https://doi.org/10.48048/tis.2022.5768

Analysis of Mutational Profiles of SARS-CoV-2 Structural and Non-Structural Proteins with Emphasis on Spike Protein Variants

Saptarshi Bhattacharyya

KIIT School of Biotechnology, Kalinga Institute of Industrial Technology, Odisha 751024, India

(Corresponding author's e-mail: saptarshibhattacharyya84@gmail.com)

Received: 26 May 2021, Revised: 23 July 2021, Accepted: 30 July 2021

Abstract

SARS-CoV-2 has very recently posed a potential threat to humanity due to its very rapid rate of mutations and repairing mechanism. The spread of this virus is considered to have occurred in Wuhan, China in December 2019. Characterized by high rates of transmission, the virus is constantly evolving towards attaining higher rates of stability and transmissibility through acquiring mutations in its genome. Therefore, this study aims to analyse the mutational profiles of SARS-CoV-2 isolates. Analysis of the mutational profiles in individual SARS-CoV-2 proteins will allow us to look into the rates of mutations associated with each protein. Frequently mutated residues have been identified in this research by aligning 688 SARS-CoV-2 nucleotide sequences, which were downloaded from NCBI (National Center For Biotechnology Information) repository. Further, mutational frequencies of these mutated residues have been studied, which is instrumental in identifying the proteins that are resistant to changes, as well as the ones that have a greater proclivity towards incorporating mutations.

Keywords: D614G spike mutation, MEGA X (10.0), Phylogenetic tree, Parsimony informative sites, Synonymous and non-synonymous mutations

Introduction

Coronavirus Disease 2019 (COVID-19) is caused by a novel coronavirus called Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2). It was designated as a pandemic by the WHO (World Health Organization) on April 12, 2020. The primary symptoms of the disease include fever, sore throats, lack of taste and smell among others, which often aggravate into more severe symptoms like dyspnea and SARS-CoV-2 induced pneumonia. The rapid spread of COVID-19 has given rise to a question whether SARS-CoV-2, a large enveloped class of ssRNA viruses with genome of positive sense orientation, has become more transmissible and infectious than before.

To explore this possibility, a short knowledge on the viral activities after its entry into the host cell is a prerequisite. Firstly, the genomic RNA is translated to form non-structural proteins (nsps) from 2 open reading frames (ORFs). The viral genome is also used as a template for replication and transcription via RNA-dependent RNA polymerase activity [13]. While this process continues, negative stranded RNA intermediates are produced to serve as a template for positive sense genomic RNA and sub-genomic RNA synthesis (Figure 2). The shorter sub-genomic RNAs code for structural proteins viz. Spike, envelope, membrane, nucleocapsid etc. (**Figure 1**) [4].

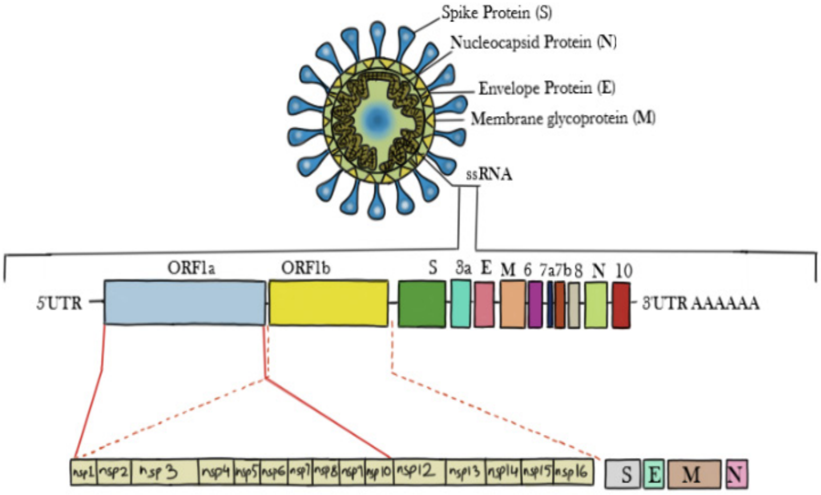

Figure 1 SARS-CoV-2 genome depicting the viral structural and non-structural proteins.

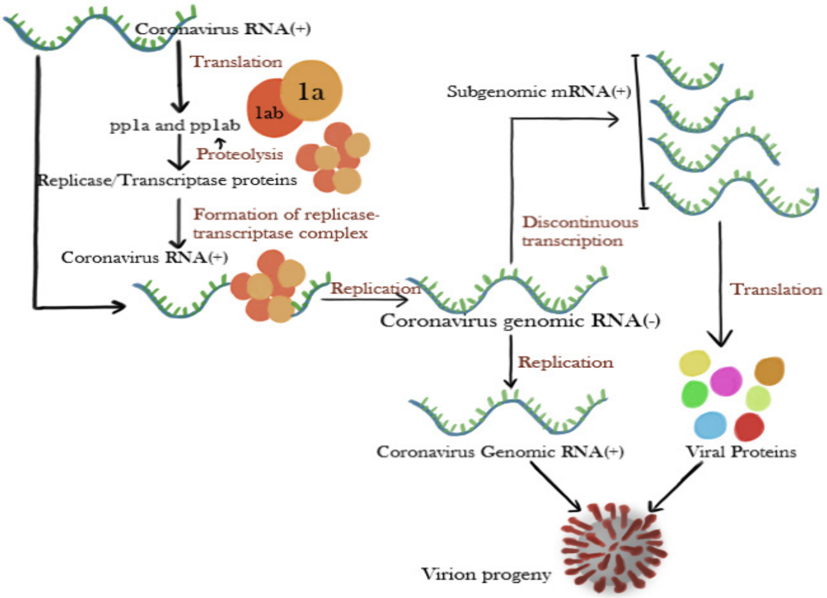

Figure 2 Cyclic representation of SARS-CoV-2 Replication.

In this project, the mutational profiles of these above-mentioned structural and non-structural proteins of SARS-CoV-2 isolates have been analyzed, with special attention to the highly dangerous D614G spike mutation, using sequences obtained from the NCBI. Alignment of the genomic sequences for the individual proteins have been performed separately with systematic tabulation of the synonymous and non-synonymous mutations both at the gene and protein levels, accompanied by a phylogenetic tree. Since, this virus' mutational rate is extraordinarily high, needless to say, its mutational profiles are subjected to change in the future [4]. But, the methodologies followed in this research can lay the base to investigate many recently identified mutational patterns like T85I, P323L, Q57H, L84S, R203K, G204R [18] etc. which are reported to be more infective than the previously identified mutational profiles. Lastly, Studying the evolution of SARS-CoV-2 strains over time and analyzing the mutations could have implications in effective vaccine development.

Materials and methods

Downloading SARS-CoV-2 isolates

A total of 688 SARS-CoV-2 nucleotide sequences were downloaded from NCBI available as of April 15, 2020 (https://www.ncbi.nlm.nih.gov/labs/virus). Partial sequences were excluded prior to downloading, ensuring only complete sequences. Apart from these, sequences with ambiguous sites were removed. The sequences corresponding to each structural and non-structural protein were extracted for downstream analysis. These included sequences from various countries including China (57), USA (585), Australia (1), South Africa (1), Greece (4), Turkey (1), Italy (2), France (1), Iran (1), South Korea (4), Spain (11), Israel (2), Pakistan (2), Peru (1), Colombia (1), Japan (3), Taiwan (3), Vietnam (1), India (2), Brazil (1), Sweden (1) , Nepal (1), and Finland (1) .

Alignment of SARS-CoV-2 protein sequences

MEGA-X (10.0) was used to perform Multiple Sequence Alignment (MSA) of the downloaded sequences. MUSCLE was employed to run the MSA [25]. The SARS-CoV-2 isolate corresponding to GenBank accession ID 'NC_045512' was considered to be the reference sequence with respect to which the alignment was carried out. Each of the viral proteins, including the 4 structural proteins and the open reading frames, were analysed both at the gene and protein levels.

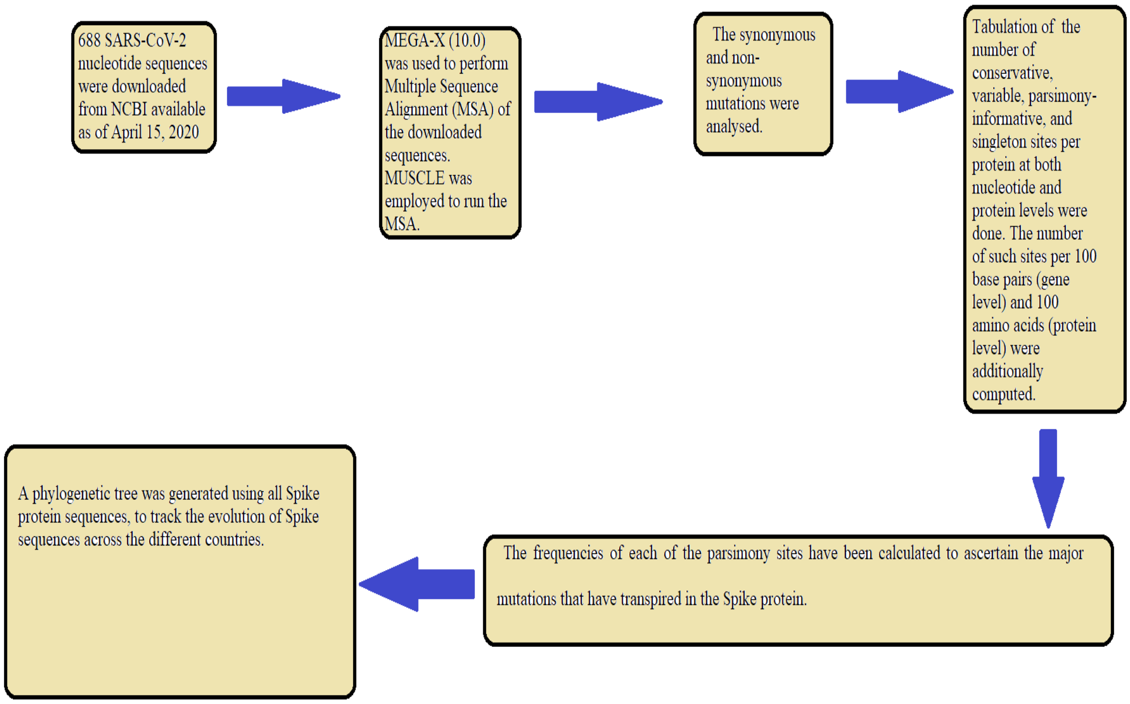
688 SARS-CoV-2 nucleotide sequences were downloaded from NCBI available as of April 15, 2020
MEGA-X (10.0) was used to perform Multiple Sequence Alignment (MSA) of the downloaded sequences. MUSCLE was employed to run the MSA.
The synonymous and non-synonymous mutations were analysed.
Tabulation of the number of conservative, variable, parsimony-informative, and singleton sites per protein at both nucleotide and protein levels were done. The number of such sites per 100 base pairs (gene level) and 100 amino acids (protein level) were additionally computed.
The frequencies of each of the parsimony sites have been calculated to ascertain the major mutations that have transpired in the Spike protein.
A phylogenetic tree was generated using all Spike protein sequences, to track the evolution of Spike sequences across the different countries.

Figure 3 A systematic, pictorial workflow of the procedure followed.

Identifying non-synonymous mutations in SARS-CoV-2 proteins

After alignment of the individual proteins, the synonymous and non-synonymous mutations were analysed. A synonymous mutation is a change in DNA sequence that codes for amino acids in a protein sequence, but does not change the encoded amino acid [2,17]. Similarly, non-synonymous mutations change the protein sequences and are frequently subjected to natural selection [16]. Since non-synonymous mutations only cause change in protein sequences, therefore more focus has been given to its analysis. Tabulation of the number of conservative, variable, parsimony-informative, and singleton sites per protein at both nucleotide and protein levels were done. The number of such sites per 100 base pairs (gene level) and 100 amino acids (protein level) were additionally computed.

It has to be noted that 0-Fold degenerate sites are those at which all changes are non synonymous whereas 2-Fold degenerate sites are those at which one out of 3 changes is synonymous. Also, all sites at which 2 out of 3 changes are synonymous are included in this category. 4-Fold degenerate sites are those at which all changes are synonymous. Conserved sites signify that the sequences have undergone little variation in spite of evolutionary differences [14,15,5]. A variable site contains at least 2 types of nucleotides or amino acids and can be of 2 types, namely singleton and parsimony informative. A singleton site contains at least 2 types of nucleotides or amino acids with at most, one occurring multiple times. Parsimony informative refers to a characteristic that can usefully distinguish between samples at a gene level. Basically, 0-fold degenerate sites signify non-synonymous mutation i.e. a nucleotide mutation that alters the amino acid sequence of the protein. 2-fold degenerate sites are those at which one out of 3 changes is synonymous. Also, all sites at which 2 out of 3 changes are synonymous are included in this category. So it must be mentioned here that synonymous mutations are nucleotide mutations that do not alter the protein sequence; they can be called silent mutation. The codons encoding 1 amino acid may differ in any of their 3 positions. Only the third position of same codons may be fourfold degenerate. At 4-fold degenerate sites, every possible mutation is synonymous.

Parsimony informative sites in spike glycoprotein

Since Spike protein has been the preferred target for vaccine engineering, we have explored the parsimony informative non-synonymous mutations that have occurred. Singleton informative sites were relegated to the background as these are less reliable and could have occurred due to sequencing errors instead of being true signals. The frequencies of each of the parsimony sites have been calculated to ascertain the major mutations that have transpired in the Spike protein.

Phylogenetic tree analysis taking the spike protein sequences

A phylogenetic tree was generated using all Spike protein sequences, to track the evolution of Spike sequences across the different countries [26]. Instead of using all sequences to depict the tree succinctly, identical sequences belonging to a particular country were collapsed into a single sequence. These sequences comprised the set of representative sequences for each country. These were used to generate the phylogenetic tree.

Results

Non-synonymous mutations of SARS-CoV-2 proteins

After aligning the individual structural proteins and open reading frames of SARS-CoV-2 genome, mutations were identified both at gene and protein levels. The number of variable sites, singleton and parsimony informative sites were tabulated. Special focus has been given on the mutations at the protein level, i.e. non-synonymous mutations, as these lead to an alternation in the resulting translated protein. Singleton informative sites are less reliable as these can be attributed to sequencing errors [19]. Therefore, the research is concentrated on the parsimony sites with view to drawing comparisons among the mutabilities of the different proteins. Those proteins having lower values of parsimony informative sites per 100 amino acids are consequently more stable and less prone to mutations.

At both the gene and protein levels, the Envelope and Membrane proteins were the most stable, which indicates that these may have housekeeping functions which make them more resistant to acquiring mutations. At the gene level, the number of mutations per 100 bases were found to be relatively high in Nucleocapsid, ORF6, ORF3a and ORF8. Highest number of mutations per 100 bases were observed in ORF6 and lowest in Envelope proteins. This paints a picture that the Nucleocapsid protein and ORF6 are more prone to mutations. Coming to the protein level, Nucleocapsid, ORF3a and ORF8 exhibited the highest rates of mutations. The complete information regarding the variable sites, singleton and parsimony informative sites at the gene and protein levels are provided in **Table 1**.

Table 1 Alignment of the SARS-CoV-2 isolates, showing the number of variable sites, singleton informative and parsimony informative sites at The gene level and The protein level.

Gene level

Name	Number of samples	Number of nucleotides	Variable sites	Variable sites per 100 bp	Singleton sites	Singleton sites per 100 bp	Parsimony informative sites	Parsimony informative sites per 100 bp
Envelope protein	687	228	4	1.75	3	1.32	1	0.44
Membrane protein	685	669	8	1.196	4	0.598	4	0.598
Nucleocapsid protein	687	1,260	52	4.13	33	2.62	19	1.51
Spike protein	688	3,823	77	2.01	58	1.52	19	0.49
ORF1ab	683	2,1291	391	1.84	294	1.38	97	0.46
ORF3a	683	828	30	3.62	19	2.29	11	1.33
ORF6	684	186	13	6.98	10	5.38	3	1.61
ORF7a	683	366	12	3.27	9	2.46	3	0.82
ORF8	684	366	9	2.46	5	1.36	4	1.09
ORF10	683	117	4	3.42	3	2.56	1	0.85

Protein level

Name	Number of samples	Number of amino acids	Variable sites	Variable sites per 100 bp	Singleton sites	Singleton sites per 100 bp	Parsimony informative sites	Parsimony informative sites per 100 bp
Envelope protein	687	76	2	2.63	2	2.63	0	0.00
Membrane protein	685	223	5	2.24	3	1.35	2	0.89
Nucleocapsid protein	687	420	28	6.66	19	4.52	9	2.14
Spike protein	688	1,274	44	3.45	34	2.67	10	0.78
ORF1ab	683	7,097	239	3.37	179	2.52	60	0.85
ORF3a	683	276	22	7.97	13	4.71	11	3.26
ORF6	684	62	7	11.29	6	9.67	1	1.61
ORF7a	683	122	6	4.92	6	3.28	2	1.64
ORF8	684	122	8	6.55	8	3.28	4	3.28
ORF10	683	39	2	5.13	2	5.13	0	0.00

Phylogenetic representation of spike protein sequences

The representative sequences corresponding to the Spike protein from each country were used to construct a phylogenetic tree, to visualize the evolution and divergence of the sequences over time. The phylogenetic tree is provided as **Figure 4**.

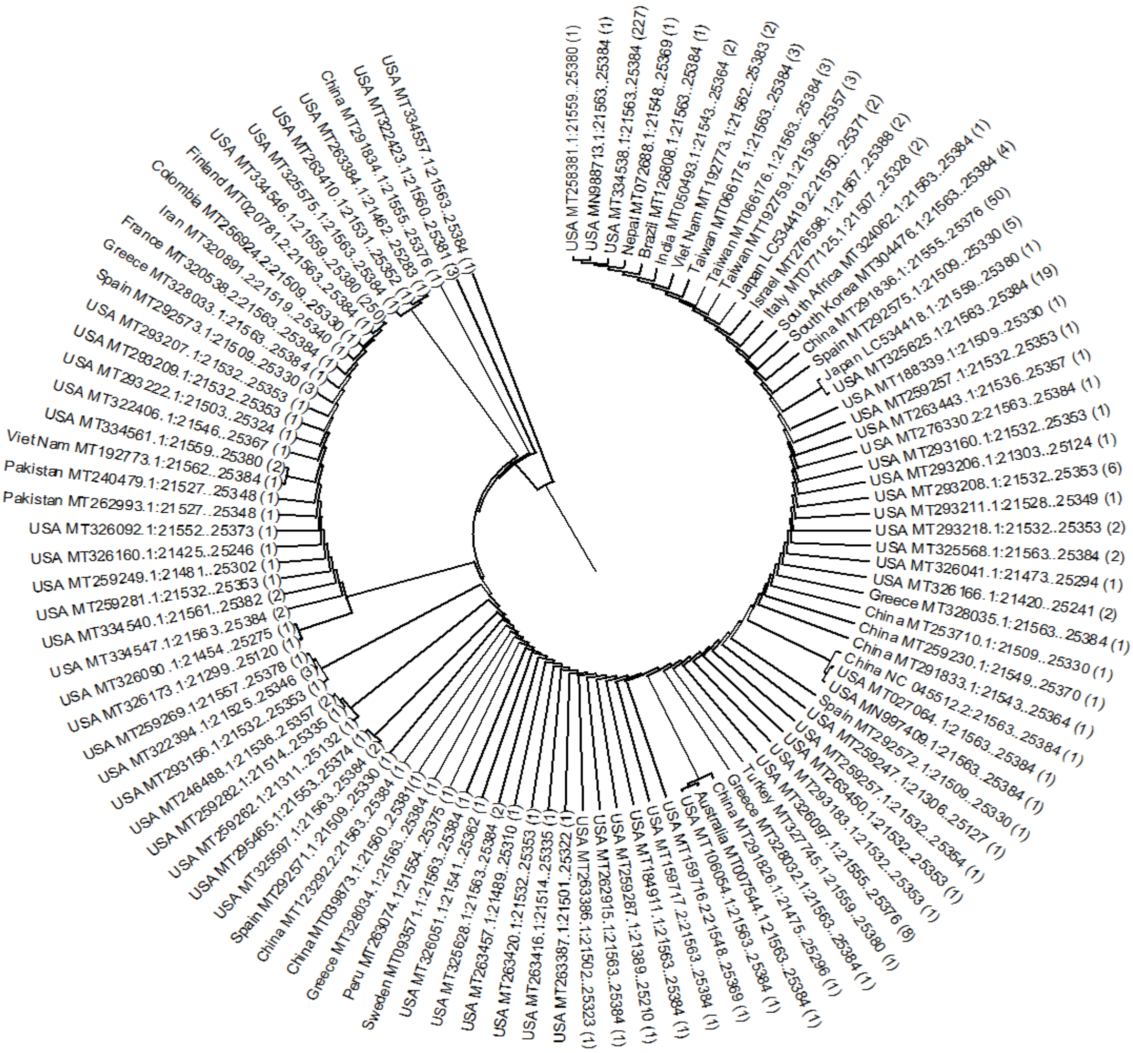
USA MT258381.1:21559..25380 (1)
USA MN988713.1:21563..25384 (1)
USA MT334538.1:21563..25384 (227)
Nepal MT072688.1:21548..25369 (1)
Brazil MT126808.1:21563..25384 (1)
India MT050493.1:21543..25364 (2)
Viet Nam MT192773.1:21562..25383 (2)
Taiwan MT066175.1:21563..25384 (3)
Taiwan MT066176.1:21563..25384 (3)
Taiwan MT192759.1:21536..25357 (3)
Japan LC534419.2:21550..25371 (2)
Israel MT276598.1:21567..25388 (2)
Italy MT077125.1:21507..25328 (2)
South Africa MT324062.1:21563..25384 (1)
South Korea MT304476.1:21563..25384 (4)
China MT291836.1:21555..25376 (50)
Spain MT292575.1:21509..25330 (5)
Japan LC534418.1:21559..25380 (1)
USA MT325625.1:21563..25384 (19)
USA MT188339.1:21509..25330 (1)
USA MT259257.1:21532..25353 (1)
USA MT263443.1:21536..25357 (1)
USA MT276330.2:21563..25384 (1)
USA MT293160.1:21532..25353 (1)
USA MT293206.1:21303..25124 (1)
USA MT293208.1:21532..25353 (6)
USA MT293211.1:21528..25349 (1)
USA MT293218.1:21532..25353 (2)
USA MT325568.1:21563..25384 (2)
USA MT326041.1:21473..25294 (1)
USA MT326166.1:21420..25241 (2)
Greece MT328035.1:21563..25384 (1)
China MT253710.1:21509..25330 (1)
China MT259230.1:21549..25370 (1)
China MT291833.1:21543..25364 (1)
China NC 045512.2:21563..25384 (1)
USA MT027064.1:21563..25384 (1)
USA MN997409.1:21563..25384 (1)
Spain MT292572.1:21509..25330 (1)
USA MT259247.1:21306..25127 (1)
USA MT259257.1:21532..25354 (1)
USA MT263450.1:21532..25353 (1)
USA MT293183.1:21532..25353 (1)
USA MT326097.1:21555..25376 (8)
Turkey MT327745.1:21559..25380 (1)
Greece MT328032.1:21563..25384 (1)
China MT291826.1:21475..25296 (1)
Australia MT007544.1:21563..25384 (1)
USA MT106054.1:21563..25384 (1)
USA MT159716.2:21548..25369 (1)
USA MT159717.2:21563..25384 (1)
USA MT184911.1:21563..25384 (1)
USA MT259287.1:21389..25210 (1)
USA MT262915.1:21563..25384 (1)
USA MT263386.1:21502..25323 (1)
USA MT263387.1:21501..25322 (1)
USA MT263416.1:21514..25335 (1)
USA MT263420.1:21532..25353 (1)
USA MT263457.1:21489..25310 (1)
USA MT325628.1:21563..25384 (2)
USA MT326051.1:21541..25362 (1)
Sweden MT093571.1:21563..25384 (1)
Peru MT263074.1:21554..25375 (1)
Greece MT328034.1:21563..25384 (1)
China MT039873.1:21560..25381 (1)
China MT123292.2:21563..25384 (1)
Spain MT292571.1:21509..25330 (1)
USA MT325597.1:21563..25384 (2)
USA MT295465.1:21553..25374 (1)
USA MT259262.1:21311..25132 (1)
USA MT259282.1:21514..25335 (1)
USA MT246488.1:21536..25357 (1)
USA MT293156.1:21532..25353 (2)
USA MT322394.1:21525..25346 (1)
USA MT259269.1:21557..25378 (3)
USA MT326173.1:21299..25120 (1)
USA MT326090.1:21454..25275 (1)
USA MT334547.1:21563..25384 (1)
USA MT334540.1:21561..25382 (2)
USA MT259281.1:21532..25353 (2)
USA MT259249.1:21481..25302 (1)
USA MT326160.1:21425..25246 (1)
USA MT326092.1:21552..25373 (1)
Pakistan MT262993.1:21527..25348 (1)
Pakistan MT240479.1:21527..25348 (1)
VietNam MT192773.1:21562..25384 (1)
USA MT334561.1:21559..25380 (2)
USA MT322406.1:21546..25367 (1)
USA MT293222.1:21503..25324 (1)
USA MT293209.1:21532..25353 (1)
USA MT293207.1:21532..25353 (1)
Spain MT292573.1:21509..25330 (3)
Greece MT328033.1:21563..25384 (1)
France MT320538.2:21563..25384 (1)
Iran MT320891.2:21519..25340 (1)
Colombia MT256924.2:21509..25330 (1)
Finland MT020781.2:21563..25384 (1)
USA MT334546.1:21559..25380 (250)
USA MT325575.1:21563..25384 (1)
USA MT263410.1:21531..25352 (1)
USA MT263384.1:21462..25283 (1)
China MT291834.1:21555..25376 (1)
USA MT322423.1:21560..25381 (3)
USA MT334557.1:21563..25384 (1)

Figure 4 Phylogenetic tree of the Spike protein representative sequences of each country. The frequency of occurrence of each of the sequences for the corresponding country is present in (brackets) at the end of the sequence id.

Parsimony informative spike protein mutations

A total of 10 parsimony informative sites were obtained in the spike protein. The frequencies of the residues occurring at each of these sites have been computed and present in **Table 2**. Among the ten mutations, nine of them occurred at less than 1 % frequency which indicates that these mutations are not prospering as of the recorded date. A reason for their dwindling status could be attributed to lesser stabilities of the mutant strains. On the other hand, the mutation D614G was observed to be abundant at a frequency of more than 40 %, which clearly indicates that this mutation is crucial in imparting greater stability and transmissibility of the virus.

Table 2 Parsimony informative sites of the Spike protein of SARS-CoV-2 showing the occurrences of wild-type and mutant residues. Frequencies of the mutant amino acids have been calculated.

Spike protein parsimony informative sites										
Amino acid position	**5**	**49**	**258**	**476**	**483**	**614**	**655**	**675**	**939**	**1,078**
Wild-type residue	L (683)	H (686)	W (686)	G (684)	V (684)	D (404)	H (686)	Q (685)	S (686)	A (685)
Mutant residue	F (5)	Y (2)	L (2)	S (4)	A (4)	G (284)	Y (2)	H (3)	F (2)	S (2), V (1)
Frequency of mutant residues	0.73	0.29	0.29	0.58	0.58	41.28	0.29	0.44	0.29	0.44

(L = Leucine, H = Histidine, W = Tryptophan, G = Glycine, V = Valine, D = Aspartate, Q = Glutamine, S = Serine, A = Alanine, F = Phenylalanine, Y = Tyrosine).

Discussion

The onset of the SARS-CoV-2 pandemic has brought the coronaviridae family back in the limelight. The virus enters the host cell through a receptor-mediated mechanism where the virus adheres with the host receptor. The attachment process is a crucial first step towards establishing a successful infection. The severity of COVID-19 and the enhanced transmissibility of SARS-CoV-2 is related to its increased potential to bind to host ACE-2. Certain exceptions in the binding of the mutant SARS-CoV-2 S protein to ACE-2 compared to the wild type were reported. Also, 2 extra residues 473 and 475 of S proteins were observed, while R393 of ACE-2 was missing from the interaction [22]. However, the influence of these interactions on virus internalization and pathogenesis needs to be evaluated. Most of the anti-SARS-CoV-2 vaccines are designed targeting the S-protein to inhibit the entry of the virus and further transmission in the body. To check the effect of distinct mutations against the vaccine efficacy is very important. The widespread mutation D614G lies at the S1-S2 junction of S protein and doesn't lead to any change on epitopes or surface structure of S protein and as a result may not get impacted by vaccines [19,20]. However, the mutations in the RBD (Receptor binding domain) region of S protein could affect the efficacy of vaccines. A study by Starr *et al.* [21], mapped viral mutations which enabled the virus to escape the antibodies against SARS-CoV-2. Both the recently reported UK and South African variants carry N501Y mutation lying in the RBD region of S protein. Detailed study has demonstrated that this mutation has maintained the interaction with ACE-2 with involvement of an additional residue (Q498) of S protein [22]. Currently various investigations are going on to check the efficacy of the vaccines under trials against variants carrying this mutation. Further, the crystal structure of S protein generated by Wrapp *et al.* [23] has revealed that D614 residue interacted with T859 of an adjacent chain when any one of the chains has RBD in up conformation. Another recent report [24] has shown that D614 forms salt bridges with K854 of the fusion peptide proximal region (FPPR). This analysis suggested that the D614-K854 interaction supported the role of FPPR in membrane fusion. Also, further studies have denoted the interaction of D614 with T859 and K854, which further led to the conclusion that D614 interaction with both residues might be reinforcing the role of FPPR [22]. The found results also depicted the elimination of this intramolecular interaction on D614G mutation. However, relevant structure-based studies are required for further investigation and validation.

The spread of such a virus arises from mutations in the viral genome and functional proteins that help in the adaptation of the virus to a new host. Mutations include the variations upon which natural selection acts, and often result in novelty. The data reported from this research suggests that the rate of D614G spike mutation surpasses all the others, and thus it is the most crucial reason for the virus to become ever-evolving. Viruses with mutation are much more infectious than those without. The mutation has the effect of markedly increasing the number of functional spikes on the viral surface. Those spikes allow the viruses to bind better and infect cells. The number or density of functional spikes on the virus is 4 to 5 times greater due to mutation [2].

This mutation provides greater flexibility to the spike's backbone. More flexible spikes allow newly made viral particles to navigate the journey from producer cell to target cell fully intact with less tendency to fall apart prematurely. Rate of D614G mutation is the highest, thus it is the major mutation. But other mutational profiles like S939F, Q675H, H655Y, V483A, G476S, W218L, H49Y, L5F [17] and others are less prominent; these are minor mutations.

Conclusions and future prospects

Through this research, several unique non-synonymous mutational profiles have been found. At both the gene and protein levels, the Envelope and Membrane proteins were the most stable. At gene level, the highest number of mutations per 100 bases were observed in ORF6 and lowest in Envelope proteins. At protein level, Nucleocapsid, ORF3a and ORF8 exhibited the highest rates of mutations.

Also, in this research, a total of 10 parsimony informative sites in the spike protein have been obtained. Among them, the mutational frequency of nine of them is 1 %; whereas one of them (D614G) shows a mutational frequency of 40 %. Thus, based on this result, it can be concluded that this mutational profile plays a very crucial role in enhancing the virus' stability and havoc transmissibility.

Based on this research, in future, more investigation can be done to determine the viability of the mutant. Analysing mutational profiles of SARS-CoV-2 can lead to identification of mechanisms that derive the SARS-CoV-2 evolution. Also, studying the mutational profiles can be instrumental to check transmissibility of the virus [7,9]. The phylogenetic tree also has a great research oriented prospect, considering that virus genomes are evolutionarily linked with each other [6]. Counting all the mutations in the sequences w.r.t a reference genome will create a mutation bias towards the most abundant or frequently sequenced isolates, that can be further studied in the future.

In this research, we have identified mutations at positions 5, 49, 258, 476, 483, 614, 655, 675, 939, 1078 and 936 and have verified that these variants are enduring among the worldwide general population over time and D614G is the most viable among them. Although previous studies depict some of the mutated residues of spike protein to have shown a greater reduction of total free energy as compared to D614G substitution, their Spatio-temporal distribution and number of isolates are comparatively lower than the substitution at 614 [10]. Thus, It clearly indicates that spike protein alone is not the

determining factor of stability, adaptability, and transmissibility of SARS-CoV-2. The specific combination of all frequently mutated variants will possibly be necessary for the prediction of the viability of the viral variants.

Finally, with view to the disparity in transmissibility of the spike protein variants, there are some crucial suggestions we have come to infer; First, the mutational profile of a COVID-19 positive patient must be analysed, specifically at these key positions that has been found in this research either by Sanger sequencing or designing probes corresponding to these regions. Then, a model must be predicted using severity and transmission of infection of the patient among the contacts for each combination of frequently mutated residues. In conclusion, it can be firmly expected that if the above suggestion can be executed effectively, it will bring positive results in curbing the COVID-19 pandemic.

References

[1] S Alai, N Gujar, M Joshi, M Gautam and S Gairola. Pan-India novel coronavirus SARS-CoV-2 genomics and global diversity analysis in spike protein. *Heliyon* 2020; **7**, e06564.

[2] F Begum, D Mukherjee, D Thagriki, S Das, PP Tripathi, AK Banerjee and U Ray. Analyses of spike protein from first deposited sequences of SARS-CoV2 from West Bengal, India. *F1000Research* 2020; **9**, 371.

[3] JR Byrnes, XX Zhou, I Lui, SK Elledge, JE Glasgow, SA Lim, RP Loudermilk, CY Chiu, TT Wang, MR Wilson, KK Leung and JA Wells. Competitive SARS-CoV-2 serology reveals most antibodies targeting the spike receptor-binding domain compete for ACE2 binding. *mSphere* 2020; **5**, e00802-20.

[4] Y Chen, Q Liu and D Guo. Emerging coronaviruses: Genome structure, replication, and pathogenesis. *J. Med. Virol.* 2020; **92**, 418-23.

[5] AM Davidson, J Wysocki and D Batlle. Interaction of SARS-CoV-2 and other coronavirus with ACE (Angiotensin-Converting Enzyme)-2 as their main receptor: Therapeutic implications. *Hypertension* 2020; **76**, 1339-49.

[6] V Gupta, S Haider, M Verma, N Singhvi, K Ponnusamy, MZ Malik, H Verma, R Kumar, U Sood, P Hira, S Satija, Y Singh and R Lal. Comparative genomics and integrated network approach unveiled undirected phylogeny patterns, Co-mutational hot spots, functional cross talk, and regulatory interactions in SARS-CoV-2. *mSystems* 2021; **6**, e00030-21.

[7] SW Huang, SO Miller, Y Chia-Hung and W Sheng-Fan. Impact of genetic variability in ACE2 expression on the evolutionary dynamics of SARS-CoV-2 spike D614G mutation. *Genes* 2020; **12**, 16.

[8] A Hussain, A Hasan, MMN Babadaei, SH Bloukh, MEH Chowdhury, M Sharifi, S Haghighat and M Falahati. Targeting SARS-CoV2 Spike protein receptor binding domain by therapeutic antibodies. biomed pharmacother. *Biomed. Pharmacother.* 2020; **130**, 110559.

[9] MR Islam, MN Hoque, MS Rahman, ASMRU Alam, M Akther, JA Puspo, S Akter, M Sultana, KA Crandall and MA Hossain. Genome-wide analysis of SARS-CoV-2 virus strains circulating worldwide implicates heterogeneity. *Sci. Rep.* 2020; **10**, 14004.

[10] S Laha, J Chakraborty, S Das, SK Manna, S Biswas and R Chatterjee. Characterizations of SARS-CoV-2 mutational profile, spike protein stability and viral transmission. *Infect. Genet. Evol.* 2020; **85**, 104445.

[11] S Lee, L Mi-Kyeong, H Na, J Ahn, G Hong, Y Lee, J Park, Y Kim, K Yun-Tae, K Chang-Ki, L Hwan-Sub and L Kyoung-Ryul. Comparative analysis of mutational hotspots in the spike protein of SARS-CoV-2 isolates from different geographic origins. *Gene Rep.* 2021; **23**, 101100.

[12] AV Letarov, VV Babenko and EE Kulikov. Free SARS-CoV-2 spike protein S1 particles may play a role in the pathogenesis of Covid-19 infection. *Biochemistry* 2021; **86**, 257-61.

[13] I Mercurio, V Tragni, F Busto, AD Grassi and CL Pierri. Protein structure analysis of the interactions between SARS-CoV-2 spike protein and the human ACE2 receptor: From conformational changes to novel neutralizing antibodies. *Cell. Mol. Life. Sci.* 2021; **78**, 1501-22.

[14] A Shah, F Rashid, A Aziz, AU Jan and M Suleman. Genetic characterization of structural and open reading Fram-8 proteins of SARS-CoV-2 isolates from different countries. *Gene Rep.* 2020. **21**: 100886.

[15] Y Wan, J Shang, R Graham, RS Baric and F Li. Receptor recognition by the novel coronavirus from wuhan: An analysis based on decade-long structural studies of SARS coronavirus. *J. Virol.* 2020; **94**, e0012720.

[16] Y Weisblum, F Schmidt, F Zhang, J DaSilva, D Poston, JC Lorenzi, F Muecksch, M Rutkowska, H Hans-Heinrich, E Michailidis, C Gaebler, M Agudelo, A Cho, Z Wang, A Gazumyan, M Cipolla, L Luchsinger, CD Hillyer, M Caskey, DF Robbiani, ..., PD Bieniasz. Escape from neutralizing antibodies by SARS-CoV-2 spike protein variants. *Elife* 2020; **9**, e61312.

[17] L Zhang, CB Jackson, H Mou, A Ojha, H Peng, BD Quinlan, ES Rangarajan, A Pan, A Vanderheiden, MS Suthar, W Li, T Izard, C Rader, M Farzan and H Choe. SARS-CoV-2 spike-protein D614G mutation increases virion spike density and infectivity. *Nat. Comm.* 2020; **11**, 6013.

[18] R Wang, J Chen, K Gao, Y Hozumi, C Yin and W Guo-Wei. Analysis of SARS-CoV-2 mutations in The United States suggest presence of four sub-strains and novel variants. *Comm. Biol.* 2021; **4**, 228.

[19] DC Groves, SL Rowland-Jones and A Angyal. The D614G mutations in the SARS-CoV-2 spike protein: Implications for viral infectivity, disease severity and vaccine design. *Biochem. Biophys. Res. Comm.* 2021; 538, 104-7.

[20] AJ McAuley, MJ Kuiper, PA Durr, MP Bruce, J Barr, S Todd, GG Au, K Blasdell, M Tachedjian, S Lowther, GA Marsh, S Edwards, T Poole, R Layton, R Sarah-Jane, TW Drew, JD Druce, TRF Smith, KE Broderick and SS Vasan. Experimental and *in silico* evidence suggests vaccines are unlikely to be affected by D614G mutation in SARS-CoV-2 spike protein. *npj Vaccines* 2020; **5**, 96.

[21] TN Starr, AJ Greaney, A Addetia, WW Hannon, MC Choudhary, AS Dingens, JZ Li and JD Bloom. Prospective mapping of viral mutations that escape antibodies used to treat Covid-19. *Science* 2021; **371**, 850-4.

[22] S Jakhmola, O Indari, D Kashyap, N Varshney, A Das, E Manivannan and HC Jha. Mutational analysis of structural proteins of SARS-CoV-2. *Heliyon* 2021; **7**, e06572.

[23] D Wrapp, N Wang, KS Corbett, JA Goldsmith, H Ching-Lin, O Abiona, BS Graham and JS McLellan. Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. *Science* 2020; **367**, 1260-63.

[24] Y Cai, J Zhang, T Xiao, H Peng, SM Sterling, RMW Jr, S Rawson, S Rits-Volloch and B Chen. Distinct conformational states of SARS-CoV-2 spike protein. *Science* 2020; **369**, 1586-92.

[25] S Kumar, G Stecher, M Li, C Knyaz and K Tamura. MEGA X: Molecular evolutionary genetics analysis across computing platforms. *Mol. Biol. Evol.* 2018; **35**, 1547-9

[26] J Hadfield, C Megill, SM Bell, J Huddleston, B Potter, C Callender, P Sagulenko, T Bedford and RA Neher. Nextstrain: Real-time tracking of pathogen evolution. *Bioinformatics* 2018; **34**, 4121-3.