Ministry of Higher Education and Scientific Research

University of sulaimani

College of science

Biology department

Phylogenetic Analysis of the Spike Region of SARS-CoV2

Lanya Ahmad Osman

Roz Kamaran Fatah

Azhin Hawzhin Jalal

Soma Mohamed Othman

Supervisored by

Dr.Paywast Jamal Jalal

2021

Phylogenetic Analysis of the Spike Region of SARS-CoV2

Lanya Ahmad Osman, Roz Kamaran Fatah, Azhin Hawzhin Jalal, Soma Mohammed Othman, Paywast Jamal Jalal

Department of Biology, College of Science, University of Sulaimani, Sulaimani City, Kurdistan region - IRAQ

4

*Correspondence: paywast.jalal@univsul.edu.iq

Abstract

Since the novel coronavirus appeared in December 2019, it has brought global attention to this new virus due to its high infectivity rate and increasing fatality rate to more than 7%. For this purpose, in this study, we aimed to address the evolution and the genetic variation of the spike region of the SARS-CoV2 extensively. In the current study, successive computational work performed starting with multiple sequence alignment and phylogenetic tree construction followed by alignment prediction with some cell receptors. The alignment analysis based on the spike glycoprotein of SARS-CoV2 showed that this novel coronavirus is closer to Bat and Pangolin coronavirus, and it is quite different from other human coronaviruses. Interestingly, the sequence and structural alignments show that 23 amino acid residues are inserted in the S1 subunit of the five spike regions and most of them in the receptor-binding domain (RBD). None of these proteins has been detected in any previously identified human coronaviruses, and some of them are not shown in the sequence of Bat and Pangolin coronaviruses. They could be considered unique for this novel coronavirus. Despite the need for more analysis, this study revealed that the hypothetical binding between the novels inserted amino acid in the SARS-CoV2 Spike RBD might be the right candidate for the development of antiviral treatment against COVID-19.

KEYWORDS: SARS-CoV2, RBD and COVID-19.

Introduction

In 2019, a new coronavirus (2019-nCoV) infecting humans has emerged in Wuhan, China (Coutard *et al.*, 2020). This novel coronavirus outbreak is currently the most serious threat to humanity and this pandemic viral infection has affected the majority of countries, with as many as 225 countries affected (Choudhury and Mukherjee, 2020). Coronaviruses cause widespread respiratory, gastrointestinal, and central nervous system diseases in humans and other animals, threatening human health and causing economic loss (Enjuanes *et al.*, 2006). Considering the high sequence similarity with SARS-CoV, the International Committee on Taxonomy of Viruses (ICTV) has renamed this newly emerged coronavirus as the Severe Acute Respiratory Syndrome Coronavirus-2 (SARS-CoV2) (Wang *et al.*, 2020). Coronaviruses (CoV) are enveloped positive-stranded RNA viruses belonging to the family Coronaviridae in the order Nidovirales (Kirchdoerfer *et al.*, 2016a) Alphacoronavirus, Betacoronavirus, Gammacoronavirus, and Deltacoronavirus are the four genera that they belong to the alpha- and beta coronaviruses infect mammals, while the gammacoronaviruses infect avian species and the delta coronaviruses infect both mammals and avian species(Kirchdoerfer *et al.*, 2016b). Human coronaviruses are primarily responsible for infections of the upper respiratory tract and the gastrointestinal tract. Among them from 2003, severe acute respiratory syndrome coronavirus (SARS-CoV) infected 8,000 people, with a fatality rate of ~10% (Rota *et al.*, 2003)And in 2012, Middle East respiratory syndrome coronavirus (MERS-CoV) has infected more than 1,700 people, with a fatality rate of ~36% (Osterhaus *et al.*, 2012).

The new CoV (2019-nCoV) was discovered in Wuhan in December 2019, and this emerging viral infection was linked to severe human respiratory disease with a 7% fatality rate(Tong *et al.*, 2020). SARS-CoV-2 is the seventh coronavirus to infect humans so far (Zheng, 2020). The virus was presumed to have initially been transmitted from an animal reservoir to humans possibly via an amplifying host. Patients were mainly reported with pneumonia-like symptoms as fever, nonproductive cough, dyspnea, myalgia, fatigue, lymphopenia, and organ dysfunction (eg, shock, acute respiratory distress syndrome [ARDS], acute cardiac injury, and acute kidney injury). In addition, patients had difficulty in breathing where chest radiographs showing bilateral patchy shadows, or ground-glass opacity in all patients including invasive pneumonic infiltrates in few cases (Kumar *et al.,* 2020; Outcomes, 2020). Among newly emerging clinical features, a large number of patients have experienced the loss of taste or smell, tremors, headaches, rashes, and muscle pain (Lescure *et al.*, 2020)

SARS-CoV-2 are enveloped with an approximate diameter of 60–140 nm in size, positive-stranded RNA viruses with a genome size of approximately 30,000 bases in length (Chen, Liu and Guo, 2020). The viral genome encoding *9860 amino acids long polyprotein, with a G+C content of 38%, in total consisting of six major open reading frames (ORFs) common to coronaviruses and several other accessory genes (Saxena *et al.,* 2020; Zheng, 2020). The nucleocapsid protein (N) forms a helical capsid around the genome, which is further enclosed by an envelope (Li, 2016). Coronaviruses encode 16 nonstructural proteins (nsp1-nsp16) in 10 ORF (Figure 1) as well as four structural proteins: spike glycoprotein (S-glycoprotein), membrane (M), envelope (E), nucleocapsid (N), and hemagglutinin (HE) (Kumar *et al.*, 2020)Virus assembly is mediated by membrane protein and envelope protein, while virus entry into host cells is mediated by spike protein. The spike, which is one of these structural proteins, protrudes from the virus surface in huge protrusions, giving coronaviruses the appearance of crowns (hence their name; corona in Latin means crown) (Li, 2016).

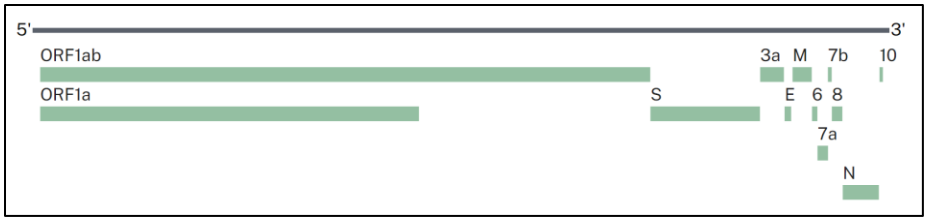

Figure 1: Diagrammatic representation of the SARS-CoV2 ORF. S (spike glycoprotein), M (membrane), E (envelope), N (nucleocapsid).

The coronavirus spike protein is a multifunctional molecule that allows coronaviruses to enter host cells. S-glycoprotein of coronaviruses consists of two functional subunits S1 and S2 subunit. The S1subunit is required for host cell attachment by recognizing cell surface sugar molecules and binding to specific cellular receptors(Belouzard *et al.*, 2012). Therefore, the S1 subunit, especially its receptor-binding domain (RBD) specifically recognizes angiotensin-converting enzyme 2 (ACE2) as its receptor and is critical in determining host range and cell tropism. The S2 subunit mediates membrane fusion and it contains a hydrophobic fusion loop and two heptad repeat regions (HR1 and HR2), which suggest a coiled helix structure of the S2 subunit (Gui, *et al.,* 2017; Kirchdoerfer, *et al.,* 2016). The S glycoprotein exists as a metastable prefusion trimer and undergoes structural rearrangement from a prefusion to a postfusion conformation upon S-protein receptor binding and cleavage (Walls, *et al.,* 2017; Bangaru, *et al.,* 2020). The RBD of the S1 subunit constantly switches between a standing-up position for receptor binding and a lying-down position for immune evasion (Yuan, *et al.,* 2017). Moreover, to fuse membranes, the SARS-CoV spike needs to be proteolytically activated at the S1/S2 boundary, such that S1 dissociates and S2 undergoes a dramatic structural change (Shang *et al.,* 2020). The spike (S) glycoprotein on the coronavirus envelope is crucial for receptor binding, internalization of the virus, tissue tropism, membrane fusion, and host range (Saxena *et al.,* 2020).

The spike is a vital determinant of viral host range and tissue tropism, as well as a significant inducer of host immune responses, in addition to mediating virus entry. For addressing the different variation, therefore, this study aimed to show and address a significant sequence variance in the Spike glycoprotein of SARS-CoV2 and their relation to the entry and pathogenicity of this virus.

Methods and Material

Multiple sequence alignment and Phylogenetic analysis

The NCBI database (www.ncbi.nlm.nih.gov) was the source for retrieving the human and animal coronavirus sequence. The full-length sequences of the novel coronavirus retrieved since it is publishing to the present day. From the full-length sequences of the SARS-CoV2, we only picked those from each country with a high mortality rate, and we excluded the sequences that

were 100% similar detected among their sequences. Finally, the start codon of each sequence spike was obtained from NCBI graphics of the following selected sequence, Human SARS-CoV2/2021 [MW527392.1 (IRQ/ Erbil), MW642249.1 (ITA/ABR), MW881790.1 (IND), MW828655.1 (IND)], Human SARS-CoV2/2020 [MW059036.1, MW255832.1 and MT873892.1 (England), MN908947.3 (Wuhan-Hu), MT066156.1 (ITA-INMI1), MT126808.1 (BRA-SP02), MN994467.1 (USA-CA), MT233523.1 (ESP- Valencia), LC528232.1 (Kng-Hu), MT007544.1 (Australia-VIC01), MT093571.1 (SWE), MT240479.1 (PAK), MT012098.1 (IND), MT072688.1 (NPL), MT020781.2 (FIN,) LC534419.1 (Kng); Bat coronavirus [MN996532.1 (RaTG13/2020); Pangolin coronavirus [MT040333.1 (GX-P4L)].

The alignment was performed based on the amino acid using MultAlin software(Sarokin and Carlson, 1984). The genetic relationships between the virus isolate were examined using phylogenetic analysis, carried out by MEGA7(Kumar, Stecher and Tamura, 2016). The tree was assessed by the neighbour-joining method to estimate the genetic evolution of the generated datasets. The tree distance was computed under a Poisson correction model and was in the units of the number of amino acid substitutions per site. Statistical analysis was estimated using the bootstrap approach via 100 replication. For each replicate, the tree generated and the method was repeated 100 times and a percentage score giving for the same grouping.

Result

Bioinformatics Analysis of the SARS-CoV-2 Spike Variability

The analysis of the spike region's multiple sequence alignment between different human coronavirus strains from (SARS-CoV2), Bat and Pangolin showed a total of 120 mutations. Pangolin coronavirus is known to be the first that most of the mutations occur in and around 92 nucleotides have been mutated (71 of them in the S1 subunit & the rest 21 are in the S2 subunit) compared to the Erbil strain. After that comes bat coronavirus in the second place with 28 mutations(24 of them are in the S1 subunit and 4 of them are in the S2 subunit) followed by (ITA/ABR) with 13 mutations (11 in the S1 subunit & 2 in the S2 subunit) and England strain with 12 mutations (7 in S1 subunit and 5 of them in the S2 subunit) Three sequences were meant to be mutated nearly in all strains like 349-479 (except in bat coronavirus)-615 (except in ITA/ABR, India, MW&MT-GBR amino acid variation among these strains. Among this total recorded data as mentioned in the table show that most of the mutations are placed in the S1 subunit, 78% of the variant located in the S1 region and 22% located in the S2 subunit. There is only some variation in the spike glycoprotein among the novel SARS-CoV2 at position 251 and 801 in both Australia and Sweden strains. Interestingly, the lowest degree of difference between the SARS-CoV2 spike region and Bat is less than 3% and with Pangolin is less than 10%. In contrast, the highest percentage of amino acid variation, approximately 29% are seen between SARS-CoV2 and SARS-CoV strains in their spike region.

On the other hand, the alignment outputs also showed four amino acids deleted in the former part of the N-terminal domain and 23 amino acids are inserted in five different N-terminal and Receptor-binding positions domain (Figure 2). Surprisingly, all the inserted amino acid are found in all the novel coronavirus strains, but one cannot find any of these inserts in the SARS-CoV strains. Moreover, the same amino acid inserts are detected in the Bat coronavirus and Pangolin coronavirus except for 4 and 6 amino acid, respectively. Our analysis showed that the four

deleted amino acid and the five insertions groups of the amino acid are unique to the novel SARS-CoV2 and could not be detected in any previously published coronavirus strains.

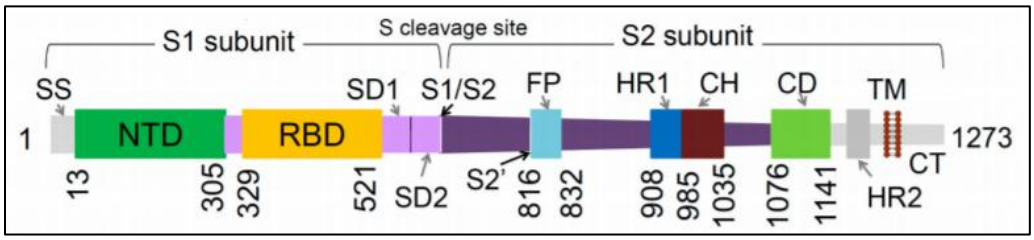

Figure 2: Different domains of the spike protein that includes; signal sequence (SS), the N-terminal domain (NTD), receptor-binding domain (RBD), subdomain 1 and 2 (SD1&2), protease cleavage sites (S1/S2/S20), fusion peptide (FP), heptad repeat 1 and 2 (HR1&2), central helix (CH), connector domain (CD), transmembrane domain (TM), and cytoplasmic tail (CT)(Kalathiya *et al.*, 2020).

Surprisingly, the comparative analysis of the new variants from England (MW059036.1) has 24 nucleotide deletion (8 amino acid) at the position of 680-686 which is located at the very beginning of the S2 subunit when compared to the SARS-CoV, SARS-CoV and MERS. However, this deletion not found in the Indian (MW881790.1) new variant SARS-CoV2 strain and another three amino acid deleted in the N-terminal region of the S1 subunit. Moreover, when analyzing the sequence data between the different strain of SARS-CoV2 with Bat and Pangolin coronavirus (Table 1), we can declare that the strain isolated in Erbil city during January 2021 has no mutation. In contrast strain, MW642249.1 from Italy has approximately 13 mutations in comparison to the Erbil SARS-CoV2 and the Strain from India has 12 mutations as well. The other selected strain have a limited number of mutation between 2 to 3 in comparison to Erbil strain.

Table 1: Sequence analysis data showing the number and site of the mutations in the spik region of various selected SARS-CoV2.

Name of the sequence	No. of Mutation	Site of the mutation
MW527392.1 (IRQ/NN Erbil)	-	-
MW642249.1 (ITA/ABR)	13	19_21_27_139_191_349_418_479_485_502_556_1028_1177
MW059036.1 (England)	3	349_479_615
MW881790.1 (IND)	12	69_70_71_349_479_502_571_682_717_944_983_1119
MW255832.1 (GBR)	2	349_479

MT873892.1 (GBR)	2	349_479
MN908947.3 (Wuhan-Hu)	3	349_479_615
MT066156.1 (ITA)	3	349_479_615
MT126808.1 (BRA)	3	349_479_615
MW828655.1 (IND)	3	349_479_615
MN994467.1 (USA-CA)	3	349_479_615
MT233523.1 (ESP-Valencia)	3	349_479_615
LC528232.1 (Kng-Hu)	3	349_479_615
MT007544.1 (Australia-VIC)	4	248_349_479_615
MT093571.1 (SWE)	4	349_479_615_798
MT240479.1 (PAK)	3	349_479_615
MT012098.1 (IND)	4	349_409_479_615
MT072688.1 (NPL)	3	349_479_615
MT020781.2 (FIN)	4	50_349_479_615
LC534419.1 (Kng)	3	349_479_615
MN996532.1 Bat coronavirus/RaTG13	28	33_219_325_347_349_373_404_440_441_442_444_446_450_460_484_485_487_491_494_495_499_502_506_520_605__615_1126_1229
MT040333.1 Pangolin coronavirus/GX-P4L	92	8_9_24_25_28_33_48_51_67_68_70_73_77_79_113_115_138_148_152_154_156_172_198_212_219_222_254_255_256_261_273_279_282_293_307_325_346_347_349_373_403_404_418_440_441_442_444_445_446_451_461_479_484_485_487_491_494_495_499_502_505_510_520_521_530_533_537_555_557_559_571_615_623_628_633_639_641_652_667_678_679_680_681_689_690_691_784_1085_1105_1126_1198_1229

Molecular Phylogenetic analysis by Maximum Likelihood method

The phylogenetic tree analysis of the full length of SARS-CoV2 showed a close relation and a high similarity with SARS-CoV, Bat and Pangolin coronavirus. Therefore, we repeated the analysis by comparing different coronavirus strains based on the spike glycoprotein genes [Figure 3]. The tree showed that all the novel coronavirus strains are approximately similar to each other and in close relation with Bat coronavirus (RaTG13). Moreover, the 2019 novel coronavirus's close genetic relationship is observed with Pangolin coronavirus.

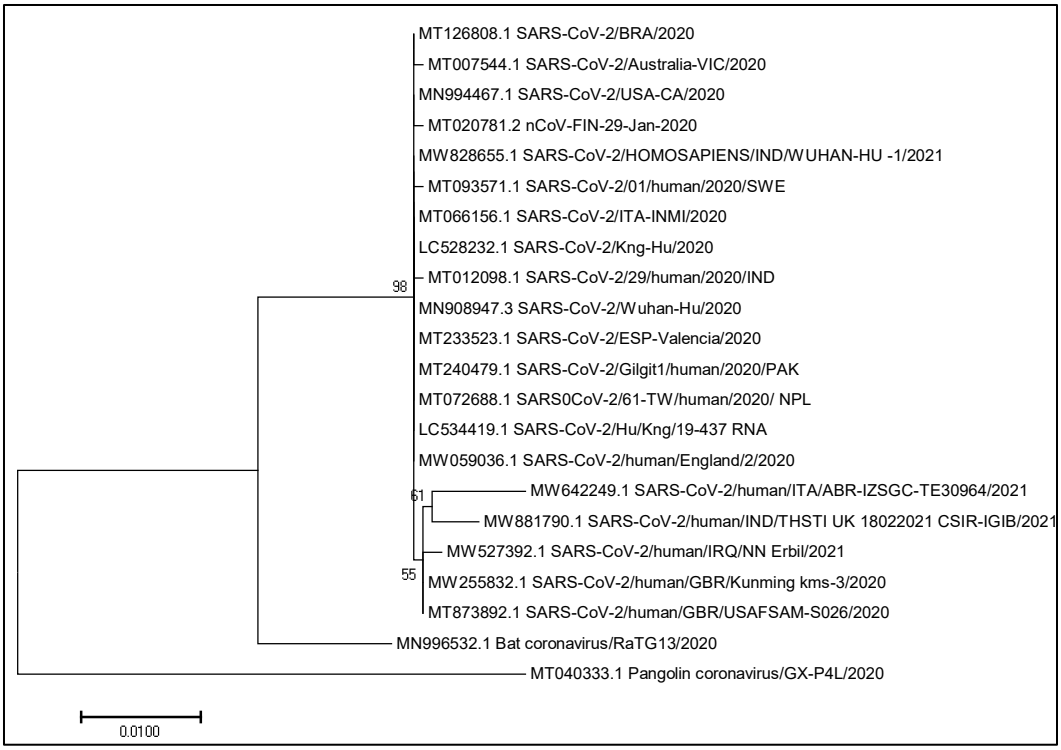

Figure 3: Phylogenetic analysis based on the spike glycoprotein. The evolutionary history was inferred by using the Maximum Likelihood method based on the JTT matrix-based model. The tree with the highest log likelihood (4624.8877) is shown. The percentage of trees in which the associated taxa clustered together is shown next to the branches. Initial tree(s) for the heuristic search were obtained by applying the Neighbor-Joining method to a matrix of pairwise distances estimated using a JTT model. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 22 amino acid sequences. The coding data was translated assuming a Standard genetic code table. There were a total of 1274 positions in the final dataset. Evolutionary analyses were conducted in MEGA7.

Discussion

Coronaviruses are capable of adapting to new environments through mutation and recombination with relative ease and hence are programmed to alter host range and tissue tropism efficiently. Therefore, health threats from coronaviruses are constant and long-term (Graham and Baric, 2010). Coronavirus has established its name in biology through the three pandemic outbreaks, namely, MERS-CoV, SARS-CoV, and the SARS- CoV-2. Out of these, the recent outbreak of COVID-19 pandemic by SARS-CoV-2 has proved to be significantly virulent due to its modifications, and adaptive mutations and re-emergence of the previously emerged SARS-CoV/MERS. In comparison to SARS-CoV, a significant sequence variance in the S-glycoprotein of SARS-CoV-2 was discovered (Saxena *et al.,* 2020). These differences enable SARS-CoV-2 RBD to have a significantly higher hACE2 binding affinity than SARS-CoV RBD does (Shang *et al.,* 2020).

SARS-CoV2 pathophysiology typically involves entry of the virus through respiratory droplets into the respiratory system, principally the alveoli of the lungs, through the airways. The viral glycoprotein on its capsid, known as "spike protein," binds to the angiotensin-converting enzyme-2 (ACE2) receptor which is a dipeptidyl carboxypeptidase type I integral membrane protein (Choudhury and Mukherjee, 2020; Rothan and Byrareddy, 2020). Upon interaction of S-glycoprotein with the hACE2 receptor, SARS-CoV2 gets internalized via an endosomal pathway that results in the release of viral RNA. Following internalization, viral RNA gets translated into long polyproteins (pp1a and pp1b) (Choudhury and Mukherjee, 2020). When viral antigens interact with host immune cells, proinflammatory reactions occur, resulting in vasodilation, increased vascular permeability, and the aggregation of humoral factors. All of these factors work together to cause fever, which disrupts gaseous exchange and makes breathing difficult (Raoult *et al.,* 2020). Upon infection with SARS-CoV2, host cells upregulate various host genes, including chemokines and proinflammatory cytokines, at both the transcriptional and translational stages (Costela-Ruiz *et al.*, 2020) more specifically, the major mediators of SARS-CoV-2 pathogenesis have been identified as IL-8, IL-6, TNF-a, and IFN.

According to previous studies, the effect of structural changes on the properties of the spike protein has sparked concern about transmissibility and pathogenicity (Frampton *et al.*, 2021). In our study, we found that most of the variations in the spike protein are located in the S1 region which is 78%. The currently identified Indian strain possesses 12 nucleotides mutations at the S1 region which has been reported to be associated with high virulence and high transmission rate. Also for the UK strain, there have been 3 nucleotides mutations and according to the previous studies, the pathogenicity of the UK strain is associated with higher viral loads in respiratory samples (Frampton *et al.*, 2021). Previous studies in this direction indicated that pangolin and bat could be the possible origin/source for SARS-CoV-2 and COVID-19 outbreak. In the present study, our data suggest that the spike protein of SARS-CoV-2 is phylogenetically closer to bat SARS-CoV than that of pangolin SARS-CoV.

Spike glycoprotein of coronavirus is divided into two subunits during cleavage also known as primming, N-terminal S1, which recognizes and binds the angiotensin-converting enzyme 2 (ACE2) (Figure 4) on the surface of host cell surface and C-terminal S2, which mediates membrane fusion (Coutard *et al.*, 2020; Örd, Faustova and Loog, 2020). It is likely that both fusion peptide (FP) which is located in S2-protein, and internal fusion peptide (IFP) which is following a second proteolytic site (S2') of the spike, participate in the viral entry process so that the S-protein must be cleaved at both S1/S2 and S2' cleavage sites to facilitate virus

entry(Coutard *et al.*, 2020; Johnson *et al.*, 2021; Tang *et al.*, 2021)S-protein of SARS-CoV2 contains 12 additional nucleotides upstream to the single Arg cleavage site(Coutard *et al.*, 2020)Early studies showed that the S1/S2 junction of spike protein contains an insert with two additional basic residues, P−R−R−A (P, proline; R, arginine; A, alanine), these four amino acids form an RXXR cleavage motif for serine proteases when added to S1/S2 cleavage site (PRRAR), which corresponds to a canonical furin-like cleavage site(Coutard *et al.*, 2020)Spike protein of SARS-COV-2 undergoes Proteolytic cleavage by furin or other cellular proteases (TMPRSS2) at S1/S2 site because this site is containing conserved putative motifs for several cellular proteases depending on genomic sequences, so it is initial to the infection(Amrun *et al.*, 2020)(Örd, Faustova and Loog, 2020). Furin cleavage site at spike protein is lacking in other group-2B coronaviruses such as SARS-CoV, so the S protein remains uncleaved after biosynthesis(Coutard *et al.*, 2020). There are pieces of evidence show that variations around the viral envelope glycoprotein cleavage site have a role in pathogenesis and cellular tropism, for instance, the insertion of such cleavage site to S protein of the infectious bronchitis virus (IBV) (Gamma-coronavirus group 3), higher pathogenicity has occurred in infected chickens(Coutard *et al.*, 2020). Similarly, higher pathogenicity has been reported in the case of the influenza virus, those viruses have a furin-like cleavage site which is cleaved by cellular proteases such as furin, which are expressed in a wide variety of cell types of the host(Coutard *et al.*, 2020).

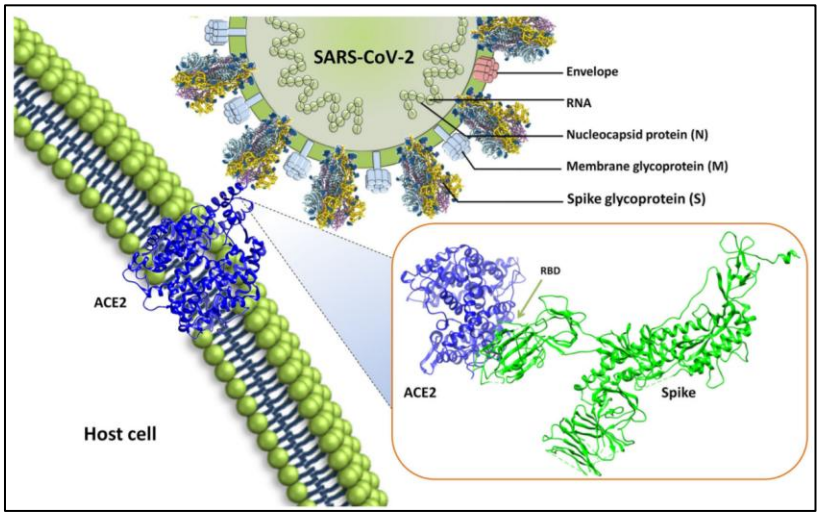

Figure 4: Structure of the severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2) and interaction with host cellular receptor ACE2. SARS-CoV-2 encodes for four structural proteins, spike glycoprotein, membrane glycoprotein (M), Envelope (E), nucleocapsid (N). Spike glycoprotein is embedded in the host-derived membrane which binds with the host cell receptor ACE2. The inset is showing the protein-protein interaction of SARS-CoV-2 S-glycoprotein with host cell receptor ACE2 mediated via receptor-binding domain (RBD) present in spike glycoprotein and peptidase domain of ACE2.

In conclusion, we report the comprehensive mutation landscape of more than 200 individual SARS- CoV-2 viral genome sequences isolated from COVID-19 patients or primary contacts. The data forms a starting point for the state government machinery to conduct further studies on virus transmission helping in making informed public health decisions. The genomic mutations also inform on potential mechanisms being employed during evasion of the host's immune response and traces of higher or lower pathogenicity, if any, being developed over the period, thus significantly impacting efforts in vaccine development.

References

Amrun, S. N. *et al.* (2020) 'Linear B-cell epitopes in the spike and nucleocapsid proteins as markers of SARS-CoV-2 exposure and disease severity', *EBioMedicine*, 58, pp. 1–9. doi: 10.1016/j.ebiom.2020.102911.

Bangaru, S. *et al.* (2020) 'Structural analysis of full-length SARS-CoV-2 spike protein from an advanced vaccine candidate', *bioRxiv*, 1094(November), pp. 1089–1094. doi: 10.1101/2020.08.06.234674.

Belouzard, S. *et al.* (2012) 'Mechanisms of coronavirus cell entry mediated by the viral spike protein.', *Viruses*, 4(6), pp. 1011–1033. doi: 10.3390/v4061011.

Chen, Y., Liu, Q. and Guo, D. (2020) 'Emerging coronaviruses: Genome structure, replication, and pathogenesis', *Journal of Medical Virology*, 92(4), pp. 418–423. doi: 10.1002/jmv.25681.

Choudhury, A. and Mukherjee, S. (2020) 'In silico studies on the comparative characterization of the interactions of SARS-CoV-2 spike glycoprotein with ACE-2 receptor homologs and human TLRs', *Journal of Medical Virology*, 92(10), pp. 2105–2113. doi: 10.1002/jmv.25987.

Costela-Ruiz, V. J. *et al.* (2020) 'SARS-CoV-2 infection: The role of cytokines in COVID-19 disease', *Cytokine and Growth Factor Reviews*, 54(May), pp. 62–75. doi: 10.1016/j.cytogfr.2020.06.001.

Coutard, B. *et al.* (2020) 'The spike glycoprotein of the new coronavirus 2019-nCoV contains a furin-like cleavage site absent in CoV of the same clade', *Antiviral Research*, 176(February), p. 104742. doi: 10.1016/j.antiviral.2020.104742.

Enjuanes, L. *et al.* (2006) 'Biochemical aspects of coronavirus replication and virus-host interaction', *Annual Review of Microbiology*, 60, pp. 211–230. doi: 10.1146/annurev.micro.60.080805.142157.

Frampton, D. *et al.* (2021) 'Genomic characteristics and clinical effect of the emergent SARS-CoV-2 B.1.1.7 lineage in London, UK: a whole-genome sequencing and hospital-based cohort study', *The Lancet Infectious Diseases*, 3099(21), pp. 1–11. doi: 10.1016/s1473-3099(21)00170-5.

Graham, R. L. and Baric, R. S. (2010) 'Recombination, Reservoirs, and the Modular Spike: Mechanisms of Coronavirus Cross-Species Transmission', *Journal of Virology*, 84(7), pp. 3134–3146. doi: 10.1128/jvi.01394-09.

Gui, M. *et al.* (2017) 'Cryo-electron microscopy structures of the SARS-CoV spike glycoprotein reveal a prerequisite conformational state for receptor binding', *Cell Research*, 27(1), pp. 119–129. doi: 10.1038/cr.2016.152.

Johnson, B. A. *et al.* (2021) 'Loss of furin cleavage site attenuates SARS-CoV-2 pathogenesis', *Nature*, 591(7849), pp. 293–299. doi: 10.1038/s41586-021-03237-4.

Kalathiya, U. *et al.* (2020) 'Highly Conserved Homotrimer Cavity Formed by the SARS-CoV-2 Spike Glycoprotein: A Novel Binding Site', *Journal of Clinical Medicine*, 9(5), p. 1473. doi: 10.3390/jcm9051473.

Kirchdoerfer, R. N. *et al.* (2016a) 'Pre-fusion structure of a human coronavirus spike protein', *Nature*, 531(7592), pp. 118–121. doi: 10.1038/nature17200.

Kirchdoerfer, R. N. *et al.* (2016b) 'Pre-fusion structure of a human coronavirus spike protein', *Nature*, 531(7592), pp. 118–121. doi: 10.1038/nature17200.

Kumar, S. *et al.* (2020) 'Structural , glycosylation and antigenic variation between 2019 novel coronavirus (2019-nCoV) and SARS coronavirus', *VirusDisease*, 31(1), pp. 13–21. doi: 10.1007/s13337-020-00571-5.

Kumar, S., Stecher, G. and Tamura, K. (2016) 'MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets', *Molecular biology and evolution*, 33(7), pp. 1870–1874. doi: 10.1093/molbev/msw054.

Lescure, F. X. *et al.* (2020) 'Clinical and virological data of the first cases of COVID-19 in Europe: a case series', *The Lancet Infectious Diseases*, 20(6), pp. 697–706. doi: 10.1016/S1473-3099(20)30200-0.

Li, F. (2016) 'Structure, Function, and Evolution of Coronavirus Spike Proteins', *Annual Review of Virology*, 3(August), pp. 237–261. doi: 10.1146/annurev-virology-110615-042301.

Örd, M., Faustova, I. and Loog, M. (2020) 'The sequence at Spike S1/S2 site enables cleavage by furin and phospho-regulation in SARS-CoV2 but not in SARS-CoV1 or MERS-CoV', *Scientific Reports*, 10(1), pp. 1–10. doi: 10.1038/s41598-020-74101-0.

Osterhaus, A. D. M. E. *et al.* (2012) 'Isolation of a Novel Coronavirus from a Man with Pneumonia in Saudi Arabia'. doi: 10.1056/NEJMoa1211721.

Outcomes, M. (2020) 'Clinical Characteristics of 138 Hospitalized Patients With 2019 Novel Coronavirus–Infected Pneumonia in Wuhan, China', pp. 1–9. doi: 10.1001/jama.2020.1585.

Raoult, D. *et al.* (2020) 'Coronavirus infections: Epidemiological, clinical and immunological features and hypotheses', *Cell Stress*, 4(4), pp. 66–75. doi: 10.15698/cst2020.04.216.

Rota, P. A. *et al.* (2003) 'Characterization of a novel coronavirus associated with severe acute respiratory syndrome', *Science*, 300(5624), pp. 1394–1399. doi: 10.1126/science.1085952.

Rothan, H. A. and Byrareddy, S. N. (2020) 'The epidemiology and pathogenesis of coronavirus disease (COVID-19) outbreak', *Journal of Autoimmunity*, 109(February), p. 102433. doi: 10.1016/j.jaut.2020.102433.

Sarokin, L. and Carlson, M. (1984) 'Nucleic Acids Research Nucleic Acids Research', *Methods*, 12(21), pp. 8235–8251.

Saxena, S. K. *et al.* (2020) 'Chasing COVID-19 through SARS-CoV-2 spike glycoprotein', *VirusDisease*. doi: 10.1007/s13337-020-00642-7.

Shang, J. *et al.* (2020) 'Cell entry mechanisms of SARS-CoV-2', *Proceedings of the National Academy of Sciences of the United States of America*, 117(21). doi: 10.1073/pnas.2003138117.

Tang, T. *et al.* (2021) 'Proteolytic Activation of SARS-CoV-2 Spike at the S1/S2 Boundary: Potential Role of Proteases beyond Furin', *ACS Infectious Diseases*, 7(2), pp. 264–272. doi: 10.1021/acsinfecdis.0c00701.

Tong, Y. *et al.* (2020) 'new england journal', pp. 1199–1207. doi: 10.1056/NEJMoa2001316.

Walls, A. C. *et al.* (2017) 'Tectonic conformational changes of a coronavirus spike glycoprotein promote membrane fusion', *Proceedings of the National Academy of Sciences of the United States of America*, 114(42), pp. 11157–11162. doi: 10.1073/pnas.1708727114.

Wang, Qihui *et al.* (2020) 'Structural and Functional Basis of SARS-CoV-2 Entry by Using Human ACE2', *Cell*, 181(4), pp. 894-904.e9. doi: 10.1016/j.cell.2020.03.045.

Yuan, Y. *et al.* (2017) 'Cryo-EM structures of MERS-CoV and SARS-CoV spike glycoproteins reveal the dynamic receptor binding domains', *Nature Communications*, 8(China CDC), pp. 1–9. doi: 10.1038/ncomms15092.

Zheng, J. (2020) 'SARS-coV-2: An emerging coronavirus that causes a global threat', *International Journal of Biological Sciences*, 16(10), pp. 1678–1685. doi: 10.7150/ijbs.45053.