

SARS-CoV19-2 spike protein mutation patterns: a global scenario

Renukaradhya K. Math^{1*}, Mallikarjun Goni¹, Palaksha K.J.¹ Ajay S.K.¹ Ajaykumar Oli¹

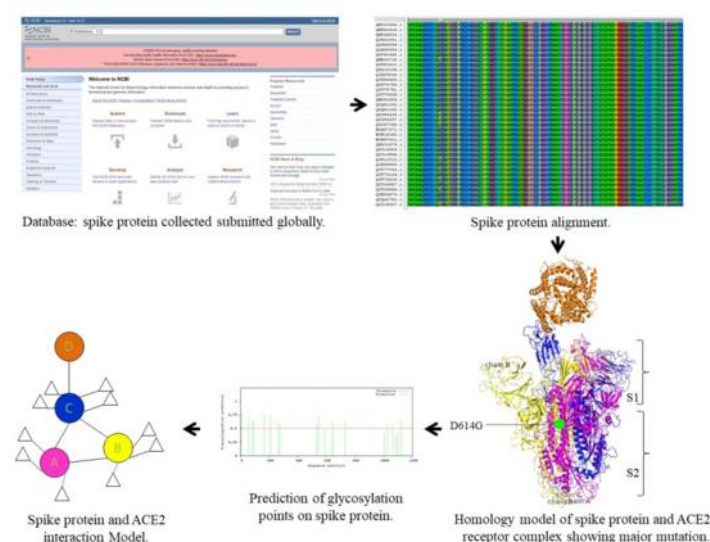
¹SDM Research Institute for Biomedical Sciences, 5th Floor, Manjushree Building, Shri Dharmasthala Manjunateshwara University, Sattur, Dharwad, India

***Corresponding author address:** ¹SDM Research Institute for Biomedical Sciences, 5th Floor, Manjushree Building, SDM Medical College & Hospital campus, Shri Dharmasthala Manjunateshwara University, Sattur, Dharwad-580009, India
E.mail: aradhya@sdmuniversity.edu.in

Abstract

Analysis of SARS-CoV-2 spike protein sequences of over 19 countries from biological databases submitted around the globe was carried out with help of bioinformatics tools and structure prediction databases. Initial data analysis showed entry of virus into different geographic regions started in the month of January 2020. Meanwhile, alignment of spike protein sequences of SARS-CoV-2 isolates from China and other countries revealed a critical mutation of D614G. Surprisingly, mutation D614G was not seen in early samples submitted in the month of January but gradually it started appearing globally from the month of March 2020. However, the mutations of amino acids in the spike protein other than D614G exhibiting similar pI and altered polarity were found to be specific to geographical regions. Besides, prediction of homology model for interaction of spike protein showed predominant role of chain C of trimeric spike protein in adhering receptor binding domain (RBD) of human ACE2 receptor. Furthermore, the prediction of glycosylation points has revealed that there are about 20 N-glycosylation potential sites on spike protein. We believe that the information present here would not only help in thorough understanding of infectivity but also enhance the knowledge of the scientific community in developing prophylactics and/or therapeutics for SARS-CoV19-2 virus.

Keywords: Coronavirus, spike protein, database, sequence alignment, mutation, homology model, hydrophobic amino acids.



Introduction

SARS-CoV-2 is a RNA virus consisting of positive-sense single-strand RNA of approximately 27-32 kb. This virus, a member of the Coronaviridae *family* is classified into four genera of CoVs: alpha, beta, delta and gamma. This virus appears like a crown under an electron microscope, hence it is named as CORONA. It is known to infect a wide range of hosts including humans, other mammals, and birds. Infected people will have clinical symptoms like asymptomatic to severe symptoms in their respiratory and digestive tracts etc. Interestingly, before it was reported/assumed that asymptomatic person can contribute up to 80% of SARS-CoV-2 transmission (Cascella et al. 2020). However, recently, (9th June 2020) WHO has mentioned that spread of virus by asymptomatic patients is rare but very next day retracted their statement since data was based on few studies. Therefore more studies are required to conclude the potential source and route of transmission of SARS-CoV-2.

The infectious SARS-CoV-2 penetrates human cells via a transmembrane spike (S) glycoprotein, which is a trimeric class I fusion protein and consists of two domains, S1 and S2. The S1 subunit mediates the attachment of the virus, and subsequently the S2 subunit mediates downstream membrane fusion of the viral and human cellular membranes (Hoffmann et al. 2020, Walls et al. 2020, Zhou et al. 2020). The receptor binding domain (RBD) for SARS-CoV-2 has been identified as the human angiotensin-converting enzyme 2 (hACE2), and recent studies determined a high binding affinity to hACE2 (Shang et al. 2020, Walls et al. 2020). Considering its key role, the S protein is one of the major targets for the study and development of preventive and therapeutic modalities. Meanwhile, presence of glycans on proteins serve several purposes starting from binding/attachment interaction, protein folding, masking epitopes to evade the host defense system. Therefore, it is also crucial to understand their location on the protein for the development of vaccines, which has been ignored previously (Wolfert and Boons 2013). However, recently many researchers are taking this fact into consideration (Watanabe et al. 2020).

Bioinformatics tools and biological databases have tremendous potential in extracting the information on the causative agent which helps in monitoring spread of disease and discovery of any new drugs or preventive strategies (Bianco et al. 2012; Sosa et al. 2017). Biological databases play a central role in Bioinformatics, especially, structural databases help to analyse 3 dimensional structures, and studying protein ligand interactions which facilitates screening and identification of candidate targets for drug discovery. Meanwhile, databases are vital for storage and updating information and not only provide a platform for

banking but also guard scientific data, and also key elements for the progress of scientific research.

As of June 19th 2020, a total of 84.6 million confirmed cases and 4.53 million deaths along with a total global recovered number 41.42 million were reported. First reporting of SARS-CoV-2 virus ever happened since 27th December 2019 and transmission of virus from human to human on 19th January 2020 by China. Ever since, SARS-CoV-2 has become pandemic, lockdown, social distancing, wearing a mask and staying and working from home has become new normal around the globe. Besides, it has also kept the scientific community, governments, healthcare workers, police, and diagnostic kit manufacturing industries on their toes for development of appropriate monitoring strategies to control the spreading of disease. Many countries have developed their own diagnostic kits, which helps them in early detection and to contain virus from spreading. Despite stringent measures taken to control the disease, the constantly changing nature of the virus has created panic not only among the common man but also in the scientific and healthcare community, globally. However, the only way to understand it better is to share information from everyone's experience and surveying available databases. Therefore, we have made an attempt to compile the information available in the biological databases and share the critical information which shall be useful to the scientific community. This might help in monitoring and understanding of disease as well as developing preventive and therapeutic modalities for SARS-CoV-2.

Methods

NCBI data survey: NCBI database was searched for protein using the keyword "Covid19 and respective country" (<https://www.ncbi.nlm.nih.gov>). Once the list is displayed sequences, protein sequences (surface glycoprotein) will be collected in fasta format based on date of sequence publication, else date of submission or sampling. Sequences published on the same date are considered as one to avoid duplication of sequences or reduce the numbers. Each sequence was collected and stored based on category as on the date of publication/sampling/submission. Later on, information was extracted and tabulated (Table 1) like, country of origin, date of publication, patient travel history (if available), mutation site, charge and position of mutant amino acids etc.

Multiple sequence alignment and domain prediction: Sequences collected were grouped based on country of origin in fasta format in a text file. The same files were used for multiple

alignment using ClustalX software version 2.1 (Higgins 1994; Chenna et al. 2003). Subsequently, alignment was analysed to look for mutation in the sequences and the same were tabulated (Table 1). For prediction of domains in spike protein of SARS-Covid19-2, the following sequence accession no. of NCBI QHU36864 was used, which was isolated from a 61 year old aged male person from Hubei, Wuhan, China. After downloading the sequence in fasta format the same was used for domain prediction using online web-tool prosite scan of expasy.org (De Castro et al. 2006).

Homology modelling: Homology model of SARS-CoV-2 spike protein (AN: QHU36864) was constructed using SWISS-MODEL homology modelling online tool in expasy.org. (Guex et al. 2009; Bertoni et al. 2017; Bienert et al. 2017; Waterhouse et al. 2018; Studer et al. 2020,). Interaction of SARS-CoV-2 spike protein and human ACE2 receptor model was also constructed using SWISS-MODEL online tool. Template used was 1R42_1 (chain A).

Protein structure view: Homology models of SARS-CoV-2 spike protein and human ACE2 Receptor Binding Domain (RBD) were viewed and analysed using pyMOL software version 2.4. (Schrodinger, 2010). Protein chain colouring and locating position of amino acids was also done using pyMOL.

Prediction of glycosylation sites: Glycosylation points on spike protein for SARS-CoV-2 were predicted by NetNGlyc.

Results and discussion

Analysis of collected spike protein sequences from over 19 countries revealed that the entry of virus into respective geography (except China) started in the month of January 2020 since the official publication of protein sequence by China. Countries like, India, Republic of Korea, Italy, Australia and USA showed entry in the month of January (based on date of sample collection in Table 1). Whereas, entries in countries like Japan, Brazil and Israel took place in the month of February. Followed by in countries like Nepal, South Africa, Bangladesh, Pakistan, Iran, Sri Lanka, and Turkey in the month of March and April 2020. However, these conclusions are based on data available in the NCBI database, which might not contain all the information. Though, the entry of virus in India, USA and Italy started in January, the virus spread and fatality rate was low in India (3.3%,

/https://coronavirus.jhu.edu/data/mortality), this could be due to change in geographical region physical factors, ethnicity, different genetic determinants and phenome of local population, food habit and relative herd immunity etc., of an individual. However, interestingly, a major mutation was seen in sequences which were submitted outside China indicating that the virus started mutating in the new host at different geographical regions or territory. These were evident in our multiple alignment sequences results as described in Table 1, Figure 1 & 2 and supplementary Figure 1. However, irrespective of temperature, race, ethnicity and diet at different geographical regions major mutation was still seen in every country trailing many doubts that how could virus spread so rapidly?. It is suspected that other specific mutations found in different isolates might play a key role in infectivity of the virus by stabilizing spike protein through replaced hydrophobic amino acids.



Figure 1. Multiple alignment of global SARS-CoV19-2 spike protein sequences showing mutation site pattern at amino acid position 614. Where, CH, China; BR, Brazil; IT, Italy; IR, Iran; SP, Spain; USA, United States of America; JP, Japan; KR, Republic of Korea; BD, Bangladesh; F, France; IS, Israel; N, Nepal; P, Pakistan; A, Africa.

Multiple alignments of protein sequences originating from 19 countries revealed one major and common mutation site and a few other mutations as listed in Table 1. The major mutation noticed is at the amino acid position D614 of the spike protein i.e. change of amino acids from aspartic acid (D) to glycine (G) D614G. Surprisingly, the mutation at D614G was not noticed in the Chinese virus isolated from the very first coronavirus in Hubei, Wuhan,

China until the last reported on 24th May 2020. This data also indicates that genetics drift of the virus started occurring when entered the new host at a different geographical region. However, some specific mutations (different mutation apart from D614G) were seen in Chinese protein sequences and all were hydrophobic in nature (Table 1). Hydrophobic amino acids like valine, leucine, isoleucine, phenylalanine and methionine have nonpolar side chains and these are considered as essential for folding of polypeptide chains in a protein and to maintain its globular structure (Kauzmann, 1959; Perutz et al. 1965; Dill, 1990; Gowder et al. 2014). Therefore, it is supposed that the majority of hydrophobic nature of amino acids on chain C surface might provide internal stability of protein chains to maintain globular structure and in maintaining trimeric structure of the spike protein (Dyson et al. 2006).

Interestingly, D614G mutation was not seen at amino acid position D614 among Indian, Italian, Japanese and American isolates during the months of January and February however, the mutation was found in the isolates collected from March 2020 onwards. This clearly demonstrated the gradual/faster adaptation of virus and subsequent genetic drift with enhanced infectivity in the new host at different geographical regions. Mutations were also observed in other positions of spike protein too, however, they were specific to particular strain and not commonly seen in all other protein sequences as listed in table 1. Again, mutated amino acids were hydrophobic in nature, except in two samples, where it was acidic. Surprisingly, none of these mutations altered pI of the protein (Table 1) as that of D614G mutation. However, the majority of mutations were observed in the S1 domain and on the amino acids present on the surface of the protein. Similarly, the probable reason for change of amino acid from acidic to neutral in mutation D614G was explained recently by Zhang et al. (2020), where they say introduction of glycine provides greater backbone flexibility and orientation to adjacent amino acids at that position and helps in greater stability of the protein.

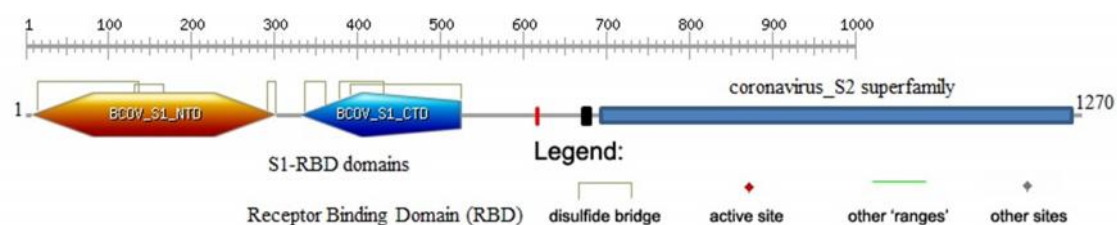


Figure 2. Graphical representation of SARA CoV-2 spike protein predicted domains generated by Prosite.expasy.org using protein sequence QHU36864. Red box indicates mutation site (D614G) and black box indicated Furin cleavage site.

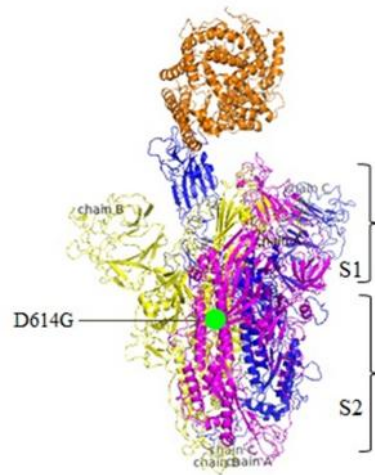


Figure 3. SARS-CoV-2 and human ACE2 RBD complex. Graphical representation of SARS-CoV-2 protein mutation point which was seen frequently on amino acid position D614G, Aspartic acid (D) to Glycine (G).

Protein sequence scan for domain analysis revealed two domains S1 and S2 (Figure 2) but when we look at the position of mutation site D614G on protein it was located outside of RBD domain and has no role in interaction with human ACE2 receptor (Figure 3). The homology model constructed showed protein has three chains A, B and C which gives it a trimer structure, similar to SARS spike protein (Krichdoerfer et al. 2018, Figure 4A,B & C, supplementary Figure 2). Chain C of the protein was mainly involved in interaction with human ACE2 receptors (Figure 4B & D).

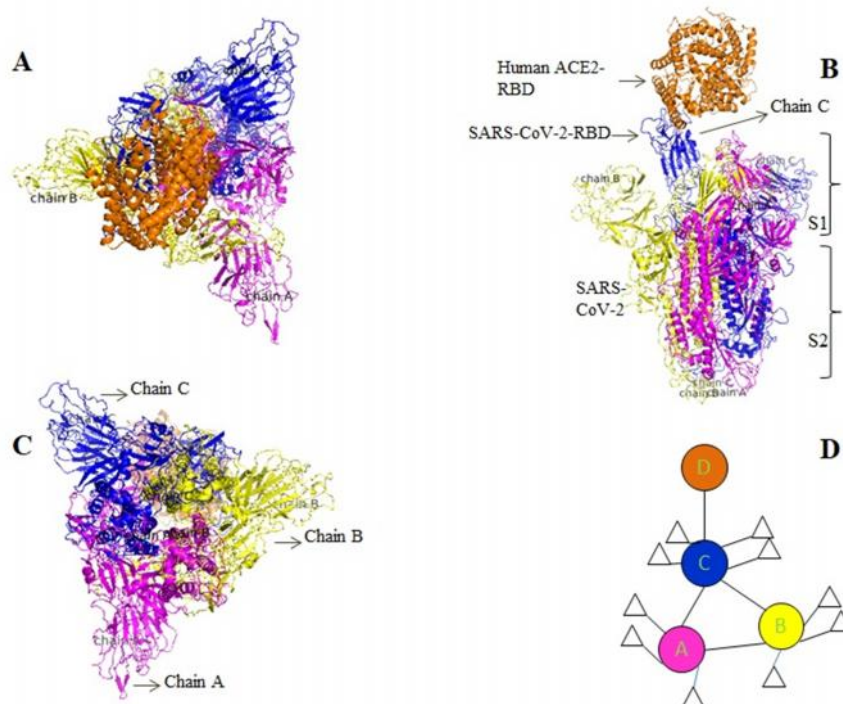


Figure 4. Homology Model of SARS-CoV-2 (PDB ID: 6acg) and human ACE2 receptor complex. Panel A, depicts front view, B, side view and C, back view of the complex. Model of and human ACE2 RBD complex. Letters in circle A,B,C are chains of SARS-CoV-2 protein and D is ACE2 receptor. , N-acetyl-D-glucosamine points.

Glycosylation prediction revealed there were 20 N-glycosylation potential sites present on the spike protein (Figure 5). Glycosylation of glycoprotein is important since many viruses use them to mask surface peptides epitopes that would otherwise elicit antibody response (Zhou et al. 2020). Therefore, it is essential to consider their presence on protein while designing vaccines.

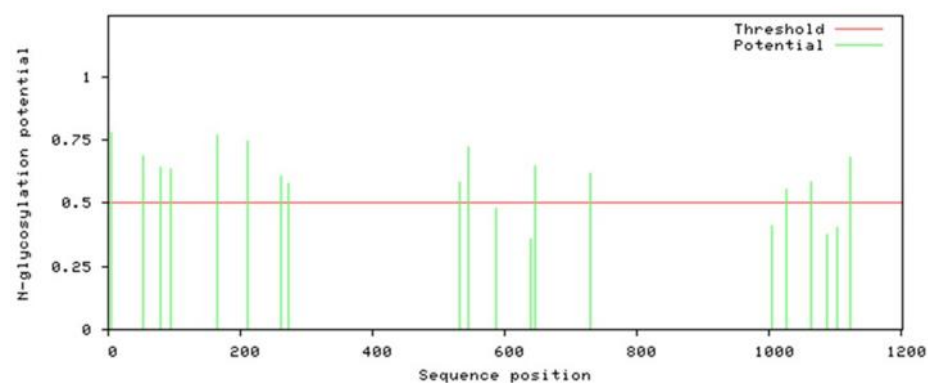


Figure 5. Graphical representation of predicted glycosylation sites on spike protein (QHU36864). Prediction shows 20 potential glycosylation sites and among them 15 are above the threshold (0.5) indicating higher probability.

A recent study by Sino Biosciences which developed a recombinant S1 protein with a D614G mutation showed comparable binding efficiency as the original 614D with ACE2 protein based on ELISA analysis (<https://www.sinobiological.com/research/virus/2019-ncov-antigen?utm>). They also claim that mutation D614G was initially predominantly found in European countries, which is converse to data analysed in this study. Besides, they claim this mutation strain is more transmissible. The study was carried out by a US-based Los Alamos National Laboratory in collaboration with Duke University and University of Sheffield, England. Author Bette Korber studied a total of 6000 sequences around the globe and found 14 mutations. They also say that this predominant strain migrated to the US East coast in mid-March 2020. But at this point of time it is not clear whether strain was carried to the US or it got mutated in the USA? However, in search of their claim (mutant D614G is more transmissible) a literature survey was carried out and found a recent study by Zhang et al 2020 investigated this fact. They showed that mutation D614G indeed enhances viral transmission

based on their experimental evidence using retroviruses pseudotyped with S^{G614} infected ACE2-expressing cells markedly more efficiently than those with S^{D614}.

In conclusion, major mutation was seen at amino acid position 614 and this mutation D614G changed amino acid from acidic to neutral. Mutation D614G did change pI of the protein from 6.05 to 6.13 but surprisingly other mutations did not affect/change pI of the protein. Based on data collected, entry of SARS-CoV-2 could be predicted, data indicates entry started in early month of January 2020 in countries like India, Republic of Korea, Italy, and USA. In Japan, Israel and Brazil it started in the month of February. Critical mutation was observed only when viruses moved outside from Chinese geography. None of the spike proteins collected from China till the end of May month 2020 showed critical mutation. Mutation D614G was seen in chain C and in the S1/S2 domain of spike protein. It is still unclear why and how this mutation has occurred? It is important to investigate both proteins (D615 and 614G) binding ability with human ACE2 receptors. Considering these factors might help in understanding virus behaviour and also in designing and finding medicine for virus inhibition.

Acknowledgement

We thank Management of Shri Dharmasthala Manjunatheshwara University for providing infrastructure facility and support.

References

- Bertoni, M., Kiefer, F., Biasini, M., Bordoli, L., Schwede, T. Modeling protein quaternary structure of homo- and hetero-oligomers beyond binary interactions by homology. *Scientific Reports* 7 (2017).
- Bianco AM., Marcuzzi A., Zanin V., Girardelli M., Vuch J., Crovella S. Database tools in genetic diseases research. *Genomics*, 101, (2013) 75–85.
- Bienert, S., Waterhouse, A., de Beer, T.A.P., Tauriello, G., Studer, G., Bordoli, L., Schwede, T. The SWISS-MODEL Repository - new features and functionality. *Nucleic Acids Res.* (2017) 45, 313-319.

- Cascella M., Rajnik M., Cuomo A., Dulebohn SC., Napoli RD. Features, Evaluation and Treatment Coronavirus (COVID-19). StatPearls Publishing LLC, May 18, (2020).
- Chenna R., Sugawara H., Koike T., Lopez R., Gibson TJ., Higgins DG., Thompson JD. Multiple sequence alignment with the Clustal series of programs. Nucleic Acids Res., (2003) 31, 3497-3500.
- De Castro E., Sigrist CJA., Gattiker A., Bulliard V., Langendijk-Genevaux PS., Gasteiger E., Bairoch A., Hulo N. *ScanProsite: detection of PROSITE signature matches and ProRule-associated functional and structural residues in proteins.* Nucleic Acids Res. (2006) Jul 1;34(Web Server issue):W362-5.
- Dill K. A., "Dominant forces in protein folding," Biochemistry, (1990) 29, 7133–7155.
- Dyson HJ, Wright PE, Scheraga HA. The role of hydrophobic interactions in initiation and propagation of protein folding. PNAS, (2006) 29, 13057–13061.
- Gowder SM., Chatterjee J., Chaudhuri T., Paul K. Prediction and Analysis of Surface Hydrophobic Residues in Tertiary Structure of Proteins. eScientific World Journal, 2014 <http://dx.doi.org/10.1155/2014/971258>.
- Guex N., Peitsch MC., Schwede T. Automated comparative protein structure modeling with SWISS-MODEL and Swiss-PdbViewer: A historical perspective. Electrophoresis (2009) 30, 162-173.
- Higgins DG. CLUSTAL V: multiple alignment of DNA and protein sequences. Methods Mol Biol., (1994) 25, 307-318.
- Hoffmann M., Kleine-Weber H., Schroeder S., Kruger N., Herrler T., Erichsen S., Schiergens TS., Herrler G., Wu NH., Nitsche A, *et al.* SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2 and Is Blocked by a Clinically Proven Protease Inhibitor. Cell (2020) 181, 1-10.
- Kauzmann W. "Some factors in the interpretation of protein denaturation," Advan. Prot. Chem. (1959) 14, 1–63.
- National Center for Biotechnology Information (NCBI), <https://www.ncbi.nlm.nih.gov/>
- Perutz MF., Kendrew JC., Watson HC. "Structure and function of haemoglobin: II. some relations between polypeptide chain configuration and amino acid sequence," J. Mol. Biol. (1965) 13, 669–678.
- Studer G., Rempfer C., Waterhouse AM., Gumienny G, Haas J, Schwede T. QMEANDisCo - distance constraints applied on model quality estimation. Bioinformatics (2020) 36, 1765-1771.
- Schrodinger, LLC. 2010. The PyMOL Molecular Graphics System, Version 2.4.

- 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).
- Sosa EJ, Burguener G, Lanzarotti E, Defelipe L, Radusky L, Pardo AM, Marti M, Turjanski AG, Dario Fernández DP. a structural bioinformatic approach to prioritize drug targets in pathogens, *Nucleic Acids Res.* Volume 46, 2018, 413–418.
- 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) 180, 1-12.
- Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R., Heer, F.T., de Beer, T.A.P., Rempfer, C., Bordoli, L., Lepore, R., Schwede, T. SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Res.* (2018) 46, 296-303.
- Watanabe et al., Site-specific glycan analysis of the SARS-CoV-2 spike. *Science* 10.1126/science.abb9983 (2020).
- Wolfert MA, Boons GJ. Adaptive immune activation: glycosylation does matter. *Nat Chem Biol.* (2013) 9, 776-784.
- Zhang L, Jackson CB, Mou H, Ojha A, Rangarajan ES, Izard T, Farzan M, Choe H. The D614G mutation in the SARS-CoV-2 spike protein reduces S1shedding and increases infectivity. doi: <https://doi.org/10.1101/2020.06.12.148726>.
- Zhou P, Yang XL, Wang XG, Hu B, Zhang L, Zhang W, Si HR, Zhu Y, Li B, Huang CL, *et al.*. A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature.* (2020) 579, 270-273.

Table 1. Details of survey data of SARS-CoV19-2 spike protein sequences submitted from around the globe to NCBI database.

Sl. No.	Country of origin	Sequence published/ Sample collected	NCBI Accession no.	Length of Protein/pI	Mutation position/chain A, B, C	Number of Mutations occurred	Change of Amino acid (AA)	Charge of mutated AA	Position of AA on protein surface/domain	Remarks
1	India	*28 April	QJF11812	1213/6.13	614/C 1250/C	2	D G C F	Neutral Hydrophobic (HP)	Interface/S1-S2 Bottom surface/S2	Total seq. submitted on this date (TS) 7
		Karnataka								
		6 May	QJQ27842	1213/6.05	Nil	0	D	--	--	TS 4
		7 May	QJQ39968	1213/6.13	614/C	1	D G	Neutral	Interface /S1-S2	TS 3
		6 April/27 January	QHS34546	1212/5.98	Nil	0	D	--	--	--
		Kerala								
		6 April	QIA98583	1213/6.05	614/C 930/C	2	D G A V	Neutral Hydrophobic	Interface /S1-S2 Side surface/S2	--
		6 May	QJQ28345	1213/6.13	614/C	1	D G	Neutral	Interface /S1-S2	TS 3
		11 May	QJR84513	1213/6.13	614/C	2	D G	Neutral	Interface /S1-S2	--
					55/C		L F	hydrophobic	Interface /S1	--
		19 May	QJW69139	Issues with sequence	614/C 78/C	2	D G R M	Neutral Hydrophobic	Interface /S1-S2 Side surface/S1	--
		Gujarat								
		23 May	QJY40601	Issues with sequence	Nil	0	D	--	--	--
		24 May	QJY51384	1213/6.13	614/C	1	D G	Neutral	Interface /S1-S2	--
		25 May	QJY77055	Issues with sequence	Nil	0	D	--	--	--
		30 April/11 March	QJF77846	1213/6.05	614/C 28/C	1	D Y H	-- Acidic (-ve)	-- Side surface/S1	--
		Telangana								
		30 April/11 March	QJF77870	1213/6.05	Nil	0	D	--	--	--
		12 May/25 March	QJS39651	1213/6.05	Nil	0	D	--	--	Age 63, Male
		12 May/24 March	QJS39639	1213/6.13	614/C	1	D G	Neutral	Interface /S1-S2	--
2	Nepal	6 April	QIB84673	1213/6.05	Nil	0	D	--	--	--
3	South Africa	13 April/7 March	QIZ15537	1213/6.13	614/C	1	D G	Neutral	Interface /S1-S2	--
4	Bangladesh	15 May/14 April	QJU11812	1213/6.13	614/C	1	D G	Neutral	Interface /S1-S2	--
		25 May	QJY77043	1213/6.13	614/C	1	D G	Neutral	Interface /S1-S2	--
5	China	26 May	QJX59599	1213/6.05	Nil	0	D	--	--	--
		11 February/1 January	QHU36864	1213/6.05	Nil	0	D	--	--	Age 61, Male
		Hubei								
		18 March	QHD43416	1213/6.05	Nil	0	D	--	--	--
		Wuhan								
		30 March	YP9724390	1213/6.05	Nil	0	D	--	--	--
		6 April	QIS30018	1213/6.05	Nil	0	D	--	--	--
		6 April/10 March	QIU80913	1213/6.05	Nil 50/C	1	D S L	-- Hydrophobic	--	--
		Fuyang								
		11 April	QHZ00369	1213/6.05	Nil	0	D	--	--	--
		30 April	QJG65958	1213/5.98	Nil		D	--	--	--
		Beijing								
					273/C, 275/C 276/C, 750/C	4	R M F Y L I S I	HP, Neutral HP, HP	All Interface /S1	--
		14 May	QIU81873	1213/6.05	Nil	0	D	--	--	--
6	Pakistan	15 May	QIU81974	1213/6.05	Nil	0	D	--	--	--
		Guangzhou								
		8 May	QJQ84088	1213/6.05	Nil	0	D	--	--	--
		Hangzhou								
		21 May	QIQ68654	1213/6.05	Nil	0	D	--	--	--
7	Republic of Korea	4 March	QIQ22760	1213/6.05	Nil	0	D	--	--	Returned from Iran
7	Republic of Korea	11 February/January 2020	QHZ00379	1213/6.05	221/C	1	S W	Neutral	Side surface/S1	Returned from

*sequences submitted on the same data are considered as one sequence