Clinical and Experimental Health Sciences

Mutations in the SARS CoV2 Spike Gene and Their Reflections on the Spike Protein

Elif Caglayan[1], Kadir Turan[2]

[1] University of Health Sciences, Kartal Kosuyolu High Specialty Educational and Research Hospital, Department of Medical Microbiology, Istanbul, Türkiye.
[2] Marmara University, Faculty of Pharmacy, Department of Basic Pharmaceutical Sciences, Istanbul, Türkiye.

Correspondence Author: Kadir Turan
E-mail: kadirturan@marmara.edu.tr
Received: 12.08.2021 **Accepted:** 03.11.2021

ABSTRACT

Objective: In this study, it was aimed to determine the mutation frequency in spike *(S)* genes of SARS CoV2 from six different regions of the world, their distribution on the gene and reflections of these mutations to the S protein.

Methods: SARS CoV2 *S* gene sequences originating from Asia, Africa, Europe, South America, Oceania and North America were obtained from NCBI virus database. The sequences were analyzed with *Geneious* and *BioEdit* multiple sequence alignment programs.

Results: 865 distinct mutations on the *S* genes were detected in the virus samples. Among these, 59 variants with numbers of 10 and above in the virus population were detected. The D614G(A1841G) substitution was found to be the most common with an average of 88.6%. Furthermore, it was determined that S477N(G1430A) substitution in the viruses of Oceania differed from other regions with a rate of 86.7%. The average mutation frequency of the *S* genes from different regions was calculated as $3{,}93x10^{-5}$.

Conclusion: The significant differences among the mutation frequencies in SARS CoV2 *S* genes isolated from different regions was identified. At least five distinct amino acid substitutions with high ratios in the population were detected in the RBD domain, which is involved in the binding of the viruses to the ACE2 receptor. These substitutions are T1355G (L452R), G1430A (S477N), C1433A (T478K), G1450A (E484K) and A1501T (N501Y). Among these, the S477N is the most predominant in the population. However, the importance of these mutations needs to be demonstrated both *in silico* and experimental studies.

Keywords: SARS CoV2, Coronavirus, Covid-19, Spike protein, Mutation

1. INTRODUCTION

Coronaviruses, which are usually the cause of mild respiratory tract infections, did not attract much attention until the outbreaks of SARS CoV(1) and MERS CoV(2,3) that caused the severe acute respiratory disease in China in 2002 and Saudi Arabia in 2012. In December 2019, a new type of coronavirus, SARS CoV2, emerged in Wuhan, China Hubei province (4), spread all over the world in a very short time, and caused a global pandemic. The SARS CoV2 causing Coronavirus Disease-2019 (COVID-19) in humans have been classified in the order *Nidovirales*, family *Coronaviridae* (5,6). According to the World Health Organization (WHO), more than 168 million people were infected with this virus in one and a half years, and about 3.5 million people died due to COVID-19 (https://covid19.who.int/ access: 27.05.21). The SARS CoV2 has a single-stranded, positive-sense RNA genome with an average size of 30 kb. This genome encodes four structural proteins including spike (S), envelope (E), membrane (M), and nucleocapsid (N) proteins (7). Generally, viruses that carry the RNA genome mutate at higher frequencies than DNA viruses. The RNA viruses typically have mutation rates that range between 10^{-6} and 10^{-4}. The mutation rates of DNA viruses are about 10^{-8} to 10^{-6} substitutions per nucleotide site per cell infection (8,9). The high rates of mutation in the RNA viruses can be explained by RdRP (RNA dependent RNA polymerase) that replicates the viral genome. Unlike many DNA polymerases, the RdRP enzyme has no proofreading activity and therefore, cannot correct the errors that occurred during replication. For this reason, influenza A viruses among the RNA viruses, which cause epidemics much more commonly than the coronaviruses in humans, have a very high mutation frequency and show a wide variation (10). Due to these high mutation rates and the variation, these viruses have a wide range of hosts (11,12). Unlike other RNA viruses, the viruses classified in the order *Nidovirales*, including coronaviruses, have proofreading capabilities that are independent of RdRP enzymes (13). This proofreading is thought to be an important factor in explaining how these viruses have much larger genomes (average 30 kb) when compared to other

DOI: 10.33808/clinexphealthsci.981816

RNA viruses (14). Although the mutation frequency is lower than the other RNA viruses, mutations in the SARS CoV2 cause great concerns due to the possibility of the emergence of new variants that can be transmitted more easily from person to person and have higher mortality. One of the most important factors affecting the spreading rate of the SARS CoV2 is the spike glycoprotein, which is responsible for the attachment of the virus to the host cell receptors and its entry into the cells (15). SARS CoV2 spike protein consists of extracellular domains, a transmembrane domain (TM) and a short intracellular tail (CP). This protein is located as a homotrimeric structure on the surface of the viral envelope. The extracellular domain of the spike protein consists of S1 and S2 subunits responsible for binding to the host cell receptor and membrane fusion, respectively. The S1 subunit of the protein contains an NH_2-terminal domain (NTD) and the carboxyl terminal domain (CTD), also called the receptor binding domain (RBD) (16). The RBD of spike protein plays a crucial role in the binding of the viruses to the host cell receptor, the angiotensin converting enzyme 2 (ACE2) that is required for viral entry (17). The mutations causing amino acid substitution in the spike protein, which has a crucial role in both the attachment and entry to the cells, can significantly alter the rate of the transmission and pathogenesis of the virus.

The COVID-19 pandemic has massively accelerated the whole-genome analysis of SARS CoV2, and increased the virus genome database. In this study, the cDNAs encoding the spike proteins of the SARS CoV2 originating from Asia, Africa, Europe, South America, Oceania and North America were analyzed in terms of mutations based on the reference SARS CoV2 spike gene, which was first identified in Wuhan, China, and the amino acid substitutions caused by these mutations in the spike protein the distribution of the variations and their possible effects on virus pathogenesis are discussed.

2. METHODS

The full-length spike gene sequences of SARS CoV2 isolates from 6 different geographic regions in the world were obtained in the FASTA format from the NCBI virus database (https://www.ncbi.nlm.nih.gov/labs/virus/vssi/) on 21 April 2021. The multiple sequence alignment of the cDNAs was achieved using *Geneious* and *BioEdit* software. The spike gene cDNA of the Wuhan-Hu-1 variant (NCBI accession code: NC_045512.2/YP_009724390.1) was used as the reference sequence. The samples with uncertainty in the nucleotide sequence were excluded from the analysis. Based on the Wuhan-Hu-1 variant, the nucleotide differences and positions, and the mutation frequencies in the genes, were determined. Similarly, amino acid substitutions in the spike proteins encoded by the genes were compared with the reference protein (NCBI accession code: YP_009724390.1). The ratio of the amino acid substitutions and the positions of these changes in the spike proteins encoded by the viruses that were isolated in different geographic regions were determined. The possible effects of amino acid replacement in the spike proteins encoded by different SARS CoV2 variants on the spread of the virus in humans were evaluated.

3. RESULTS

3.1. The ratios of single nucleotide changes in SARS CoV2 Spike genes

The mutations in the SARS CoV2 spike gene that cause amino acid substitutions in the protein have the potential to affect the transmission rate and pathogenicity of the virus. In this regard, the SARS CoV2 spike genes isolated in different geographical regions of the world were aligned using multiple sequence alignment programs, and both single nucleotide changes in the genes and amino acid substitution in the spike glycoproteins were analyzed. As of 21 April 2021, the sequencing data of the 5760 SARS CoV2 spike genes from six geographical regions of the world collected in genome databases were examined in terms of nucleotide changes. The mutation rates in the spike genes of SARS CoV2 isolates in different regions were determined (Table 1). Among the SARS CoV2 variants isolated in Africa, South America, or Europe until 21 April 2021, all samples except those with errors in the nucleotide sequences were evaluated. 2265, 9900 and 75630 samples isolated in Asia, Oceania and North America were defined in the SARS CoV2 genome database collected between 13 January 2020-21 April 2021, respectively. A total of 1000 samples within the SARS CoV2 isolates from Asia were evaluated in proportion to the sample number collected over 16 months. No data was found in Oceania between January and April 2020. In this region, a total of 1150 samples determined in proportion to the data collected between May 2020-April 2021 were analyzed. A total of 149 genome sequences of the SARS CoV2 isolates in North America were found in January-March 2020. In this region, 1649 samples, including the 149-genome sequence, and randomly selected 100 samples among the collected sequences in each month were evaluated. Regardless of the number of variants in the virus populations, the number of distinct nucleotide changes in the spike genes isolated from each region, was determined, and the mutation frequencies in the SARS CoV2 spike gene were calculated relative to the total number of nucleotides sequenced (Table 1).

Table 1. *The number of SARS CoV1 spike genes analyzed from different geographical regions in the world, the mutation frequencies and the number of substitutions in the genes.*

Geographic Regions	Number of Samples	Mutation Frequencies	Number of Nucleotide Substitution
Africa	721	$8{,}27x10^{-5}$	228
N.America	1649	$7{,}20x10^{-5}$	454
Europe	800	$5{,}56x10^{-5}$	170
Asia	1000	$5{,}94x10^{-5}$	227
S.America	440	$4{,}10x10^{-5}$	69
Oceania	1150	$2{,}37x10^{-5}$	104
Average (All Geographic Regions)	**5760**	**$3{,}93x10^{-5}$**	**865**

Table 2. *The average distribution rates of spike variants with numbers 10 or more in virus populations of different geographical regions in the world.*

Nucleotide Changes	Amino Acid Changes	The rates of amino acid substitutions in The SARS CoV2 Spike proteins from different geographic regions (%)							Phenotypic Effect
		Africa	Asia	Europe	North America	South America	Oceania	Total	
C13T	L5F	-	1,4	-	2,12	-	-	0,85	Missense
C35T	S12F	5,55	-	-	-	-	-	0,69	Missense
G38T	S13I		-	-	1,52	-	-	0,43	Missense
C52T	L18F	-	-	-	1,09	-	-	0,31	Missense
C59T	T20N	-	-	-	0,91	-	-	0,26	Missense
C76T	P26S	-	-	-	1,15	-	-	0,33	Missense
G162T	L54F	-	7	-	-	-	-	1,22	Missense
C249A	V83V	-	-	-	0,85	-	-	0,24	Silent
T363C	N121N	-	-	-	-	-	1,65	0,33	Silent
G412T	D138Y	-	-	-	1,03	-	-	0,30	Missense
G456T	W152C	-	-	-	1,33	-	-	0,38	Missense
G570T	R190S	-	-	-	0,73	-	-	0,21	Missense
T600C	Y200Y	-	-	-	1,33	-	-	0,38	Silent
C665T	A222V	2,08	-	2,38	-	-	-	0,59	Missense
A693T	I231I	-	-	-	2,79	-	-	0,80	Silent
A758G	D253G	-	-	-	2,06	-	-	0,59	Missense
C882T	D294D	-	22,5	-	-	-	-	3,91	Silent
G906T	T302T	2,22	1,6	1,63	-	-	-	0,78	Silent
C915T	S305S	-	1,7	-	-	-	-	0,30	Silent
C918T	F306F	-	-	-	-	-	17	3,40	Silent
C1062T	N354N	5,55	-	-	-	-	-	0,69	Silent
C1151T	P384L	1,94	-	-	-	-	-	0,24	Missense
A1250C	K417T	-	-	-	0,73	-	-	0,21	Missense
T1355G	L452R	1,94	-	-	1,58	-	-	0,69	Missense
A1358T	Y453F	-	-	1,63	-	-	-	0,23	Missense
G1430A	S477N	3,88	-	-	-	-	86,7	17,80	Missense
C1433A	T478K	-	-	-	1,7	-	-	0,49	Missense
G1450A	E484K	-	-	-	0,91	-	-	0,26	Missense
A1501T	N501Y	-	-	-	0,85	-	-	0,24	Missense
C1565T	A522V	-	1,5	-	-	-	-	0,26	Missense
C1629T	F543F	-	-	-	0,79	-	-	0,23	Silent
G1839A	Q613Q	-	-	-	-	-	86,7	17,31	Silent
A1841G	D614G	90,6	89,7	83,5	82,8	95,2	95,6	88,56	Missense
C1895A	T632N	-	-	-	-	-	1,39	0,28	Missense
C1963T	H655Y	-	-	-	0,91	-	-	0,26	Missense
T1986C	C662C	-	-	-	-	-	1,48	0,30	Silent
A2030C	Q677P	-	-	-	0,67	-	-	0,19	Missense
G2031T	Q677H	22,5	-	-	2,12	-	-	3,42	Missense
C2042G	P681R	6,8	-	-	-	-	-	0,85	Missense
C2042A	P681H	-	-	-	3,64	-	-	1,04	Missense
C2093T	S698L	-	-	-	-	-	0,96	0,19	Missense
G2101A	A701T	-	1,3	-	-	-	-	0,23	Missense
T2133A	S711S	-	-	1,5	-	-	-	0,21	Silent
C2169T	T723T	30,2	-	5,38	-	11,4	-	5,40	Silent
A2194G	T732A	-	-	-	3,09	-	-	0,89	Missense
C2367T	Y789Y	-	1,8	-	-	-	-	0,31	Silent
C2435T	P812L	-	3,3	-	-	-	-	0,57	Missense
C2472T	N824N	-	-	-	1,52	-	1,04	0,64	Silent
T2514C	G838G	-	-	-	1,58	-	-	0,45	Silent
G2515T	D839Y	-	-	2,13	-	-	-	0,30	Missense
T2787C	S929S	-	-	-	1,09	-	-	0,31	Silent
C3080T	T1027I	-	-	-	1,15	-	-	0,33	Missense
A3132T	G1044G	-	-	-	0,97	-	-	0,28	Silent
T3249A	H1083Q	-	1,7	-	-	-	-	0,30	Missense
C3342T	I1114I	-	-	-	1,64	-	-	0,47	Silent
G3371T	G1124V	-	-	-	-	-	3,39	0,68	Missense
G3526T	V1176F	-	-	-	1,15	-	-	0,33	Missense
C3782T	S1261F	-	-	-	-	4,09	-	0,31	Missense
G3790T	V1264L	5,96	-	-	-	-	-	0,75	Missense

The highest mutation frequency was defined in the spike genes with a value of $8.27x10^{-5}$ in the SARS CoV2 isolates from Africa when compared to other regions. The SARS CoV2 population from North America followed the African population with a mutation rate of $7.20x10^{-5}$. The lowest mutation frequency ($2.37x10^{-5}$) in the spike genes among the SARS CoV2 populations evaluated as defined in the viruses from Oceania. This mutation ratio was followed by mutations in the spike gene of the virus populations originating from South America with $4.10x10^{-5}$, Europe with $5.63x10^{-5}$ and Asia with $5.94x10^{-5}$. The average mutation frequency in the spike genes of the 5760 SARS CoV2 samples collected from different geographic regions was calculated as $3.93x10^{-5}$.

3.2. The amino acid substitutions and distribution in the SARS CoV2 Spike protein due to mutations in the gene

We performed multiple alignments of the amino acid sequence of the spike proteins encoded by the genes collected from different geographic regions in the world, defined the amino acid substitutions and compared them with each other. The nucleotide changes of the variants with numbers 10 or more in different virus populations and the corresponding amino acid substitution rates in the spike proteins are given in Table 1. Among the spike genes analyzed in the virus populations, the total distinct 59-nucleotide changes at different positions were detected in the variants with numbers 10 and above. It was defined that 20 of these nucleotide changes were silent mutations and did not cause amino acid substitution in the spike protein. Among all the virus populations, the most common amino acid substitution in the spike proteins is the D614G substitution comparing the reference Wuhan-HU1 (Spike cDNA: NC_045512.2 / Spike Protein: YP_009724390.1) virus. The ratio of variants having this substitution in the virus populations from different geographic regions varies between 82.8% and 95.6%. The average of the D614G substitution in spike proteins in all the viruses analyzed is about 89.6%. It draws attention to the variants having G1430A (S477N) substitution in the spike gene as high as 86.7% in the Oceania virus population. This mutation was defined with a rate of 3.88% only in the virus population originating from Asia, and was not found in the viruses from other regions. Another mutation with a high rate of 86.7% in the Oceania virus population is the G1839A (Q613Q), which is a silent mutation. Both of these mutations

likely emerged in the early stages of the pandemic. The most common mutations on the spike gene in the African-origin virus population are the G2031T (Q677H) and C2169T (T723T) with 22.5% and 30.2% respectively. The ratio of the C2169T silent mutation is 11.4% in the South American virus population, while it is 5.38% in Europe. This mutation was not defined in the virus populations from the other three geographic regions. The C882T (D294D) silent mutation has the second highest mutation rate in the Asian population with 22,4%. This mutation has not been found in other virus populations in the world.

In SARS CoV2 viruses, as a result of missense mutations in virus genes, especially caused by the viral RdRP, a large number of amino acid substitutions occur in the viral protein. The viral surface antigens such as the spike and influenza HA proteins have the most flexibility to change with missense mutations. These changes offer a wider range of host organisms to RNA viruses for replication. Therefore, the amino acid substitutions that occurred in the SARS CoV2 surface antigens have greater importance for the spread of the viruses and increasing their host diversity. The amino acid substitutions resulting from the missense mutations and their distributions in the spike protein consisting of 1273 amino acid residues are given in Figures 1-7. Figure 1 shows the distinct amino acid substitutions, in total 5760 spike proteins of SARS CoV2 viruses from different geographic regions analyzed in this study. The distribution of the amino acid substitutions in the spike protein caused by the missense mutations detected at least once in all virus populations is shown in Figure 1A. Although the homogeneous amino acid substitutions covering almost all parts of the spike protein is observed, it appears that the variations are slightly more intense in some regions. For instance, more amino acid substitutions are seen in regions spanning 1-100., 225-300., and 650-725. amino acid residues are found on the amino terminal half of the spike proteins. There are also some intense variations in regions corresponding to the 1150-1273 amino acid residues involving the HR2, TM and CP domains at the carboxyl-terminal end of the protein. On the other hand, most of the spike gene variants seen in the SARS CoV2 populations from different geographical regions consist of one or two samples. The number of variants with a rate of 0.25% and above in the virus populations and their positions on the spike protein is given in Figure 1B. Among the functional regions on the spike protein, the intensity of the amino acid substitutions in the amino terminal end of the spike including signal peptide draws attention. In this region, C13T (L5F), C35T (S12F), G38T (S13I), C52T (L18F), C59T (T20N) and C76T (P26S) substitutions with a ratio of 0.25% and above were detected.

At least five distinct amino acid substitutions with high ratios in the population were detected in the RBD domain, which is involved in the binding of the SARS CoV2 to the ACE2 receptor on the host cells. These are T1355G (L452R), G1430A (S477N), C1433A (T478K), G1450A (E484K) and A1501T (N501Y) substitutions. Among these, the S477N is the most predominant in the population. The mutations that cause the amino acid substitutions in the RBD of the spike protein may have a negative effect on the replication and spread of the viruses as well as facilitate the virus replication. The prevalence of G1430A (S477N) substitution in the virus population suggests that this mutation positively affects virus replication.

The amino acid substitutions in the spike proteins of SARS CoV2 were also evaluated regionally within themselves. In the viruses isolated from Asia, there is no amino acid substitution with a ratio of 0.5% or more in the spike proteins, affecting the RBD located in the amino terminal half, and the CP, TM, HR1 and HR2 domains at the carboxyl terminal of the protein (Figure 2). In contrast, four distinct changes in the RBD are seen in the spike proteins of viruses originating in Africa. Three amino acid substitutions with a ratio of over 0.5% in the SP located at the amino terminal end of the spike proteins of the viruses from Africa were defined (Figure 3). The number of SARS CoV2 spike gene sequencing data from South America was less than from the other regions. In the virus samples examined, there are no amino acid substitutions in the RBD of spike protein with a significant ratio comparing to the reference protein (Figure 4). Three amino acid substitutions with a ratio of more than 0.5% in the RBD of the proteins were determined within the SARS CoV2 isolates from Europe. In this virus population, two amino acid substitutions in the SP domain and one in the CP domain were determined in the spike proteins (Figure 5).

It was observed that the mutation resulting in the replacement of aspartic acid (D) to glycine (G) at the 614th position (D614G) of the spike protein had the highest ratio in all the virus populations. In addition to A1841G (D614G) missense mutation, a high ratio of the S477N substitution was detected in the virus isolates from Oceania. This amino acid substitution occurred in the RBD of the protein and was detected in 997 of the 1150 samples examined (Figure 6). However, the S477N substitution in the spike protein was not found within the viruses isolated in Europe, America, and Africa.

Most of the genome sequence data for the SARS CoV2 originated from North America. As of April 21, 2021, the SARS CoV2 genome sequencing data was determined as 75630. Here, the spike gene/protein belonging to 1649 samples of this data was evaluated. Four different amino acid substitutions with a ratio of 0.5% and more were detected in both the SP and RBD of the spike protein originating from North America (Figure 7).

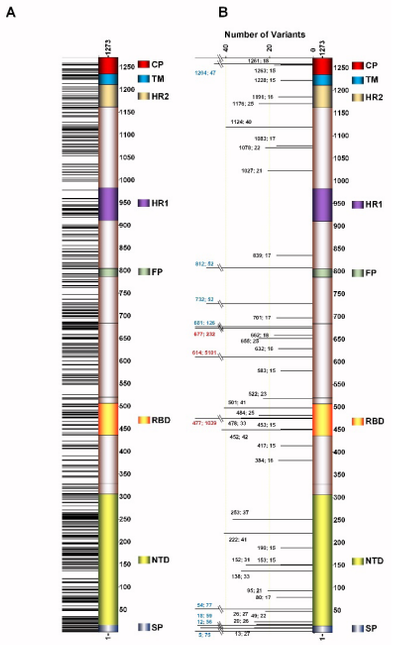

Figure 1. *Distribution of the amino acid substitutions on the spike proteins of SARS CoV2 isolates in the six geographic regions of the world and the number of distinct variants in the virus population. A. Distribution of amino acid substitutions on the spike protein, which is determined in all virus samples regardless of their number in the population. B. The number of variants with a ratio of higher than 0.25% (15 and more) in the virus population and the positions of amino acid substitutions. CP: cytoplasmic domain, TM: transmembrane domain; HR1 and HR2: heptad repeat, FP: fusion peptide RBD: receptor-binding domain NTD: N-terminal domain; SP: signal peptide.*

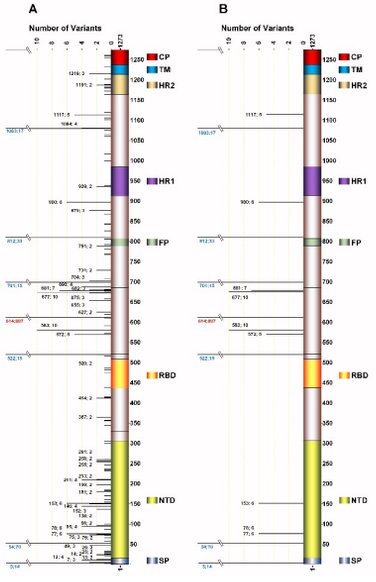

Figure 2. *Distribution of the amino acid substitutions on the spike proteins of SARS CoV2 isolates in Asia and the number of distinct variants in the virus population. A. Distribution of amino acid substitutions on the spike protein, which is determined in all virus samples regardless of their number in the population. B. The number of variants with a ratio of ≥ 0.5% (5 and more) in the virus population and the positions of amino acid substitutions.*

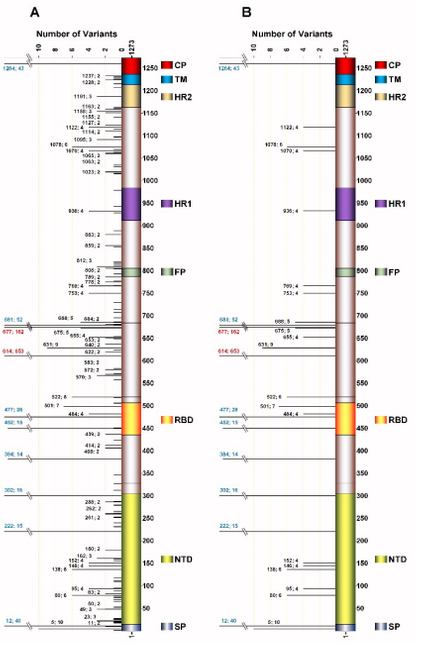

Figure 3. *Distribution of the amino acid substitutions on the spike proteins of SARS CoV2 isolates in Africa and the number of distinct variants in the virus population. A. Distribution of amino acid substitutions on the spike protein, which is determined in all virus samples regardless of their number in the population. B. The number of variants with a ratio of ≥ 0.5% (4 and more) in the virus population and the positions of amino acid substitutions.*

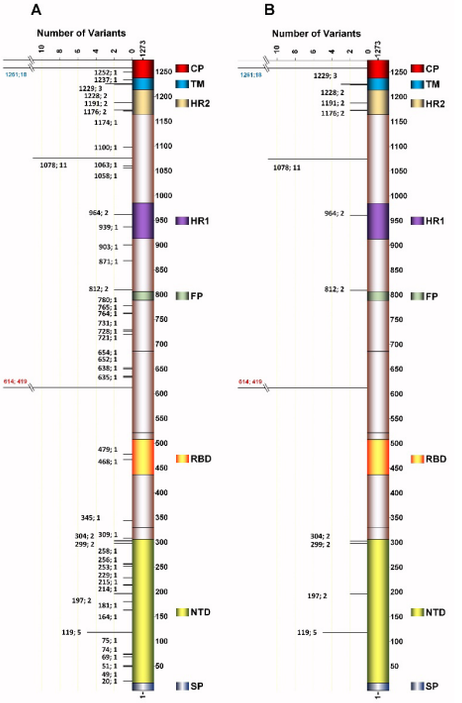

Figure 4. *Distribution of the amino acid substitutions on the spike proteins of SARS CoV2 isolates in South America and the number of distinct variants in the virus population. A. Distribution of amino acid substitutions on the spike protein, which is determined in all virus samples regardless of their number in the population. B. The number of variants with a ratio of ≥ 0.5% (2 and more) in the virus population and the positions of amino acid substitutions.*

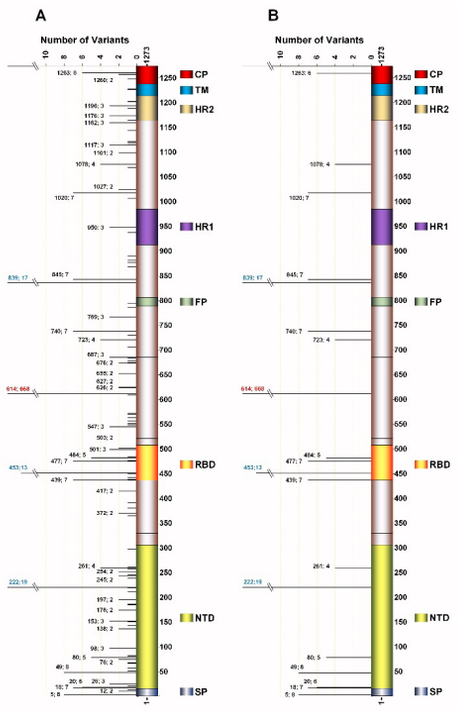

Figure 5. *Distribution of the amino acid substitutions on the spike proteins of SARS CoV2 isolates in Europe and the number of distinct variants in the virus population. A. Distribution of amino acid substitutions on the spike protein, which is determined in all virus samples regardless of their number in the population. B. The number of variants with a ratio of ≥ 0.5% (4 and more) in the virus population and the positions of amino acid substitutions.*

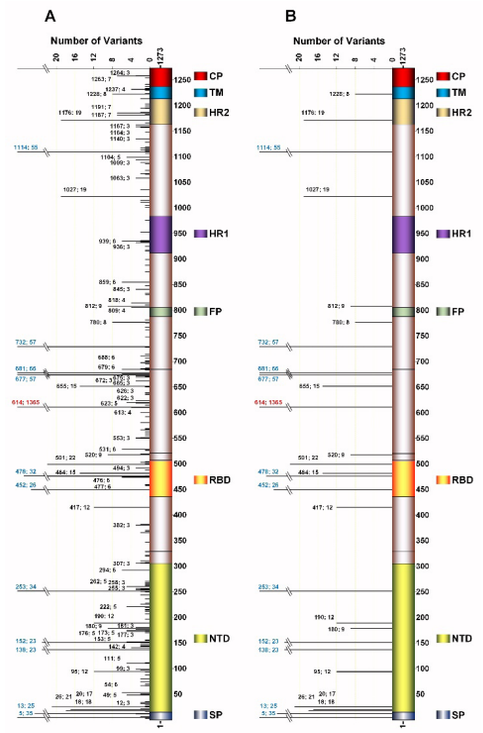

Figure 7. *Distribution of the amino acid substitutions on the spike proteins of SARS CoV2 isolates in North America and the number of distinct variants in the virus population. A. Distribution of amino acid substitutions on the spike protein, which is determined in all virus samples regardless of their number in the population. B. The number of variants with a ratio of ≥ 0.5% (8 and more) in the virus population and the positions of amino acid substitutions.*

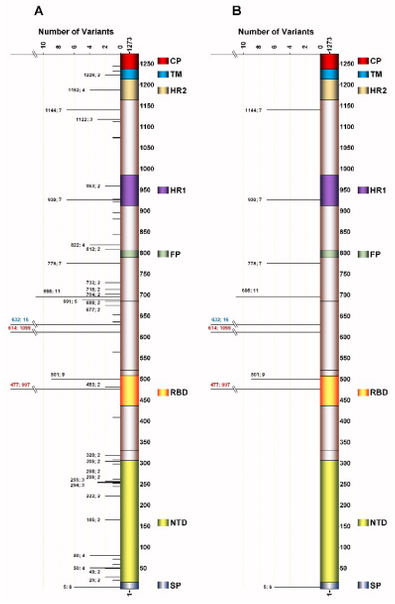

Figure 6. *Distribution of the amino acid substitutions on the spike proteins of SARS CoV2 isolates in Oceania and the number of distinct variants in the virus population. A. Distribution of amino acid substitutions on the spike protein, which is determined in all virus samples regardless of their number in the population. B. The number of variants with a ratio of ≥ 0.5% (6 and more) in the virus population and the positions of amino acid substitutions.*

4. DISCUSSION

The spike protein, which allows the SARS CoV2 to attach to the ACE2 receptor, is of great importance for the transmission of the virus and immune response in humans. In this study, the nucleotide/amino acid substitutions seen in the spike gene/protein were evaluated. It was found the significant differences between the mutation frequencies in the virus populations isolated from different geographical regions suggesting that the reason may be caused by the deficiencies in the experimental data rather than the ability of the virus to mutate. Considering the possible changes in the nucleotide sequences that occur during genome analyses, it can be asserted that the mutation frequency in the SARS CoV2 spike gene, which is largely due to the errors of the viral RdRP enzyme, is lower than $3.93x10^{-5}$. It can be argued that much lower rates of nucleotide changes occur in the genes of the SARS CoV2 virus compared to the $1.8x10^{-4}$ – $2.5x10^{-4}$ mutation frequency of the influenza A viruses carrying a single-stranded segmented RNA genome (18). It is suggested that the low mutation ratios in the SARS CoV2 genes may be the result of the viral repair mechanism that is independent of the RdRP enzyme (19,20). The mutations in the SARS CoV2 spike gene result in a lower rate of phenotypic changes in the spike protein due to silent mutations. Therefore, it would not be wrong to say that the immune response to the SARS CoV2 virus in humans will be longer than expected.

5. CONCLUSION

The mutation rate of the SARS CoV2 is lower compared to the other RNA viruses with high mutation frequency, such as the influenza A virus. However, the mutations in the spike genes causing the amino acid changes in the RBD of the protein can affect the pathogenesis and spreading rate of the SARS CoV2. At least six mutations causing amino acid substitutions in the RBD of spike protein with a ratio of over 0.5% were detected in the SARS CoV2 isolates from different geographic regions. These mutations resulted in the L452R, Y453F, S477N, T478K, E484K, and N501Y substitutions in the RBD of the proteins. The S477N substitution in the spike protein in the Oceania virus population draws attention with a high ratio of 86,7%. Whether these amino acid substitutions in the RBD of the spike protein are important in terms of their effect on the virus adsorption to the host cell via ACE2 receptor, remains the subject of both *in silico* and experimental research. However, the results will be useful in predicting the spread of the virus in the human population, the persistence of acquired immunity against these viruses, and the protective effects of vaccines.

Conflict of interest

All the authors declare no conflict of interest.

Acknowledgements

This work was supported by the Health Institutes of Turkey (TUSEB) (Grant No: 8608)

REFERENCES

[1] Rota PA, Oberste MS, Monroe SS, Nix WA, Campagnoli R, Icenogle JP, Penaranda S, Bankamp B, Maher K, Chen MH, Tong S, Tamin A, Lowe L, Frace M, DeRisi JL, Chen Q, Wang D, Erdman DD, Peret TC, Burns C, Ksiazek TG, Rollin PE, Sanchez A, Liffick S, Holloway B, Limor J, McCaustland K, Olsen-Rasmussen M, Fouchier R, Gunther S, Osterhaus AD, Drosten C, Pallansch MA, Anderson LJ, Bellini WJ. Characterization of a novel coronavirus associated with severe acute respiratory syndrome. Science 2003; 300: 1394-1399.

[2] Azhar EI, El-Kafrawy SA, Farraj SA, Hassan AM, Al-Saeed MS, Hashem AM, Madani TA. Evidence for camel-to-human transmission of MERS coronavirus. N Engl J Med 2014; 370: 2499-505.

[3] Zaki AM, van Boheemen S, Bestebroer TM, Osterhaus AD, Fouchier RA. Isolation of a novel coronavirus from a man with pneumonia in Saudi Arabia. N Engl J Med 2012; 367: 1814-1820.

[4] Zhang X, Tan Y, Ling Y, Lu G, Liu F, Yi Z, Jia X, Wu M, Shi B, Xu S, Chen J, Wang W, Chen B, Jiang L, Yu S, Lu J, Wang J, Xu M, Yuan Z, Zhang Q, Zhang X, Zhao G, Wang S, Chen S, Lu H. Viral and host factors related to the clinical outcome of COVID-19. Nature 2020; 583: 437-440.

[5] Bertram S, Dijkman R, Habjan M, Heurich A, Gierer S, Glowacka I, Welsch K, Winkler M, Schneider H, Hofmann-Winkler H, Thiel V, Pohlmann S. TMPRSS2 activates the human coronavirus 229E for cathepsin-independent host cell entry and is expressed in viral target cells in the respiratory epithelium. J Virol 2013; 87: 6150-6160.

[6] Fehr AR, Perlman S. Coronaviruses: an overview of their replication and pathogenesis. Methods Mol Biol 2015; 1282: 1-23.

[7] Chan JF, Kok KH, Zhu Z, Chu H, To KK, Yuan S, Yuen KY. Genomic characterization of the 2019 novel human-pathogenic coronavirus isolated from a patient with atypical pneumonia after visiting Wuhan. Emerg Microbes Infect 2020; 9: 221-236.

[8] Duffy S. Why are RNA virus mutation rates so damn high? PLoS Biol 2018; 16: e3000003.

[9] Lauring AS, Andino R. Quasispecies theory and the behavior of RNA viruses. PLoS Pathog 2010; 6: e1001005.

[10] Boni MF, Gog JR, Andreasen V, Feldman MW. Epidemic dynamics and antigenic evolution in a single season of influenza A. Proc Biol Sci 2006; 273: 1307-1316.

[11] Taubenberger JK, Kash JC. Influenza virus evolution, host adaptation, and pandemic formation. Cell Host Microbe 2010; 7: 440-451.

[12] O'Donnell CD, Subbarao K. The contribution of animal models to the understanding of the host range and virulence of influenza A viruses. Microbes Infect 2011; 13: 502-515.

[13] Minskaia E, Hertzig T, Gorbalenya AE, Campanacci V, Cambillau C, Canard B, Ziebuhr J. Discovery of an RNA virus 3'->5' exoribonuclease that is critically involved in coronavirus RNA synthesis. Proc Natl Acad Sci U S A 2006; 103: 5108-5113.

[14] Gorbalenya AE, Enjuanes L, Ziebuhr J, Snijder EJ. Nidovirales: evolving the largest RNA virus genome. Virus Res 2006; 117: 17-37.

[15] Walls AC, Park YJ, Tortorici MA, Wall A, McGuire AT, Veesler D. Structure, Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein. Cell 2020; 181: 281-92 e6.

[16] Xia S, Zhu Y, Liu M, Lan Q, Xu W, Wu Y, Ying T, Liu S, Shi Z, Jiang S, Lu L. Fusion mechanism of 2019-nCoV and fusion inhibitors targeting HR1 domain in spike protein. Cell Mol Immunol 2020; 17: 765-767.

[17] Letko M, Marzi A, Munster V. Functional assessment of cell entry and receptor usage for SARS-CoV-2 and other lineage B betacoronaviruses. Nat Microbiol 2020; 5: 562-569.

[18] Pauly MD, Procario MC, Lauring AS. A novel twelve class fluctuation test reveals higher than expected mutation rates for influenza A viruses. Elife 2017; 6.

[19] Ogando NS, Ferron F, Decroly E, Canard B, Posthuma CC, Snijder EJ. The curious case of the nidovirus exoribonuclease: Its role in RNA synthesis and replication fidelity. Front Microbiol 2019; 10: 1813.

[20] Romano M, Ruggiero A, Squeglia F, Maga G, Berisio R. A Structural view of SARS-CoV-2 RNA replication machinery: RNA synthesis, proofreading and final capping. Cells 2020; 9.

How to cite this article: Caglayan E, Turan K. Mutations in the SARS CoV2 Spike Gene and Their Reflections on the Spike Protein. Clin Exp Health Sci 2022; 12: 472-478. DOI: 10.33808/clinexphealthsci.981816