



# Genetic Path of the Emergence of SARS-CoV-2

Taslima Nasrin <sup>1</sup> and Safdar Ali <sup>1,\*</sup>

<sup>1</sup>Aliah University, Kolkata, India

\*Corresponding author: Department of Biological Sciences, Aliah University, 700160, Kolkata, India. Email: safdar\_mgl@live.in

Received 2021 August 16; Revised 2021 September 18; Accepted 2021 September 20.

## Abstract

**Context:** Severe acute respiratory syndrome corona virus 2 (SARS-CoV-2) is the seventh coronavirus that has humans as the host. Because of its highly infectious nature, toward the end of January 2020, the WHO declared it a public health emergency of international concern. The present review is about understanding the journey of SARS-CoV-2 to its present form with an attempt to assess the genetic basis of its pandemic-causing abilities.

**Evidence Acquisition:** The data for the present review were accessed through different publications and preprint repositories.

**Results:** SARS-CoV-2 is a beta-coronavirus, and is approximately 60 - 140 nm in size. The appearance of its structure as a crown shape under an electron microscope led to the coining of its name 'Coronavirus'. Comparative genome and proteome analysis exhibits similarities and differences with reference to SARS-CoV. The open reading frames (ORFs) found on the SARS-CoV-2 genome, and their corresponding proteins have been discussed. Bats may act as reservoir hosts but not exclusively. The possibility of snakes as the host, as well as other intermediate hosts, before reaching humans seems plausible. This has been supported by ACE2 receptor diversity and conservation across different tissues and organisms. The role of spike glycoprotein and its interaction with the receptor through specific residues for invading host cells makes a perfect therapeutic target, but the variations therein and the resulting impact on interactions pose challenges for the same.

**Conclusions:** Though the differences between the MERS, SARS-CoV, and SARS-CoV-2 genomes indicate amino acid changes, leading to the present pandemic situation, the fact that new variants are still emerging signifies that the journey is an ongoing one, which requires monitoring.

**Keywords:** Genome, Interaction, Lineage, Mutation

## 1. Context

The emergence of SARS-CoV-2 began with many people reporting pneumonia in the Wuhan district of China sometime in December 2019. The outbreak was thereon found to be triggered by a new human coronavirus, which was subsequently named severe acute respiratory syndrome corona virus 2 (SARS-CoV-2) (1). The WHO declared SARS-CoV-2 as a public health emergency of international concern on 30 January 2020, and the world is still struggling to combat its menace (2). As of 11 July 2021, there were 187,419,263 reported cases and 4,045,647 deaths worldwide ([www.worldometers.info/coronavirus/](http://www.worldometers.info/coronavirus/)) due to SARS-CoV-2 of which, 3,08,37,222 cases and 4,08,040 deaths were reported from India ([www.mygov.in/covid-19](http://www.mygov.in/covid-19)). The journey of coronaviruses and human diseases has many prior incidents of which SARS-CoV and MERS-CoV have been known to cause severe diseases. Others, including OC43, HKU1, NL63, and 229E were reported to be responsible for minor indicators (3, 4). In the present review, we discuss the origin of SARS-CoV-2 by comparing the genome, host range,

its target receptor, viral spike protein, and strength of interaction of receptor binding domain (RBD) with its host receptor with those of earlier reported coronaviruses.

## 2. Evidence Acquisition

The data for the present review were accessed through different publications and preprint repositories.

## 3. Results

### 3.1. Genomic Variation of SARS-CoV-2

As known, SARS-CoV-2 belongs to beta-coronaviruses. Morphologically, it may be round or oval of about 60 - 140 nm. Its characteristic crown-shaped appearance under an electron microscope led to the basis of the name 'Coronavirus' (5). Besides beta ( $\beta$ ), the *Coronaviridae* family also has other genera, namely  $\alpha$ ,  $\gamma$ , and  $\delta$ . Among them, members of  $\alpha$  and  $\beta$  genera are known to have mammals and

humans as the host. The coronaviruses harbor a single-stranded RNA genome of about 26 kb to 32 kb (6). Among those infecting humans, the MERS-CoV genome has been reported to be of ~ 30.1 kb RNA (+) (7), while the SARS-CoV-2 reference genome is 29.9 kb (8). The SARS-CoV-2 genome has been reported to harbor around 26 proteins whose details and comparative mutation profiles across deceased and asymptomatic patients have been reported by our group recently (9). The 3' end of the genome encodes four structural proteins: (1) spike (S), (2) envelope (E), (3) membrane (M), and (4) nucleocapsid (N), of which the S protein is known to interact and attach to the host cell aiding in viral entry.

The genomes of SARS-CoV-2 and SARS-CoV have more than 80% identity to human coronavirus (SARS-like bat CoV) (10). The common entity includes ORF1ab with 16 non-structural proteins (NSPs) and four structural proteins. However, SARS-CoV-2 ORF3b and ORF10 have little similarity to those of SARS-CoV. Further, ORF8 is undivided in the former and split as ORF8a and 8b in the latter (7). A comparative illustration of the genomes of SARS-CoV, SARS-CoV-2, and MERS-CoV is shown in Figure 1. There are similarities and dissimilarities at the protein level. For instance, there are 154 amino acid 3b proteins in SARS-CoV as compared to just 22 amino acids in SARS-CoV-2 (10). Contrastingly, NSP7, NSP13, envelope, matrix, and accessory proteins have no amino acid substitutions, whereas there are variations in NSP2, NSP3, and S protein (11). Of these, NSP2 and NSP3 mutations are known to affect infectivity (12). The new emerging variants are discussed later.

### 3.2. Definitive and Intermediate Host Range

The SARS-Cov-2 pandemic owes its origin to a seafood market, and efforts have been made to study the animals in the market acting as hosts. Snakes have been one of the probable candidates supported by similar codon usage (14). The high level of similarity between SARS-CoV-2 and SARS-like bat coronavirus at the genome level supports the bats acting as reservoirs (15). This was also supported by phylogenetic studies that not only placed the bat-derived coronaviruses across all the five subgenera of beta-coronavirus but also found SARS-CoV-2 was evolutionarily closest to bat-SL-CoVZC45 and bat-SL-CoVZXC21 (16). Further evidence is provided by over 96% genome similarity between bat CoV RaTG13 and SARS-CoV-2 (2). Moreover, similar receptor sequences across species indicate that turtles, pangolin, and snakes can also act as intermediate hosts (17). It is noteworthy to mention that SARS-CoV and MERS-CoV also have bats as natural reservoirs and masked palm civets or dromedary camels as intermediate hosts. Needless to say that humans are the terminal hosts (18). The transmission of SARS-CoV-2 from bat to human may

have happened through animals being sold at the Wuhan market.

Metagenomic analysis of samples from pangolins reported the presence of  $\beta$ -CoV28 in most. One of the studied samples showed almost 99% similarity to SARS-CoV-2. This supports the candidature of pangolin as an intermediate host (19). It was also supported by several pangolin Coronaviruses from pangolin with identical sequences at important positions of RNA binding domain (RBD) of the S protein, implying the ability to bind to human ACE2 receptor, facilitating their movement to humans (20).

### 3.3. Angiotensin-Converting Enzyme 2 (ACE2) as Target Receptor

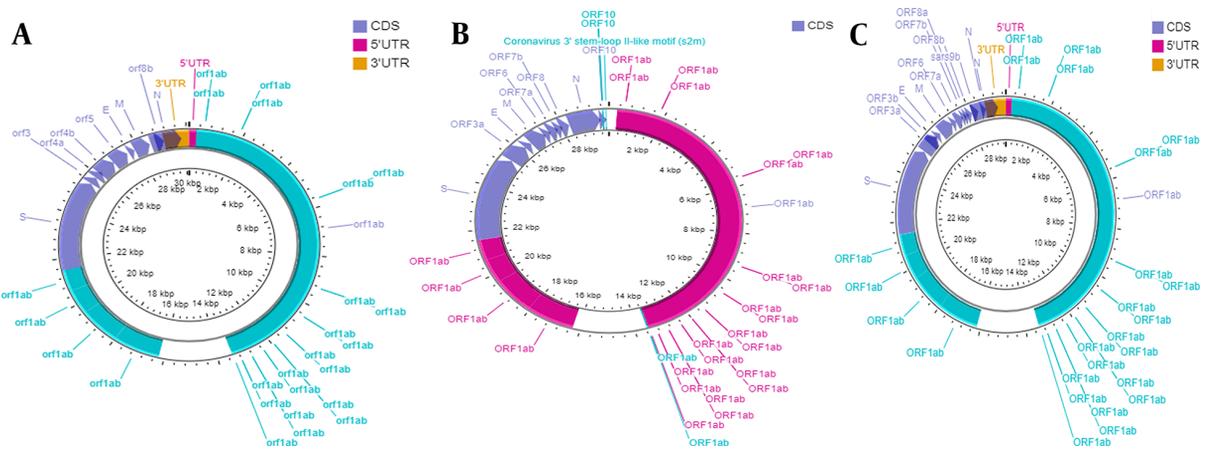
The S protein is a glycosylated protein found on the surface of coronaviruses, appearing as a spike. It mediates viral entry by binding RBD to ACE2 receptors on the host cell (21). As known, ACE2 encodes for an exopeptidase enzyme, which is responsible for angiotensin conversion to different forms (22). Besides, SARS-CoV-2 is believed to employ a similar method to enter the host cell as uses SARS-CoV (23).

The implication of ACE2 receptor in viral entry makes the cells expressing this receptor susceptible to SARS-CoV-2 infection. Diverse cells acting as hosts would imply the severity of symptoms. The major cells known to express this receptor include the epithelial cells of lung alveoli, the smooth muscle cells of the artery, the epithelium of the kidney tubule, and the epithelium of the small intestine (23). Though SARS-CoV-2 has been known to be primarily associated with respiratory problems, the presence of ACE2 receptors across tissues makes the body more vulnerable. Corresponding physiological manifestations have been reported as well, with diarrhea being present in the sample followed by the presence of virus in feces (22).

Expectedly, there are variations in the expression levels of ACE2 across tissues adding to the dynamics. Its expression is normally low in the lungs but can be enhanced under some physiological manifestations (24). The possibility of co-receptors aiding ACE2 is also very much plausible. Aminopeptidase N (ANPEP) and dipeptidyl-peptidase 4 (DPP4) are candidates for the same (25).

### 3.4. ACE2 Receptor Diversity in Nature

The ACE2 expression is reported across fish to mammals. Moreover, it shows structural conservation, as well. Human ACE2, when analyzed with that of the civet, bat, bird, snake, frog, and fish, showed amino acid sequence identities of 83, 81, 83, 61, 60, and 59%, respectively (24). Thus, ACE2 from these species can potentially interact with the RBD of SARS-CoV-2, making these animals not only prone to infection but also as reservoirs. This also highlights the host adaptability of the virus (24).



**Figure 1.** Genome organization of three different beta coronaviruses. A, SARS-CoV; B, SARS-CoV-2; and C, MERS-CoV. Figure were developed by using CGView DNA Plotter (13).

### 3.5. Spike Glycoprotein

The role of spike (S) glycoprotein in mediating the entry of SARS-CoV-2 in human host cells is well documented. At the protein level, it is comprised of an ectodomain at the N terminal, followed by a transmembrane region and a C-terminal intracellular tail (21). This ectodomain region makes up subunits S1 (receptor-binding) and S2 (membrane-fusion). The RBD, which is critical for interacting with host receptors, is localized in the S1 region. Interestingly, the S2 region is greatly conserved with 99% identity to not only SARS-CoV but also several bat coronaviruses (24). Once it is fused to the target receptor, S1 is thereon acted on by S2, as well as host protein cleaving enzymes (26). The induced changes in conformation can lead to the activation of membrane fusion proteins (27).

In terms of variation, the RBD of the S protein is highly diverse, but there are six amino acids, which, if affected, will directly impact the interaction with the host receptor. They are summarized in Table 1. Structural comparison of RBD reveals that it is almost superimposing for SARS-CoV and SARS-CoV-2. The only exception is a loop. The structure is primarily of beta-sheets with no breakers and multiple disulfide bonds forming cysteinyl residues: C366/C418 and C467/C474 (24). The S protein is also conserved concerning 22 glycosylation sites, with the exception of N370 being present only in SARS-CoV-2 (28).

Also, SARS-CoV-2 has a polybasic cleavage site (RRAR) which on cleavage influences viral infectivity and host range. For instance, if it changes to PRRA, it creates a cleavage site exclusive to SARS-CoV-2. This mutation has helped the spike protein evolve to be able to bind to human ACE2 (19).

### 3.6. RBD and ACE2 Interaction

The binding energy for the interaction between RBD and ACE2 receptors has been shown to be positive and involves two b-sheets and three loop structures. The residues directly associated with this interaction, if altered, will have implications in viral entry and hence pathogenesis (24). There are a total of 14 residues in the ACE2/RBD interface. Of these, nine are conserved, and four are variable (33). At the 479-position, amino acid aspergine (N) was mutated to lysine (K) & arginine (R) in civet and bat respectively. This phenomenon determines the species barrier of host preference in SARS-CoV infections. The N479K mutation reduces binding affinity drastically (34). Contrastingly, the S487T mutation enhances RBD/ACE2 interaction (33). Thus, mutations favoring this interaction would support transmission to humans and vice versa. As an example, SARS-CoV with T487S mutation, which hugely decreases binding affinity, exhibited no human to human transmission, implying the importance of the hydroxyl group (34).

The presence of hydrophobic amino acid L472 is crucial for RBD/ACE binding, which is substituted by proline and phenylalanine in SARS-CoV and SARS-CoV-2 (F486), respectively. It is localized in the loop region formed by C467/C474. This loop differs from CTPPALNC in SARS-CoV to CNGVEGFNC in SARS-CoV-2 (Table 1). The substitution of prolines by glycines alters the structure flexibility. Aromatic residues, when present in the region, further increases binding efficiency (24). Another substitution at P499 in SARS-CoV-2 has been reported to lead to better adaptability (32). Another important aspect of the S protein is the glycosylation sites. Of the reported 23 sites of this type, only two are present on RBD, but their impor-

**Table 1.** Amino Acid Changes in Spike Protein Between SARS-CoV and SARS-CoV-2

S. No.	SARS-CoV	SARS-CoV-2	Effect on RBD/ACE2 Interaction	References
1	N442	L455	RBD of SARS-CoV-2 has better efficiency than that of SARS-CoV	(29)
2	L472	F486	Hot spot 3; Adds flexibility to the interaction	(24)
3	N479	Q493	Hot spot 3; Promotes the civet to human transmission	(30)
4	D480	S494	Enhances viral binding to human ACE2	(18)
5	T487	N501 (hot spot 353)	RBD/ACE2 interaction is more efficient in SARS-CoV-2	(24)
6	Y491	Y505	This Y491/Y505 alteration has the enhanced RBD/ACE2 interaction at various temperature.	(31)
7	T499	P499	P499 forms better adaptation for host binding in SARS-CoV-2	(32)
8	CTPPALNC; 68 - 471	CNGVEGFNC; 482 - 485	Flexibility to RBD/ACE2 interaction	(24)
9	RRAR (cleavage site)	PRRA	The addition of Proline makes the cleavage site more unique and affects transmission and pathogenesis in the animal model	(19)

tance in RBD/ACE2 interaction, if any, needs to be ascertained.

### 3.7. Various New Strains of SARS-CoV-2 Lineage

Since November 2020, a new strain of SARS-CoV-2 has manifested in Britain. It is believed that 60% of recent infections happened in London by this new strain. As of 20 December 2020, a large number of areas in London, as well as South-East and East England, showed confirmed cases of this new variant. Scientists have named the new strain as “VUI-202012/01” (originally termed as B.1.1.7 by Public Health England), which became dominant over existing versions of SARS-CoV-2 (35). The 3D images have been shown in Figure 2, indicating the locations of key mutations of SARS-CoV, SARS-CoV-2, B.1.1.7 lineage, and B.1.617 lineage. High rates of mutation were shown in a short period in immunodeficient and immunosuppressed people infected by SARS-CoV-2 (36). This strain inhabited nonsynonymous mutations and deletion (36).

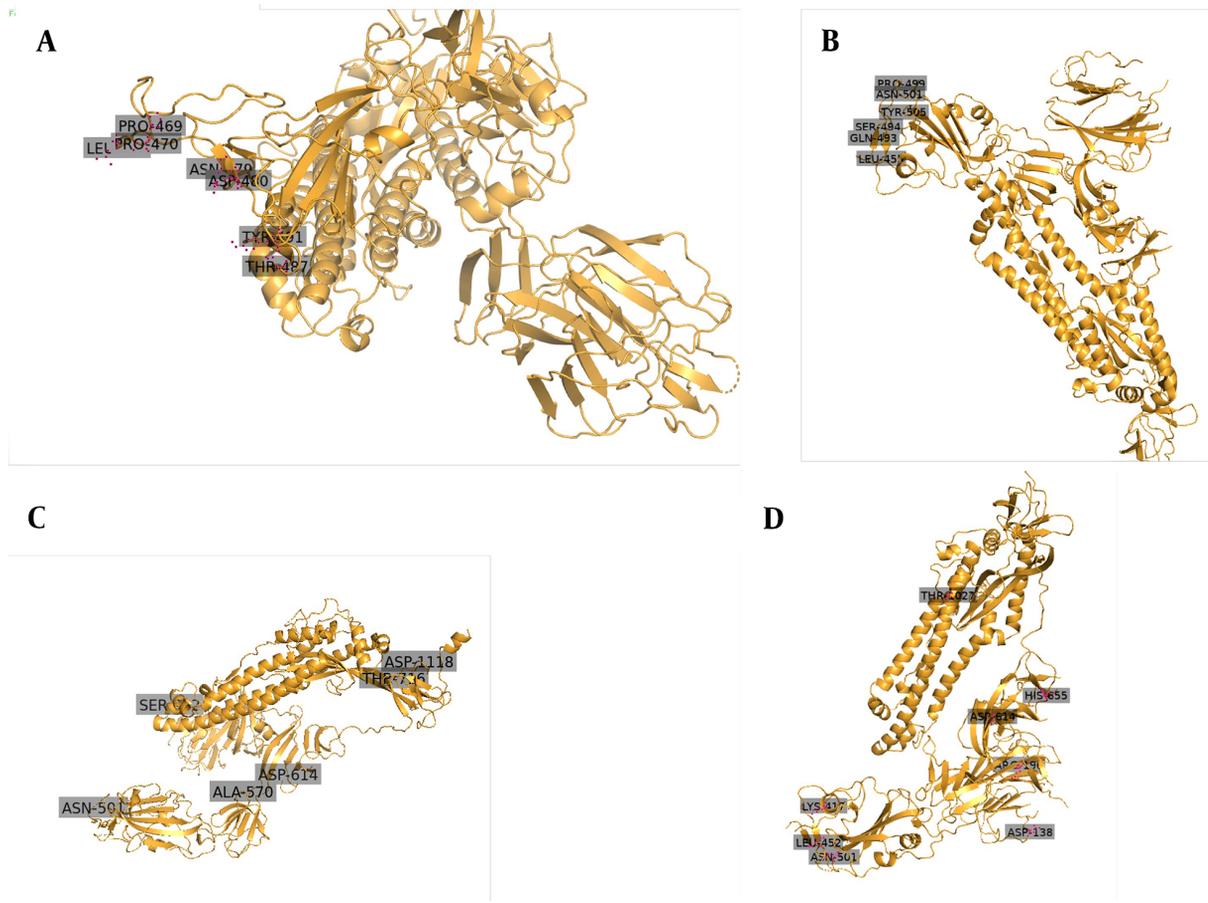
A study showed eminent changes in the Spike protein, importantly position 501 in RBD changes N (asparagine) to Y (tyrosine). That enhances the binding to human ACE2. This N501Y mutation can increase its ACE2 receptor binding affinity (37). The N501Y mutation also enhances virus adaption in the animal model (38). This mutation raises virus transmissibility by 70% (39). Another important mutation, P681H, occurs in the S1/S2 furin cleavage site (40). This furin cleavage site has been shown to enhance access to respiratory epithelial cells and also helps in virus transmission in the animal model (41). A deletion at 69-70HV in the N terminal of the spike protein has also been reported (42). Other important nonsynonymous mutations in the Spike protein include A570D, P681H, T716I, S982A, D1118H, and Y144 deletion (40).

A study suggested that SARS-CoV-2 lineage aggregate one to two mutations per month (43). Various SARS-CoV-2

lineages emerge throughout the world, and a few important mutations are listed in Table 2. Some mutations help viruses escape from the immune system and resist antibodies. Multiple mutations give rise to a new strain. Mutations in viruses are often a result of low polymerase fidelity and a survival mechanism to adjust a new host tropism (44). Recently an Indian variant named B.1.617 emerged with two mutations in the spike protein (L452 and E484Q mutations) separately (45). These two E484K and L452R mutations coexist and enhance the affinity for binding to ACE2 receptors (46). As per the World Health Organization, lineage B.1.617.2 was named as ‘delta variant’ with substitutions T478K, P681R, and L452R. Further, the B.1.617 variant also carries another unusual combinational mutation L452R & E484Q (designated as B.1.617.1 or kappa variant) and V383L mutation along with L452R & E484Q (known as B.1.63.3) [Coronavirus disease (COVID-19) situation reports (who.int)].

## 4. Conclusions

The biggest scientific challenge in combating the present pandemic has been the interpretation of constantly accumulating data from all over the world. The spike glycoprotein has emerged as the frontrunner therapeutic candidate, but its mutations and their impacts make it highly unpredictable. The widespread presence of ACE2 receptors in the body has led to varying physiological manifestations of SARS-CoV-2 infection. With long-term effects of the infection still unknown, a comprehensive understanding of the disease remains elusive, but the strong correlation at both genome and proteome levels with other coronaviruses should help us manage the pandemic.



**Figure 2.** The 3D images indicating the location of the key mutations. A, SARS-CoV; B, SARS-CoV-2; C, B.1.1.7 lineage; and D, B.1.617 lineage (images are from PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC).

**Table 2.** Important Mutations in SARS-CoV-2 Lineages

S.No.	Lineage B.1.1.7 or 501Y.V1/Alpha (UK)	Lineage B.1.351 or 501Y.V2 /Beta (South Africa)	Lineage B.1.1.24 or P.1/ Gamma Variant (Brazil)	B.1.617.2/Delta (India)	B.1.617.1/Kappa (India)	B.1.617.3	Effect of Amino Acid Changes in the New Strain	References
1	HV 69-70 del	L18F; D80A	L18F; T20N; P26S	T19R; 156-148 del	T19R; 156-148 del	T19R; 156-148 del	The H69/V70 single mutant showed two-fold more infectivity compared to the wild-type virus.	(47)
2	Y144 del	D215G; 242-245 del; R246I	D138Y; R190S	L452R mutation (without E484Q)	L452R.	V383I mutation along with L452R	Also present in California B.1.427/B.1.429. Affecting resistance to antibodies	(46)
3	E484K <sup>a</sup> , N501Y <sup>a</sup>	K417N; E484K <sup>a</sup> ; N501Y <sup>a</sup>	K417I; E484K <sup>a</sup> ; N501Y <sup>a</sup>	T478K	E484Q <sup>a</sup>	E484Q <sup>a</sup>	1.N501Y mutation was associated with increased transmissibility of the virus; 2.The E484K was reported to be an escape mutation from a monoclonal antibody, which neutralizes SARS-CoV-2; 3.E484 in RBM interacts with the K31 interaction hotspot, enhancing binding affinity; 4. K417N mutation has a functional significance in the receptor-binding domain; 5. E484K mutation in B.1.351, P.1, and P.3 imparts partial resistance to antibodies.	(31, 37, 45, 47, 48)
4	D614G <sup>a</sup> , A570D	D614G <sup>a</sup> , A701V	D614G <sup>a</sup> , H655Y	D614G <sup>a</sup>	D614G <sup>a</sup>	D614G <sup>a</sup>	Scientists showed that the 614 position of spike protein has a serine protease elastase 2 proteolytic site. This D614G mutation participates in proteolytic activity, which enhances viral entry into 293T:ACE2 cells.	(49)
5	P681H; T716L	-	-	P681R	P681R	P681R	P681H mutation is involved in creating a furin cleavage site at S1/S2, which promotes the entry of coronavirus to respiratory epithelial cells and animal model; 2.P681R arrived due to an increase in the amount of cleaved spike protein (S) on virions.	(40, 41, 46)
6	Q27Stop	-	-	-	-	-	Located in ORF8, it is a stop mutation that truncates the protein making it inactive.	(40)

<sup>a</sup> The conserved ones.

## Acknowledgments

The authors thank the Department of Biological Sciences, Aliah University, Kolkata, India, for all the financial and infrastructural support provided. The authors acknowledge all the authors associated with originating and submitting laboratories of the sequences from GISAID's EpiFlu™ (www.gisaid.org) database on which most of the research cited herein is based. The authors also acknowledge the ICMR-JRF grant with reference file no. 3/1/3/JRF-2019/HRD-64 (27200).

## Footnotes

**Authors' Contribution:** Taslima Nasrin, database search, data analysis, and draft preparation; Safdar Ali, conceptualization, supervision, draft editing, and finalization.

**Conflict of Interests:** The authors declare they have no competing interests.

**Funding/Support:** The authors thank the Department of Biological Sciences, Aliah University, Kolkata, India, for all the financial and infrastructural support provided.

## References

- Tan W, Zhao X, Ma X, Wang W, Niu P, Xu W, et al. A Novel Coronavirus Genome Identified in a Cluster of Pneumonia Cases - Wuhan, China 2019-2020. *China CDC Wkly.* 2020;**2**(4):61-2. [PubMed: 34594763]. [PubMed Central: PMC8393069].
- Zhou P, Yang XL, Wang XG, Hu B, Zhang L, Zhang W, et al. A pneumonia outbreak associated with a new Coronavirus of probable bat origin. *Nature.* 2020;**579**(7798):270-3. doi: 10.1038/s41586-020-2012-7. [PubMed: 32015507]. [PubMed Central: PMC7095418].
- Corman VM, Muth D, Niemeyer D, Drosten C. Hosts and Sources of Endemic Human Coronaviruses. *Adv Virus Res.* 2018;**100**:163-88. doi: 10.1016/bs.aivir.2018.01.001. [PubMed: 29551135]. [PubMed Central: PMC7112090].
- Fazeli Nasab B. [Biological Evaluation of Coronaviruses and the Study of Molecular Docking, Linalool, and Thymol as orfiab Protein Inhibitors and the Role of SARS-CoV-2 Virus in Bioterrorism]. *J Ilam Uni Med Sci.* 2021;**28**(6):77-96. Persian. doi: 10.29252/sjimu.28.6.77.
- Zhu N, Zhang D, Wang W, Li X, Yang B, Song J, et al. A Novel Coronavirus from Patients with Pneumonia in China, 2019. *N Engl J Med.* 2020;**382**(8):727-33. doi: 10.1056/NEJMoa2001017. [PubMed: 31978945]. [PubMed Central: PMC7092803].
- Shereen MA, Khan S, Kazmi A, Bashir N, Siddique R. COVID-19 infection: Origin, transmission, and characteristics of human coronaviruses. *J Adv Res.* 2020;**24**:91-8. doi: 10.1016/j.jare.2020.03.005. [PubMed: 32257431]. [PubMed Central: PMC7113610].
- de Wit E, van Doremalen N, Falzarano D, Munster VJ. SARS and MERS: Recent insights into emerging coronaviruses. *Nat Rev Microbiol.* 2016;**14**(8):523-34. doi: 10.1038/nrmicro.2016.81. [PubMed: 27344959]. [PubMed Central: PMC7097822].
- Wu F, Zhao S, Yu B, Chen YM, Wang W, Song ZG, et al. A new coronavirus associated with human respiratory disease in China. *Nature.* 2020;**579**(7798):265-9. doi: 10.1038/s41586-020-2008-3. [PubMed: 32015508]. [PubMed Central: PMC7094943].
- Laskar R, Ali S. Differential mutation profile of SARS-CoV-2 proteins across deceased and asymptomatic patients. *Chem Biol Interact.* 2021;**347**:109598. doi: 10.1016/j.cbi.2021.109598. [PubMed: 34303694]. [PubMed Central: PMC8299203].
- Wu A, Peng Y, Huang B, Ding X, Wang X, Niu P, et al. Genome Composition and Divergence of the Novel Coronavirus (2019-nCoV) Originating in China. *Cell Host Microbe.* 2020;**27**(3):325-8. doi: 10.1016/j.chom.2020.02.001. [PubMed: 32035028]. [PubMed Central: PMC7154514].
- Lu R, Zhao X, Li J, Niu P, Yang B, Wu H, et al. Genomic characterization and epidemiology of 2019 novel Coronavirus: Implications for virus origins and receptor binding. *Lancet.* 2020;**395**(10224):565-74. doi: 10.1016/S0140-6736(20)30251-8. [PubMed: 32007145]. [PubMed Central: PMC7159086].
- Angeletti S, Benvenuto D, Bianchi M, Giovanetti M, Pascarella S, Ciccozzi M. COVID-2019: The role of the nsp2 and nsp3 in its pathogenesis. *J Med Virol.* 2020;**92**(6):584-8. doi: 10.1002/jmv.25719. [PubMed: 32083328]. [PubMed Central: PMC7228367].
- Stothard P, Wishart DS. Circular genome visualization and exploration using CGView. *Bioinformatics.* 2005;**21**(4):537-9. doi: 10.1093/bioinformatics/bti054. [PubMed: 15479716].
- Ji W, Wang W, Zhao X, Zai J, Li X. Cross-species transmission of the newly identified coronavirus 2019-nCoV. *J Med Virol.* 2020;**92**(4):433-40. doi: 10.1002/jmv.25682. [PubMed: 31967321]. [PubMed Central: PMC7138088].
- Chan JF, Kok KH, Zhu Z, Chu H, To KK, Yuan S, et al. Genomic characterization of the 2019 novel human-pathogenic coronavirus isolated from a patient with atypical pneumonia after visiting Wuhan. *Emerg Microbes Infect.* 2020;**9**(1):221-36. doi: 10.1080/22221751.2020.1719902. [PubMed: 31987001]. [PubMed Central: PMC7067204].
- Hu D, Zhu C, Ai L, He T, Wang Y, Ye F, et al. Genomic characterization and infectivity of a novel SARS-like coronavirus in Chinese bats. *Emerg Microbes Infect.* 2018;**7**(1):154. doi: 10.1038/s41426-018-0155-5. [PubMed: 30209269]. [PubMed Central: PMC6135831].
- Liu Z, Xiao X, Wei X, Li J, Yang J, Tan H, et al. Composition and divergence of Coronavirus spike proteins and host ACE2 receptors predict potential intermediate hosts of SARS-CoV-2. *J Med Virol.* 2020;**92**(6):595-601. doi: 10.1002/jmv.25726. [PubMed: 32100877]. [PubMed Central: PMC7228221].
- Alagaili AN, Briese T, Mishra N, Kapoor V, Sameroff SC, Burbelo PD, et al. Middle East respiratory syndrome Coronavirus infection in dromedary camels in Saudi Arabia. *mBio.* 2014;**5**(2):e00884-14. doi: 10.1128/mBio.00884-14. [PubMed: 24570370]. [PubMed Central: PMC3940034].
- Andersen KG, Rambaut A, Lipkin WI, Holmes EC, Garry RF. The proximal origin of SARS-CoV-2. *Nat Med.* 2020;**26**(4):450-2. doi: 10.1038/s41591-020-0820-9. [PubMed: 32284615]. [PubMed Central: PMC7095063].
- Zhang T, Wu Q, Zhang Z. Pangolin homology associated with 2019-nCoV. *bioRxiv.* 2020; **Preprint.** doi: 10.1101/2020.02.19.950253.
- Li F. Structure, Function, and Evolution of Coronavirus Spike Proteins. *Annu Rev Virol.* 2016;**3**(1):237-61. doi: 10.1146/annurev-virology-110615-042301. [PubMed: 27578435]. [PubMed Central: PMC5457962].
- Holshue ML, DeBolt C, Lindquist S, Lofy KH, Wiesman J, Bruce H, et al. First Case of 2019 Novel Coronavirus in the United States. *N Engl J Med.* 2020;**382**(10):929-36. doi: 10.1056/NEJMoa2001191. [PubMed: 32004427]. [PubMed Central: PMC7092802].
- Xu X, Chen P, Wang J, Feng J, Zhou H, Li X, et al. Evolution of the novel Coronavirus from the ongoing Wuhan outbreak and modeling of its spike protein for risk of human transmission. *Sci China Life Sci.* 2020;**63**(3):457-60. doi: 10.1007/s11427-020-1637-5. [PubMed: 32009228]. [PubMed Central: PMC7089049].
- Chen Y, Guo Y, Pan Y, Zhao ZJ. Structure analysis of the receptor binding of 2019-nCoV. *Biochem Biophys Res Commun.* 2020. doi: 10.1016/j.bbrc.2020.02.071. [PubMed: 32081428]. [PubMed Central: PMC7092824].

25. Peck KM, Scobey T, Swanstrom J, Jensen KL, Burch CL, Baric RS, et al. Permissivity of Dipeptidyl Peptidase 4 Orthologs to Middle East Respiratory Syndrome Coronavirus Is Governed by Glycosylation and Other Complex Determinants. *J Virol.* 2017;**91**(19). doi: [10.1128/JVI.00534-17](https://doi.org/10.1128/JVI.00534-17). [PubMed: [28747502](https://pubmed.ncbi.nlm.nih.gov/28747502/)]. [PubMed Central: [PMC5599747](https://pubmed.ncbi.nlm.nih.gov/PMC5599747/)].
26. Sims AC, Baric RS, Yount B, Burkett SE, Collins PL, Pickles RJ. Severe acute respiratory syndrome Coronavirus infection of human ciliated airway epithelia: role of ciliated cells in viral spread in the conducting airways of the lungs. *J Virol.* 2005;**79**(24):15511-24. doi: [10.1128/JVI.79.24.15511-15524.2005](https://doi.org/10.1128/JVI.79.24.15511-15524.2005). [PubMed: [16306622](https://pubmed.ncbi.nlm.nih.gov/16306622/)]. [PubMed Central: [PMC1316022](https://pubmed.ncbi.nlm.nih.gov/PMC1316022/)].
27. Walls AC, Park YJ, Tortorici MA, Wall A, McGuire AT, Veesler D. Structure, Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein. *Cell.* 2020;**181**(2):281-292 e6. doi: [10.1016/j.cell.2020.02.058](https://doi.org/10.1016/j.cell.2020.02.058). [PubMed: [32155444](https://pubmed.ncbi.nlm.nih.gov/32155444/)]. [PubMed Central: [PMC7102599](https://pubmed.ncbi.nlm.nih.gov/PMC7102599/)].
28. Chuck CP, Wong CH, Chow LM, Fung KP, Wayne MM, Tsui SK. Expression of SARS-Coronavirus spike glycoprotein in *Pichia pastoris*. *Virus Genes.* 2009;**38**(1):1-9. doi: [10.1007/s11262-008-0292-3](https://doi.org/10.1007/s11262-008-0292-3). [PubMed: [18958613](https://pubmed.ncbi.nlm.nih.gov/18958613/)]. [PubMed Central: [PMC7088578](https://pubmed.ncbi.nlm.nih.gov/PMC7088578/)].
29. Shang J, Ye G, Shi K, Wan Y, Luo C, Aihara H, et al. Structural basis of receptor recognition by SARS-CoV-2. *Nature.* 2020;**581**(7807):221-4. doi: [10.1038/s41586-020-2179-y](https://doi.org/10.1038/s41586-020-2179-y). [PubMed: [32225175](https://pubmed.ncbi.nlm.nih.gov/32225175/)]. [PubMed Central: [PMC7328981](https://pubmed.ncbi.nlm.nih.gov/PMC7328981/)].
30. Li F. Structural analysis of major species barriers between humans and palm civets for severe acute respiratory syndrome Coronavirus infections. *J Virol.* 2008;**82**(14):6984-91. doi: [10.1128/JVI.00442-08](https://doi.org/10.1128/JVI.00442-08). [PubMed: [18448527](https://pubmed.ncbi.nlm.nih.gov/18448527/)]. [PubMed Central: [PMC2446986](https://pubmed.ncbi.nlm.nih.gov/PMC2446986/)].
31. Ku Z, Xie X, Davidson E, Ye X, Su H, Menachery VD, et al. Author Correction: Molecular determinants and mechanism for antibody cocktail preventing SARS-CoV-2 escape. *Nat Commun.* 2021;**12**(1):4177. doi: [10.1038/s41467-021-24440-x](https://doi.org/10.1038/s41467-021-24440-x). [PubMed: [34210985](https://pubmed.ncbi.nlm.nih.gov/34210985/)]. [PubMed Central: [PMC8246437](https://pubmed.ncbi.nlm.nih.gov/PMC8246437/)].
32. Othman H, Bouslama Z, Brandenburg JT, da Rocha J, Hamdi Y, Ghedira K, et al. Interaction of the spike protein RBD from SARS-CoV-2 with ACE2: Similarity with SARS-CoV, hot-spot analysis and effect of the receptor polymorphism. *Biochem Biophys Res Commun.* 2020;**527**(3):702-8. doi: [10.1016/j.bbrc.2020.05.028](https://doi.org/10.1016/j.bbrc.2020.05.028). [PubMed: [32410735](https://pubmed.ncbi.nlm.nih.gov/32410735/)]. [PubMed Central: [PMC7221370](https://pubmed.ncbi.nlm.nih.gov/PMC7221370/)].
33. 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. *J Virol.* 2020;**94**(7). doi: [10.1128/JVI.00127-20](https://doi.org/10.1128/JVI.00127-20). [PubMed: [31996437](https://pubmed.ncbi.nlm.nih.gov/31996437/)]. [PubMed Central: [PMC7081895](https://pubmed.ncbi.nlm.nih.gov/PMC7081895/)].
34. Li W, Zhang C, Sui J, Kuhn JH, Moore MJ, Luo S, et al. Receptor and viral determinants of SARS-Coronavirus adaptation to human ACE2. *EMBO J.* 2005;**24**(8):1634-43. doi: [10.1038/sj.emboj.7600640](https://doi.org/10.1038/sj.emboj.7600640). [PubMed: [15791205](https://pubmed.ncbi.nlm.nih.gov/15791205/)]. [PubMed Central: [PMC1142572](https://pubmed.ncbi.nlm.nih.gov/PMC1142572/)].
35. Centers for Disease Control and Prevention. *What is Genomic Surveillance?* Georgia, USA: Centers for Disease Control and Prevention; 2021, [cited 9th Jun 2021].
36. Avanzato VA, Matson MJ, Seifert SN, Pryce R, Williamson BN, Anzick SL, et al. Case Study: Prolonged Infectious SARS-CoV-2 Shedding from an Asymptomatic Immunocompromised Individual with Cancer. *Cell.* 2020;**183**(7):1901-1912 e9. doi: [10.1016/j.cell.2020.10.049](https://doi.org/10.1016/j.cell.2020.10.049). [PubMed: [33248470](https://pubmed.ncbi.nlm.nih.gov/33248470/)]. [PubMed Central: [PMC7640888](https://pubmed.ncbi.nlm.nih.gov/PMC7640888/)].
37. Starr TN, Greaney AJ, Hilton SK, Ellis D, Crawford KHD, Diggins AS, et al. Deep Mutational Scanning of SARS-CoV-2 Receptor Binding Domain Reveals Constraints on Folding and ACE2 Binding. *Cell.* 2020;**182**(5):1295-1310 e20. doi: [10.1016/j.cell.2020.08.012](https://doi.org/10.1016/j.cell.2020.08.012). [PubMed: [32841599](https://pubmed.ncbi.nlm.nih.gov/32841599/)]. [PubMed Central: [PMC7418704](https://pubmed.ncbi.nlm.nih.gov/PMC7418704/)].
38. Gu H, Chen Q, Yang G, He L, Fan H, Deng YQ, et al. Adaptation of SARS-CoV-2 in BALB/c mice for testing vaccine efficacy. *Science.* 2020;**369**(6511):1603-7. doi: [10.1126/science.abc4730](https://doi.org/10.1126/science.abc4730). [PubMed: [32732280](https://pubmed.ncbi.nlm.nih.gov/32732280/)]. [PubMed Central: [PMC7574913](https://pubmed.ncbi.nlm.nih.gov/PMC7574913/)].
39. Fratev F. The SARS-CoV-2 S1 spike protein mutation N501Y alters the protein interactions with both hACE2 and human derived antibody: A Free energy of perturbation study. *BioRxiv.* 2020;**Preprint**.
40. Rambaut A, Loman N, Pybus O, Barclay W, Barrett J, Carabelli A. *Preliminary genomic characterisation of an emergent SARS-CoV-2 lineage in the UK defined by a novel set of spike mutations.* USA: Virological; 2020, [cited 10th Jun 2021]. Available from: <https://virological.org/t/preliminary-genomic-characterisation-of-an-emergent-sars-cov-2-lineage-in-the-uk-defined-by-a-novel-set-of-spike-mutations/563>.
41. Zhu Y, Feng F, Hu G, Wang Y, Yu Y, Zhu Y, et al. The S1/S2 boundary of SARS-CoV-2 spike protein modulates cell entry pathways and transmission. *bioRxiv.* 2020;**Preprint**. doi: [10.1101/2020.08.25.266775](https://doi.org/10.1101/2020.08.25.266775).
42. McCarthy KR, Rennick LJ, Nambulli S, Robinson-McCarthy LR, Bain WG, Haidar G, et al. Natural deletions in the SARS-CoV-2 spike glycoprotein drive antibody escape. *bioRxiv.* 2021;**Preprint**. doi: [10.1101/2020.11.19.389916](https://doi.org/10.1101/2020.11.19.389916).
43. Duchene S, Featherstone L, Haritopoulou-Sinanidou M, Rambaut A, Lemey P, Baele G. Temporal signal and the phylodynamic threshold of SARS-CoV-2. *Virus Evol.* 2020;**6**(2):veaa061. doi: [10.1093/ve/veaa061](https://doi.org/10.1093/ve/veaa061). [PubMed: [33235813](https://pubmed.ncbi.nlm.nih.gov/33235813/)]. [PubMed Central: [PMC7454936](https://pubmed.ncbi.nlm.nih.gov/PMC7454936/)].
44. Sanjuan R, Domingo-Calap P. Mechanisms of viral mutation. *Cell Mol Life Sci.* 2016;**73**(23):4433-48. doi: [10.1007/s00018-016-2299-6](https://doi.org/10.1007/s00018-016-2299-6). [PubMed: [27392606](https://pubmed.ncbi.nlm.nih.gov/27392606/)]. [PubMed Central: [PMCS075021](https://pubmed.ncbi.nlm.nih.gov/PMCS075021/)].
45. Wang P, Nair MS, Liu L, Iketani S, Luo Y, Guo Y, et al. Antibody Resistance of SARS-CoV-2 Variants B.1.351 and B.1.1.7. *bioRxiv.* 2021;**Preprint**. doi: [10.1101/2021.01.25.428137](https://doi.org/10.1101/2021.01.25.428137). [PubMed: [33532778](https://pubmed.ncbi.nlm.nih.gov/33532778/)]. [PubMed Central: [PMC7852271](https://pubmed.ncbi.nlm.nih.gov/PMC7852271/)].
46. Tada T, Zhou H, Dcosta BM, Samanovic MI, Mulligan MJ, Landau NR. The Spike Proteins of SARS-CoV-2 B.1.617 and B.1.618 Variants Identified in India Provide Partial Resistance to Vaccine-elicited and Therapeutic Monoclonal Antibodies. *BioRxiv.* 2021;**Preprint**.
47. Weisblum Y, Schmidt F, Zhang F, DaSilva J, Poston D, Lorenzi JC, et al. Escape from neutralizing antibodies by SARS-CoV-2 spike protein variants. *Elife.* 2020;**9**. doi: [10.7554/eLife.61312](https://doi.org/10.7554/eLife.61312). [PubMed: [33112236](https://pubmed.ncbi.nlm.nih.gov/33112236/)]. [PubMed Central: [PMC7723407](https://pubmed.ncbi.nlm.nih.gov/PMC7723407/)].
48. Tegally H, Wilkinson E, Giovanetti M, Iranzadeh A, Fonseca V, Giandhari J, et al. Emergence and rapid spread of a new severe acute respiratory syndrome-related coronavirus 2 (SARS-CoV-2) lineage with multiple spike mutations in South Africa. *medRxiv.* 2020;**Preprint**. doi: [10.1101/2020.12.21.20248640](https://doi.org/10.1101/2020.12.21.20248640).
49. Hu J, He C, Gao Q, Zhang G, Cao X, Long Q, et al. D614G mutation of SARS-CoV-2 spike protein enhances viral infectivity. *Microbiology.* 2020;**Preprint**. doi: [10.1101/2020.06.20.161323](https://doi.org/10.1101/2020.06.20.161323).