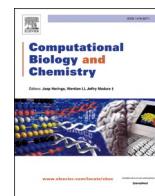




Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.



## Genetic analysis of SARS-CoV-2 isolates collected from Bangladesh: Insights into the origin, mutational spectrum and possible pathomechanism

Md. Sorwer Alam Parvez<sup>a</sup>, Mohammad Mahfujur Rahman<sup>a</sup>, Md. Niaz Morshed<sup>a</sup>,  
Dolilur Rahman<sup>a</sup>, Saeed Anwar<sup>b</sup>, Mohammad Jakir Hosen<sup>a,\*</sup>

<sup>a</sup> Department of Genetic Engineering & Biotechnology, Shahjalal University of Science & Technology, Sylhet, 3114, Bangladesh

<sup>b</sup> Department of Medical Genetics, Faculty of Medicine and Dentistry, University of Alberta, 8440 112 St. NW, Edmonton, AB, T6G 2R7, Canada

### ARTICLE INFO

#### Keywords:

COVID-19  
SARS-CoV-2  
Bangladeshi isolates  
Spike protein  
Mutation  
ACE2 receptor

### ABSTRACT

As the coronavirus disease 2019 (COVID-19), caused by the severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2), rages across the world, killing hundreds of thousands and infecting millions, researchers are racing against time to elucidate the viral genome. Some Bangladeshi institutes are also in this race, sequenced a few isolates of the virus collected from Bangladesh. Here, we present a genomic analysis of these isolates. The analysis revealed that SARS-CoV-2 isolates sequenced from Dhaka and Chittagong were the lineage of Europe and India, respectively. Our analysis identified a total of 42 mutations, including three large deletions, half of which were synonymous. Most of the missense mutations in Bangladeshi isolates found to have weak effects on the pathogenesis. Some mutations may lead the virus to be less pathogenic than the other countries. Molecular docking analysis to evaluate the effect of the mutations on the interaction between the viral spike proteins and the human ACE2 receptor, though no significant difference was observed. This study provides some preliminary insights into the origin of Bangladeshi SARS-CoV-2 isolates, mutation spectrum and its possible pathomechanism, which may give an essential clue for designing therapeutics and management of COVID-19 in Bangladesh.

### 1. Introduction

The coronavirus disease 2019 (COVID-19) is an infectious disease caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). Human to human transmission of this viral infection occurs via the droplets expelled during the face to face talking, coughing and sneezing. The time of exposure is very crucial factors for the transmission of infection from an infected person to a healthy person. Prolonged exposure has a high risk of transmission while short exposure is less likely to transmit. It takes an average of 11.5 days to develop the symptoms of the disease after the successful transmission to a healthy person (Tang et al., 2020; Wiersinga et al., 2020). Common symptoms include fever, cough, fatigue, shortness of breath, nausea, vomiting, and diarrhoea. The disease has emerged as a critical, rapidly evolving global health crisis (Yin et al., 2020; Zheng et al., 2020). More than 6.5 million people have contracted the virus, and nearly 400 thousand have died (CSSE, 2020). In Bangladesh, the COVID-19 was first reported on 7 March by the Institute of Epidemiology Disease Control and Research (IEDCR) (Paul, 2020). Until the end of March, the infection rate was sort of low;

however, as the non-therapeutic prevention measures enforced by the government faced enormous challenges, the infection rate raised drastically in April and kept on rising (Nabi and Shovon, 2020). The people did not maintain the social distancing enforced by the government and trend to gather in crowded places (The Business Standard, 2020). Moreover, an inadequacy of testing for COVID-19 diagnosis is a common criticism in Bangladesh (Tithila, 2020). As of 31 August 2020, nearly 313,000 confirmed cases were reported, with a total of 4281 deaths in Bangladesh (IEDCR, 2020).

SARS-CoV-2 is a positive-stranded RNA virus with a genome of ~ 30 kb, encodes structural and non-structural proteins. SARS-CoV-2 is a spherical-shaped enveloped virus and characterized by spike proteins projecting from the virion surface. Generally, the viral structure is formed with some structural proteins such as spike (S), membrane (M), envelope (E) and nucleocapsid (N) protein where S, M and E proteins are embedded in the viral envelope and N protein is located in the core regions (Ashour et al., 2020). Like other RNA viruses, the SARS-CoV-2 is prone to frequent mutations, which makes it challenging to develop therapeutics and vaccines against the virus (Ruan et al., 2003; Wang

\* Corresponding author.

E-mail addresses: [sorwersust@yahoo.com](mailto:sorwersust@yahoo.com) (Md.S.A. Parvez), [sanwar@ualberta.ca](mailto:sanwar@ualberta.ca) (S. Anwar), [jakir-gen@sust.edu](mailto:jakir-gen@sust.edu) (M.J. Hosen).

et al., 2020). Sequence information of both the pathogen and the host would greatly facilitate an effective therapeutic strategy or vaccine development (Seib et al., 2009). Analysis of the genome sequences obtained from a vast array of isolates collected from different regions could provide an idea about the efficacy of the vaccines being developed (Amanat and Krammer, 2020). Henceforth, researchers across the world are running against time to unravel the genomic insights into the virus.

Till the end of August 2020, more than 80,000 genome sequences of SARS-CoV-2 has been submitted from different countries, where most of the sequences have come from European countries (~46,000). About 20,000 complete genome sequences have been submitted from the USA while China has submitted about 1000 genome sequences. In Bangladesh, more than 300 isolates of the virus have been sequenced and deposited in GSAID (Global Initiative on Sharing Avian Influenza Data) database until the end of August 2020. Unfortunately, there is yet a study on the genomics of the SARS-CoV-2 in Bangladeshi isolates.

This study aimed to provide some preliminary insights into the genetic structure of all isolates reported in Bangladesh along with the mutational spectrum. It presents the first study on SARS-CoV-2 genomes obtained from Bangladesh, which, in broader terms, would help the therapeutic strategy development and vaccination programs against the virus in the country.

## 2. Materials and methods

### 2.1. Retrieval of the genome sequences of SARS-CoV-2

Till the end of August, more than 300 genome sequences of SARS-CoV-2 isolates were deposited from Bangladesh in the GSAID database (<https://www.gisaid.org/>) and we retrieved all of them from the database. As many of the Bangladeshi people return during the COVID-19 outbreak mainly from China, India, Saudi Arabia, Spain, Italy, Japan, Qatar, Canada, Kuwait, USA, France, Sweden, and Switzerland, the first deposited genome sequence of those countries were also retrieved. Sequence information of the first isolate collected from China was considered as a reference for further analysis.

### 2.2. Multiple sequence alignment and phylogenetic tree reconstruction

Multiple sequence alignment of all genome sequences of Bangladeshi isolates along with other countries was done using MUSCLE alignment program (Edgar, 2004). This alignment file was further proceeded for the reconstruction of the phylogenetic tree with Maximum Likelihood (ML) method using IQ-TREE (Nguyen et al., 2015). Model test was performed by ModelFinder tools to select the best-fit substitution model for the tree reconstruction (Kalyaanamoorthy et al., 2017). A total number of 88 models were tested and the best-fit model (GTR + F+I + G4) was selected according to Bayesian Information Criterion (BIC). Besides, to assess the branch supports, the ultrafast bootstrap was performed adopting UFBoot2 and the number of bootstrap replicates was set to 1000 (Hoang et al., 2018). Finally, the reconstructed phylogenetic tree was visualized and analyzed by iTOL online tool (Letunic and Bork, 2019).

### 2.3. Identification of nucleotide variations in Bangladeshi Strain

To identify the nucleotide variations, we performed multiple sequence alignment using Clustal Omega (Sievers and Higgins, 2014; Madeira et al., 2019), and the sequence of the strain China [EPI\_ISL\_402124] was used as a reference genome. The alignment file was analyzed using MVIEW program of Clustal Omega (Brown et al., 1998). Till 20th May, only 14 complete genome sequences of SARS-CoV-2 isolates were deposited in the database from Bangladesh and our further analysis was done using these 14 sequences.

### 2.4. Prediction and identification of the viral genes

FGENESV of SoftBerry (<http://linux1.softberry.com/berry.phtml>), which is a Trained Pattern/Markov chain-based viral gene prediction tools, was adopted for the prediction of the genes as well as the proteins from the viral genomes. Each predicted protein (for each viral genomes) was identified using the Basic Local Alignment Search Tool (BLAST), at the interface of the National Center for Biotechnology Information (NCBI) (Madden, 2013). The identity of each protein was evaluated compared to the proteins of the reference strain.

### 2.5. Detection of mutation Spectrum

Again, Clustal Omega was used for the multiple sequence alignment of each protein, which further analyzed by MVIEW. The amino acid variations were identified in each protein comparing to the protein of the reference strain. Further, both nucleotide variations and amino acid variations were compared to study the types of mutations.

### 2.6. Prediction of mutational effects

The structural and functional effects of the missense variants, along with the stability change, were analyzed using different prediction tools. I-mutant was employed to analyze the stability change where all the parameters were kept in default (Capriotti et al., 2005). Additionally, Mutpred2 was adopted to predict the molecular consequences and functional effect of these mutations (Pejaver et al., 2017).

### 2.7. Homology modeling of spike proteins and model validation

The BLASTp program at the NCBI interface (link) was used to find the most suitable template for homology modeling. Blasting against the protein databank reservoir (PDB) identified spike protein (Human) with PDB ID: 6VSB as a suitable template, as it has 99.59 % sequence similarity and 94 % coverage with the target sequence. The homology modeling of all mutant spike proteins along with the spike protein of the reference was done using SWISS-MODEL (Schwede et al., 2003). The validation of the predicted model was done by adopting Rampage and ERRAT (Colovos and Yeates, 1993; Lovell, 2002).

### 2.8. Molecular docking of Spike Protein with ACE2 receptor

The molecular docking approach was employed to investigate the interaction of mutant spike protein with the human ACE2 receptor. First, the crystal structure of human ACE2 (PDB ID: 6D0G) was obtained from the Protein Data Bank (PDB), and PyMOL was used to clean the structure to remove all the complex molecules and water (Berman et al., 2000; DeLano, 2002). The HDOCK webserver was used for prediction of the interaction between Spike protein and human ACE2 receptor through the protein-protein molecular docking (Yan et al., 2017). PyMOL was also used for the visualization of docking interactions.

## 3. Results

### 3.1. Retrieved genome sequence of the SARS-CoV-2

A total number of 311 complete genome sequences of the SARS-CoV-2 isolates from Bangladesh and 12 genome sequence from the isolates of other countries (China, India, Saudi Arabia, Spain, Italy, Japan, Qatar, Canada, Kuwait, USA, France, Sweden, and Switzerland) have been retrieved from GSAID. The strain of Wuhan accession number with EPI\_ISL\_402124 was considered as the reference strain.

### 3.2. Phylogenetic tree analysis

Phylogenetic tree analysis revealed that Bangladeshi isolates which

initially collected from Dhaka were very close to Spain as well as Switzerland whereas the isolates collected from Chittagong were found to share a common ancestor with India and USA (Fig. 1). The isolates collected from Chittagong also centroid with the Middle East countries such as Kuwait and Saudi Arabia. Moreover, all the isolates initially collected before the pandemic started in Bangladesh were clustered with the strain of China indicating the same lineage of the virus. However, the phylogenetic distance of the isolates from this lineage increased over time.

### 3.3. Predictions of the genes and proteins

FGENESV predicted the presence of 12 genes in the reference. Interestingly, all except five isolates (EPI\_ISL\_445213, EPI\_ISL\_445214, EPI\_ISL\_450342, EPI\_ISL\_450343, and EPI\_ISL\_450344) of Bangladesh also showed a similar result. Both isolates EPI\_ISL\_445213 and EPI\_ISL\_445217 found to have ten genes (missing of ORF7a and ORF10 genes) and isolate EPI\_ISL\_450343 and EPI\_ISL\_450344 have 11 genes (missing ORF8 gene). Multiple sequence alignment revealed that most of the variation in Bangladeshi isolates occurred in the ORF1a polyprotein, surface glycoproteins, and nucleocapsid phosphoprotein. Remarkably, envelope glycoprotein, ORF6, ORF8, and ORF10 were found 100 % identical in most of the isolates compared to the reference sequence (Table 1).

### 3.4. Mutation Spectrum of bangladeshi SARS-CoV-2 isolates

Analysis of all 14 Bangladeshi isolates revealed a total of 42 single nucleotide variants (Fig. 2); 24 of them were nonsynonymous missense in character. Besides, three large deletions were also found in those isolates (Table 2). Among the deletions, two deletions were responsible for the deletion of ORF7a in EPI\_ISL\_445213 and EPI\_ISL\_445217 isolates. Another large deletion from nucleotide 27911–28254, occurred in EPI\_ISL\_450343 and EPI\_ISL\_450344 isolates, responsible for the deletion of ORF8 in both isolates. Surprisingly, three consecutive mutations were found at nucleotide position 28882–28884; resulted in two amino acids substitution in nucleocapsid phosphoprotein.

### 3.5. Mutational effects

Mutational effects analysis of the 24 missense mutations found that 18 mutations were responsible for decreasing structural stability. Mutations located in the ORF1a polyprotein and surface glycoprotein were predicted to decrease the structural stability of both proteins (Table 3). Additionally, three mutations occurring in surface glycoprotein, ORF3a and ORF6 were predicted to alter the molecular consequences, including loss of sulfation in surface glycoprotein and loss of proteolytic cleavage in ORF3a and loss of allosteric site in ORF6 (Table 4 and Supplementary Table 1).

### 3.6. Prediction and validation of the homology models

In total, three models were generated using the template PDB ID: 6VSB; one model for the spike protein of reference strain, and the two others were for two different mutant isolates from Bangladesh (Fig. 3). Two types of mutations were found in the spike proteins of all Bangladeshi isolates, where most of the isolates were found to contain a substitution of D623 G. Only one strain, EPI\_ISL\_445214, found to have two substitutions; one was similar to the previous substitution, and the other was F1118 L. The validation assessment scores of these three models were mostly similar to the template, which provided the reliability of these models (Table 5).

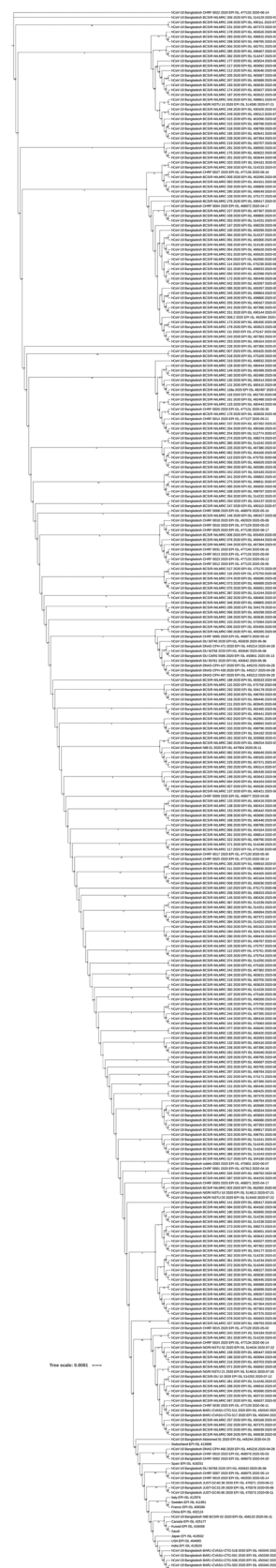
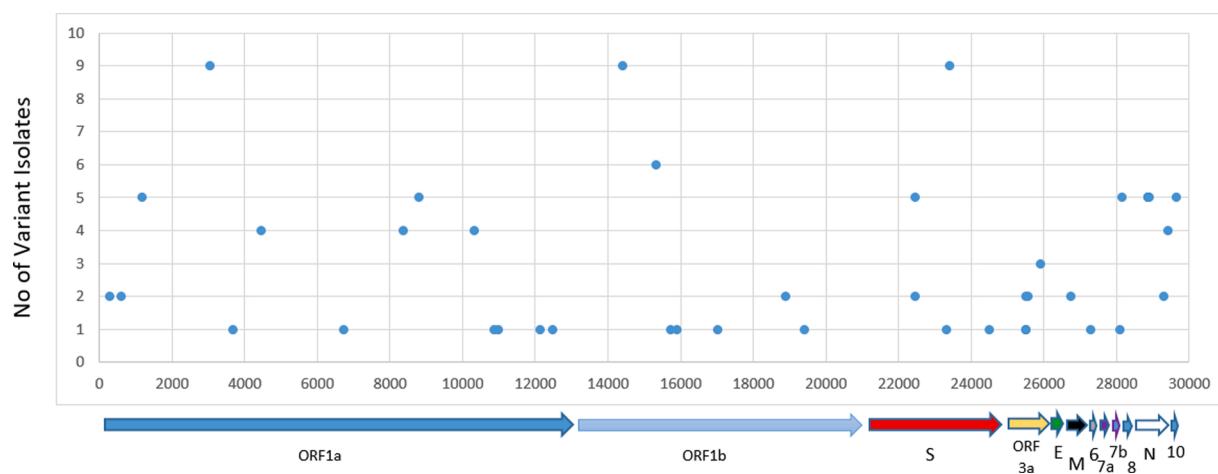


Fig. 1. Maximum likelihood phylogenetic tree reconstructed with the sequences of all Bangladeshi isolates and other countries. The value in the nodes represents the bootstrap value of the branches where the branch length represents the evolutionary distance.

**Table 1**

Predicted number of genes and identity compared to the reference strain. (Legends: S1: EPI\_ISL\_437912; S2: EPI\_ISL\_445213; S3: EPI\_ISL\_445214; S4: EPI\_ISL\_445215; S5: EPI\_ISL\_445216; S6: EPI\_ISL\_445217; S7: EPI\_ISL\_445244; S8: EPI\_ISL\_450339; S9: EPI\_ISL\_450340; S10: EPI\_ISL\_4503441; S11: EPI\_ISL\_450342; S12: EPI\_ISL\_450343; S13: EPI\_ISL\_450344; S14: EPI\_ISL\_450345; M: Missing).

No	Protein	S1	S2	S3	S4	S5	S6	S7	S8	S9	S11	S11	S12	S13	S14
1	ORF1a Polyprotein	99.98	99.93	99.95	99.95	100	99.95	100	99.95	99.98	99.98	99.95	99.98	99.98	99.98
2	ORF1b Polyprotein	99.96	100	100	100	100	100	100	100	100	100	99.96	100	100	100
3	Surface Glycoprotein	99.92	100	99.84	99.92	99.92	99.92	99.92	100	100	100	100	99.92	99.92	100
4	ORF3a protein	100	99.64	100	99.64	100	99.64	100	100	100	100	100	99.27	99.64	99.64
5	envelope protein	100	100	100	100	100	100	100	100	100	100	100	100	100	100
6	Membrane Glycoprotein	100	100	100	100	100	100	100	100	100	100	100	100	100	100
7	ORF6 protein	100	100	100	100	100	100	100	100	100	100	100	99.36	100	100
8	ORF7a protein	100	M	100	100	100	M	100	100	100	100	100	100	100	100
9	ORF7b	100	100	100	100	100	100	100	100	100	100	100	100	100	100
10	ORF8	100	100	100	100	100	100	100	99.17	100	99.17	99.17	M	M	99.35
11	Neucleocapsid phosphoprotein	99.52	99.76	99.28	99.28	100	99.28	100	99.76	99.52	99.52	99.76	99.76	99.76	99.76
12	ORF10	100	M	100	100	100	M	100	100	100	100	M	100	100	100



**Fig. 2.** Variations Plot of SARS-CoV-2 in Bangladeshi isolates.

### 3.7. Analysis of the interaction between spike proteins and human ACE2 receptor

HDOCK server was used to predict the interaction between the above-mentioned 3D models of reference spike proteins along with mutant models and the human ACE2 receptor. Interestingly, this molecular docking analysis revealed that the docking score for the three models against the human ACE2 receptor was similar, and it was -244.42 (Table 6); mutation in the spike proteins do not hamper binding with ACE2 receptor. For three spike protein models, this study found that a domain of spike protein instead of whole protein, amino acid ranging from 345 to 527, was involved in the interactions. This domain was conserved in all isolates resulting in similar interactions with ACE2 (Fig. 4).

## 4. Discussion

COVID-19 has become a global challenge for the scientific communities affecting millions of people and taking thousands of lives every day. Scientists worldwide are working hard to combat against SARS-CoV-2, but no significant outcome is obtained (Lake, 2020; Yuen et al., 2020). Along with other studies, genetic studies can give a significant clue to understanding the pathogenesis of COVID-19. Together with the critical therapeutic target, the genomic sequence data may provide insights into the pattern of global spread, the diversity during the epidemics, and the dynamics of evolutions, which are crucial to unwind the molecular mechanism of COVID-19 (Khailany et al., 2020). This study gives insights into the transmission of SARS-CoV-2, genetic diversity of the isolates, and predicts the impacts of mutations in

Bangladesh.

It has been reported that during the COVID-19 outbreak about 600,000 people had entered into Bangladesh from the other countries including Spain and Switzerland (wsws, 2020). The phylogenetic study revealed that the Bangladeshi isolates found in Dhaka were descendent from Europe, and most of the isolates from Chittagong are descendent from India. India is the neighbour country of Bangladesh and a lot of people crosses the border between Bangladesh-India every day for business, education and treatment purposes. So, the chances of India for being the origin of the virus which caused the COVID-19 pandemic in Bangladesh is very high. Besides, Middle East could also be a potential source of the virus as they were very close to the isolates collected from Chittagong. However, some isolates of Chittagong were close to the isolates from Dhaka. Dhaka is the capital city of Bangladesh and the sixth most densely populated city in the world. This virus may spread to other regions of the country from this city as it is the central hub of Bangladesh for financial, political, entertainment, and education. The SARS-CoV-2 isolates collected from Chittagong are close to the strain from the Middle East is not surprising. As most of the migrants from Bangladesh live in Middle East are from Chittagong, and during the COVID-19 outbreak, thousands of them returned to their home city (Dastider, 2018; Ullah, 2020). Moreover, the phylogenetic distance from the initially collected isolates increased over time which indicates about the extensive mutation that the virus had gone during the human to human transmission in Bangladesh.

Mutation in the viral genome is a ubiquitous phenomenon for the viruses to escape the host defence. But the mutation rate in SARS-CoV-2 much lower than the other RNA viruses, including seasonal flu viruses (Oberemok et al., 2020). In this study, there was found some variations

**Table 2**

All mutations found in the coding regions of the 14 isolates compared to the reference strain. (Legends: S1: EPI\_ISL\_437912; S2: EPI\_ISL\_445213; S3: EPI\_ISL\_445214; S4: EPI\_ISL\_445215; S5: EPI\_ISL\_445216; S6: EPI\_ISL\_445217; S7: EPI\_ISL\_445244; S8: EPI\_ISL\_450339; S9: EPI\_ISL\_450340; S10: EPI\_ISL\_450341; S11: EPI\_ISL\_450342; S12: EPI\_ISL\_450343; S13: EPI\_ISL\_450344; S14: EPI\_ISL\_450345).

Strain	Mutation	Protein	Amino Acid Changes	Mutation Types
S11, 14	283:C > T	ORF1a	No change	Synonymous
S9, 10	602:C > T	Polyprotein	No Change	Synonymous
S1,2,3, 4,6	1164:A > T	ORF1a	I300F	Missense
S1,2,3, 4, 5, 6, 7, 12, 13	3038:C > T	ORF1a	No Change	Synonymous
S5	3689:C > T	Polyprotein	No Change	Synonymous
S2,3, 4, 6	4445:G > T	ORF1a	No Change	Synonymous
S8	6730:A > G	ORF1a	N2155S	Missense
S2, 3, 4, 6	8372:G > T	ORF1a	Q2702H	Missense
S8, 9, 10, 11, 14	8783:C > T	ORF1a	No change	Synonymous
S8, 9, 10, 11	10330:A > G	ORF1a	D3355G	Missense
S14	10871:G > T	ORF1a	K3353R	Missense
S2	10980:G > A	ORF1a	V3572M	Missense
S11	12120:C > T	ORF1a	P3952S	Missense
S8	12485:C > T	ORF1a	No Change	Synonymous
S1, 2, 3, 4, 5, 6, 7, 12, 13	14409:C > T	ORF1ab	P214L	Missense
S5, 8, 9, 10, 11, 14	15325:C > T	ORF1ab	No Change	Synonymous
S8	15739:C > T	ORF1ab	No change	Synonymous
S4	15896:C > T	ORF1ab	No Change	Synonymous
S1	17020:G > T	ORF1ab	E1084D	Missense
S12, 13	18878:C > T	ORF1ab	No Change	Synonymous
S11	19405:G > A	ORF1ab	V1883T	Missense
S12, 13	22445:C > T	Surface Glycoprotein	No change	Synonymous
S14	23321:C > T	Surface Glycoprotein	No change	Synonymous
S8, 9, 10, 11, 14	22469:G > T	Surface Glycoprotein	No change	Synonymous
S1,2, 3, 4, 5, 6, 7, 12, 13	23404:A > G	Surface Glycoprotein	D623G	Missense
S3	24488:T > C	Surface Glycoprotein	F1118L	Missense
S12, 13	25495:G > T	ORF3a protein	No change	Synonymous
S14	25506:A > T	ORF3a protein	Q38L	Missense
S12	25512:C > T	ORF3a protein	S40L	Missense
S12, 13	25564:G > T	ORF3a protein	Q57H	Missense
S2, 4, 6	25907:G > T	ORF3a protein	G172C	Missense
S12, 13	26736:C > T	Membrane Glycoprotein	No Change	Synonymous
S12	27282:G > T	ORF6 protein	W27L	Missense
S2	27432–27651:DEL	ORF7a protein	Whole protein deletion	Deletion

**Table 2 (continued)**

Strain	Mutation	Protein	Amino Acid Changes	Mutation Types
S6	27486–27613:DEL	ORF7a protein	Whole protein deletion	Deletion
S12, 13	27911–28254:DEL	ORF8	Whole protein deletion	Deletion
S14	28098:C > T	ORF8	A65V	Missense
S8, 9, 10, 11, 14	28145:T > C	ORF8	L84S	Missense
S8, 9, 10, 11, 14	28879:G > A	Neucleocapsid phosphoprotein	S202N	Missense
S1,2,3, 4, 6	28882:G > A	Neucleocapsid phosphoprotein	R203K	Missense
S1,2,3, 4, 6	28883:G > A	Neucleocapsid phosphoprotein	R203K	Missense
S1,2,3, 4, 6	28884:G > C	Neucleocapsid phosphoprotein	G204R	Missense
S9, 10	29293:G > T	Neucleocapsid phosphoprotein	K373N	Missense
S2,3, 4, 6	29404:A > G	Neucleocapsid phosphoprotein	D377G	Missense
S8, 9, 10, 11, 14	29643:G > A	ORF10	No Change	Synonymous

**Table 3**

Prediction of the mutational effects on the structural stability.

Protein	Amino Acid Changes	SVM2 Prediction Effect	DDG Value (kcal/mol)
ORF1a Polyprotein	I300F	Decrease	-1.79
ORF1a Polyprotein	N2155S	Decrease	-0.60
ORF1a Polyprotein	Q2702H	Decrease	-0.68
ORF1a Polyprotein	D3355G	Decrease	-0.95
ORF1a Polyprotein	K3353R	Increase	-0.13
ORF1a Polyprotein	V3572M	Decrease	-0.88
ORF1a Polyprotein	P3952S	Decrease	-1.21
ORF1b Polyprotein	P214L	Decrease	-0.83
ORF1b Polyprotein	E1084D	Decrease	-0.75
ORF1b Polyprotein	V1883T	Decrease	-1.46
Surface Glycoprotein	D623G	Decrease	-0.93
Surface Glycoprotein	F1118L	Decrease	-0.81
ORF3a protein	Q38L	Increase	0.12
ORF3a protein	S40L	Increase	0.40
ORF3a protein	Q57H	Decrease	-0.90
ORF3a protein	G172C	Decrease	-0.83
ORF6 protein	W27L	Decrease	-0.96
ORF8	A65V	Increase	0.02
ORF8	L84S	Decrease	-2.29
Neucleocapsid phosphoprotein	S202N	Increase	-0.78
Neucleocapsid phosphoprotein	R203K	Decrease	-0.93
Neucleocapsid phosphoprotein	G204R	Decrease	-0.52
Neucleocapsid phosphoprotein	K373N	Increase	-0.10
Neucleocapsid phosphoprotein	D377G	Decrease	-0.44

in the SARS-CoV-2 isolated in Bangladesh, which may affect the epidemiology and pathogenicity of the virus. A total of 42 mutations were identified with a large deletion in the coding regions, where about half were synonymous. Even some isolates were found not to encode one or more accessory proteins such as ORF7a, ORF8, and ORF10 caused by a large deletion in the genome. An 80-nucleotides deletion in ORF7a was also reported by a study conducted in Arizona (Mercatelli and Giorgi, 2020). Absent of these accessory proteins may have adverse effects on the viral replication or pathogenesis and the expression of structural protein E (Keng et al., 2006).

**Table 4**  
Prediction of the mutational effects on the molecular consequences.

Protein Name	Mutation	Effects
Surface Glycoprotein	F1118L	Altered Ordered interface
		Altered Disordered interface
		Altered DNA binding
		Loss of Sulfation at Y1119
		Altered Metal binding
ORF3a	G172C	Loss of O-linked glycosylation at S171
		Gain of Disulfide linkage at G172
		Loss of Intrinsic disorder
		Altered Transmembrane protein
		Altered Ordered interface
ORF6	W27L	Gain of Loop
		Loss of Proteolytic cleavage at D173
		Altered Ordered interface
		Altered Disordered interface
		Loss of Strand
		Gain of Helix
		Loss of Allosteric site at F22
Gain of Sulfation at Y31		
		Altered DNA binding
		Altered Transmembrane protein

Moreover, ORF8 is involved in the crucial adaptation pathways of coronavirus from human-to-human. At the same time, ORF7a contributes to the viral pathogenesis in the host by inhibiting Bone Marrow Stromal Antigen 2 (BST-2), which restricts the release of coronaviruses from affected cells. Loss of ORF7a causes a much more significant restriction of the virus's spreading into the host (Taylor et al., 2015; Decaro and Lorusso, 2020). Loss of these accessory proteins may lead to the virus being less pathogenic, resulting in a meager infection rate and mortality compared to the other countries (Keng et al., 2006).

Additionally, many variations in structural and non-structural proteins caused substitutions of one or more amino acids were found in the isolates of Bangladesh compared to the reference. Most of the mutations found to affect the structural stability of the proteins rather than alter the molecular functions. Among the structural proteins, most variations were found in Surface glycoproteins (spike) and Nucleocapsid phosphoprotein. Spike proteins play a crucial role in the viral entry into the cell by interacting with the human ACE2 receptor. At the same time, Nucleocapsid phosphoprotein is essential for the packaging of viral genomes into a helical ribonucleocapsid (RNP) and fundamental for viral self-assembly (Chang et al., 2014; Hoffmann et al., 2020). These

functions may not affect much by those mutations, as Mutpred2 predicted that these mutations did not alter any molecular consequences of the proteins which are consistent with the study conducted by Wrapp and his co (Wrapp et al., 2020).

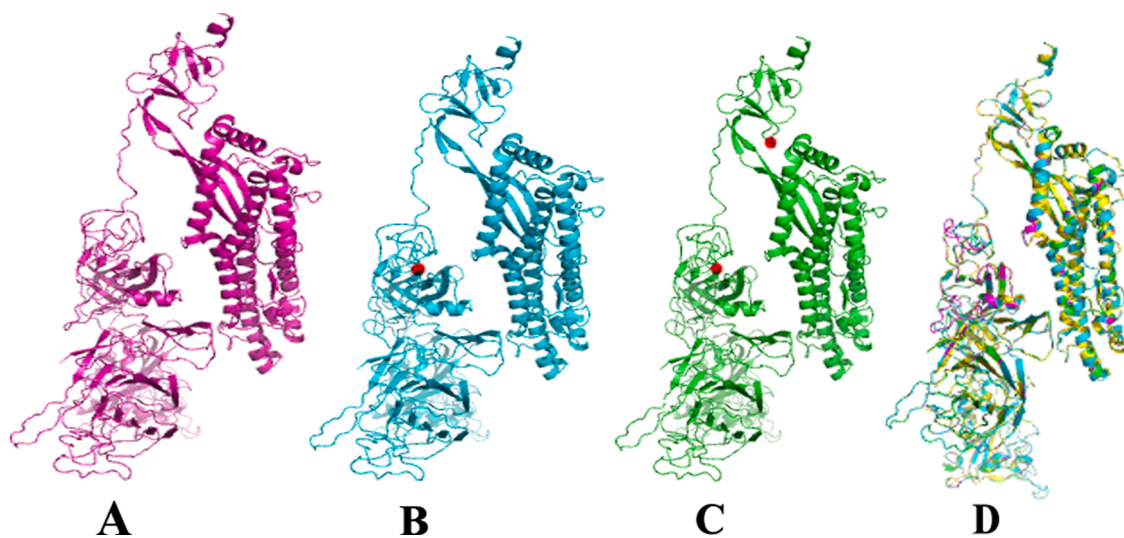
Interestingly, D623 G mutation was found in the spike protein of all isolates which is similar to the mutation D614 G of the spike protein of SARS-CoV-2 mentioned by many studies. They only differed in the amino acid numbering which occurred due to the use of predictive model in this study. This mutation in spike protein has now become the dominant genotype around the world and could boost the transmission of the virus (Grubaugh et al., 2020). However, several recent studies demonstrated that this mutation had not any differences in the hospitalization outcomes (Korber et al., 2020; Wagner et al., 2020; Lorenzo-Redondo et al., 2020). Moreover, our molecular docking analysis revealed that these mutations in spike proteins do not affect the interaction with the ACE2 receptor; give us a notion that mutation in the spike protein maybe for the better adaption of the SARS-CoV-2. This observation is also supported by two independent studies (Grubaugh et al., 2020; Isabel et al., 2020). Additionally, this study identified a domain in the spike protein (amino acid ranging from 345 to 527) involved with human ACE2 receptor interaction rather than the whole protein. This domain was conserved in all isolates reported in Bangladesh, resulting in no effect of the mutations. A recent study

**Table 5**  
Model Validation assessment score.

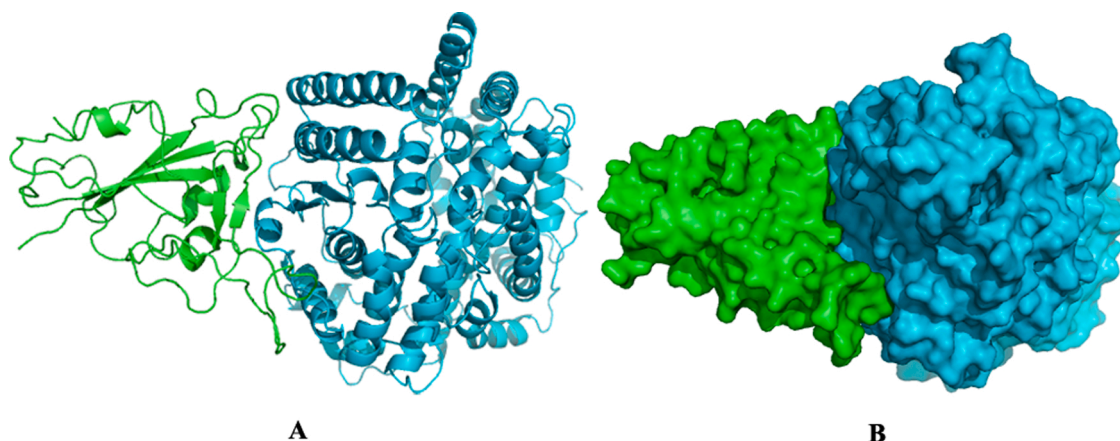
Structures	Rampage Score		ERRAT Score
	Favoured Region	Allowed Region	
Template	95.8 %	4.1 %	76 %
Wild type	92.9 %	5.7 %	83 %
Mutant Model 1	92.6%	5.3 %	84.69 %
Mutant Model 2	92.8%	5.3 %	83.78 %

**Table 6**  
Molecular docking results of human ACE2 receptor against wild-type and mutant spike protein of SARS-CoV-2.

Models	Variations	HDOCK Score
Model 1	Wild type	-244.42
Model 2	D623G	-244.42
Model 3	D623 G, F1118L	-244.42



**Fig. 3.** Homology model of the spike proteins; (A) wildtype (B) Model with one mutation: D623 G (C) Model with two mutations: D623 G and F1118 L (D) Superimpose of all models. Here, in B and C, red dot represents the mutation site. In D, purple color represents the wildtype model; the cyan represents a model with one mutation, and the green represents a model with two mutations.



**Fig. 4.** Interaction of Spike protein with ACE2: (A) cartoon model and (B) Surface model. Here, green represents the receptor binding domain (RBD) of spike protein, and cyan represents human ACE2.

identified the receptor-binding domain of spike protein, amino acid ranging from 319 to 541, to interact with the ACE2 receptor, which is similar to our findings (Lan et al., 2020).

## 5. Conclusion

SARS-CoV-2 isolates from Dhaka and Chittagong were close to European and Mideast lineage. A large deletion in the EPI\_ISL\_445213, EPI\_ISL\_445214, EPI\_ISL\_450343, and EPI\_ISL\_450344 isolates may explain the less pathogenic result of COVID-19 compared to other countries. Mutations in the spike protein of SARS-CoV-2 may induce more adaptation of this fetal virus; can cause less effective therapeutics if targeted. Our study gives novel insights to understand the SARS-CoV-2 epidemiology in Bangladesh.

## Ethical approval

Not required.

## Data availability

All data supporting the findings of this study are available within the article and its supplementary materials.

## Funding

SUST Research Center funds for MJH. SA is supported by the (1) Alberta Innovates Graduate Student Scholarship (AIGSS), and the (2) Maternal and Child Health (MatCH) Scholarship programs.

## CRedit authorship contribution statement

**Md. Sorwer Alam Parvez:** Conceptualization, Methodology, Software, Data curation, Formal analysis, Visualization, Validation, Writing - original draft. **Mohammad Mahfujur Rahman:** Formal analysis, Validation, Investigation. **Md. Niaz Morshed:** Formal analysis, Validation, Investigation. **Dolilur Rahman:** Formal analysis, Writing - original draft. **Saeed Anwar:** Data curation, Writing - review & editing. **Mohammad Jakir Hosen:** Supervision, Conceptualization, Writing - review & editing.

## Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:<https://doi.org/10.1016/j.compbiolchem.2020.107413>.

## References

- Amanat, F., Krammer, F., 2020. SARS-CoV-2 vaccines: status report. *Immunity*.
- Ashour, H.M., Elkhatib, W.F., Rahman, M., Elshabrawy, H.A., 2020. Insights into the recent 2019 novel coronavirus (SARS-CoV-2) in light of past human coronavirus outbreaks. *Pathogens* 9 (3), 186.
- Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., et al., 2000. The protein data bank. *Nucleic Acids Res.* 28 (1), 235–242.
- Brown, N.P., Leroy, C., Sander, C., 1998. MView: a web-compatible database search or multiple alignment viewer. *Bioinformatics (Oxford, England)* 14 (4), 380–381.
- Capriotti, E., Fariselli, P., Casadio, R., 2005. I-Mutant2.0: predicting stability changes upon mutation from the protein sequence or structure. *Nucleic Acids Res.* 33 (suppl\_2), W306–W310.
- Chang, C.K., Hou, M.H., Chang, C.F., Hsiao, C.D., Huang, T.H., 2014. The SARS coronavirus nucleocapsid protein-forms and functions. *Antiviral Res.* 103, 39–50.
- Colovos, C., Yeates, T.O., 1993. ERRAT: an empirical atom-based method for validating protein structures. *Protein Sci.* 2 (9), 1511–1519.
- CSSE, 2020. COVID-19 Dashboard by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University (JHU). ArcGIS. Johns Hopkins University. Retrieved 1 June 2020.
- Dastider, P., 2018. Manpower Export From Chittagong Regions Rises in 2017, 10 June. Available at: The Financial Express <https://thefinancialexpress.com.bd/economy/bangladesh/manpower-export-from-chittagong-region-rises-in-2017-1515563258>.
- Decaro, N., Lorusso, A., 2020. Novel human coronavirus (SARS-CoV-2): a lesson from animal coronaviruses. *Vet. Microbiol.*, 108693.
- DeLano, W.L., 2002. Pymol: an open-source molecular graphics tool. *CCP4 NewsL.* *Protein Crystallogr.* 40 (1), 82–92.
- Edgar, R.C., 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 32 (5), 1792–1797.
- Grubaugh, N.D., Hanage, W.P., Rasmussen, A.L., 2020. Making sense of mutation: what D614G means for the COVID-19 pandemic remains unclear. *Cell*.
- Hoang, D.T., Chernomor, O., Von Haeseler, A., Minh, B.Q., Vinh, L.S., 2018. UFBoot2: improving the ultrafast bootstrap approximation. *Mol. Biol. Evol.* 35 (2), 518–522.
- Hoffmann, M., Kleine-Weber, H., Schroeder, S., Krüger, N., Herrler, T., Erichsen, S., et al., 2020. SARS-CoV-2 cell entry depends on ACE2 and TMPRSS2 and is blocked by a clinically proven protease inhibitor. *Cell*.
- IEDCR, 2020. Institute of Epidemiology, Disease Control and Research (IEDCR). Available at: [www.corona.gov.bd](http://www.corona.gov.bd).
- Isabel, S., Grana-Miraglia, L., Gutierrez, J.M., Bundalovic-Torma, C., Groves, H.E., Isabel, M.R., et al., 2020. Evolutionary and structural analyses of SARS-CoV-2 D614G spike protein mutation now documented worldwide. *Sci. Rep.* 10, 14031.
- Kalyaanamoorthy, S., Minh, B.Q., Wong, T.K., von Haeseler, A., Jermini, L.S., 2017. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat. Methods* 14 (6), 587–589.
- Keng, C.T., Choi, Y.W., Welkers, M.R., Chan, D.Z., Shen, S., Lim, S.G., et al., 2006. The human severe acute respiratory syndrome coronavirus (SARS-CoV) 8b protein is distinct from its counterpart in animal SARS-CoV and down-regulates the expression of the envelope protein in infected cells. *Virology* 354 (1), 132–142.
- Khailany, R.A., Safdar, M., Ozaslan, M., 2020. Genomic characterization of a novel SARS-CoV-2. *Gene Rep.*, 100682.



- Korber, B., Fischer, W.M., Gnanakaran, S., Yoon, H., Theiler, J., Abfalterer, W., et al., 2020. Tracking changes in SARS-CoV-2 Spike: evidence that D614G increases infectivity of the COVID-19 virus. *Cell* 182 (4), 812–827.
- Lake, M.A., 2020. What we know so far: COVID-19 current clinical knowledge and research. *Clin. Med.* 20 (2), 124.
- Lan, J., Ge, J., Yu, J., Shan, S., Zhou, H., Fan, S., et al., 2020. Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor. *Nature* 1–6.
- Letunic, I., Bork, P., 2019. Interactive Tree of Life (iTOL) v4: recent updates and new developments. *Nucleic Acids Res.* 47 (W1), W256–W259.
- Lorenzo-Redondo, R., Nam, H.H., Roberts, S.C., Simons, L.M., Jennings, L.J., Qi, C., et al., 2020. A unique clade of SARS-CoV-2 viruses is associated with lower viral loads in Patient Upper Airways. *medRxiv*.
- Lovell, S.C., 2002. AWadBP, et al. Structure validation by  $\alpha$  geometry: phi, psi and  $\beta$ -deviation. *J. Proteins* 50, 437–450.
- Madden, T., 2013. The BLAST sequence analysis tool. In *The NCBI Handbook*, 2nd edition. National Center for Biotechnology Information (US) [Internet].
- Madeira, F., Park, Y.M., Lee, J., Buso, N., Gur, T., Madhusoodanan, N., et al., 2019. The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic Acids Res.* 47 (W1), W636–W641.
- Mercatelli, D., Giorgi, F.M., 2020. Geographic and genomic distribution of SARS-CoV-2 mutations. *Front. Microbiol.* 11.
- Nabi, M.S., Shovon, F.R., 2020. 20-fold Rises in COVID-19 Cases in Bangladesh Since April 1, 14 April. Available at: Dhaka Tribune <https://www.dhakatribune.com/health/coronavirus/2020/04/14/20-fold-rise-of-covid-19-cases-in-bangladesh-since-april-1>.
- Nguyen, L.T., Schmidt, H.A., Von Haeseler, A., Minh, B.Q., 2015. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* 32 (1), 268–274.
- Oberemok, V.V., Laikova, K.V., Yurchenko, K.A., Fomochkina, I.I., Kubyshkin, A.V., 2020. SARS-CoV-2 will continue to circulate in the human population: an opinion from the point of view of the virus-host relationship. *Inflamm. Res.* 1–6.
- Paul, R., 2020. Bangladesh Confirms Its First Three Cases of Coronavirus, 8 March. Available at: Reuters <https://www.reuters.com/article/us-health-coronavirus-bangladesh/bangladesh-confirms-its-first-three-cases-of-coronavirus-health-officials-idUSKBN20V0FS>.
- Pejaver, V., Urresti, J., Lugo-Martinez, J., Pagel, K.A., Lin, G.N., Nam, H.J., et al., 2017. MutPred2: inferring the molecular and phenotypic impact of amino acid variants. *BioRxiv*, 134981.
- Ruan, Y., Wei, C.L., Ling, A.E., Vega, V.B., Thoreau, H., Thoe, S.Y.S., et al., 2003. Comparative full-length genome sequence analysis of 14 SARS coronavirus isolates and common mutations associated with putative origins of infection. *Lancet* 361 (9371), 1779–1785.
- Schwede, T., Kopp, J., Guex, N., Peitsch, M.C., 2003. SWISS-MODEL: an automated protein homology-modeling server. *Nucleic Acids Res.* 31 (13), 3381–3385.
- Seib, K.L., Dougan, G., Rappuoli, R., 2009. The key role of genomics in modern vaccine and drug design for emerging infectious diseases. *PLoS Genet.* 5 (10).
- Sievers, F., Higgins, D.G., 2014. Clustal omega. *Curr. Protoc. Bioinf.* 48 (1), 3–13.
- Tang, D., Comish, P., Kang, R., 2020. The hallmarks of COVID-19 disease. *PLoS Pathog.* 16 (5), e1008536.
- Taylor, J.K., Coleman, C.M., Postel, S., Sisk, J.M., Bernbaum, J.G., Venkataraman, T., et al., 2015. Severe acute respiratory syndrome coronavirus ORF7a inhibits bone marrow stromal antigen 2 virion tethering through a novel mechanism of glycosylation interference. *J. Virol.* 89 (23), 11820–11833.
- The Business Standard, 2020. So Much Social Distancing, 25 March Available online at: <https://tbsnews.net/coronavirus-chronicle/coronavirus-bangladesh/so-much-social-distancing-60973>.
- Tithila, K.K., 2020. Bangladesh Expands Covid-19 Testing, 3 (2020)April Available online at: Dhaka Tribune <https://www.dhakatribune.com/bangladesh/2020/04/03/bangladesh-expands-covid-19-testing>.
- Ullah, A., 2020. Coronavirus: Bangladesh Imposes 14-day Quarantine on Gulf Workers, 17 March. Available at: Middle East Eye <https://www.middleeasteye.net/news/coronavirus-bangladesh-quarantine-workers-gulf-kuwait-saudi-arabia-qatar>.
- Wagner, C., Roychoudhury, P., Hadfield, J., Hodcroft, E.B., Lee, J., Moncla, L.H., et al., 2020. Comparing Viral Load and Clinical Outcomes in Washington State Across D614G Mutation in Spike Protein of SARS-CoV-2.
- Wang, H., Li, X., Li, T., Zhang, S., Wang, L., Wu, X., Liu, J., 2020. The genetic sequence, origin, and diagnosis of SARS-CoV-2. *Eur. J. Clin. Microbiol. Infect. Dis.* 1.
- Wiersinga, W.J., Rhodes, A., Cheng, A.C., Peacock, S.J., Prescott, H.C., 2020. Pathophysiology, transmission, diagnosis, and treatment of coronavirus disease 2019 (COVID-19): a review. *JAMA*.
- World Socialist Web Site (wsws), 2020. Bangladesh Government Downplays COVID-19 Threat As Jobless Mount. Available at: <https://www.wsws.org/en/articles/2020/03/16/bang-m16.html>.
- Wrapp, D., Wang, N., Corbett, K.S., Goldsmith, J.A., Hsieh, C.L., Abiona, O., et al., 2020. Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. *Science* 367 (6483), 1260–1263.
- Yan, Y., Zhang, D., Zhou, P., Li, B., Huang, S.Y., 2017. HDock: a web server for protein–protein and protein–DNA/RNA docking based on a hybrid strategy. *Nucleic Acids Res.* 45 (W1), W365–W373.
- Yin, W., Mao, C., Luan, X., Shen, D.D., Shen, Q., Su, H., et al., 2020. Structural basis for inhibition of the RNA-dependent RNA polymerase from SARS-CoV-2 by remdesivir. *Science*.
- Yuen, K.S., Ye, Z.W., Fung, S.Y., Chan, C.P., Jin, D.Y., 2020. SARS-CoV-2 and COVID-19: the most important research questions. *Cell Biosci.* 10 (1), 1–5.
- Zheng, Y.Y., Ma, Y.T., Zhang, J.Y., Xie, X., 2020. COVID-19 and the cardiovascular system. *Nat. Rev. Cardiol.* 17 (5), 259–260.