoV-2 spike protein reduces S1 shedding and increases infectivity.* BioRxiv, 2020.
19. Motozono, C., et al., *SARS-CoV-2 spike L452R variant evades cellular immunity and increases infectivity.* Cell host & microbe, 2021. **29**(7): p. 1124-1136. e11.
20. Jangra, S., et al., *SARS-CoV-2 spike E484K mutation reduces antibody neutralisation.* The Lancet Microbe, 2021.
21. Pierobon, A., et al., *Outbreak of SARS-CoV-2 B. 1.617. 2 (Delta) variant in a Nursing Home 28 weeks after two doses of mRNA anti-Covid-19 vaccines: evidence of a waning immunity.* Clinical Microbiology and Infection, 2021.
22. Andrews, N., et al., *Effectiveness of COVID-19 vaccines against the Omicron (B. 1.1. 529) variant of concern.* medRxiv, 2021.
23. Eslami, S., M.C. Glassy, and S. Ghafouri-Fard, *A comprehensive overview of identified mutations in SARS CoV-2 spike glycoprotein among Iranian patients.* Gene, 2021: p. 146113.
24. Fattahi, Z., et al., *SARS-CoV-2 outbreak in Iran: The dynamics of the epidemic and evidence on two independent introductions.* Transboundary and emerging diseases, 2021.
25. Moradi, J., et al., *Variation analysis of SARS-CoV-2 complete sequences from Iran.* bioRxiv, 2021.
26. Mazhari, S., et al., *SARS-CoV-2 NSP-12 mutations survey during the pandemic in the world.* 2021.
27. Rahimian, K., et al., *SARS2Mutant: SARS-CoV-2 Amino-Acid Mutation Atlas Database.* bioRxiv, 2022: p. 2022.06.12.495856.
28. Fattahi, Z., et al., *Disease waves of SARS-CoV-2 in Iran closely mirror global pandemic trends.* medRxiv, 2021.
29. Wang, R., et al., *Analysis of SARS-CoV-2 mutations in the United States suggests presence of four substrains and novel variants.* Communications biology, 2021. **4**(1): p. 1-14.
30. Mou, K., et al., *Emerging Mutations in Nsp1 of SARS-CoV-2 and Their Effect on the Structural Stability.* Pathogens, 2021. **10**(10): p. 1285.

31. Zheng, Y.-X., et al., *Nsp2 has the potential to be a drug target revealed by global identification of SARS-CoV-2 Nsp2-interacting proteins.* Acta Biochimica et Biophysica Sinica, 2021.
32. Mariano, G., et al., *Structural characterization of SARS-CoV-2: Where we are, and where we need to be.* Frontiers in molecular biosciences, 2020: p. 344.
33. Cornillez-Ty, C.T., et al., *Severe acute respiratory syndrome coronavirus nonstructural protein 2 interacts with a host protein complex involved in mitochondrial biogenesis and intracellular signaling.* Journal of virology, 2009. **83**(19): p. 10314-10318.
34. von Brunn, A., et al., *Analysis of intraviral protein-protein interactions of the SARS coronavirus ORFeome.* PloS one, 2007. **2**(5): p. e459.
35. Sarkar, R., et al., *Comprehensive analysis of genomic diversity of SARS-CoV-2 in different geographic regions of India: an endeavour to classify Indian SARS-CoV-2 strains on the basis of co-existing mutations.* Archives of Virology, 2021. **166**(3): p. 801-812.
36. Gupta, M., et al., *CryoEM and AI reveal a structure of SARS-CoV-2 Nsp2, a multifunctional protein involved in key host processes.* bioRxiv, 2021.
37. Vilar, S. and D.G. Isom, *One year of SARS-CoV-2: How much has the virus changed?* Biology, 2021. **10**(2): p. 91.
38. Bui, N.-N., et al., *Haplotype distribution of SARS-CoV-2 variants in low and high vaccination rate countries during ongoing global COVID-19 pandemic in early 2021.* Infection, Genetics and Evolution, 2022. **97**: p. 105164.
39. Bitew, M., et al., *SARS-CoV-2 Genome Sequence Obtained from Ethiopia.* Microbiology Resource Announcements, 2022. **11**(2): p. e01182-21.
40. Miljanovic, D., et al., *The first molecular characterization of Serbian SARS-CoV-2 isolates from a unique early second wave in Europe.* Frontiers in microbiology, 2021. **12**: p. 1526.
41. Sakai, Y., et al., *Two-amino acids change in the nsp4 of SARS coronavirus abolishes viral replication.* Virology, 2017. **510**: p. 165-174.
42. Tiwari, M. and D. Mishra, *Investigating the genomic landscape of novel coronavirus (2019-nCoV) to identify non-synonymous mutations for use in diagnosis and drug design.* Journal of Clinical Virology, 2020. **128**: p. 104441.
43. Fooladinezhad, H., et al., *SARS-CoV-2 NSP3, NSP4 and NSP6 mutations and Epistasis during the pandemic in the world: Evolutionary Trends and Natural Selections in Six Continents.* medRxiv, 2022.
44. Timmers, L.F.S.M., et al., *SARS-CoV-2 mutations in Brazil: from genomics to putative clinical conditions.* Scientific reports, 2021. **11**(1): p. 1-14.
45. Yashvardhini, N., A. Kumar, and D.K. Jha, *Analysis of SARS-CoV-2 mutations in the main viral protease (NSP5) and its implications on the vaccine designing strategies.* Vacunas, 2021.
46. Yuan, F., et al., *Global SNP analysis of 11,183 SARS-CoV-2 strains reveals high genetic diversity.* Transboundary and emerging diseases, 2021. **68**(6): p. 3288-3304.
47. Fung, S.-Y., et al., *SARS-CoV-2 main protease suppresses type I interferon production by preventing nuclear translocation of phosphorylated IRF3.* International Journal of Biological Sciences, 2021. **17**(6): p. 1547.
48. Li, W., et al., *SARS-CoV-2 Nsp5 Activates NF-κB Pathway by Upregulating SUMOylation of MAVS.* Frontiers in immunology, 2021. **12**: p. 750969-750969.
49. Angelini, M.M., et al., *Severe acute respiratory syndrome coronavirus nonstructural proteins 3, 4, and 6 induce double-membrane vesicles.* MBio, 2013. **4**(4): p. e00524-13.
50. Wang, R., et al., *Decoding asymptomatic COVID-19 infection and transmission.* The journal of physical chemistry letters, 2020. **11**(23): p. 10007-10015.
51. Sun, X., et al., *SARS-CoV-2 non-structural protein 6 triggers NLRP3-dependent pyroptosis by targeting ATP6AP1.* Cell Death & Differentiation, 2022: p. 1-15.

52. Reshamwala, S.M., et al., *Mutations in SARS-CoV-2 nsp7 and nsp8 proteins and their predicted impact on replication/transcription complex structure.* Journal of medical virology, 2021. **93**(7): p. 4616-4619.
53. Rosas-Lemus, M., et al., *The crystal structure of nsp10-nsp16 heterodimer from SARS-CoV-2 in complex with S-adenosylmethionine.* BioRxiv, 2020.
54. Chen, Y., et al., *Biochemical and structural insights into the mechanisms of SARS coronavirus RNA ribose 2′-O-methylation by nsp16/nsp10 protein complex.* PLoS pathogens, 2011. **7**(10): p. e1002294.
55. Rogstam, A., et al., *Crystal structure of non-structural protein 10 from severe acute respiratory syndrome coronavirus-2.* International journal of molecular sciences, 2020. **21**(19): p. 7375.
56. Kozielski, F., et al., *Identification of fragments binding to SARS-CoV-2 nsp10 reveals ligand-binding sites in conserved interfaces between nsp10 and nsp14/nsp16.* RSC Chemical Biology, 2022.
57. Wu, C., et al., *Analysis of therapeutic targets for SARS-CoV-2 and discovery of potential drugs by computational methods.* Acta Pharmaceutica Sinica B, 2020. **10**(5): p. 766-788.
58. Vazquez, C., et al., *SARS-CoV-2 viral proteins NSP1 and NSP13 inhibit interferon activation through distinct mechanisms.* PLoS One, 2021. **16**(6): p. e0253089.
59. Sui, C., et al., *SARS-CoV-2 NSP13 Inhibits Type I IFN Production by Degradation of TBK1 via p62-Dependent Selective Autophagy.* The Journal of Immunology, 2022. **208**(3): p. 753-761.
60. Farooq Rashid, M.S., et al., *Structural Analysis on the Severe Acute Respiratory Syndrome Coronavirus 2 Non-structural Protein 13 Mutants Revealed Altered Bonding Network With TANK Binding Kinase 1 to Evade Host Immune System.* Frontiers in Microbiology, 2021. **12**.
61. Pachetti, M., et al., *Emerging SARS-CoV-2 mutation hot spots include a novel RNA-dependent-RNA polymerase variant.* Journal of translational medicine, 2020. **18**(1): p. 1-9.
62. Chen, Y., et al., *Functional screen reveals SARS coronavirus nonstructural protein nsp14 as a novel cap N7 methyltransferase.* Proceedings of the National Academy of Sciences, 2009. **106**(9): p. 3484-3489.
63. Hsu, J.C.-C., et al., *Translational shutdown and evasion of the innate immune response by SARS-CoV-2 NSP14 protein.* Proceedings of the National Academy of Sciences, 2021. **118**(24).
64. Fan, L.-q., et al., *Biological Significance of the Genomic Variation and Structural Dynamics of SARS-CoV-2 B. 1.617.* Frontiers in Microbiology, 2021: p. 2837.
65. Hackbart, M., X. Deng, and S.C. Baker, *Coronavirus endoribonuclease targets viral polyuridine sequences to evade activating host sensors.* Proceedings of the National Academy of Sciences, 2020. **117**(14): p. 8094-8103.
66. Deng, X., et al., *Coronavirus endoribonuclease activity in porcine epidemic diarrhea virus suppresses type I and type III interferon responses.* Journal of virology, 2019. **93**(8): p. e02000-18.
67. Umair, M., et al., *Genomic surveillance reveals the detection of SARS-CoV-2 delta, beta, and gamma VOCs during the third wave in Pakistan.* Journal of Medical Virology, 2022. **94**(3): p. 1115-1129.
68. Snijder, E.J., et al., *Unique and conserved features of genome and proteome of SARS-coronavirus, an early split-off from the coronavirus group 2 lineage.* Journal of molecular biology, 2003. **331**(5): p. 991-1004.
69. Decroly, E., et al., *Crystal structure and functional analysis of the SARS-coronavirus RNA cap 2′-O-methyltransferase nsp10/nsp16 complex.* PLoS pathogens, 2011. **7**(5): p. e1002059.
70. Yurkovetskiy, L., et al., *Structural and functional analysis of the D614G SARS-CoV-2 spike protein variant.* Cell, 2020. **183**(3): p. 739-751. e8.

71. Korber, B., et al., *Tracking changes in SARS-CoV-2 spike: evidence that D614G increases infectivity of the COVID-19 virus.* Cell, 2020. **182**(4): p. 812-827. e19.
72. Plante, J.A., et al., *Spike mutation D614G alters SARS-CoV-2 fitness.* Nature, 2021. **592**(7852): p. 116-121.
73. Korber, B., et al., *Spike mutation pipeline reveals the emergence of a more transmissible form of SARS-CoV-2.* BioRxiv, 2020.
74. Volz, E., et al., *Evaluating the effects of SARS-CoV-2 spike mutation D614G on transmissibility and pathogenicity.* Cell, 2021. **184**(1): p. 64-75. e11.
75. Singh, A., et al., *Serine 477 plays a crucial role in the interaction of the SARS-CoV-2 spike protein with the human receptor ACE2.* Scientific reports, 2021. **11**(1): p. 1-11.
76. Choi, J.Y. and D.M. Smith, *SARS-CoV-2 variants of concern.* Yonsei medical journal, 2021. **62**(11): p. 961.
77. Chakraborty, S., *E484K and N501Y SARS-CoV 2 spike mutants Increase ACE2 recognition but reduce affinity for neutralizing antibody.* International Immunopharmacology, 2022. **102**: p. 108424.
78. Garcia-Beltran, W.F., et al., *Multiple SARS-CoV-2 variants escape neutralization by vaccine-induced humoral immunity.* Cell, 2021. **184**(9): p. 2372-2383. e9.
79. Zhao, S., et al., *Quantifying the transmission advantage associated with N501Y substitution of SARS-CoV-2 in the UK: an early data-driven analysis.* Journal of travel medicine, 2021. **28**(2): p. taab011.
80. Weisblum, Y., et al., *Escape from neutralizing antibodies by SARS-CoV-2 spike protein variants.* Elife, 2020. **9**: p. e61312.
81. Sengupta, A., S.S. Hassan, and P.P. Choudhury, *Clade GR and clade GH isolates of SARS-CoV-2 in Asia show highest amount of SNPs.* Infection, Genetics and Evolution, 2021. **89**: p. 104724.
82. Yang, X.-J., *SARS-COV-2 γ variant acquires spike P681H or P681R for improved viral fitness.* bioRxiv, 2021.
83. Wang, R., et al., *Characterizing SARS-CoV-2 mutations in the United States.* Research square, 2020.
84. Azad, G.K. and P.K. Khan, *Variations in Orf3a protein of SARS-CoV-2 alter its structure and function.* Biochemistry and biophysics reports, 2021. **26**: p. 100933.
85. Rahman, M.S., et al., *Mutational insights into the envelope protein of SARS-CoV-2.* Gene reports, 2021. **22**: p. 100997.
86. Mukhtar, F., et al., *Emerging mutations in spike and other structural proteins of SARS-CoV-2.* Authorea Preprints, 2021.
87. Troyano-Hernáez, P., R. Reinosa, and Á. Holguín, *Evolution of SARS-CoV-2 envelope, membrane, nucleocapsid, and spike structural proteins from the beginning of the pandemic to September 2020: a global and regional approach by epidemiological week.* Viruses, 2021. **13**(2): p. 243.
88. Ghosh, P., et al., *High diversity in Delta variant across countries revealed via genome-wide analysis of SARS-CoV-2 beyond the Spike protein.* bioRxiv, 2021.
89. Lee, J.-G., et al., *Characterization of SARS-CoV-2 proteins reveals Orf6 pathogenicity, subcellular localization, host interactions and attenuation by Selinexor.* Cell & bioscience, 2021. **11**(1): p. 1-12.
90. Zhou, Z., et al., *Structural insight reveals SARS-CoV-2 Orf7a as an immunomodulating factor for human CD14+ monocytes.* Iscience, 2021. **24**(3): p. 102187.
91. Addetia, A., et al., *Identification of multiple large deletions in ORF7a resulting in in-frame gene fusions in clinical SARS-CoV-2 isolates.* Journal of Clinical Virology, 2020. **129**: p. 104523.
92. Hassan, S.S., et al., *The importance of accessory protein variants in the pathogenicity of SARS-CoV-2.* Archives of Biochemistry and Biophysics, 2022: p. 109124.

93. Lau, B.T., et al., *Profiling SARS-CoV-2 mutation fingerprints that range from the viral pangenome to individual infection quasispecies.* Genome medicine, 2021. **13**(1): p. 1-23.
94. Yashvardhini, N., et al., *Genetic variations in the Orf7a protein of SARS-CoV-2 and its possible role in vaccine development.* Biomedical Research and Therapy, 2021. **8**(8): p. 4497-4504.
95. Rosenthal, S.H., et al., *Identification of eight SARS-CoV-2 ORF7a deletion variants in 2,726 clinical specimens.* bioRxiv, 2020.
96. Zhang, J., et al., *A systemic and molecular study of subcellular localization of SARS-CoV-2 proteins.* Signal transduction and targeted therapy, 2020. **5**(1): p. 1-3.
97. Stukalov, A., et al., *Multilevel proteomics reveals host perturbations by SARS-CoV-2 and SARS-CoV.* Nature, 2021. **594**(7862): p. 246-252.
98. Yang, R., et al., *SARS-CoV-2 accessory protein ORF7b mediates tumor necrosis factor-α-induced apoptosis in cells.* Frontiers in Microbiology, 2021. **12**.
99. Nguyen, T.T., et al., *Genomic mutations and changes in protein secondary structure and solvent accessibility of SARS-CoV-2 (COVID-19 virus).* Scientific Reports, 2021. **11**(1): p. 1-16.
100. Rangaiah, A., et al., *Whole-Genome Sequencing of SARS-CoV-2 Strains from Asymptomatic Individuals in India.* Microbiology Resource Announcements, 2022. **11**(2): p. e00850-21.
101. Rahman, M.M., S.B. Kader, and S.S. Rizvi, *Molecular characterization of SARS-CoV-2 from Bangladesh: Implications in genetic diversity, possible origin of the virus, and functional significance of the mutations.* Heliyon, 2021. **7**(8): p. e07866.
102. Collier, D.A., et al., *Sensitivity of SARS-CoV-2 B. 1.1. 7 to mRNA vaccine-elicited antibodies.* Nature, 2021. **593**(7857): p. 136-141.
103. Faria, N.R., et al., *Genomics and epidemiology of the P. 1 SARS-CoV-2 lineage in Manaus, Brazil.* Science, 2021. **372**(6544): p. 815-821.
104. Wu, H., et al., *Nucleocapsid mutations R203K/G204R increase the infectivity, fitness, and virulence of SARS-CoV-2.* Cell host & microbe, 2021. **29**(12): p. 1788-1801. e6.
105. Wu, S., et al., *Effects of SARS-CoV-2 mutations on protein structures and intraviral protein–protein interactions.* Journal of medical virology, 2021. **93**(4): p. 2132-2140.
106. Barona-Gómez, F., et al., *Phylogenomics and population genomics of SARS-CoV-2 in Mexico during the pre-vaccination stage reveals variants of interest B. 1.1. 28.4, B. 1.1. 222 or B. 1.1. 519 and B. 1.243 with mutations in the Spike protein and the Nucleocapsid.* medRxiv, 2021.
107. Singh, J., et al., *Structure-function analyses of new SARS-CoV-2 variants B. 1.1. 7, B. 1.351 and B. 1.1. 28.1: clinical, diagnostic, therapeutic and public health implications.* Viruses, 2021. **13**(3): p. 439.
108. Singh, J., et al., *Structure-function investigation of a new VUI-202012/01 SARS-CoV-2 variant.* 2021.
109. Jian, M.-J., et al., *SARS-CoV-2 variants with T135I nucleocapsid mutations may affect antigen test performance.* International Journal of Infectious Diseases, 2022. **114**: p. 112-114.
110. Zheng, X., et al., *Interactome analysis of the nucleocapsid protein of SARS-CoV-2 virus.* Pathogens, 2021. **10**(9): p. 1155.
111. Gordon, D.E., et al., *A SARS-CoV-2 protein interaction map reveals targets for drug repurposing.* Nature, 2020. **583**(7816): p. 459-468.
112. Shi, C.-S., et al., *SARS-coronavirus open reading frame-9b suppresses innate immunity by targeting mitochondria and the MAVS/TRAF3/TRAF6 signalosome.* The Journal of Immunology, 2014. **193**(6): p. 3080-3089.
113. Jiang, H.-w., et al., *Global profiling of SARS-CoV-2 specific IgG/IgM responses of convalescents using a proteome microarray.* MedRxiv, 2020.
114. Andres, A.D., et al., *SARS-CoV-2 ORF9c is a membrane-associated protein that suppresses antiviral responses in cells.* bioRxiv, 2020.

115. Velasco, J.M., et al., *Coding-complete genome sequences of 11 SARS-CoV-2 B. 1.1. 7 and B. 1.351 variants from metro Manila, Philippines.* Microbiology Resource Announcements, 2021. **10**(28): p. e00498-21.
116. Anand, N.M., et al., *A comprehensive SARS-CoV-2 genomic analysis identifies potential targets for drug repurposing.* Plos one, 2021. **16**(3): p. e0248553.
117. Basit, A., et al., *Designing short peptides to block the interaction of SARS-CoV-2 and human ACE2 for COVID-19 therapeutics.* Frontiers in pharmacology, 2021: p. 2310.
118. Kiryanov, S.A., et al., *Identification of Hotspot Mutations in the N Gene of SARS-CoV-2 in Russian Clinical Samples That May Affect the Detection by Reverse Transcription-PCR.* Diagnostics, 2022. **12**(1): p. 147.
119. Cubuk, J., et al., *The SARS-CoV-2 nucleocapsid protein is dynamic, disordered, and phase separates with RNA.* Nature communications, 2021. **12**(1): p. 1-17.
120. Rahman, M.S., et al., *Evolutionary dynamics of SARS-CoV-2 nucleocapsid protein and its consequences.* Journal of medical virology, 2021. **93**(4): p. 2177-2195.
121. Chang, C.-k., et al., *Transient oligomerization of the SARS-CoV N protein–implication for virus ribonucleoprotein packaging.* PloS one, 2013. **8**(5): p. e65045.
122. Wu, C., et al., *Characterization of SARS-CoV-2 nucleocapsid protein reveals multiple functional consequences of the C-terminal domain.* Iscience, 2021. **24**(6): p. 102681.
123. Raman, R., K.J. Patel, and K. Ranjan, *COVID-19: Unmasking emerging SARS-CoV-2 variants, vaccines and therapeutic strategies.* Biomolecules, 2021. **11**(7): p. 993.
124. Cerutti, G., et al., *Potent SARS-CoV-2 neutralizing antibodies directed against spike N-terminal domain target a single supersite.* Cell host & microbe, 2021. **29**(5): p. 819-833. e7.
125. Greaney, A.J., et al., *Complete mapping of mutations to the SARS-CoV-2 spike receptor-binding domain that escape antibody recognition.* Cell host & microbe, 2021. **29**(1): p. 44-57. e9.
126. Liu, Z., et al., *Identification of SARS-CoV-2 spike mutations that attenuate monoclonal and serum antibody neutralization.* Cell host & microbe, 2021. **29**(3): p. 477-488. e4.
127. Zhou, L., et al., *Predicting spike protein NTD mutations of SARS-CoV-2 causing immune escape by molecular dynamics simulations.* Physical Chemistry Chemical Physics, 2022.
128. Chi, X., et al., *A neutralizing human antibody binds to the N-terminal domain of the Spike protein of SARS-CoV-2.* Science, 2020. **369**(6504): p. 650-655.
129. Saxena, S.K., et al., *Characterization of the novel SARS-CoV-2 Omicron (B. 1.1. 529) Variant of Concern and its global perspective.* Journal of medical virology, 2021.
130. Kirchdoerfer, R.N. and A.B. Ward, *Structure of the SARS-CoV nsp12 polymerase bound to nsp7 and nsp8 co-factors.* Nature communications, 2019. **10**(1): p. 1-9.
131. Yoshimoto, F.K., *The proteins of severe acute respiratory syndrome coronavirus-2 (SARS CoV-2 or n-COV19), the cause of COVID-19.* The protein journal, 2020. **39**(3): p. 198-216.
132. Bianchi, M., et al., *Sars-CoV-2 envelope and membrane proteins: structural differences linked to virus characteristics?* BioMed Research International, 2020. **2020**.
133. Zekri, A.-R.N., et al., *Characterization of the SARS-CoV-2 genomes in Egypt in first and second waves of infection.* Scientific Reports, 2021. **11**(1): p. 1-11.
134. Kaushal, N., et al., *Mutational frequencies of SARS-CoV-2 genome during the beginning months of the outbreak in USA.* Pathogens, 2020. **9**(7): p. 565.
135. Guruprasad, K., *Mutations in human SARS-CoV-2 spike proteins, potential drug binding and epitope sites for COVID-19 therapeutics development.* Current research in structural biology, 2022. **4**: p. 41-50.
136. Lan, J., et al., *Structural insights into the SARS-CoV-2 Omicron RBD-ACE2 interaction.* Cell research, 2022: p. 1-3.

137. Li, Q., et al., *SARS-CoV-2 501Y. V2 variants lack higher infectivity but do have immune escape.* Cell, 2021. **184**(9): p. 2362-2371. e9.
138. Yi, C., et al., *Key residues of the receptor binding motif in the spike protein of SARS-CoV-2 that interact with ACE2 and neutralizing antibodies.* Cellular & molecular immunology, 2020. **17**(6): p. 621-630.
139. Saito, A., et al., *Enhanced fusogenicity and pathogenicity of SARS-CoV-2 Delta P681R mutation.* Nature, 2022. **602**(7896): p. 300-306.
140. Liu, Y., et al., *Delta spike P681R mutation enhances SARS-CoV-2 fitness over Alpha variant.* Cell Reports, 2022: p. 110829.
141. Khan, A., et al., *Structural insights into the mechanism of RNA recognition by the N-terminal RNA-binding domain of the SARS-CoV-2 nucleocapsid phosphoprotein.* Computational and structural biotechnology journal, 2020. **18**: p. 2174-2184.
142. Woo, H.G. and M. Shah, *Omicron: A heavily mutated SARS-CoV-2 variant exhibits stronger binding to ACE2 and potently escape approved COVID-19 therapeutic antibodies.* bioRxiv, 2021.
143. Di Giacomo, S., et al., *Preliminary report on severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) Spike mutation T478K.* Journal of medical virology, 2021. **93**(9): p. 5638-5643.
144. Suratekar, R., et al., *High diversity in Delta variant across countries revealed by genome-wide analysis of SARS-CoV-2 beyond the Spike protein.* Molecular systems biology, 2022. **18**(2): p. e10673.
145. Anwar, M.Z., et al., *Coronavirus Genomes and Unique Mutations in Structural and Non-Structural Proteins in Pakistani SARS-CoV-2 Delta Variants during the Fourth Wave of the Pandemic.* Genes, 2022. **13**(3): p. 552.
146. Kannan, S.R., et al., *Infectivity of SARS-CoV-2: there is something more than D614G?* Journal of Neuroimmune Pharmacology, 2020. **15**(4): p. 574-577.
147. Adebali, O., et al., *Phylogenetic analysis of SARS-CoV-2 genomes in Turkey.* Turkish Journal of Biology, 2020. **44**(SI-1): p. 146-156.
148. Biswas, S.K. and S.R. Mudi, *Spike protein D614G and RdRp P323L: the SARS-CoV-2 mutations associated with severity of COVID-19.* Genomics & informatics, 2020. **18**(4).