Journal of Advanced Zoology
ISSN: 0253-7214
Volume 43 Issue 01 Year 2022 Page 199:219

Chrono-Geographical Analysis Of Sars-Cov2 Genome Wide Mutations

Manprit Kaur[1], Arvind Sharma[2], Ashwani Kumar[3]*

[1, 3]*Department of Biotechnology, Chaudhary Bansi Lal University, Bhiwani, Haryana, India;
[2]National Institute of Malaria Research, New Delhi

***Corresponding Author:** - Ashwani Kumar
*Department of Biotechnology, Chaudhary Bansi Lal University, Bhiwani, Haryana, India;

Abstract

Since the reporting of the first cases of coronavirus in China and the publication of the first sequence of SARS-CoV-2 in December 2019, the virus has undergone numerous mutations. In the present study, we gathered 1,404 SARS-CoV-2 complete genomes from the NCBI and detected the mutations via GISAID. We analysed and annotated all SARS-CoV-2 mutations compared with the reference Wuhan genome NC_045512.2. The $S1^B$ domain (333-527) of the spike protein was found the highest mutating region in the entire genome whereas NS6 protein was found the lowest mutating region. Interestingly, no any mutation was detected from NSP11 protein. The D614G from spike protein, T81I from NSP2, A890D from NSP3, L37F from NSP6, P323L from NSP12, Q57H from NS3 and Y73C from NS8 were identified in maximum numbers of SARS-CoV-2 populations from all six continents. Many co-occurring mutations were detected in spike proteins, N proteins and NSP12 proteins. The deletions were only found in S, N, NSP1, NSP2, NSP3, NSP4, NSP6, NS7a and NS7b proteins. The co-occurring deletions were identified only in N, NSP1 and NSP6 proteins. A few insertion mutations were identified in spike proteins and NSP6 proteins. But the high prevalence of stop-codon mutations was detected in spike, NSP6, NS7b and NS8 proteins. Our results provide an in-depth analysis of SARS-COV-2 whole genome which we believe, can shed light in the understanding of SARS-COV-2 pathogenesis and mutation pattern which can aid in the development of prevention methods as well as future research into the pathogenesis of SARS-CoV-2 and therapeutic development.

Keywords: SARS-CoV-2, COVID-19, mutations, D614G.

1. Introduction

Coronaviruses (CoVs) are enveloped viruses having a positive sense, single-stranded RNA genome. CoVs have the largest genomes for RNA viruses with size ranging from 26 to 32 kilobases (kb) in length. They belong to Coronaviridae family and order *Nidovirales*. Coronavirinae and Torovirinae subfamilies are divided from the family Coronaviridae. The subfamily Coronavirinae is further divided into four genera: Alpha, Beta, Gamma and Deltacoronavirus (1). The α coronavirus and β coronavirus mainly infect mammalians while γ coronavirus, and δ coronavirus affect most of the bird species (2). Previously six coronaviruses were identified that infect human population, namely Human Coronavirus-229E and Human Coronavirus-NL63 belonging to

α coronaviruses and Human Coronavirus-HKU1, Human Coronavirus-OC43, Severe Acute Respiratory Syndrome (SARS-CoV) and Middle East Respiratory Syndrome (MERS-CoV) belonging to β coronaviruses (3). In March, 2003 an outbreak was reported in China and spread rapidly in 29 countries, affecting more than 8000 people, causing many deaths due to respiratory diseases which was named as SARS (Severe acute respiratory syndrome) by the World Health Organization (WHO). A similar outbreak was caused by Middle East Respiratory Syndrome Coronavirus (MERS-CoV) in September, 2012 in Arabian Peninsula and widespread mainly in the Middle East, affected above 2000 people (4, 5). After the emergence of SARS, the MERS was the second coronavirus resulting in a major global public health crisis. These two outbreaks demonstrated the high transmissibility and pathogenicity of emerging coronaviruses (CoVs). Most human coronaviruses commonly cause relatively mild to moderate respiratory disease, however two coronaviruses, SARS- CoV (6) and MERS-CoV (7) can cause severe illness and death.

In December 2019, an outbreak of fatal pneumonia, the Coronavirus Disease 2019 (COVID- 19), caused by a novel Betacoronavirus, Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) was reported from the Wuhan City of China's Hubei Province after SARS- CoV and MERS-CoV. The WHO had declared COVID-19 a global public health emergency on 30th January and an epidemic on 11th March 2020. In short span of time, the disease was spread to most parts of the world (8), and has already affected over 346 million people worldwide, resulting in 5.5 million deaths up to 23 Jan, 2022) (9). In comparison to previously reported coronaviruses, SARS-CoV-2 is considerably more infectious and hazardous. As the epidemic progressed, human-to-human transmission by droplets and fomites became a major source of infection. It is a novel infection which greatly affects our respiratory system with other symptoms such as dry cough, elevated body temperature, fatigue, headache, diarrhea, sore throat and pneumonia (10). The risk of developing symptoms depends on the individual's overall health, level of transmission, and pathogenicity of the strain. Clearly, the severity of these infections and the lack of efficacious and licensed treatments for CoV infections strengthen the need for a detailed and complete understanding of coronaviral molecular biology, with a specific emphasis on their structural proteins.

SARS-CoV-2 genome is made up of a positive-sense, single-stranded enveloped RNA, about 30 kb in size. The entire genome of SARS-CoV-2 showed 80% homology with genome of SARS-CoV and 96% identity to the BatCoV-RaTG13 (11). At the 5'UTR (terminal region), more than two-thirds of the genome comprises ORF1ab that encodes sixteen non-structural proteins (NSP1-NSP16). NSPs are mostly enzymes or functional proteins that play a role in viral replication and methylation and may induce host responses to infection. These genes are encoded in several groups, namely ORF1a (NSP1- 11), ORF1b (NSP12-16), ORF3a, ORF6, ORF7a, ORF7b, ORF8 and ORF10. At the 3'UTR, one third consists of genes that encode structural proteins (spike glycoprotein (S), envelope protein (E), membrane protein (M), and nucleocapsid protein (N)) (12).

The coronaviruses spike (S) protein is a trimeric glycoprotein belonging to class I fusion proteins with 1273 amino acids length. The spike protein of SARS-CoV-2 shares 76-78% homology with that of SARS-CoV. S-protein comprises 3 domains: a large ectodomain, a single-pass transmembrane anchor, and a short intracellular tail. The ectodomain is made up of two subunits: subunit S1 and subunit S2. S1 subunit itself contains an amino-terminus domain of 14–305 residues and a receptor-binding domain (RBD) consisting of 319– 541 amino-acid residues. This RBD interacts with receptor angiotensin converting enzyme 2 (ACE2) present on host cell. On the other side, the S2 subunit possesses the fusion peptide (FP) (788–806 residues), heptapeptide repeat sequence 1 (HR1) (912–984 residues), HR2 (1163–1213 residues), transmembrane (TM) domain (1213–1237 residues), and cytoplasmic domain (1237–1273 residues) (13, 14). S2 subunit plays a prominent role in fusion of SARA CoV-2 with target cells. Interestingly, the SARS-CoV-2 S protein is 10- to 20 times more potent for host ACE2 receptor as compared to S protein of SARS-CoV (15). The homology between sequences of SARS-CoV-2 and SARS-CoV spike proteins demonstrates their binding affinity to the same ACE2 receptor in the host cell (16). Due to its presence at the top of the surface and its major role in the host cell infection, the spike protein becomes an attractive therapeutic target for the antibodies that play role in spike protein and ACET2 receptor interaction. Transmembrane Serine Protease 2 (TMPRSS2) is a serine protease that cleaves spike proteins at S1/S2 and S2 sites and activates it (17). The activated S-protein now plays a key role in virus and host membrane fusion (18). Recent research has shown that SARS-CoV-2 requires both ACE2 receptor and serine protease TMPRSS2 for protein priming in order to penetrate into the host cell (17).

Unlike the S protein, structural information on the remaining three structural proteins (E, M, and N) is limited. E-protein is a smallest structural transmembrane protein which regulates ion channel activity and therefore facilitates the various processes of virus life cycle such as virion assembly, budding, envelope formation, pathogenesis, and release of the virus (19). The E protein may not be responsible for viral replication, but plays major role in pathogenesis. Membrane (M) protein is most abundantly found structural protein in the viral

genome. It has been considered that it gives shape to the virus. Recent research suggests that M protein exists in the dimerization form and it can acquire two different conformations. These conformations help in membrane bending which promotes the binding of M protein with nucleocapsid (N) protein (20). The M protein consists of a small N-terminal glycosylated ectodomain with three transmembrane domains (TM) and a large C-terminal CT domain (21). This protein performs a major role in viral assembly. Moreover, the protein also promotes apoptosis (22).

The nucleocapsid (N) protein of SARS-CoV-2 is an alkaline protein, made of 422 amino-acid residues with size of 46-kDa. A short lysine-rich region is present in this protein which functions as nuclear localization signal (23). The N protein can be divided into five regions; a predicted intrinsically disordered N-terminal arm (1–40 aa), amino-terminal domain (NTD, e.g. an RNA binding domain, 41–186 aa), a predicted disordered central linker (187–257 aa), carboxy-terminal domain (CTD, e.g. a dimerization domain, 258–361 aa) (24). Among all these regions, the two N-terminal and C-terminal domains can tie up with SARS-CoV-2 RNA and forms the long, flexible, helical virion nucleocapsid. It is considered to be a versatile protein as needed for SARS-CoV-2 replication, improves viral transcription efficiency and viral pathogenicity and also provoke the host immune responses towards viral infectivity (25). The synthesis of virion nucleocapsid composed of RNA and N protein is always occurred within the cytoplasm, while the translation of remaining structural proteins (S, M, and E proteins) is occurred in the ER and the resulting vesicles moves to the Golgi apparatus for further processing (26). This multifunctional protein has been also used in the formation of specific antibodies in natural infection as it possesses antigenic action (27).

The genomes of RNA viruses usually exhibit greater mutation rates as compared to DNA viruses. These mutations are considered as a natural by-product arising during viral replication (28). Due to high mutations, these RNA viruses possess an adaptive property to a new environment. Therefore, they affect host tropism and the threats are constant and prolonged (29, 30, 31). The mutations arising at the time of viral replication also influence the viral infectivity, transmission ability, and virulence. Novel mutations are rapidly detected in the whole-genome of SARS-CoV-2 but only selected mutations greater affect the viral infectivity. Different strains of SARS-CoV-2 have been detected from various parts of the globe, for instance Alpha (B.1.1.7) from United Kingdom; Beta (B.1.351) from South Africa; and Gamma (P.1) from northern Brazil; Kappa (B.1.617.1) from India, Delta (B.1.617.2) from India, and Omicron (B.1.1.529) from South Africa. These different strains are called variants. A genetic variant is a new version of the reference strain that has attained one or more novel mutations which serve as the founder for further genetic diversity and evolution. The Alpha and Beta variants are associated with many mutations in spike protein along with D614G which is a worldwide prevalent mutation (32). N501Y mutation of RBD is shared by both the variants (33). E484K is also the mutation of RBD which is shared by all strains of Beta variant but few strains of Alpha variant (32). N501Y and E484K mutations are also associated with Gamma variant in addition to K417T mutation. The L452R, E484Q and P681R mutations shared by Kappa variant are associated with enhanced the hACE2 receptor-binding affinity (34). Delta variant harbours T478K mutation along with two L452R and P681R mutations (35). Most of these mutations greatly affect the host immune signals that are pointed toward the receptor-ligand attachment site. Among all these mutations, D614G, is the most frequently occurring mutation shared by more than 99% of variants since Jan, 2020 (34, 36).

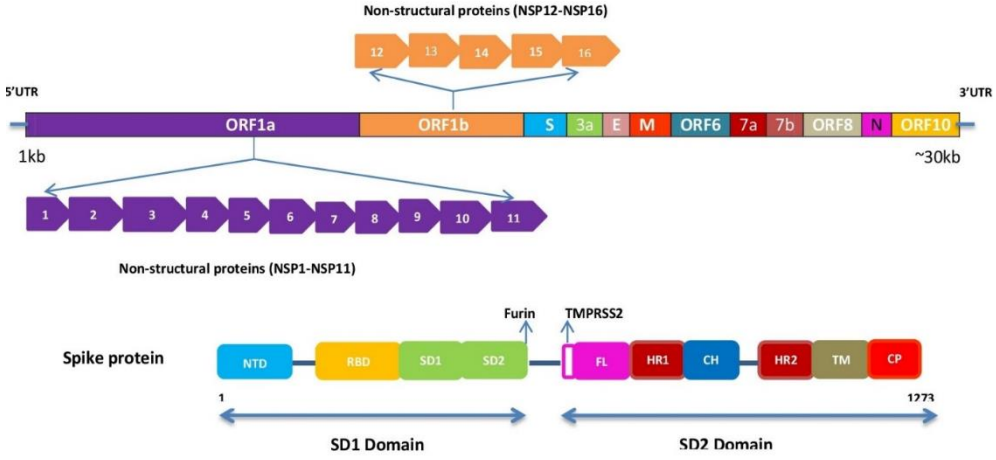

Figure 1: Genome organization of SARS-CoV2

Although extensive research has been done to highlight the alterations in the SARS-CoV-2 genome, but systematic data on insertion and deletion mutations are still absent, which may help to understand the evolving pathogenicity of the virus over time. Our investigation involves the amino-acid sequences coding for the structural and non- structural proteins of SARS-CoV-2 from different geographical regions for detecting the most frequently occurring mutations, insertions, deletions and stop-codon mutations at specific sites in the genome.

2.1 Materials and Methods

2.1 Sequence retrieval

From 58 countries, amino acid sequence data for 1,404 complete SARS-CoV-2 genomes were collected from NCBI database. The sequences submitted from January, 2020 to January 2022 were analyzed. The data with incomplete amino acid gap sequences for respective proteins were not included.

2.2 Sequence alignment and mutation analysis

As a reference genome, we used the genomic sequence of the isolate Wuhan-Hu-1, accession number NC_045512. The CoVsurver mutations program of GISAID was used to align the query sequence with the reference sequence in order to find mutations.

3. Results

3.1 Mutations in structural proteins

Spike (S) protein mutations

Among the total complete genome sequences retrieved from different geographical regions, the majority of the mutations were identified from the spike proteins. Many novel mutations were found to have emerged after March 2020 and were seen to be on the highest rate in specific countries. By the use of CoVsurver mutations app at GISAID, all these mutations were identified. In our analyses, the total numbers of mutations detected in spike proteins were 5,908 (742 from year 2020, 4033 from 2021 and 1137 in Jan, 2022).

The D614G was found the highly variable mutation in SARS-CoV-2 populations from all six continents. This mutation was identified in 935 SARS-CoV-2 populations. From year 2020, the D614G mutation was identified in 477 SARS-CoV-2 populations. The frequency of this mutation was higher in SARS-CoV-2 populations from Asia (255 times) followed by Europe (108 times), South America (54 times), North America (21 times), Africa (20 times) and Oceania (19 times). The D614G mutation was found at greater frequency in some specific countries including Taiwan, Hong Kong, India and Pakistan from Asia, Germany, Spain, Russia and Finland from Europe, Brazil, Peru and Chile from South America, Mexico from North America, Egypt and Ghana from Africa and New Zealand from Oceania. The H49Y (12), P681H (12), N501Y (11), A570D (8), D1118H (8), S982A (8) and T716I (8) mutations were also detected in the SARS-CoV-2 populations in year 2020. In 504 SARS-CoV-2 populations from year 2021, the D614G mutation was detected 244 times from Asia, 92 times from Europe, 73 times from Africa, 49 times from North America, 29 times from South America and 17 times from Oceania. The mutation spread faster in countries like Bangladesh, Japan and India from Asia, Italy and Slovakia from Europe, Brazil, Peru and Venezuela from South America, Mexico and Jamaica from North America, Kenya, Ghana and Nigeria from Africa and New Zealand from Oceania. After D614G, the other highly variable mutations were detected in SARS-CoV-2 populations in year 2021 are N501Y (233) followed by P681H (185), T478K (164), P681R (163), D1118H (162), T716I (162), A570D (160), S982A (160), T19R (141) and D950N (136). The N501Y, P681H, D1118H and T716I mutations were identified in maximum numbers of SARS-CoV-2 populations from Kenya, Bangladesh, Mexico, Japan, New Zealand and Ghana whereas T478K, P681R, A570D, S982A, D950N and T19R mutations from Bangladesh, Japan, India, Jamaica and Slovakia. From Jan, 2022, a total of 977 spike protein mutations were identified in SARS-CoV-2 populations from Asia whereas 160 were detected from Europe. The D614G mutation was identified 54 times from Japan (20), Slovakia (11), Hong Kong (10), Pakistan (9) and Malaysia (4). The T95I, T478K, T19R, P681R, L452R, G142D, Y505H, L212I, L981F, N440K, K477N, N679K, N501Y, N856K, N969K, P681H, Q954H, R346K, S373P, S375P, T547K and G446S mutations were commonly detected.

Table 1: Distribution of mutations in the different regions of human SARS-CoV-2 spike proteins.

Regions	Name of distinct Mutations	Total number of mutations
$S1^A$ domain (1-302)	L5F, L7S, V6A, L8V, P9L, V11I, S12F, S13I, Q14R, V16I, L18F, T19I, T19K, T19R, T20N, T22I, P26F, P26H, P26S, A27S, A27V, Y28H, T29A, T29I, F32L, Y38C, S46L, H49Y,	1,224

	S50L, Q52R, L54F, F59Y, V70I, N74K, T76I, A67S, A67V, H69S, I68R, D80A, D80Y, F86S, T95I, E96G, S98F, R102G, R102I, R102S, V102I, V102L, D111N, I119V, D138H, D138Y, L141F, G142D, V143C, V143F, Y144F, Y144V, Y145H, H146R, S151G, W152C, W152L, W152R, M153I, M153T, E154K, S155I, E156G, F157C, F157L, F157S, R158G, Q173H, P174S, L176F, D178N, L179F, E180D, G181V, Q183H, N185K, L189F, R190S, I197T, G199R, H207Y, L212I, V213G, D215A, D215G, D215H, D215Y, Q218H, S221L, S221W, A222V, V227I, G232C, I233V, R237S, R246G, G252S, D253A, D253N, S254F, S256P, W258L, G261D, G261R, A262S, Y265C, P272L, R273S, V289L, L293F	
$S1^{A}$ -$S1^{B}$ linker (303-332)	F306I, T307S, V308C, E309K, E309R, T323I, P330S	7
$S1^{B}$ domain (333-527)	F338L, A344S, A344T, G339D, G339S, R346K, R346S, A348S, A352S, V367F, S371F, S371L, S373P, S375F, V382L, P384L, V401L, R408I, Q414R, K417N, V433I, N439K, N440K, V445A, G446S, G446V, L452Q, L452R, Y453F, I468T, E471Q, S477I, S477N, T478K, V483I, E484A, E484D, E484G, E484K, E484Q, F490S, Q493R, S494L, Q493R, G496S, Q498R, N501T, N501Y, Y505H, Y508H, S514F, A520S, A522V	1,482
$S1^{B}$ - $S1^{C}$ (528-533)	**No mutation observed**	-
$S1^{C}$ domain (534-589)	T547I, T547K, A570D, T572I, T573I**,** D574Y, E583D	228
$S1^{C}$ - $S1^{D}$ (590-593)	No mutation observed	-
$S1^{D}$ domain (594-674)	Q613H, D614G, V622I, S640F, F643L, A653V, H655Y, E661D, V662F, S673T	982
Protease cleavage site (675-692)	Q675H, Q677H, T678I, T678S, N679K, P681H, P681L, P681R, A684V, R682W, R685H, S686G, Q690R	508
S1-S2 subunits linker (693-710)	A701V, L699I	41
Central β-strand (711-737)	T716I, T732A	180
Downward helix (738-782)	M740I, D745G, T747I, L754F, Q762P, L763A, N764K, R765L, A766S, T768K, G769V, V772I, E780K	69
S2’ cleavage site (783-815)	T791I, D796G, D796Y, F797C, D808G, P812L, P812R, P812S, S813I	66
Fusion peptide (816-828)	L822F	2
Connecting region (829-911)	A829S, A829T, A831V, Y837W, A852S, N856K, T859N, A871V, A879S, A903V, D839Y, S884F, F888L, A899S, M900I	101
Heptad repeat region (912-983)	S929T, A930V, D936Y, L938F, S939F, S943I, L948I, D950B, D950H, D950N, Q954H, S968A**,** N969K, D979E, L981F, S982A	329
Central helix (984-1034)	V1104L, Q1106P, A1020S, A1020V, T1027I,	37
β-hairpin (1035-1068)	V1040A, D1041E, F1042Y, C1043S, G1044R, F1062L, T1066N	10
β-sheet domain (1069-1133)	P1069S, Q1071H, K1073N, A1078S, G1085R, A1087S, R1091H, E1092K, T1117I, D1118H, V1122L, G1124C, V1129E	185
Heptad repeat region (1134-1213)	P1143L, L1145J, D1146Y, S1147L, D1153A, D1153Y, Y1155F, P1162L, P1162S, G1167V, S1170A, V1176F, N1192S, Q1201H, E1202Q, Q1208H,	62
Transmembrane (1214-1236)	G1219V, V1228L, M1229I, C1235F, M1237I	9
Cytoplasmic region (1237-1273)	C1247F, G1251V, S1252F, D1259Y, V1264L	13

Mutations in Receptor binding domain (RBD) of S protein

According to continents, the total numbers of different mutations in RBD detected were; Asia (926), Europe (202), Africa (90), Oceania (32), South America (53) and North America (94).

From year 2020, a total of 58 mutations were detected 23 times from Asia, 23 times Europe, 5 times from Africa, 4 times from South America, 2 times from North America and 1 time from Oceania. The mutations were observed at positions; 338, 344, 367, 384, 433, 439, 440, 453, 477, 478, 488, 484, 490, 501 and 514. The 501 amino acid position was found the most variable in the RBD of the spike protein. Two mutations were identified at this position N501T and N501Y. The N501T was identified in 5 SARS-CoV-2 populations from Saudi Arabia and Mexico whereas N501Y was identified in 14 SARS-CoV-2 populations from South Africa, Ghana, Hong Kong, Taiwan, Finland, Germany, Netherlands, Poland, Switzerland and Brazil. The Y453F mutation was identified 6 times from Denmark (3), Netherlands (3) and Russia (1). The S477N mutation was detected in 7 SARS-CoV-2 populations from Germany (4), Hong Kong (1), Iran (1) and France (1). From year

2021, a total of 970 mutations were identified in 553 SARS-CoV-2 populations from Asia, 132 from Europe, 93 from North America, 84 from Africa, 73 from South America and 35 from Oceania. The mutations were identified at amino acid positions; 323, 339, 346, 348, 352, 367, 371, 373, 375, 382, 401, 414, 417, 439, 440, 445, 446, 452, 471, 477, 478, 479, 483, 484, 490, 493, 494, 496, 498, 501, 505, 520, 521 and 522. The 501Y mutation was identified in 233 whereas N501T was identified only from 7 SARS-CoV-2 populations. Other frequently occurring RBD mutations were identified as T478K (164 times), L452R (162 times), E484K (83 times) and K417N (73 times). In Jan 2022, a total of 487 mutations were detected in 452 SARS-CoV-2 populations from Asia whereas 35 from Europe. The highly variable mutations were detected as E484A, G339D, G446S, G496S, K417N, N440K, N501Y, Q493R, Q498R, R346K, S371L, S373P, S375F, S477N, T478K and Y505H.

Envelope (E) protein mutations

A total of 23 mutations were detected in SARS-CoV-2 populations from year 2020. The maximum numbers of envelope protein mutations were identified from Asia (12 times) followed by Europe (6 times) then from Africa (3) and South America (2). The names of the identified mutations were F4S, L28P, L37H, L73F, L73I, P71L, S50G, S68F, S6L, T30I, V58F, V5A, V5I and V62F. The V5A mutation was found in 7 SARS-Cov-2 populations from India (4), Cambodia (1), Germany (1) and Brazil (1). From year 2021, the E protein mutations were detected in 77 SARS-CoV-2 populations from Asia (45 times), Europe (20 times), Africa (8 times), North America (3 times) and Oceania (1). The detected mutations were as C43R, F26L, K53R, L21F, P71L, R61L, T9I, V58F, V62F and V70F. The highly variable mutation was found P71L (31) followed by L21F (19) and T9I (15). From Jan, 2022, a total of 35 envelope protein mutations were identified in 27 SARS-CoV-2 populations from Asia whereas 8 from Europe. Only two distinct mutation sites i.e. T9I (Japan (17), Pak (9) and Slovakia (8)) and V62F (Japan) were detected.

A single mutation (T9I) was identified in 4 SARS-CoV-2 populations from year 2021 from USA.

Membrane (M) protein mutations

The A2S, A2V, D3G, F28S, H125Y, I73M, S214R, T130P, T175M and V70L mutations were identified in 13 SARS-CoV-2 populations from year 2020. Accordingly, the total numbers of mutations in membrane protein observed were; Asia (7), South America (3) Europe (2) and Africa (1). In 252 SARS-CoV-2 populations from year 2021, the maximum numbers of M protein mutations were detected from Asia (162) followed by Europe (52) then North America (24), Africa (10), South America (3) and Oceania (1). The identified mutations from year 2021 were as A104S, A194T, A2V, A63T, A85V, C86F, D3G, F28L, H125Y, H155Y, I201V, I48V, I73M, I82S, I82T, K15N, L16I, L17I, L29F, N41S, Q19E, S212G, S212N, S4P, T208I, V66L and W31C. Among these mutations, the I82T (158) was detected as highly variable mutation followed by Q19E (15), A63T (13) and D3G (13). From Jan, 2022, the membrane protein mutations were identified in 94 SARS-CoV-2 populations from Asia whereas 20 from Europe. The identified mutations were as A63T (17 from Japan, 9 from Pakistan and 8 from Slovakia), Q19E (17 from Japan, 9 from Pakistan and 2 from Slovakia), D3G (16 from Japan and 9 from Pakistan and 7 from Slovakia) and I82T (10 from Hong Kong, 4 from Malaysia, 9 from Pakistan and 3 from Slovakia).

In the 20 sequences of year 2021 from USA, only two mutations were detected i.e. A63X and I82T. The A63X was identified only single time whereas I82T was detected 11 times. But from Jan, 2022, a total of 92 mutations were detected.

Nucleocapsid (N)protein mutations

In case of Nucleocapsid (N) proteins, a total of 271 mutations were detected in SARS-CoV-2 populations from Asia from year 2020. The R203K and G204R mutations were detected as co-occurring mutations. The R203K mutation was detected 68 times whereas G204R detected 67 times. Both the mutations were identified in SARS-CoV-2 populations from Bangladesh, Hong Kong, India, Japan, Lebanon, Pakistan, Philippines, Saudi Arabia, South Korea and Taiwan. The maximum numbers of each of the mutations were detected from Japan (18), India (15), Taiwan (15) and Hong Kong (7). The S194L was found the second most variable mutation in 25 N proteins from China (2), Hong Kong (6), India (3), Iran (3), Japan (1), Malaysia (3), Pakistan (4) and Taiwan (3). The P13L was found 19 times in N proteins from India (3), Japan (2), Malaysia (4), Myanmar (1), Pakistan (1), Philippines (1), Saudi Arabia (1), South Korea (3) and Taiwan (3). But in Asian SARS-CoV-2 populations from year 2021 up to Jan, 2022, a total of 951 mutations were detected. The mutations that were found in maximum numbers are R203M (122), D377Y (122), R203K (117), D63G (114), G204L (114), G215C (75), D3L (49), S235F (49) and P13L (39). In most of the cases, the G204R and R203K, the D63G, R203M and D377Y and the D3L and S235F mutations were found accompanied with each other. In SARS-CoV-2

populations from Europe from year 2020, a total of 117 mutations were identified. The G204R (34) and R209I (34) mutations from year 2020, the R203K (32), D3L (31), R203M (29), D63G (29) and D377Y (29) mutations from year 2021 and P13L (8), G204R (7) and R203K (7) from Jan, 2022 were detected as highly variable mutations. From North America, a total of 211 amino acid substitutions were detected in N proteins. Each G204R and R203K mutations were found 32 times in N proteins from Belize 2020 (1), Canada 2020 (1), Mexico 2020 (7), Mexico 2021 (21) and Canada 2021 (2). Similarly, each D377Y and R203M mutations were found 20 times in N proteins from Canada 2021 (2) and Jamaica 2021 (18). These two mutations were also found associated with each other. The D63G was found 18 times in N proteins from Canada 2021 (1) and Jamaica 2021 (17). From South America, the N protein mutations were detected in 202 SARS-CoV-2 populations. The identified mutations were A119S, A134V, D288N, D3L, G204R, G214C, E290D, I292T, M234I, P13L, P383L, P6H, P80R, R203K, S197L, S202T, S235F and T366I. Each G204R and R203K mutations were found 65 times in N proteins from Argentina 2020 (1), Brazil 2020 (1), Brazil 2020 (21), Chile 2020 (4), Peru 2020 (9), Brazil 2021 (12), Peru 2021 (9) and Venezuela 2021 (8). The P80R was found 19 times in N proteins from Brazil 2020 (1), Brazil 2021 (8), Peru 2021 (2) and Venezuela 2021 (8). The II292T was found 17 times in N proteins from Argentina 2020 (1), Brazil 2020 (1), Brazil 2020 (14) and Chile 2020 (1). The A156S, A211V, A220V, A398S, D3L, G204R, P13L, P199L, P365S, P67S, P80R, Q389L, R203K, R209I, S183Y, S194L, S202N, S235F and T205I mutations were detected in N proteins from Oceania. Each G204R and R203K mutations were identified 25 times from New Zealand 2020 (10), New Zealand 2021 (14) and Australia 2021 (1). Similarly, both D3L and S235F mutations were found 12 times in SARS-CoV-2 populations from New Zealand from year 2021. From Africa, a total of 304 mutations were detected in N proteins. Both G204R (68) and R203K (69) mutations were identified in SARS-CoV-2 populations from Egypt 2020, Ghana 2020, Nigeria 2020, Egypt 2021, Ghana 2021, Kenya 2021 and Nigeria 2021 (5). The maximum numbers of these two mutations were detected from Kenya (39) and Ghana (12). Each D3L and S235F were found 55 times in N proteins from Ghana 2020 (1), Ghana 2021 (11), Kenya 2021 (39) and Nigeria 2021 (4). A total of 59 mutations were detected in SARS-CoV-2 sequences retrieved from year 2020 from USA. In the sequences of year 2021, a total of 54 mutations were detected. The D377Y (11), D63G (11), N203M (11) and G215C (8) mutations were detected in more numbers. From Jan 2022, D343G, D377Y, D63G, R203K, R203M, G204R, G215C and P13L mutations were identified.

3.2 Mutations in non-structural proteins

Mutations in NSP1 protein

Nsp1 is the NH2-terminus region of ORF1ab which is cut off by a papain-like protease (PLpro) from the polyprotein. This protein is also called leader protein. It degrades the host mRNA molecules, blocks the translation process and also affects the innate immune defense system (Chaudhuri, 2020; Thoms et al., 2020). Many mutations were identified in NSP1 proteins. In the SARS-CoV-2 populations from Asia from year 2020, the D33E (1) (Iran), D75E (1) (Taiwan), E57K (1) (Bangladesh), L27F (1) (India), M85V (1) (South Korea), R24C (1) (Bangladesh), R29H (1) (Taiwan) and W161L (1) (Iran) mutations were detected in NSP1 proteins. The R24C mutation was also observed in a single isolate from India submitted in year 2021. Ten other distinct types of mutations (E2G, E36D, V54I, L61F, L92F, E102K, G112S, E148G and G150C) were identified in 12 NSP1 proteins from Asia from year 2021. The two mutations i.e. E2G and E148G were also identified in a single SARS-Co-2 population from Japan from year Jan, 2022. The S135R, a unique mutation was detected in a single SARS-CoV-2 isolate from Japan from year Jan, 2022. From Europe, the three distinct types of mutations i.e. H110Y (2) (Germany 2020, Netherlands 2021), H45Y (1) (Georgia 2020) and P80L (1) (Finland 2021) were identified in NSP1 proteins. From Africa, the E102Q (Kenya 2021), I114T (Egypt 2020 & 2021), R73H (Kenya 2021), V116L (Kenya 2021) and V86F (Nigeria 2021) were detected in NSP1 proteins. The I114T mutation was also identified in two SARS-CoV-2 populations from North America (1) and South America (1). From North America, the I114T mutation was detected from Mexico 2021 whereas from South America the same mutation was identified from Venezuela 2020. One another mutation i.e. T78A was found two times in NSP1 proteins from Mexico from year 2021. The R24C (Peru 2020) and E57K (Peru 2021) mutations were also identified in 2 SARS-CoV-2 populations from South America. Four other mutations i.e. G112D (Brazil 2020), G49V (Brazil 2020), K120N (Venezuela 2021) and T78I (Venezuela 2021) were also identified in NSP1 proteins from South America. The T78I (New Zealand 2021) mutation was also identified in a single SARS-CoV-2 isolate from Oceania. In the sequences of year 2020 from USA, the L75E, G146S and D75N mutations were detected whereas only a single mutation (G150S) was identified from Jan 2022.

Mutations in NSP2 protein

In case of NSP2 protein, a total of 112 mutations were detected in SARS-CoV-2 populations from Asia from year 2020 whereas 138 mutations from year 2021. The A306V, K500N, M141V, S122F, T85I, V157F, and V381A mutations were found common to both the years. The highly variable mutations were found in SARS-CoV-2 populations from year 2021 are T85I (27), P129L (23), K81N (10) and V381A (9). Four distinct types of mutations (H145Y, L266F, R218H and T429I) were found common in SARS-CoV-2 populations from year 2021 and Jan, 2022. Three other mutations i.e. I514T, K492N and L266F were also detected in the in SARS-CoV-2 populations from Jan, 2022. In SARS-CoV-2 populations from Europe, a total of 36 mutations were identified in NSP2 proteins from year 2020 whereas 28 from year 2021. Only two mutations (T85I and N280Y) were reported common from both the years. In the in SARS-CoV-2 populations from year 2020, the T85I mutation was found 15 times from Czech Republic (1), Denmark (1), France (1), Georgia (4), Germany (3), Netherlands (1), Russia (2), Sweden (1) and Turkey (1). The D144G mutation was identified in a single SARS-CoV-2 isolate from Slovakia from Jan, 2022. From Africa, a total of 34 mutations were identified in NSP2 proteins. The L550F was found 12 times in NSP2 proteins from Kenya 2021. The T85I was found 6 times in NSP2 proteins from Kenya 2020 (1), Kenya 2021 (2), South Africa 2020 (2) and Ghana 2021 (1). From North America, the A476V, E167V, E460G, E467G, E587K, G423C, I349T, I559V, K500N, K81N, L270F, L270F, L550F, N195S, P181L, P585S, Q376K, R4C, T175N, T580I, T634I, T85I and V571L mutations were identified in NSP2 proteins. The T85I was found 6 times from Mexico 2020 (2), Mexico 2021 (3) and Canada 2021 (1). From South America, the D511Y, E172K, G88E, H208Y, I484F, K534N, L289F, L550R, S138L, S533A, S99F, T429I and T85I mutations were identified in NSP2 proteins. From Oceania, a total of 10 mutations were detected in NSP2 proteins and the distinct types of identified mutations were G339S, L550F, N168D, T580I, T85I, V198I and V381A. All the detected mutations were observed in SARS-CoV-2 populations from New Zealand from Jan, 2020 upto December, 2021.
The T85I was identified in 13 SARS-CoV-2 populations from year 2020 from USA. The E57X, G465X, S358A, T85I and T85X mutations were found in SARS-CoV-2 genomes from year 2021 whereas a single mutation V157F was identified in only one sequence from Jan, 2022.

Mutations in NSP3 protein

The A890D, D218N, I1412T, K1211N, N1778S, P1044S, P109L, P1261L, P874S, S1206L, S126L, S1670F, T1004I, T1198I, T1306I, T1365I, T1456I, T183I, T428I and T749I mutations were identified common to both the years. In the SARS-CoV-2 populations from year 2020, the T1198I mutation was identified in maximum numbers from India (2), Japan (1), Malaysia (4), Myanmar (1), Pakistan (1), Philippines (1), Saudi Arabia (1) and Taiwan (3). In the SARS-CoV-2 populations from year 2021 till Jan, 2022, the highly variable mutations were identified as P1228L (80), A488S (73), P1469S (77), T183I (49), I1412T (49), A890D (47), K38R (38), P822L (38), A1892T (38), L1266I (31), K837N (24) and V932A (22). From Europe, the A890D, G301S, I1412T, I1683T, T1189I, T1306I and T183I mutations were detected common to years 2020 and 2021. In the SARS-CoV-2 populations from year 2021, the mutations were identified in maximum numbers are A890D (31), I1412T (30), T183I (30), P1469S (25), A488S (24) and P1228L (24). The A890D was identified one time in NSP3 proteins from Denmark, three times from Finland, four times from Germany, two times from Italy, eight times from Netherlands, eight times from Slovakia, five times from Spain and one time from Turkey. The I1412T was found in 30 SARS-CoV-2 populations from Denmark (1), Finland (3), Germany (3), Italy (2), Netherlands (8), Slovakia (7), Spain (5) and Turkey (1). The T183I was also detected 30 times from Netherlands (8), Slovakia (7), Spain (5), Finland (3), Germany (3), Italy (2), Denmark (1) and Turkey (1). A total of 35 mutations i.e. A1711V (1), A1736V (1), A1892T (8), A488S (3), K38R (8), L1266I (8), P1228L (3) and S660Y (1) were identified in SARS-CoV-2 populations from Slovakia from Jan, 2022. From North America, the mutations that found in maximum numbers are A488S (17) Jamaica 2021, P1228L (17) (Jamaica 2021), P1469S (17) (Jamaica 2021), A890D (14) (Mexico2021), I1412T (14) (Mexico 2021) and T183I (14) (Mexico 2021). From South America, the A146S, A1711V, A54T, A890D, E122D, E177D, F1569V, I1412T, K589R, K977Q, L1791F, L689F, P1044T, P1469S, Q764R, S1735F, S370L, T1063I, T1127S, T1303I, T183I, T237I, T428I, T725I, V1272G and V393F mutations were detected in NSP3 proteins. Both S370L and K977Q mutations were identified 19 times each in NSP3 proteins from Brazil 2020 (1), Brazil 2021 (8), Peru 2021 (2) and Venezuela 2021(8). The T428I was found 15 times in NSP3 proteins from Chile 2020 (1), Peru 2020 (8) and Peru 2021 (7). From Africa, the A890D (55) and I1412T (55) were found in NSP3 proteins from Ghana 2020(1), Ghana 2021(11), Kenya 2021(39) and Nigeria 2021 (4). The T183I (56) was also detected from Ghana 2020(1), Ghana 2021(11), Kenya2021 (39) and Nigeria 2021 (5). A total of 63 mutations were identified in NSP3 proteins from Oceania. Among them, the A890D (12), I1412T (12) and T183I (12) mutations were found 36 times in NSP3 proteins from New Zealand 2021.
In case of NSP3 protein, a total of 52 mutations were identified in sequences from year 2020 from USA. All

the detected mutations were found single times. A total of 32 mutations were detected in SARS-CoV-2 genomes from year 2021. Among them, the P1228L (8) and P1469S (8) were identified in multiple times. On the other side, only two mutations (A1892T and K38R) were found in 8 SARS-CoV-2 populations from Jan, 2022.

Mutations in NSP4 protein

A total of 30 mutations were detected in SARS-CoV-2 populations from Asia year 2020. The identified mutations were A231V, A307V, A380V, A446V, D279N, E219K, G108S, G309C, H313Y, H470Y, I284T, L398I, M324I, Q9L, R222G, S209P, S481L, T143P, T173N, T204I, T237I, T28I and V233A mutations. In SARS-CoV-2 populations from year 2021 up to Jan, 2022, the A446V, T492I, T54I, T60I, V167L, V20F and V498I mutations were identified in NSP4 proteins. The maximum numbers of the mutations were T492I (114), V167L (75) and A446V (43). The D70G (1) (Poland), F17L (1) (Germany), F308Y (2) (Spain), M324I (5) (France and Germany), M33I (1) (Georgia), T265A (1) (Finland), T350I (2) (Russia and Spain) and T492I (1) (Georgia) mutations were detected in the SARS-CoV-2 populations from Europe from year 2020. The mutations that were detected in NSP4 proteins from year 2021 are A307V (1), A380V (1), A446V (3), F17L (1), H36Y (1), K86N (1), L206F (1), M324I (1), S336L (2), T204I (1), T295I (2), T327I (1), T492I (30), V13F (1), V167L (26) and Y205H (1). The T492I was found 30 times in NSP4 proteins from Belgium (1), Portugal (2), Russia (2), Slovakia (20), Switzerland (4) and Turkey (1). The V167L was identified 26 times in NSP4 proteins from Portugal (2), Slovakia (19), Switzerland (4) and Turkey (1). Two distinct types of mutations i.e. T492I (11) and V167l (3) were identified in SARS-CoV-2 populations from Slovakia from Jan, 2022.From North America, the A446V, A58T, I23T, L353F, L438P, P166S, S395T, S432G, T492I, T73I and V167L mutations were detected in NSP4 proteins. The T492I was found 23 times in NSP4 proteins from Mexico 2020 (1), Canada 2021 (1), Jamaica 2021 (17) and Mexico 2021 (4). The V167L was found 17 times in NSP4 proteins from Jamaica 2021 (17). From South America, the A307V, F308Y, L438P, N131D, S137L, S184N and T492I mutations were identified in NSP4 proteins. The L438P and T492I both were identified 7 times in NSP4 proteins from Peru 2021. Only five mutations D279N, T492I, V294L, V328I and V334A were detected in NSP4 proteins from Oceania. The A446V, F326L, L15F, L438P, P29S, S455N, T492I, V334A and Y300S mutations were identified in NSP4 proteins from Africa. The A446V mutation was found 5 times in NSP4 proteins from Egypt 2020 (1), Egypt 2021 (1), Ghana 2021 (1) and Kenya 2021 (2).

Out of 24 mutations in SARS-CoV-2 genomes of USA from year 2021, the T492I and V167L mutations were detected multiple times. Both of these mutations were also detected in five SARS-CoV-2 populations from Jan, 2022.

Mutations in NSP5 protein

In case of NSP5 protein, the A206P (1), A234V (1), A94V (1), G15S (1), I259F (1), K90R (1), L89F (3), P108S (10), P96L (4), Q83K (1), R217K (1), T169I (1), T24A (2) andV212F (1) mutations were identified in SARS-CoV-2 populations from Asia year 2020. A total of 105 mutations were detected in NSP5 proteins from year 2021 up to Jan, 2022. The identified mutations were A191V, E47K, G15S, K100R, K90R, L220F, L89F, P132H, P132S, P96L, R279C, R60L, S267L, T169S and T24A. The P132H (39), K90R and T169S (15) mutations were found highly variable mutations. The P132H was identified in SARS-CoV-2 populations from China 2021 (2), Hong Kong 2021 (2), India 2021 (6), Japan 2021 (3), Japan 2022 (17) and Pakistan 2022 (9). The nine mutations i.e. A234V (2) (Germany 2020), E288D (1) (Russia 2020), G71S (1) (Serbia2020), I259V (3) (Netherlands 2020), L75F (1) (Switzerland 2020), V157A (1) (Germany 2020) and S301L (1) (Turkey 2020) were identified in NSP5 proteins from Europe 2020. The A193V (1) (France 2021), F8L (2) (Slovakia 2021), G15S (1) (Spain 2021), K90R (7) (Finland2021, France 2021, Germany 2021 and Slovakia 2021), L227F (1) (Slovakia 2021), P132H (1) (Belgium 2021) and R217S (1) (Poland 2021) were detected in NSP5 proteins from Europe 2021. A single mutation P132H was identified 8 times in NSP5 proteins from Slovakia in the month of Jan, 2022. The A193T, A94V, H246Y, K90R, L75F and P132H mutations were detected in NSP5 proteins from North America. The G15S, L205V, L75F, L87F, N274D, P108S and P96H mutations were detected in NSP5 proteins from South America. The G15S was identified 16 times in NSP5 proteins from Chile 2020 (1), Peru 2020 (8) and Peru 2021 (7). From Oceania, the A191V, G15S, K90R, L89F and P132H mutations were detected inn NSP5 proteins. The G15S, I106V, K88R, K90R, P184S, P252L and R60C mutations were identified in NSP5 proteins from Africa. The G15S was found 6 times in NSP5 proteins from Egypt 2021. The K90R was found 4 times in NSP5 from South Africa 2020 (2) and Kenya 2021 (2).

The L89F and K90R were detected in 5 SARS-CoV-2 populations from year 2020 from USA. The A260V, L75F, L89F and T24I mutations were detected in SARS-CoV-2 sequences of year 2021 whereas a single mutation i.e. P132H was identified in 4 SARS-CoV-2 populations from Jan, 2022.

Mutations in NSP6 protein

The mutations A88T, D102I, E39K, G218D, H64N, L37F, L260F, L67P, M100I, M164T, M183I, M47R, N40K, Q160K, T180I, V84F and Y80C were detected in SARS-CoV-2 populations from Asia from year 2020. The L37F mutation was found highly variable mutation. This mutation was detected in SARS-CoV-2 populations from China (2), Thailand (3), Lebanon (2), India (2), Iran (4), Japan (2), Malaysia (5), Pakistan (6), Philippines (3), Saudi Arabia (1), South Korea (7), Sri Lanka (1) and Taiwan (14). From year 2021, a total of 222 mutations were identified in NSP6 proteins. The highly variable mutations identified were T77A (85), V149A (38), I189V (38) and L37F (14). The T77A was found in maximum number of SARS-CoV-2 populations from Bangladesh (17), Japan (22), India (12) and Hong Kong (10). In the SARS-CoV-2 populations from Europe from year 2020, the L37F mutation was found in maximum numbers from Czech Republic (1), Finland (2), France (1), Georgia (3), Poland (1), Spain (1), Sweden (1) and Turkey (2). In the SARS-CoV-2 populations from year 2021, the T77A was found highly variable mutation from Portugal (2), Slovakia (19), Switzerland (4) and Turkey (1). A total of 12 mutations T77A (3), L185F (1) and I189V (8) were identified from Slovakia in the month of Jan, 2022. The C197F, G277S, H11Y, I189V, I202V, I49V, L260F, L37F, T181I, V149A and T77A mutations were identified in NSP6 proteins from North America. The T77A was found 18 times in NSP6 proteins from Canada 2021 (1) and Jamaica 2021 (18). From South America, the F108L, F36L, G107S, L260F, L37F, S106T and V149F mutations were identified in NSP6 proteins. The A46S, G107S, I168T, I189V, L37F, L75F, L98F, M183I and M86I mutations were identified in NSP6 proteins from Oceania. The A51V, F108L, F184V, L37F, L98F, M183I, M86I, P198L, T147I and V149A mutations were detected in NSP6 proteins from Africa. From year 2020 from USA, four mutations were detected ie., L37F, A136V, V149I and T181I. All the reported mutations were found only single times.

Mutations in NSP7 protein

In case of NSP7 proteins from Asia, the D67A (1) (South Korea 2020), E74K (1) (Iran 2020), L71F (5) (Hong Kong 2021, Myanmar 2021), M3I (1) (Taiwan 2020), S25L (10) (South Korea 2020), D77N (1) (Japan 2021), S54L (1) (Hong Kong 2022) and V33A (1) (Pakistan 2020) mutations were detected. The D67A (Germany) was identified in a SARS-CoV-2 isolate from Europe from year 2021 whereas L20F mutation (Slovakia) was identified from Jan, 2022. The S25L (Jamaica) (NCBI No. MT507794.1) mutation was found single time in a SARS-CoV-2 isolate from North America from year 2020. Both L71F (7) (Brazil 2020, Brazil 2021) and S25L (1) (Brazil 2020) mutations were detected in 8 NSP7proteins from South America.

The two distinct types of mutations (T77A and I162V) were detected in 12 SARS-CoV-2 sequences of USA from year 2021. The T77A was found 8 times whereas I162V was identified 4 times. The T77A mutation was also detected in a single SARS-CoV-2 population from Jan, 2022. One another mutation (I189V) was also identified 4 times.

Mutations in NSP8 protein

In the Asian SARS-CoV-2 populations from year 2020, a total of 16 mutations were identified in NSP8 proteins. The identified mutations were A125P (Iran), A126P (1) (Iran), A14V (1) (Pakistan), A16V (3) (Saudi Arabia), A45V (1) (Iran), E155A (1) (Taiwan), M129I (1) (Taiwan), P133S (1) (Taiwan), Q158H (1), R96T (1) (Iran), T148I (3) (China, Pakistan, Iran) and T93I (1) (Bangladesh). From year 2021, the A16V, A74V, E23G, P133S, Q24R, T148I and V26I mutations were detected in NSP8 proteins. Five mutations i.e. A74V (1) (Netherlands 2020), G144D (1) (Spain2020), L169F (1) (Russia2020), T145I (1) (Poland 2020) and T148A (1) (France 2020) were identified in SARS-CoV-2 populations from Europe from year 2020. The A14V (1) (Slovakia 2021), A21S (1) (Turkey 2021), Q24R (1) (Spain 2021) and S177T (2) (Russia 2021) mutations were detected from year 2021. The three mutations i.e. A125T (1) (Mexico 2020), T123I (1) (Mexico 2020) and T145I (1) (Mexico 2021) were identified in NSP8 proteins from North America. The I107S, N108C, N109D, N109V, P133S and R57S mutations were detected in NSP8 proteins from South America. All the identified mutations except P133S were identified from Brazil 2020 (9) whereas P133S was detected from Peru 2020 (1). Only one mutation Q158H (2) (New Zealand 2020) was identified in a single SARS-CoV-2 isolate from Oceania. The A3T, D50N, E155G, L169F, T141M, T145I and T148I mutations were detected in NSP8 proteins from Africa. The T148I was found 6 times in NSP8 proteins from Egypt 2021. The three mutations i.e. A125T (1) (Mexico 2020), T123I (1) (Mexico 2020) and T145I (1) (Mexico 2021) were identified in NSP8 proteins from North America. The I107S, N108C, N109D, N109V, P133S and R57S mutations were detected in NSP8 proteins from South America submitted in year 2020. All the identified mutations except P133S were identified from Brazil (9) whereas P133S was detected in a sequence from Peru. Only one mutation Q158H (2) (New Zealand 2020) was identified in NSP8 protein from Oceania. The A3T,

D50N, E155G, L169F, T141M, T145I and T148I mutations were detected in NSP8 proteins from Africa. The T148I was found 6 times in NSP8 proteins from Egypt from year 2021. A single mutation, M87X was detected only in a single sequence from USA from year 2020.

Mutations in NSP9 protein

From Asia, P71L (1) (India 2020), S5R (1) (Taiwan 2020), I65V (1) (Taiwan 2021), M101K (1) (Pakistan 2021), P80L (2) (Japan 2021), P83L (2) (Myanmar 2021), Q49H (1) (Taiwan 2021), T109I (2) (Hong Kong 2021), T19I (3) (Japan 2021) and T21I (1) (India 2021) mutations were detected in NSP9 proteins. Nine mutations A107V (1) (Netherlands 2020), G37E (3) (Netherlands 2020), M101I (3) (Czech Republic 2020, Germany 2020, Poland 2020), R55K (1) (Spain 2020) and T62I (1) (Germany 2020) were detected from Europe from year 2020. The mutations D47N (1) (Poland 2021), M101I (1) (Poland 2021), P80L (1) (Netherlands 2021) and T35I (1) (Spain 2021) were identified in SARS-CoV-2 populations from Europe from year 2021. The L42F (1) (Mexico 2021), T109I (1) (Jamaica 2021), T34I (1) (Mexico 2020) and T35I (6) (Mexico 2020 & 2021) mutations were identified 9 SARS-CoV-2 genomes from North America. Three mutations L42F, T19I and T35I were detected in NSP9 proteins from Oceania. All the three mutations were found in the SARS-CoV-2 populations of New Zealand from year 2021. A single mutation V7I (Kenya 2021) was detected in a single NSP9 protein from Africa.

Two mutations i.e. I65V and P83H were detected in 3 SARS-CoV-2 populations from USA from year 2021.

Mutations in NSP10 protein

The T118N (1) (Pakistan 2020), T51I (6) (India 2020, Philippines 2020, Hong Kong 2021),D82N (1) (India2020), I55T (1) (Japan 2020), P84S (2) (Japan 2022), P8S (1) (India 2020), Q36R (1) (Myanmar 2021) and T39S (1) (Philippines 2021) mutations were detected in NSP10 proteins from Asia. Only three mutations F68L (1) (Sweden 2020), T101I (1) (Turkey 2020) and T102I (1) (Germany 2020) were identified in NSP10 proteins from Europe from year 2020. The N105K (1) (Finland 2021), T102I (2) (Thailand 2020, Slovakia 2021), T12I (1) (Netherlands 2021) and T39N (1) (Germany 2021) mutations were detected in SARS-CoV-2 populations from year 2021. The D131G (Mexico 2021) was found single time in a NSP10 protein from North America. From South America, the A32S (Brazil 2020), A4V (Venezuela 2021), C103R (Venezuela 2021), L138F (Peru 2020), M63I (Peru 2020) and P86H (Brazil 2021) mutations were identified in NSP10 proteins. Only two mutations M122T (1) (Ghana 2021) and N105K (1) (South Africa 2020) were detected from Africa.

Mutations in NSP11 protein

No any mutation was identified in case of NSP11 protein.

Mutations in NSP12 protein

In the Sars-CoV-2 populations from years 2020 and 2021, a total of 586 mutations were detected from Asia. Two mutations (P323L and P323F) were reported at 323 position. P323L was the major mutation found in NSP12 protein. The P323L mutation was identified 146 times in the Asian Sars-CoV-2 populations from year 2020 whereas 182 times from year 2021. In the month of Jan, 2022, the P323L mutation was identified in NSP12 proteins from Asia. TheP323Fmutation was found in one NSP12 protein from Bangladesh from year 2021 and also one time from Malaysia in the month of Jan, 2022. From Europe, the P323L mutation was detected in 105 NSP12 proteins from year 2020, 80 from year 2021 and 11 (only from Slovakia) from Jan, 2022. The P323F mutation was identified one time in NSP12 proteins from Finland from year 2021 and 11 times from Italy from Jan, 2022. The G671S was the second highly variable mutation found 97 times from Asia and 29 times from Europe from year 2021. A total of 101 mutations were detected in NSP12 proteins from North America. The P323L was identified 70 times from Canada 2020 (1), Canada 2021 (5), Jamaica 2020 (3), Jamaica 2021 (19), Mexico 2020 (16), Mexico 2022 (25) and Belize 2020 (1). From Oceania, a total of 43 mutations were detected in NSP12 proteins. The mutations were A185V (1), K718N (1), P227L (1), P323L (36), S520N (1), T248I (1), V776L (1) and V880I (1). All the mutations were identified in NSP12 proteins from New Zealand except P323L mutation. The P323L mutation was identified in Sars-CoV-2 populations from both New Zealand and Australia. From South America, the A185V, D738Y, E254D, G678C, H613Y, L805F, P323L, R249K, T680I and V354L mutations were detected. All mutations were identified only one time except P323L mutation that was identified 12 times in NSP12proteins. A total of 137 mutations were identified in NSP12 proteins from Africa. The P323L mutation was identified 87 times from Kenya 2020 (1), Kenya 2021(43), Ghana 2020 (5), Ghana 2021 (12), Egypt 2020 (7), Egypt 2021 (7), Nigeria 2020 (4), Nigeria 2021 (5) and South Africa (3).

Two distinct types of mutations i.e. G671S and P323L were detected in 40 SARS-CoV-2 populations from

USA from Jan, 2021 till Jan, 2022. The G671S mutation was detected in 10 SARS-CoV-2 populations from year 2021 whereas only one sequence from Jan, 2022. The P323L mutation was identified in 16 SARS-CoV-2 populations from year 2021 whereas only in five sequences from Jan, 2022.

Mutations in NSP13 protein

The P77L mutation was identified 144 times in NSP13 proteins in Asia from year 2021 up to Jan, 2022. The maximum numbers of P77L mutations were found in the SARS-CoV-2 populations from Bangladesh (31), Japan (27), India (17) and Hong Kong (10) retrieved. The A379V, A598S, D260Y, E261D, H164Y, H290Y, K218R, K460R, P504L, P53L, R339L, R392C, S263P and S485L mutations were identified in SARS-CoV-2 populations from Europe from year 2020. A total of 65 mutations were detected from NSP13 proteins Europe from year 2021 up to Jan, 2022. The P77L mutation was found 29 times in NSP13 proteins from Norway 2021 (1), Portugal 2021 (2), Slovakia 2021 (21), Switzerland 2021 (4), Turkey 2021 (1) and Slovakia 2022 (3). From North America, the A296S, D260Y, E341D, H164Y, I334V, M429I, P529L, P53L, P77L, S259L, S74A, T416I and Y306C mutations were identified in NSP13 proteins. The P77L was found 18 times in NSP13 proteins from Canada 2021 (1) and Jamaica 2021 (18). From South America, E341D, H164Y, L581F and P419S mutations were detected in NSP13 proteins. The E341D was found 19 times in NSP13 proteins from Brazil 2020 (1), Brazil 2021 (8), Peru 2021 (2) and Venezuela 2021 (8). From Oceania, the A520V, E261D, E341D, I258S, K460R, P504L, P78S, P82S, S259L, T547I, V169I and Y541C mutations were detected in NSP13 proteins. All the identified mutations were from New Zealand from years 2020 and 2021. From Africa, the A446S, E261D, G222S, P77L, R155C and V157L mutations were detected in NSP13 proteins.

Out of 17 mutations in NSP13 proteins from year 2021 from USA, the P77L was found in multiple times. The P77L was also identified in a single sequence from Jan 2022. One more mutation i.e. G671S was also found in a single sequence from Jan, 2022.

Mutations in NSP14 protein

The A138V, A320V, A323S, E347Q, F233L, G44S, H26Y, L149F, M315I, M501I, M72V, N129D, N438K, P43L, Q313H, Q458H, S503L and T113I mutations were identified in SARS-CoV-2 populations from Asia from year 2020. A total of 177 mutations were identified in SARS-CoV-2 populations from Asia from year 2021 till Jan, 2022. The A394V was found 75 times in NSP14 proteins but in maximum numbers from Japan 2021 (22), Bangladesh 2021 (27) and Hong Kong 2022 (10). The I42V was found 39 times in NSP14 proteins from China 2021 (2), Hong Kong 2021 (2), India 2021 (6), Japan 2021 (3), Japan 2022 (17) and Pakistan 2022 (9). From Europe, 7 mutations i.e. M62I (1) (Germany), S357I (1) (Germany), T16I (2) (Finland), T250I (2) (Germany) and T31I (1) (Netherlands) were identified in SARS-CoV-2 populations from year 2020. In the SARS-CoV-2 populations from year 2021, the A394V mutation was found in 26 NSP14 proteins from Portugal (2), Slovakia (19), Switzerland (4) and Turkey (1). The A394V (3) and I42V (8) mutations were identified in 11 SARS-CoV-2 populations from Slovakia submitted in the month of Jan, 2022. From North America, the A281S, A320V, A394V, D375Y, I42V, M49V, N130S, S255T and T31I mutations were identified in NSP13 proteins. The A394V was found 17 times in NSP14 proteins from Jamaica 2021. The A320V, A394V, A497T, H26Y, H373Y, N256S, P140S, P46L, S357N, T16A, Y235N and Y420C mutations were detected in NSP14 proteins from South America. Five mutations i.e. D345Y (New Zealand 2020), I42V (Australia 2021), N129D (New Zealand 2021), V125F (New Zealand 2020) and V167L (New Zealand 2021) were identified in NSP14 proteins from Oceania. A total of 22 mutations were detected in NSP14 proteins from Africa. The identified mutations were A360S, G114S, K139R, L493F, P297H, P297S, P451S and P46L. The P451S was found 14 times in NSP14 proteins from Ghana 2020 (1), Ghana 2021 (10) and Kenya 2021(3).

A total of 53 mutations were detected in the sequences retrieved from USA from year 2021. A single mutation was identified 8 times but all other detected mutations were found only single-single times. From Jan, 2022, two mutations i.e. A394V (1) and I42V (4) were detected in 5 SARS-CoV-2 populations.

Mutations in NSP15 protein

In the SARS-CoV-2 populations from Asia from year 2020, the A231V, A54P, C116Y, E228K, G129V, K259R, L75S, P205L, Q19H, R206S, S273I, S288F, T33I, V148F, V172L and V35F mutations were detected in NSP15proteins. The V172L was found 11 times in NSP15 proteins from Cambodia (1), Japan (1) and Thailand (9). From year 2021 up to Jan, 2022, a total of 72 mutations were identified in NSP15 proteins. The maximum number of mutations were H234Y (16), K259R (10), P205L (10) and S261A (9). The G17R (1) (Netherlands 2020), S161I (2) (Spain 2020), T33I (1) (Serbia 2020), A81V (2) (Germany 2021), D219Y (1) (Spain 2021), G17E (1) (Spain 2021), G17V (1) (Slovakia 2021), I107V (1) (Poland 2021), P270L (2) (Germany 2021), S261P (1) (Slovakia 2021) and V66L (1) (Italy 2021) and R138C (1) (Slovakia 2022)

mutations were detected in NSP15 proteins from Europe. Atotalof8mutations (D128Y, E170A, I280V, K259R, P153L, P205L, P262L and T325I) were identified in NSP15 proteins from North America. Seven mutations i.e. D272Y (1) (Brazil 2020), D282N (3) (Venezuela 2021), H337Y (1) (Peru 2021) and K110R (2) (Brazil 2021) were detected in NSP15 proteins from South America. The mutations E223A, G129V, G286S, Q19H, S147C and V318A were identified in NSP15 proteins from Oceania. The identified mutations were only from New Zealand. The A81V, H337Y, I252V, I269T and V66L mutations were detected in NSP15 proteins from Africa.
Only a single mutation, D132Y was detected in one SARS-CoV-2 populations from USA from year 2021.

Mutations in NSP16 protein
A total of 34 mutations were identified in SARS-CoV-2 populations from Asia from year 2020. The distinct types of identified mutations were G71C, G77R, I169M, K160R, K182N, M270T, P134S, Q6L, R216C, R287I, S240C, S240Y, S261P, T234I, T35I, T56I, T91M, T93M and W189C. The A188S, I112T, K160R, K76R, L183F, P215L, R216N, V167M andV294F mutations were detected in SARS-CoV-2 populations from Asia from year 2021 up to Jan, 2022. From Europe from year 2020, only three mutations i.e. K160R (1) (Poland 2020), Q28K (1) (Netherlands 2020) and T56K (1) (Georgia 2020) were detected in NSP16 proteins. A total of 8 mutations (A116S, A34V, G39D, H186Y, K160R and T35I) were identified in SARS-CoV-2 populations from year 2021. The D269E, L126F, M17I, N227K, P134S, Q238H, T140I, T151I, T91M and V288F mutations were identified in NSP16 proteins from North America. The K160R (Brazil 2020), M270I (Peru 2020), R216H (Brazil 2020) and R216N (Brazil 2020) mutations were detected in NSP16 proteins from South America. The three mutations, M270I (New Zealand 2020), Q159K (New Zealand 2021) and R216C (New Zealand 2021) were identified in NSP16 proteins from Oceania. The F193Y, L252J, Q238H and T56I mutations were detected in NSP16 proteins from Africa.
Four mutations, M20I, P134S, R216C and T140I were detected in 4 SARS-CoV-2 populations from USA from year 2021.

Mutations in NS3 protein
In the SARS-CoV-2 populations from Asia from year 2020, the D155Y, D173G, D250T, G172V, G251V, H243R, I263M, I62T, K21N, K75E, L140F, L140V, L53F, L94P, M125I, P25S, Q38E, Q57H, R6I, S171L, S220I, T175I, T208P, T223I, T24I, V259L, W128L, W131C and W149Lmutations were detected in NS3 proteins. The Q57H was found 59 times in NS3 proteins. A total of 306 mutations were identified in NS3 proteins from 2021 up to Jan, 2022. The mutations that were found in maximum number are S26L (127), Q57H (42) and S171L (27). A total of 62 mutations i.e. A110V, E239D, G196E, G196V, G251D, H182Y, K21N, L15F, L53F, L65F, P240H, Q185H, Q213K, Q57H, S171L, S92L, T229I, V256I and W45L were detected in NS3 proteins from Europe 2020. The Q57H was found 30 times in NS3proteins, S26L was found 36 times and G251V was found 8 times in NS3 proteins. From Europe, six mutations D142G (1), D155Y (1), P25S (1) and S26L (3) were identified from Slovakia submitted in the month of Jan, 2022. The A110S, A98S, D155Y, D22Y, E239Q, G172C, G224V, G251V, G254V, I62T, P104S, Q57H, R122I, R134C, R68I, S165F, S171L, S216P, S253P, S26L, T175I, T217I, V225A, V29A and Y184C mutations were detected in NS3 proteins from North America. The S26L was found 20 times in NS3 proteins from Canada 2021 (2) and Jamaica 2021 (18). The Q57H was found 9 times in NS3 proteins from Jamaica 2020 (3), Mexico 2020 (2), Canada 2021 (1) and Mexico 2021 (3). A total of 43 mutations were identified in NS3 proteins from South America and the detected mutations were A23V, G196V, G251V, H78Y, L106F, L127F, L83F, P104L, Q57H, S171L, S253P, S74F, T151I, T223I, V80G, W149C and Y264N. The S253P was found 20 times in NS3 proteins from Brazil 2020 (2), Brazil 2021 (8), Peru 2021 (2) and Venezuela 2021 (8). The Q57H was found 6 times in NS3 proteins from Brazil 2020 (1), Chile 2020 (2), Colombia 2020 (1), Peru 2020 (1) and Venezuela 2020 (1). From Oceania, the A54V, G172V, G76S, I35T, Q185H, Q57D, S166L, S180P, S253P, T12I, T175I, V112L, V13L and V202I mutations were detected in NS3 proteins. All the identified mutations were from New Zealand 2020 and 2021. The Q57H was identified 6 times in NS3 proteins from New Zealand 2020 (5) and New Zealand 2021 (1). A total of 36 mutations were detected in NS3 proteins from Africa. The Q57H was found 14 times in NS3 proteins from Egypt 2020 (4), Kenya 2020 (1), South Africa 2020 (2), Egypt 2021 (4), Ghana 2021 (1) and Kenya 2021 (2). The S171L was found 6 times in NS3 proteins from Egypt 2020 (1), South Africa 2020 (2), Egypt 2021 (1), Ghana 2021 (1) and Kenya 2021 (1).
In the sequences of year 2021 from USA, a total of 24 mutations were identified. Among them, the S26L (11) and Q57H (4) were found in multiple times. Some other NS3 mutations i.e. M20I, P134S, R216C and T140I were detected in sequences from Jan 2022.

Mutations in NS6 protein

The mutations in NS6 protein were identified in different SARS-CoV-2 populations from Asia, South America and North America. The distinct type of mutations that were detected in NS6 are; D53Y, F2S, L4I, M19T, T10I, W27L, I33T, M58V and Q8P. From Asia, a total of 19 mutations i.e. F2S (10) (Saudi Arabia 2020 & 2021), I33T (3) (India 2020 & India 2021), M19T (1) (Japan 2022), M58V (3) (China 2020 & China 2021), Q8P (1) (Philippines 2020), T10I (1) (Pakistan 2021) and W27L (1) (Malaysia 2021) were observed in NS6 proteins. From South America, a single mutation I33T (n=17) was identified from Brazil 2020 (15), Argentina 2020 (1) and Chile 2020 (1). From North America, I33T mutation (1) was observed only from Canada 2021. From Africa, only two mutations, D53Y (1) (Nigeria2021) and L4I (1) (Ghana 2020) were detected.
In the sequences of year 2021 from USA, a total of 34 mutations were identified. Among them, the T120I (11) and V82A (11) were found in multiple times. Two other NS3 mutations i.e. T120I and V82A were detected in sequences from Jan 2022.

Mutations in NS7a protein

In NS7a, A105V, C67Y, E16D, L17F, L96F, R80I, S37F, T28I and V93F mutations were detected in SARS-CoV-2 populations from Asia from 2020. A total of 284 mutations were identified in NS7a proteins from Asia from year 2021 up to 2022. The V82A (118), T120I (111) and L116F (28) were found in maximum numbers. In most of the cases, the T120I and V82A were found associated with each other. The A13T (2) (Georgia 2020), T14I (1) (Russia 2020), V104F (1) (Turkey 2020) and V93F (1) (Russia 2020) were identified in NS7a proteins from Europe 2020. From year 2021, the E121Q, G38V, H47Y, L116F, L5F, P34L, P45L, Q62H, Q94L, Q94R, R89I, S44L, T120I and V82Amutations were detected. The T102I was found 30 times in NS7a proteins from Italy (1), Norway (1), Portugal (2), Slovakia (21), Switzerland (4) and Turkey (1). The V82A was found 27times in NS7a proteins from Norway (1), Portugal (2), Slovakia (20) and Switzerland (4). From year Jan, 2022, five mutations T120I (3) and V82A (2) were identified from Slovakia 2022 till Jan, 2022. From North America from year 2021, theA13T (1) (Jamaica), P84S (1) (Jamaica), S81L (1) (Mexico), T11I (1) (Canada), T120I (19) (Canada and Jamaica), V71I (1) (Jamaica) and V82A (20) (Canada and Jamaica) mutations were detected in NS7a proteins. The three mutations, S37T (1) (Peru 2020) and V29L (2) (Brazil 2020 and Peru 2020) were identified in NS7a proteins from South America. The L96F, P99S, Q21R, S37F, S83L and V24F mutations were detected in NS7a proteins from Oceania. All the identified mutations were from New Zealand from years 2020 (1) and 2021 (7). From Africa, the D69G, I10V, L116F, P99L, T120I, T28I, V82A and V93F were detected in NS7a proteins.

Mutations in NS7b protein

In case of NS7b protein, the A15S (1) (Pakistan), C41F (5) (Thailand), E39Q (1) (Pakistan), H37R (1) (India), L32F (1) (Bangladesh) and L4F (1) (India) mutations were detected in SARS-CoV-2 populations from Asia from year 2020. A total of 77 mutations i.e. A43S (2) and T40I (75) were identified in NS7b proteins from Asia from year 2021 up to 2022. A single mutationH37Y (Poland 2020) was found in a single SARS-CoV-2 isolate from Europe from year 2020. The D8C (1) (Russia2021), F9A (1) (Russia2021), L32F (2) (Finland2021), S5L (1) (Spain2021), T40I (26) and Y10F (1) (Russia 2021) mutations were identified in NS7b proteins from Europe from year 2021. The T40I was found 26 times in NS7b proteins from Portugal 2021 (2), Slovakia 2021 (19), Switzerland 2021 (4) and Turkey 2021 (1). From Europe from Jan, 2022, the T40I mutation was identified in 3 NS7b proteins from Slovakia submitted in the month of Jan, 2022. The T40I (Jamaica 2021) was identified 17 times in NS7b proteins from North America. From South America, only two mutations E33A and V21F were identified from Brazil from year 2020. From Africa, only 3 mutations i.e.F13L (1) (Ghana 2020) and F9S (2) (Egypt 2020 and Egypt 2021) were detected in NS7bproteins.
In the sequences of year 2021 from USA, a total of 9 mutations were identified. Among them, the T40I was found eight times. One other NS7b mutation i.e. T40I was also detected in a single sequence from Jan 2022.

Mutations in NS8 protein

In SARS-CoV-2 populations from Asia from year 2020, a total of 55 mutations i.e. A51V, C20F, H17Q, I9T, L109F, L4P, L84S, R52I, S24L, S67F, V33F, V62L and Y73C were identified in NS8 proteins. The L84S was found35timesinNS8 proteins from Cambodia (1), China (4), Hong Kong (2), India (1), Japan (1), South Korea (12), Taiwan (2), and Thailand (12). A total of 158 mutations were identified in SARS-CoV-2 populations from year 2021 up to Jan, 2022. TheY73C (49) and R52I (49) were always found accompanied with each other. A total of 22 mutations were identified in SARS-CoV-2 populations from Europe from year 2020 whereas 76 mutations from year 2021. From year 2021, the R52I (31) and Y73C (31) mutations were detected in NS8 proteins from Denmark, Finland, Germany, Italy, Netherlands, Poland, Slovakia, Spain and Turkey.

From Europe from Jan, 2022, the T40I (3) mutation was identified from Slovakia. From North America, the A65V, C61F, E92K, F120V, F16S, G66C, I121L, L60F, R52I, S67F, S69L, V32L, W45L and Y73C mutations were detected in NS8 proteins. From South America, A51S, A65S, E92K, F6S, I39T, L84S, R52I, Y73C and T26I mutations were identified in NS8 proteins. The E92K was found 19 times in NS8 proteins from Brazil 2020 (1), Brazil 2021 (8), Peru 2021 (2) and Venezuela 2021 (8). The E92K, L84S, R52I, S24L and Y73C mutations were identified in NS8 proteins from Oceania. Each R52I and Y73C mutations were found 12 times in NS8 proteins from New Zealand from year 2021. A total of 122 mutations were detected in NS8 proteins from Africa. Each R51I and Y73C mutations were found 55 times in NS8 proteins from Ghana 2020 (1), Ghana 2021 (11), Kenya 2021 (39) and Nigeria 2021(4).

3.3 Deletion analysis of SARS-CoV 2 genome

Structural proteins deletions

From year 2020, the T676del, T678del, S71del, S680del, A243del, G72del, G75del, H69del, I68del, L242del, L244del, Q677del, Q675del, N74del, N679del, N148del, K147del, H146del, F140del, T73del, T76del, V70del, Y144del and Y145del were identified in spike proteins. Among the identified deletions, the H69 (16), V70del (15), T676del (10), T678del (10), Q677del (10), N679del (10) and Y144del (9) were detected in maximum numbers of spike proteins. From year 2021, a total of 2721 deletions were identified in spike proteins from year 2021. The V70del (179), H69del (179), Y144del (178), F157del (117), R158del (114), A243del (30), L242del (30) and L244del (30) were identified multiple times. From Jan, 2022, the total numbers of the identified deletions were 235. The identified deletions were H69del (33), V70del (33), V143del (33), Y144del (33), Y145del (33), N211del (27), F157del (20), R158del (20), L24del (1), P25del (1) and P26del (1). The three deletions, H69del, V70del and Y144del were identified as co-occurring deletions. The Y144del was identified in total of 9 spike proteins from Ghana, Hong Kong, India, Taiwan, Finland, Germany, Netherlands, Poland and Switzerland from year 2020. But the same deletion was detected in 178 spike proteins from year 2021. The maximum numbers of Y144del were identified in spike proteins from Kenya (39), Italy (13), Mexico (13), Japan (13), New Zealand (12) and Ghana (10). From Jan, 2022, the Y144del was identified in 33 spike proteins from Japan (16), Pakistan (9) and Slovakia (8). The H69del was identified in 16 spike proteins from Ghana (1), Hong Kong, Taiwan (4), Belgium, Czech Republic, Finland, France, Germany (2), Netherlands, Poland, Russia and Switzerland from year 2020. The H69del deletion was detected in maximum numbers of spike proteins from Kenya (39), Taiwan (13), Mexico (13), Japan (13), New Zealand (12), Nigeria (10) and Ghana (10) from year 2021. From Jan, 2022, the H69del was identified from Japan (17), Pakistan (9) and Slovakia (8). The V70del was identified in 15 spike proteins from year 2020, 179 spike proteins from year 2021 and 33 spike proteins from Jan, 2022. In the SARS-CoV 2 populations from year 2021, the deletions were identified at 31, 32, 33, 211, 212 and 213 positions in case of N proteins. Each E31del, R32deland S33del was detected from Belgium (1) (NCBI No. OL672836.1), Canada (1) (NCBI No. OM131552.1), China (2) (NCBI No. OL988214.2, OM095411.1), Hong Kong (2) (NCBI No. OM212473.1, OM403309.1) and Japan (2) (NCBI No. BS001392.1, BS001397.1). Our analysis showed that these three deletions were found as co-occurring mutations. Each A211del, G212del and N213 del was detected only one time from in N protein (NCBI No. OK433257.1) from New Zealand from year 2021. From Jan, 2022, a total of 102 deletions were identified in N proteins at 31, 32 and 33 positions. All three deletions (E31del, R32del and S33del) were found in N proteins from Japan (51), Pakistan (27) and Slovakia (24). No any deletion was found in spike proteins from envelope proteins and membrane proteins.

Non-structural proteins deletions

In case of NSP1protein, a total of 8 types of deletions (G82del, H83del, V84del, M85del, V86del, K141del, S142del and F143del) were identified from years 2020 and 2021. Each G82del (2), H83del (2) and V84del (2) were found in NSP1 proteins from South Korea (3) (NCBI No. MT810119.1) from year 2020 and Slovakia (3) (NCBI No. OU857679.1) from year 2021. These three deletions were found as co-occurring mutations. The M85del was identified in NSP1 proteins from South Korea 2020 (1) (NCBI No. MT810119.1), Spain 2020 (1) (NCBI No. MZ902042.1) and Slovakia 2021 (1) (NCBI no. OU857679.1). The V86del was identified from Georgia 2020 (1) (NCBI No. MT786455.1), Spain 2020 (1) (NCBI No. MZ902042.1) and Slovakia 2021 (1) (NCBI No. OU857679). The three other deletions that were reported in NSP1 protein are K141del, S142del and F143del. All the three deletions were found to occur with the same frequency and also found accompanied with each other. These three deletions were identified from Spain 2020 (6) (NCBI No. MZ082987 and MZ099821), Japan 2021 (6) (NCBI No. BS002300 and BS002298) and Nigeria 2021 (3) (NCBI No. OL601833). A single deletion, D268del was found in 3 NSP2 proteins from Netherlands 2020 (MT457398.1,

MT457397.1 and MT457399). The S1265del was identified in 40 NSP3 proteins from Belgium 2021 (1) (NCBI No. OL672836.1), Canada 2021 (1) (NCBI No. OM131552.1), China 2021 (2) (NCBI No. OL988214.2 and OM095411.1), Hong Kong 2021 (2) (NCBI No. OM403309.1 and OM212473.1), Japan 2021 (2) (NCBI No. BS001392.1 and BS001397.1), Japan 2022 (15), Pakistan 2022 (9) and Slovakia 2022 (8). From Malaysia, four deletions (E206del (1), V207del (1), N208del (1) and S209del (1)) were detected in a single NSP3 protein (NCBI No. OM319603.1) from year 2021. The N1313del, S1314del, V1315del, P1316del, W1317del, D1318del, T1319del, I1320del and A1321del were identified in a single NSP3 protein (NCBI No. OM319605.1) from Malaysia from Jan, 2022. Four deletions (I205del, E206del, V207del and N208del) were reported in a single NSP3 protein (NCBI No. OM327536.1) from Pakistan from Jan, 2022. In a single NSP4 protein (NCBI No. MZ314348.1), three deletions i.e. D217del, F216del and S218del were identified from Iran from year 2020. Each F42del, G48del, H62del, H64del, I49del, I50del, K61del, K63del, S53del, N40del, M52del and M47del was identified two times whereas each A41del and A51del was identified four times in NSP6 proteins from Netherlands from year 2020. The S106del, G107del and F108del were identified as co-occurring deletions in 24 NSP6 proteins from South Africa (6), Hong Kong (3), Taiwan (3), Finland (3), Germany (3), Netherlands (3) and Poland (3). A total of 754 NSP6 deletions were identified from year 2021. Each A41del, A46del, A51del, A54del, F42del, G48del, H62del, H64del, I49del, I50del, K61del, K63del, M47del, M52del, N40del and S53 del was identified only one time in NSP6 protein from Slovakia. The S106del, G107del and F108del were identified as co-occurring deletions. These three deletions were found in maximum numbers of spike proteins from Kenya (40), Bangladesh (19), Mexico (17), Italy (14), New Zealand (14), Japan (11), Nigeria (10) and Ghana (10). A single deletion, L105del was identified in 9 spike proteins from Belgium (1), Canada (1), China (2), Hong Kong (2) and Japan (3). A total of 102 NSP6 deletions were identified in the month of January, 2022. The F108del was identified in only one NSP6 protein from Japan. Each G107del and S106del was identified in 34 NSP6 proteins from Japan (17), Pakistan (9) and Slovakia (8) whereas L105del was identified in 33 NSP6 proteins.

In case of NS7aprotein, only four deletions (F63del, F101del, L102del, and I103del) were identified. The F63del (1) was found in single NS7a protein (NCBI No. OU857679.1) from Slovakia from year 2021. Both F101del (1) and L102del (1) were also detected in a single NS7a protein (NCBI No. OV410126.1) from Slovakia from year 2021. The I103del (NCBI No. BS000708.1) was identified only one time in NS7a protein from Japan from year 2020. Two deletion mutations (L14del and E39del) were identified in two NS7b proteins at positions 14 and 39. The L14del (1) (NCBI No. OL757809.1) was identified from Myanmar from year 2021 whereas E39del (1) (NCBI No. MW425837.1) was identified from India from year 2020. Both G66del (2) (NCBI No. OM214020.1 and MZ328043.1) and S67del (2) (NCBI No. OM214020.1 and MZ328043.1) were identified in NS8 proteins from Malaysia and Pakistan from year 2021. Our results showed that these two deletions were only found in Asian continent.

3.4 Insertion mutations of SARS-CoV 2 genome

From year 2020, the ins37F (MT507793.1) was identified in a single NSP6 protein from Jamaica whereas ins79M (LR883018.1 and LR883157.1) was identified in 2 NSP6 proteins from Netherlands. One other insertion mutation i.e. ins214XXXX (MZ314996.2) was identified from Germany. From year 2021, ins79M(OU475092.1) was found in one NSP6 protein from Slovakia, ins679GIAL (MW750606.1 and MW750605.1) in 2 spike proteins from Russia and ins214EPE in 15 spike proteins from Belgium (OL672836.1), Canada (OM131552.1), China (OM095411.1 and OL988214.2), Hong Kong (OM403309.1 and OM212473.1), India (6) and Japan (3). From Jan, 2022, ins214EPE was identified in 25 spike proteins from Japan (16) and Pakistan (9).

3.5 Stop codon mutations of SARS-CoV 2 genome

A total of 17 stop codon mutations were detected from year 2020. The E39stop was identified in 2 NS7b proteins from Pakistan and Taiwan. The E106stop was identified in only one NS8 protein from Philippines. The K68stop was found in 3 NS8 proteins from Ghana, Netherlands and Poland. The Q27stop was identified in 8 NS8 proteins from Ghana, Hong Kong, Taiwan Finland, Germany, Netherlands, Poland and Switzerland. The P73stop was found in one NSP6 protein from Netherlands. The G1267stop was identified in one spike protein from Taiwan. From year 2021, a total of 307 stop codon mutations were detected. The K68stop was detected in 106 NS8 proteins. The maximum numbers of K68stop were identified from Australia (20), Kenya (20), Bangladesh (12), Ghana (11), Slovakia (6), Netherlands (5) and New Zealand (5). The Q27stop was identified in 200 NS8 proteins and maximum numbers from Australia (39), Kenya (39), Bangladesh (15),

Mexico (14), New Zealand (12), Ghana (11) and Japan (10). The P73stop was identified in one NS6 protein from Slovakia.

4. Discussion

The R24C and E57K mutations were detected in NSP1 proteins from two continents (Asia and South America). But the I114T mutation was detected in NSP1 proteins from Africa, North America and South America. The T78I mutation was identified in NSP1 proteins from South America and Oceania. Mou et al., (2021) reported the R24C, H45Y, M85V, K120N, D75E, L92F and H110Y mutations as the most common in the flexible loop of NSP1 proteins from different countries (37). The R24C, M85V and K120N mutations were identified as destabilizing mutations and H45Y, D75E, L92F and H110Y as stabilizing mutations. In our analyses, the highly variable detected mutation was T85I in NSP2 proteins. This mutation was identified in SARS-CoV-2 isolates from all the six continents. The V198I mutation was detected in NSP2 proteins from Asia, Europe and Oceania. The L550F mutation was identified in NSP2 proteins from Asia, North America and Oceania. The K81N mutation was found in NSP2 proteins from Asia and Europe. The T429I was observed in NSP2 proteins from Asia and South America. A recent study by Patro et al., (2021) showed the occurrence of T85I mutation in 47 different countries deposited on or before May 17, 2020 (38). The V378I, A206T, R207C and T265I mutations were reported in NSP2 proteins from Turkey (39). The most commonly found mutation was A890D in NSP3 proteins from all six continents. The A488S, A1469S and P1228L mutations were found in NSP3 proteins from Asia, Europe and North America. The I1412T and T183I mutations were found in NSP3 proteins from Asia, Europe, North America, Africa and Oceania. Nagy et al., (2021) reported two mutations, S1197R and T1198K in NSP3 proteins (40). Two mutations (F105F and G251V) were reported in 558 different isolates of SARS-CoV-2 genomes up to March 23, 2020 (Y). The G251V was detected as the prominent mutation. In a study by Rehman et al., (2020), two new mutations (P1213L and A1824V) were also detected from Turkey. Both of the reported mutations were missense mutations (39).

The T1492I (75) mutation was found in high frequency in case of NSP4 proteins. This mutation was identified in SARS-CoV-2 populations from Asia, Europe, North America and South America. The V167L mutation was found in 47 NSP4 proteins from three continents (Asia, Europe and North America). The A446V mutation was found 23 times in NSP4 proteins from Asia, North America and Oceania. The G15S mutation was identified in NSP5 proteins from Asia, Europe, Africa, South America and Oceania. The K90R and P132H mutations were identified in NSP5 proteins from Asia, Europe, Africa, North America and Oceania. The L75F mutation was found in NSP5 proteins from Europe, North America and South America. The L89F mutation was detected in NSP5 proteins from Asia and Oceania. The A234V mutation was identified in NSP5 proteins from Asia and Europe. The G15S and K90R mutations were detected in a study by Patro et al., (2021) (38). The G15S mutation was observed on the outer exposed portion of Nsp5 protein. It interferes in the interaction with other virus proteins and also with host proteins. Yashvardhini et al., (2021) identified a total of 33 mutations in 675 isolates of SARS-CoV-2 genomes submitted from India in March 2020 to April 2021 but selected only eight prominent mutations (K236R, N142L, K90R, A7V, L75F, C22N, H246Y and I43V) for further analysis (41). The L37F mutation was found highly variable mutation in NSP6 proteins from Asia, Europe, Africa, South America, North America and Oceania. The T77A mutation was identified in NSP6 proteins from Asia, Europe and North America. The recent studies demonstrated the presence of L37F mutation in NSP6 proteins (38, 40, 42, 43). This L37F was found a prominent mutation in Asian countries at the early pandemic of this novel virus (44). The T85I mutation was also reported in NSP6 protein (38). The L3606F mutation (45) was detected in 25 isolates from Turkey (Rehman et al., 2020). The mutation was also reported from Japan, Netherlands and Australia (42, 46). The L71F mutation was detected found 12 times in case of NSP7 proteins from Asia and South America. The S25L mutation was identified in 12 NSP7 proteins from Asia, North America and South America. In recent studies, the rapid occurrence of L71F mutation was reported in the NSP7 proteins (40). A total of 23 different mutations were detected from 218 isolates submitted up to July 14, 2020 (47). Among all, only two mutations, (S25L and S26F) were reported as prominent ones.

The A14V mutation was identified in NSP8 proteins from Asia and Europe. The 145I mutation was identified in NSP8 proteins from Europe and North America. The P133S was identified in NSP8 proteins from Asia and South America. The T148I mutation was identified in NSP8 proteins from Asia and Africa. The Q158H mutation was identified in NSP8 proteins from Asia and Oceania. A total of 34 different mutations were detected from 130 isolates submitted up to July 14, 2020 (47). Among all, only two mutations, (M129I and I156V) were reported as frequently occurring mutations. The T35I mutation was identified in NSP9 proteins from Europe, North America and Oceania. The T19I mutation was identified in NSP9 from Asia and Oceania. The P80L was detected in NSP9 proteins from Asia and Europe. The T109I was found in NSP9 proteins from

Asia and North America. The L42F mutation was identified in NSP9 proteins from North America and Oceania. The N105K mutation was identified in two NSP10 proteins from Europe and Africa. In a study by Azad, (2020), two A20V and P236L mutations were reported in NSP10 proteins from India. The A20V mutation alters the secondary structure of NSP10 proteins. A unique mutation, S6059F was reported in a study by Oulas et al., (2021) (48).

The P323L, the most variable mutation in case of NSP12 proteins was detected from all six continents. The G671, a second highly variable mutation was identified in NSP12 proteins from Asia, Europe, Africa and North America. The P227L mutation was found in NSP12 proteins from Asia, Europe, Africa, North America and Oceania. The A185V mutation was found in NSP12 proteins from Asia, South America and Oceania. The V776L mutation was identified in NSP12 proteins from Asia, Europe and Oceania. The T26I mutation was detected in NSP12 proteins from Asia, Europe and North America. Sahin et al., (2021) reported P323L, V111A, H133R, Y453C and M626K mutations in 45 different isolates from Turkey. The P323L mutation was found common to all geographic regions of India (49). The P323L, A97V and A449V mutations were reported in different SARS-CoV-2 isolates deposited up to May 17, 2020 (38). Among three mutations, the P323L mutation was detected in 71 countries (69.47%). Out of 71 countries, this mutation was found highly frequent mutation in two countries i.e. England and USA.

The P77L mutation was found highly variable mutation in case of NSP13 proteins from Asia, Europe, North America and Africa. The E341D mutation was identified in 19 NSP13 proteins from South America. The most commonly found mutation was A394V in NSP14 proteins from Asia, Europe, South America and North America. The I42V was the second most variable mutation identified in NSP14 proteins from Asia, Europe, North America and Oceania. The V172L was found the most variable mutation in NSP15 proteins from Cambodia 2020 (1), Japan 2020 (1) and Thailand 2020 (9). The maximum number of NSP15 mutations were H234Y (16), K259R (10), P205L (10) and S261A (9). The V66L was identified in NSP15 proteins from Europe and Africa. The G129V was identified in NSP15 proteins from Asia and Oceania. The K160R mutation was identified in NSP16 proteins from Asia, Europe and South America. The T56I mutation was identified in NSP16 proteins from Asia and Africa. The T35I was identified in NSP16 proteins from Asia and Europe. The Q238H mutation was found in NSP16 proteins from Africa and North America. The Q57H mutation was found the most variable mutation in 166 NS3 proteins from all six continents. The S126L was the second most variable mutation identified in NS3 proteins from Asia, Europe and North America. The G251V was identified in NS3 proteins from Asia, Europe, North America and South America. The S253P was identified in NS3 proteins from South America and Oceania. The I33T mutation was detected in 21 NS6 proteins from Asia, South America and North America. The Y73C and R52I mutations were always found accompanied with each other in NS8 proteins.

The deletions were identified from S, N, NSP1, NSP2, NSP3, NSP4, NSP6, NS7A, NS7b and NS8 proteins. Among the identified deletions from year 2020, the H69 (16), V70del (15), T676del (10), T678del (10), Q677del (10), N679del (10) and Y144del (9) were detected in maximum numbers of spike proteins. From year 2021, the V70del (179), H69del (179), Y144del (178), F157del (117), R158del (114), A243del (30), L242del (30) and L244del (30) were identified multiple times. The three deletions, H69del, V70del and Y144del were identified as co-occurring deletions in S proteins from Jan, 2022. The E31del, R32del and S33del were detected as co-occurring mutations in N proteins from year 2020 and Jan, 2022. The G82del (2), H83del (2) and V84del (2) were found as co-occurring mutations in NSP1 proteins. The S106del, G107del and F108del were identified as co-occurring deletions in 24 NSP6 proteins.

5. Conclusion

In conclusion, we compared the mutations, stop-codon mutations, deletions and insertions identified in complete genomes of SARS-CoV-2 from six continents. Constant monitoring of mutations will also be pivotal in tracking the movement of the virus between individuals and across geographical areas. These results emphasize the importance of timely monitoring of genetic variation and its effect on disease severity. The S1B domain (333-527) of the spike protein was found the highest mutating region in the entire genome whereas NS6 protein was found the lowest mutating region. Interestingly, no any mutation was detected from NSP11 protein. In structural proteins, high frequencies of the mutations were identified from spike and nucleocapsid proteins whereas less from E and M proteins. But in case of non-structural proteins, majority of the mutations were detected from NSP2, NSP3, NSP6, NSP12 and NS3 proteins. The D614G from spike protein, T81I from NSP2, A890D from NSP3, L37F from NSP6, P323L from NSP12, Q57H from NS3 and Y73C from NS8 were identified in maximum numbers of SARS-CoV-2 populations from all six continents. Many co-occurring mutations were detected in spike proteins, N proteins and NSP12 proteins. The deletions were only found in

S, N, NSP1, NSP2, NSP3, NSP4, NSP6, NS7a and NS7b proteins. The co-occurring deletions were identified only in N, NSP1 and NSP6 proteins. A few insertion mutations were identified in spike proteins and NSP6 proteins. But the high prevalence of stop-codon mutations was detected in spike, NSP6, NS7b and NS8 proteins. Thus, constant monitoring of mutations will be significant in tracking the movement of the virus between individuals and across geographical areas. To gain a better understanding of the possible function in terms of virulence, infectivity, and virus release, a complete investigation would be useful. Therefore, the present study might be of great interest to the researchers/companies working to develop therapeutics against SARS-CoV-2 as well as gaining fundamental insights into pathogenesis of the virus.

References:

1. Phan MV, Ngo Tri T, Hong Anh P, Baker S, Kellam P, Cotten M. Identification and characterization of Coronaviridae genomes from Vietnamese bats and rats based on conserved protein domains. Virus evolution. 2018 Jul;4(2):vey035.
2. Tang Q, Song Y, Shi M, Cheng Y, Zhang W, Xia XQ. Inferring the hosts of coronavirus using dual statistical models based on nucleotide composition. Scientific reports. 2015 Nov 26;5(1):1-8.
3. Bakkers MJ, Lang Y, Feitsma LJ, Hulswit RJ, De Poot SA, Van Vliet AL, Margine I, de Groot-Mijnes JD, Van Kuppeveld FJ, Langereis MA, Huizinga EG. Betacoronavirus adaptation to humans involved progressive loss of hemagglutinin-esterase lectin activity. Cell host & microbe. 2017 Mar 8;21(3):356-66.
4. Assiri A, McGeer A, Perl TM, Price CS, Al Rabeeah AA, Cummings DA, Alabdullatif ZN, Assad M, Almulhim A, Makhdoom H, Madani H. Hospital outbreak of Middle East respiratory syndrome coronavirus. New England Journal of Medicine. 2013 Aug 1;369(5):407-16.
5. Song Z, Xu Y, Bao L, Zhang L, Yu P, Qu Y, Zhu H, Zhao W, Han Y, Qin C. From SARS to MERS, thrusting coronaviruses into the spotlight. viruses. 2019 Jan 14;11(1):59.
6. Rota PA, Oberste MS, Monroe SS, Nix WA, Campagnoli R, Icenogle JP, Penaranda S, Bankamp B, Maher K, Chen MH, Tong S. Characterization of a novel coronavirus associated with severe acute respiratory syndrome. science. 2003 May 30;300(5624):1394-9.
7. Zaki AM, Van Boheemen S, Bestebroer TM, Osterhaus AD, Fouchier RA. Isolation of a novel coronavirus from a man with pneumonia in Saudi Arabia. New England Journal of Medicine. 2012 Nov 8;367(19):1814-20.
8. Zhu N, Zhang D, Wang W, Li X, Yang B, Song J, Zhao X, Huang B, Shi W, Lu R, Niu P. A novel coronavirus from patients with pneumonia in China, 2019. New England journal of medicine. 2020 Jan 24.
9. World Health Organization. COVID-19 weekly epidemiological update edition 76, published 25 January 2022 [Internet]. Geneva: WHO; 2022 [cited 31 March 2022].
10. Wu F, Zhao S, Yu B, Chen YM, Wang W, Song ZG, Hu Y, Tao ZW, Tian JH, Pei YY, Yuan ML. A new coronavirus associated with human respiratory disease in China. Nature. 2020 Mar;579(7798):265-9.
11. Zhou P, Yang XL, Wang XG, Hu B, Zhang L, Zhang W, Si HR, Zhu Y, Li B, Huang CL, Chen HD. A pneumonia outbreak associated with a new coronavirus of probable bat origin. nature. 2020 Mar;579(7798):270-3.
12. Zumla A, Chan JF, Azhar EI, Hui DS, Yuen KY. Coronaviruses—drug discovery and therapeutic options. Nature reviews Drug discovery. 2016 May;15(5):327-47.
13. 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 Apr 16;181(2):281-92.
14. Chen L, Liu W, Zhang Q, Xu K, Ye G, Wu W, Sun Z, Liu F, Wu K, Zhong B, Mei Y. RNA based mNGS approach identifies a novel human coronavirus from two individual pneumonia cases in 2019 Wuhan outbreak. Emerging microbes & infections. 2020 Jan 1;9(1):313-9.
15. Wrapp D, Wang N, Corbett KS, Goldsmith JA, Hsieh CL, Abiona O, Graham BS, McLellan JS. Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. Science. 2020 Mar 13;367(6483):1260-3.
16. Wan Y, Shang J, Graham R, Baric RS, Li F. Receptor recognition by the novel coronavirus from Wuhan: an analysis based on decade-long structural studies of SARS coronavirus. Journal of virology. 2020 Mar 17;94(7):e00127-20.
17. Hoffmann M, Kleine-Weber H, Schroeder S, Krüger N, Herrler T, Erichsen S, Schiergens TS, Herrler G, Wu NH, Nitsche A, Müller MA. SARS-CoV-2 cell entry depends on ACE2 and TMPRSS2 and is blocked by a clinically proven protease inhibitor. cell. 2020 Apr 16;181(2):271-80.
18. Matsuyama S, Nao N, Shirato K, Kawase M, Saito S, Takayama I, Nagata N, Sekizuka T, Katoh H, Kato F, Sakata M. Enhanced isolation of SARS-CoV-2 by TMPRSS2-expressing cells. Proceedings of the National Academy of Sciences. 2020 Mar 31;117(13):7001-3.

19. Surya W, Li Y, Torres J. Structural model of the SARS coronavirus E channel in LMPG micelles. Biochimica et Biophysica Acta (BBA)-Biomembranes. 2018 Jun 1;1860(6):1309-17.
20. Neuman BW, Kiss G, Kunding AH, Bhella D, Baksh MF, Connelly S, Droese B, Klaus JP, Makino S, Sawicki SG, Siddell SG. A structural analysis of M protein in coronavirus assembly and morphology. Journal of structural biology. 2011 Apr 1;174(1):11-22.
21. Ujike M, Taguchi F. Incorporation of spike and membrane glycoproteins into coronavirus virions. Viruses. 2015 Apr 3;7(4):1700-25.
22. Tsoi H, Li L, Chen ZS, Lau KF, Tsui SK, Chan HY. The SARS-coronavirus membrane protein induces apoptosis via interfering with PDK1–PKB/Akt signalling. Biochemical Journal. 2014 Dec 15;464(3):439-47.
23. Surjit M, Lal SK. The nucleocapsid protein of the SARS coronavirus: structure, function and therapeutic potential. In Molecular biology of the SARS-coronavirus 2010 (pp. 129-151).
24. Cubuk J, Alston JJ, Incicco JJ, Singh S, Stuchell-Brereton MD, Ward MD, Zimmerman MI, Vithani N, Griffith D, Wagoner JA, Bowman GR. The SARS-CoV-2 nucleocapsid protein is dynamic, disordered, and phase separates with RNA. Nature communications. 2021 Mar 29;12(1):1-7.
25. McBride R, Van Zyl M, Fielding BC. The coronavirus nucleocapsid is a multifunctional protein. Viruses. 2014 Aug 7;6(8):2991-3018.
26. Ahmadpour D, Ahmadpoor P. How the COVID-19 overcomes the battle? An approach to virus structure. Iranian journal of kidney diseases. 2020 May 1;14(3):167.
27. Hartley GE, Edwards ES, Aui PM, Varese N, Stojanovic S, McMahon J, Peleg AY, Boo I, Drummer HE, Hogarth PM, O'Hehir RE. Rapid generation of durable B cell memory to SARS-CoV-2 spike and nucleocapsid proteins in COVID-19 and convalescence. Science immunology. 2020 Dec 22;5(54):eabf8891.
28. Grubaugh ND, Petrone ME, Holmes EC. We shouldn't worry when a virus mutates during disease outbreaks. Nature microbiology. 2020 Apr;5(4):529-30.
29. Li F. Structure, function, and evolution of coronavirus spike proteins. Annual review of virology. 2016 Sep 29;3(1):237. https://doi.org/10.1146%2Fannurev-virology-110615-042301.
30. Li F. Receptor recognition and cross-species infections of SARS coronavirus. Antiviral research. 2013 Oct 1;100(1):246-54.
31. Graham RL, Baric RS. Recombination, reservoirs, and the modular spike: mechanisms of coronavirus cross-species transmission. Journal of virology. 2010 Apr 1;84(7):3134-46.
32. Boehm E, Kronig I, Neher RA, Eckerle I, Vetter P, Kaiser L. Novel SARS-CoV-2 variants: the pandemics within the pandemic. Clinical Microbiology and Infection. 2021 Aug 1;27(8):1109-17.
33. Shang J, Ye G, Shi K, Wan Y, Luo C, Aihara H, Geng Q, Auerbach A, Li F. Structural basis of receptor recognition by SARS-CoV-2. Nature. 2020 May;581(7807):221-4.
34. Starr TN, Greaney AJ, Hilton SK, Ellis D, Crawford KH, Dingens AS, Navarro MJ, Bowen JE, Tortorici MA, Walls AC, King NP. Deep mutational scanning of SARS-CoV-2 receptor binding domain reveals constraints on folding and ACE2 binding. cell. 2020 Sep 3;182(5):1295-310.
35. Jhun H, Park HY, Hisham Y, Song CS, Kim S. SARS-CoV-2 Delta (B. 1.617. 2) variant: A unique T478K mutation in receptor binding motif (RBM) of spike gene. Immune network. 2021 Oct;21(5).
36. Plante JA, Liu Y, Liu J, Xia H, Johnson BA, Lokugamage KG, Zhang X, Muruato AE, Zou J, Fontes-Garfias CR, Mirchandani D. Spike mutation D614G alters SARS-CoV-2 fitness. Nature. 2021 Apr;592(7852):116-21.
37. Mou K, Mukhtar F, Khan MT, Darwish DB, Peng S, Muhammad S, Al-Sehemi AG, Wei DQ. Emerging mutations in Nsp1 of SARS-CoV-2 and their effect on the structural stability. Pathogens. 2021 Oct 6;10(10):1285.
38. Patro LP, Sathyaseelan C, Uttamrao PP, Rathinavelan T. Global variation in SARS-CoV-2 proteome and its implication in pre-lockdown emergence and dissemination of 5 dominant SARS-CoV-2 clades. Infection, Genetics and Evolution. 2021 Sep 1;93:104973.
39. Rehman SU, Shafique L, Ihsan A, Liu Q. Evolutionary trajectory for the emergence of novel coronavirus SARS-CoV-2. Pathogens. 2020 Mar 23;9(3):240.
40. Nagy A, Pongor S, Gyorffy B. Different mutations in SARS-CoV-2 associate with severe and mild outcome. International journal of antimicrobial agents. 2021 Feb 1;57(2):106272.
41. Yashvardhini N, Kumar A, Jha DK. Immunoinformatics identification of B-and T-cell epitopes in the RNA-dependent RNA polymerase of SARS-CoV-2. Canadian Journal of Infectious Diseases and Medical Microbiology. 2021 Apr 22;2021.
42. Mercatelli D, Giorgi FM. Geographic and genomic distribution of SARS-CoV-2 mutations. Frontiers in

microbiology. 2020 Jul 22;11:1800.
43. Yin W, Mao C, Luan X, Shen DD, Shen Q, Su H, Wang X, Zhou F, Zhao W, Gao M, Chang S. Structural basis for inhibition of the RNA-dependent RNA polymerase from SARS-CoV-2 by remdesivir. Science. 2020 Jun 26;368(6498):1499-504.
44. Wang MY, Zhao R, Gao LJ, Gao XF, Wang DP, Cao JM. SARS-CoV-2: structure, biology, and structure-based therapeutics development. Frontiers in cellular and infection microbiology. 2020 Nov 25;10:587269.
45. Laamarti M, Alouane T, Kartti S, Chemao-Elfihri MW, Hakmi M, Essabbar A, Laamarti M, Hlali H, Bendani H, Boumajdi N, Benhrif O. Large scale genomic analysis of 3067 SARS-CoV-2 genomes reveals a clonal geo-distribution and a rich genetic variations of hotspots mutations. PloS one. 2020 Nov 10;15(11):e0240345.
46. Benvenuto D, Giovanetti M, Ciccozzi A, Spoto S, Angeletti S, Ciccozzi M. The 2019-new coronavirus epidemic: evidence for virus evolution. Journal of medical virology. 2020 Apr;92(4):455-9.
47. Reshamwala SM, Likhite V, Degani MS, Deb SS, Noronha SB. Mutations in SARS-CoV-2 nsp7 and nsp8 proteins and their predicted impact on replication/transcription complex structure. Journal of medical virology. 2021 Jul;93(7):4616-9.
48. Oulas A, Richter J, Zanti M, Tomazou M, Michailidou K, Christodoulou K, Christodoulou C, Spyrou GM. In depth analysis of Cyprus-specific mutations of SARS-CoV-2 strains using computational approaches. BMC Genomic Data. 2021 Dec;22:1-5.
49. Sahin E, Bozdayi G, Yigit S, Muftah H, Dizbay M, Tunccan OG, Fidan I, Caglar K. Genomic characterization of SARS-CoV-2 isolates from patients in Turkey reveals the presence of novel mutations in spike and nsp12 proteins. Journal of Medical Virology. 2021 Oct;93(10):6016-26.