



Research article

Molecular characterization of SARS-CoV-2 from Bangladesh: implications in genetic diversity, possible origin of the virus, and functional significance of the mutations

Md. Marufur Rahman^a, Shirmin Bintay Kader^b, S.M. Shahriar Rizvi^{c,*}^a Centre for Medical Biotechnology, Management Information System, Directorate General of Health Services, Mohakhali, Dhaka, 1212, Bangladesh^b International Centre for Diarrhoeal Disease Research, Bangladesh^c Communicable Disease Control, Directorate General of Health Services, Mohakhali, Dhaka, 1212, Bangladesh

ARTICLE INFO

Keywords:

SARS-CoV-2

Mutation

Phylogeny

COVID-19

Bangladesh

D614G

Genome sequence

ABSTRACT

In a try to understand the pathogenesis, evolution and epidemiology of the SARS-CoV-2 virus, scientists from all over the world are tracking its genomic changes in real-time. Genomic studies can be helpful in understanding the disease dynamics. We have downloaded 324 complete and near complete SARS-CoV-2 genomes submitted in GISAID database from Bangladesh which were isolated between 30 March to 7 September, 2020. We then compared these genomes with Wuhan reference sequence and found 4160 mutation events including 2253 missense single nucleotide variations, 38 deletions and 10 insertions. The C>T nucleotide change was most prevalent (41% of all mutations) possibly due to selective mutation pressure to reduce CpG sites to evade CpG targeted host immune response. The most frequent mutation that occurred in 98% isolates was 3037C>T which is a synonymous change that usually accompanied 3 other mutations that include 241C>T, 14408C>T (P323L in RdRp) and 23403A>G (D614G in spike protein). The P323L was reported to increase mutation rate and D614G is associated with increased viral replication and currently most prevalent variant circulating all over the world. We identified multiple missense mutations in B-cell and T-cell predicted epitope regions and/or PCR target regions (including R203K and G204R that occurred in 86% of the isolates) that may impact immunogenicity and/or RT-PCR based diagnosis. Our analysis revealed 5 large deletion events in ORF7a and ORF8 gene products that may be associated with less severity of the disease and increased viral clearance. Our phylogeny analysis identified most of the isolates belonged to the Nextstrain clade 20B (86%) and GISAID clade GR (88%). Most of our isolates shared common ancestors either directly with European countries or jointly with middle eastern countries as well as Australia and India. Interestingly, the 19B clade (GISAID S clade) was unique to Chittagong, which was originally prevalent in China. This reveals possible multiple introductions of the virus in Bangladesh via different routes. Hence, more genome sequencing and analysis with related clinical data is needed to interpret functional significance and better predict the disease dynamics that may be helpful for policy makers to control the COVID-19 pandemic.

1. Introduction

The world is suffering from COVID-19, a devastating pandemic caused by a novel coronavirus originating from Wuhan, China (Zhou et al., 2020). Meanwhile, scientists from all over the world are trying to understand the virus better using various processes including genome sequencing. The first reported complete genome sequence was identified in January 3, 2020 (Tan et al., 2020). More than 141 thousands genomic sequences of SARS-CoV-2 have been submitted in the Global Initiative on

Sharing All Influenza Data (GISAID) database (Elbe and Buckland-Merrett, 2017, GISAID-Initiative, 2020).

The genomic sequences revealed that the length of the SARS-CoV-2 viral genome is ~30kb. The longest part of the genome at 5' end encodes for ORF1ab polyprotein whereas the rest of the genome consists of genes for encoding four structural proteins namely surface (S), envelope (E), membrane (M) and nucleocapsid (N), accessory proteins and other non-structural proteins (NSP) encoded by ORF3a, ORF6, ORF7a, ORF7b, ORF8 and ORF10 genes (Khailany RA and Ozaslan, 2020). The initial

* Corresponding author.

E-mail address: shahriarrizvi@gmail.com (S.M.S. Rizvi).

assessment of 3 clades indicates distinct geographic distribution. Depending on amino acid changes Förster et al. reported three central variants (A,B,C) of SARS-CoV-2 where A and C being the most common type in Europe and USA and B being the most common type in East Asia (Forster et al., 2020). Pachetti et al. identified multiple mutation hotspots with geographic location specificity. They identified mutations in RNA dependent RNA polymerase (RdRp) gene which are important as RdRp protein is the target for some proposed antiviral drugs and mutations in the gene may facilitate the virus to escape from those drugs (Pachetti et al., 2020). Yao et al. identified pathogenic variations depending specific SNVs in the Spike glycoprotein (S) changing viral load and cytopathic effects up to 270 folds (Yao et al., 2020). Deletions in the viral genomes are also common phenomena and sometimes are related to severity of the diseases (Armengaud et al., 2020; Holmes, 2009; Pachetti et al., 2020). Still there is a lack of studies to integrate all the deletions in the whole genome of SARS-CoV-2 globally. This may contribute to understand the pathogenic dynamics of the virus over time. The genetic differences among SARS-CoV-2 strains from different locations can be linked with their geographical distributions (Islam et al., 2020).

The clade and lineage nomenclatures for SARS-CoV-2 are changing rapidly. Specific combinations of 9 genetic markers shows 95% of the hCoV-19 data in GISAID can be further classified in major 6 clades named S, L, V, G, GH, GR (Khailany RA and Ozaslan, 2020). Initially the virus was classified into 2, then further into 3 super clades (Forster et al., 2020; Pachetti et al., 2020). A team of scientists has developed an open source bioinformatics and visualization toolkit named Nextstrain (www.nextstrain.org) for real-time tracking of pathogen evaluation including SARS-CoV-2 (nextstrain-github, 2020). Their clade nomenclature is different but supplementary to Rambaut et al., 2020 (GISAID-clade, 2020).

The virus was first reported in Bangladesh on March 8, 2020 as first 3 cases were identified at the Institute of Epidemiology and Disease Research, Dhaka. Currently the country is at the community spread stage and the total number of infected is about 379,738 with 5,555 reported deaths from COVID-19 till 12 October, 2020 (IEDCR, 2020; worldometer, 2020).

It is important to get more sequences from all over the world which will help us for better understanding of the evolution pattern, disease dynamics, phylogeographic distribution of the clades, designing drugs and vaccines. In this paper we tried to determine the phylogenetic relationship of Bangladeshi isolates with other isolates from around the world. This can help us to assume the travelling routes of the virus into Bangladesh as well in other parts of the world. We also tried to know the specific mutational differences in Bangladeshi isolates compared to the reference sequence and whether there is any clinico-pathological significance associated with those mutations.

2. Materials and methods

2.1. Local sequence retrieval

For retrieval of genome sequences from Bangladesh, we have searched in the GISAID database using search term “Bangladesh” as location. We have downloaded all the relevant sequences from search result in FASTA format and also downloaded patient status metadata, sequencing technology metadata and acknowledgement table separately (accessed on 30 August, 2020).

2.2. Mutation analysis

We have used Genome Detective Coronavirus Typing Tool version 1.13 and CoVsurver enabled by GISAID to analyse our query sequences in FASTA format (Cleemput et al., 2020; A*STAR-Bioinformatics-Institute, 2020). These are freely available online based bioinformatic tools which are validated to identify and reassemble novel Corona virus isolates. Using these tools, we identified both nucleotide and amino acid

mutations and similarities compared to SARS-CoV-2 (NCBI Taxonomy ID: 2697049) reference sequence NC_045512 (NCBI) and EPI_ISL_402124 (GISAID).

For functional prediction of mutational changes, we have used two web based tools namely SIFT (Sorting Intolerant From Tolerant) and MutPred2 (Ng and Henikoff, 2003; Pejaver et al., 2017). We also have used USCS SARS-CoV-2 Genome Browser (<https://genome.ucsc.edu/cgi-bin/hgGateway?db=wuhCor1>) to align mutations along base position and functionally significant areas in the genome (Fernandes et al., 2020). A support vector machine (SVM) based tool namely i-Mutant 3.0 was used for predicting the change of protein stability change and $\Delta\Delta G$ from specific mutations (Capriotti et al., 2005).

2.3. Phylogeny analysis

We have used open source bioinformatics visualization platform Nextstrain (nextstrain.org) for phylogeny analysis of our sequences (Hadfield et al., 2018). Pairwise sequence alignment and clade assignment was done using web based Nextstrain tool Nextclade beta version 0.4.9 (Nextclade, 2020). The GISAID clade identification of the sequences was done using GISAID CoVsurver tool. Further detailed information of different clusters was derived from a preformed interactive web-based tree developed from a subsample of global sequences (~5000) by neighbour joining method in Nextstrain web interface (<https://nextstrain.org/ncov/global>, date accessed on 30 August, 2020).

3. Results

From GISAID database we have found 329 submissions from Bangladesh (accessed on 30 August, 2020). Out of the 329 submissions, 324 were complete/near genomes. Among the complete genomes, 102 were isolated from female patients, 220 were from male patients and for 2 isolates the gender was not reported. The age range of collected samples was between 8 days to 95 years and the median age was 38 years. The submissions came from a total of 9 laboratories and different sequencing platforms were used by different laboratories. A quality control (QC) analysis was done where 7 sequences were flagged as “bad” (private mutation cut-off was set at 20) and 6 sequences were reported to have more than 5 ambiguous mutations.

3.1. Mutation analysis with functional significance

Variation analysis from 324 genome isolates revealed a total of 4160 mutation events out of which 4112 are Single Nucleotide Variations (SNVs), 38 deletions (in 30 isolates) and 10 insertions (in 10 isolates). Among the SNVs, 2253 were missense mutations in coding regions, 1216 were synonymous mutations and rest were in non-coding part of the genome (Table 1). Among all the SNVs, the most common change was C>T (~41%) and the second most prevalent change was G>A (~16%). There were 5 large deletions (>50 nucleotide) among which three resulted in deletion of a large portion ORF7a gene and another two deleted the ORF8 gene. The highest number of mutation events (including all SNVs and indels) observed in one isolate (EPI_ISL_445217) was 59 and least was zero as one isolate (EPI_ISL_458133) was identical to the reference genome (NC_045512.3) with 99.4% genome coverage. The average number of mutation events was approximately 13 per isolate.

The most common mutation in non-coding region was 241C>T, observed in 96% (312) of the isolates. The mutation with highest frequency (~98%) in coding region was 3037C>T (synonymous) and 14408C>T (missense) in ORF1ab gene, and 23403A>G (missense) in S gene. The latter caused D614G amino acid change in the spike protein of the virus. Among the 19 missense SNVs that occurred more than 5 times, 12 were predicted to decrease the stability of their respective protein structure (DDG value less than -0.5 kcal/mol where DDG or Delta Delta G is a measurement for predicting the effect of SNVs on protein stability)

and six of the SNVs were predicted to alter protein function. Among these 19 frequent SNVs, 6 were in CD4+ T-Cell epitope predicted regions, 6 were in CD8+ T-Cell epitope predicted regions and 5 were in B-Cell epitope predicted regions. The T592I mutation in ORF1ab polyprotein (NSP2) was strongly predicted for CD8+ T-Cell epitope that was also predicted for altered protein function. The P4715L, one of the highest frequent SNVs, occurred in the RNA dependent RNA polymerase (RdRp) region of ORF1ab polyprotein. Beside these, two of the high frequency (86%) SNVs (R203K and G204R) occurred in COVID-19 diagnostic RT-PCR target and B-Cell predicted epitope regions, of which the later was predicted to cause altered function of the nucleocapsid protein (Figure 1). Further analysis revealed two SNVs (L3606F, H125Y) were predicted to cause altered ordered interface and altered transmembrane protein for ORF1ab polyprotein (NSP6) and Membrane (M) protein respectively where L3606F was predicted to cause gain of sulfation at Y360 position of NSP3 protein (Table 2).

3.2. Mutation analysis of S gene

A separate and detailed analysis of SARS-CoV-2 S gene was done and revealed a total of 530 mutation events among which 414 were missense events and 3 were single amino acid deletions. These mutation events comprised of 56 SNVs and 2 deletions where D614G was the mutation of highest frequency that occurred in 97.2% (315) isolates. Among the SNVs 35 were predicted to decrease protein stability (DDG less than or around -0.5 kcal/mol), 8 were predicted to alter protein function, 15 were in predicted B-Cell epitope region and 28 were in T-Cell predicted epitope region. Individual analysis of these mutations with functional significance is shown in Supplementary Table 1. Three SNVs were found in the receptor binding domain of the spike protein but none of them was in the ACE2 receptor binding part of the protein (Figures 2 and 3).

3.3. Phylogeny analysis

After phylogeny analysis we have found, our 324 isolates were distributed among all the nextstrain clades where 20B clade was the most frequent (86%) (Figure 4). The 19B clade was unique to Chittagong (5 isolates) and one root clade 19A isolate was reported from Dhaka. There were 24 isolates for which the location data was unknown and only one isolate was found in 20C clade. We also have analysed clade distribution of our sequences according to GISAID nomenclature and found more than 96% of the isolates belonged to the G clade and its two major branches GH and GR clade. The common distinctive feature of these three clades is

D614G mutation. About 88% of the sequences clustered in GR clade the distinctive feature of which is G204R mutation in the nucleocapsid protein (Figure 4 and Table 3).

The overall analysis revealed that, most of the isolates shared common ancestors with European countries. A subsampled global phylogeny analysis revealed the largest cluster of Bangladeshi isolates shared common ancestor with some Australian isolates that were reported between mid-May to mid-July. Several other clusters were formed sharing common ancestors with countries including Senegal, Morocco, Egypt, Oman, Saudi Arabia, India, Srilanka, Zhenjiang (China), Portugal, Norway, Luxembourg, Bosnia and Herzegovina, England and Italy (Supplementary Figure 1, Supplementary Figure 2, Supplementry Figure 3).

4. Discussion

The current SARS-CoV-2 pandemic has changed the world in many ways bringing devastating effects on the society and environment yet we have seen some positive changes and one of which is increased collaboration of scientists and open-source projects from all over the world. The collaborative efforts are making huge impact on research and evidence generation. In our study, we have tried to gather data on genetic evolution and mutational impacts of SARS-CoV-2 those have been isolated and sequenced in Bangladesh. Our analysis revealed multiple introductions of the virus from different regions in our country as the phylogeny tree shows isolates closely related to different countries and regions of the world. Although most of the isolates were related to isolates from Middle Eastern and European countries, this can be explained as a lot of Bangladeshi migrant workers live in those countries including Saudi Arabia, Oman, Belgium and Italy. Many of these migrant workers came back to Bangladesh during first and second quarter of 2020 as number of COVID-19 cases were very high in those regions (Siddiqui et al., 2020; IOM, 2020). Interestingly the largest cluster was formed around Australian isolates but going further back on the tree reveals last common ancestor was also related to the European isolates (Sweden and Switzerland). Hence the abundance of 20B clade is observed in Bangladesh unlike neighbouring India and Pakistan where Asian and North American Clades (19A, 19B, 20B) are more dominant (Figure 5).

In our analysis we have found the most common single nucleotide change was C to T (~41%). This phenomenon was reported earlier and can be explained by selective mutation pressure to reduce CpG sites in the presence of abundant human antiviral proteins including APOBEC3 and ZAP (Wei et al., 2020). The CpG sites are common targets of viral genome and can be recognized by Toll-like receptors (TLRs) that results in release

Table 1. Distribution of mutation events along different genomic regions of SARS-CoV-2 Bangladeshi isolates.

Genome segment*		Base Position	Total mutation	Missense Mutation	Synonymous Mutation	Deletion**	Insertion
Coding region	<i>ORF1ab</i>	266–21555	1796	1021	775	2	0
	<i>S</i>	21563–25384	528	416	112	3	0
	<i>ORF3a</i>	25393–26220	135	101	34	7	0
	<i>E</i>	26245–26472	4	3	1	0	0
	<i>M</i>	26523–27191	65	14	51	0	0
	<i>ORF6</i>	27202–27387	8	6	2	1	0
	<i>ORF7a</i>	27394–27759	50	9	41	9	0
	<i>ORF7b</i>	27756–27887	8	5	3	0	0
	<i>ORF8</i>	27894–28259	70	34	36	2	0
	<i>N</i>	28274–29533	949	635	314	0	3
	<i>ORF10</i>	29558–29674	13	0	13	0	2
Non-coding region	Intergenic	-	12	-	-	2	1
	5'-UTR	1–265	386	-	-	5	1
	3'-UTR	29675–29903	136	-	-	7	3
Total			4160	2253	1382	38**	10

* Genes in displayed in italics.

** Some deletion events continued over multiple regions.

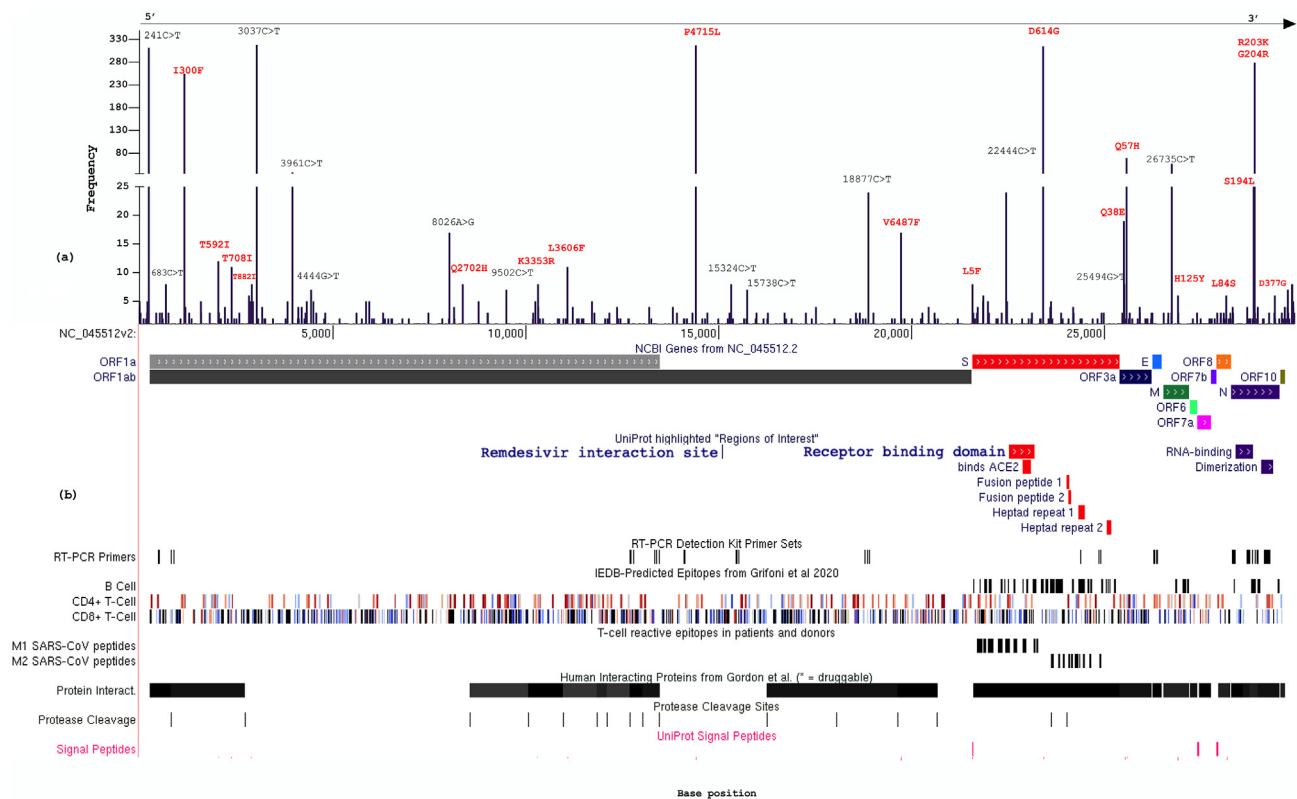


Figure 1. (a) Frequency of SNVs along the base positions of SARS-CoV-2 genome with high frequency (>5) missense SNVs marked red (b) The high frequency missense SNVs aligned (red lines) to different regions of SARS-CoV-2 genome, uniport regions of interest, RT-PCR diagnostic primer set, B-Cell and T-Cell predicted epitope regions, SARS-CoV T Cell epitope regions (M1 and M2 peptides), human protein interaction, protease cleavage and signal peptide regions derived from UCSC genome browser.

of pro-inflammatory cytokines including type- I interferon, IL-6, IL-12 and TNF- α (Arpaia and Barton, 2011) those play key role in severe COVID-19 including lung tissue damage (Costela-Ruiz et al., 2020). The reducing CpG in SARS-CoV-2 may indicate the mutational change facilitates the viral replication. Similar findings have also been reported in other studies for other RNA viruses relating the CpG motifs and viral replication (Atkinson et al., 2014). Relation of CpG with acute inflammatory response also has been mentioned in case of non-viral gene therapy vectors (Yew and Cheng, 2004). Inverse relation of disease severity with viral load has been reported in a recent study (Argyropoulos et al., 2020). Hence, we think further research is needed focusing the CpG suppression rate in SARS-CoV-2 and its relation with viral load and disease severity which may help designing more potent vaccine and therapeutics.

Large deletion events were observed among some of the isolates that resulted in deletion of most of ORF7a or ORF8 gene products. Three such deletions were between base position 27487 to 27552 (65 nucleotide), 27912 to 28256 (344 nucleotide) and 27472 to 27672 (200 nucleotide). Though these proteins are accessory proteins and not necessary for viral replication, ORF7a was found to interact with human ribosomal transport proteins MDN1 and HEATR3 (Gordon et al., 2020). In SARS-CoV the ORF7a was reported to act as cellular translation inhibitor and apoptosis inducer (Kopecky-Bromberg et al., 2006). Hence deletion of ORF7a may not change viral replication but can alter disease severity by reducing the chance of ORF7a mediated apoptosis. Similar deletions were reported earlier from USA (Addetia et al., 2020). On the other hand, ORF8 protein is least similar to its SARS-CoV homolog and reported to be associated with MHC-I downregulation that facilitates the virus for immune evasion from Cytotoxic T-Cells. This kind of immune evasion facilitates virus to

replicate without getting detected by immune cells hence producing less symptoms. Similar mechanism of immune evasion is observed in some chronic infection causing viruses including HIV-1 and Kaposi Sarcoma associated Herpes Virus (KSHV). This may be one of the reasons for the large number of asymptomatic patients, chronic viral shedding even after clinical cure and redetection of virus long after recovery (Zhang et al., 2020b). Recently study found that SARS-CoV-2 ORF8 also can potentially mediate unique immune suppression and evasion mechanisms (Flower et al., 2021).

The 241C>T mutation was one of the most frequent (96%) we have observed in our study. This mutation occurred in 5' UTR region, so it may not have any functional significance except for reducing CpG sites. Interestingly, this mutation always accompanied 3 other mutations in same isolates. Those mutations include 3037C>T (98%), 14408C>T (98%) and 23403A>G (97%) where the last two are non-synonymous mutations (P4715L in ORF1ab or P323L in RdRp and D614G in Spike protein). This co-occurrence of said mutations are not by chance rather a linkage disequilibrium that has been reported earlier by several studies (Bai et al., 2020; Demir et al., 2020; Mercatelli and Giorgi, 2020). These co-mutations are primary features of GISAID G clade that started rising since February in Europe and now include more than 70% of all SARS-CoV-2 sequences from all over the world (Mercatelli and Giorgi, 2020). The high prevalence of these 4 mutations in Bangladesh also establishes stronger linkage with European isolates. The P323L mutation in RdRp is reported to be associated with increased mutation rate (Begum et al., 2020; Eskier et al., 2020b). Among all 4 mutations the D614G is most studied and reported. Though not in the receptor binding part of Spike protein, multiple studies reported that the D614G mutation provides the SARS-CoV-2 an evolutionary advantage for replication. Studies

Table 2. Variants of SARS-CoV-2 genomes observed in more than 5 isolates.

Genomic change	Type of mutation	Gene	Amino acid change	No. of samples	Structural Prediction Effect (SVM3)	DDG Value (kcal/mol)	Functional prediction effect	Predicted molecular mechanism change	Finding from other studies
241C>T	Non-coding	5'-UTR	-	312	-	-	-	-	Frequently observed as co-mutation with 3037C>T (98%), 14408C>T (98%) and 23403A>G (97%) linked with European clade and high infection rate (Mercatelli and Giorgi, 2020; Rahimi et al., 2021)
683C>T	Synonymous	ORF1ab	None	8	-	-	-	-	-
1163A>T	Missense	ORF1ab	I300F	255	Decrease	-1.79	Tolerated	CD4+ T Cell epitope	Less likely to interact with host factors (Ul Alam et al., 2020)
2040C>T	Missense	ORF1ab	T592I	12	Neutral	-0.48	Affect function	CD8+ T Cell epitope (strong prediction)	-
2388C>T	Missense	ORF1ab	T708I	11	Neutral	-0.35	Tolerated	CD4+ and CD8+ T Cell epitope	-
2836C>T	Synonymous	ORF1ab	None	6	-	-	-	-	-
2910C>T	Missense	ORF1ab	T882I	8	Neutral	-0.10	Tolerated	-	-
3037C>T	Synonymous	ORF1ab	None	318	-	-	-	-	Frequently observed as co-mutation with 3037C>T (98%), 14408C>T (98%) and 23403A>G (97%) linked with European clade and high infection rate (Rahimi et al., 2021)
3961C>T	Synonymous	ORF1ab	None	38	-	-	-	-	Associated with manifestation of diarrhoea and sore throat in patients (Rabbi et al., 2021)
4444G>T	Synonymous	ORF1ab	None	7	-	-	-	-	Co-evolving mutation with 8371G>T and 29403A>G (Shishir et al., 2021)
8026A>G	Synonymous	ORF1ab	None	17	-	-	-	-	-
8371>T	Missense	ORF1ab	Q2702H	8	Decrease	-0.68	Affect function	-	Co-evolving mutation with 8371G>T and 29403A>G (Shishir et al., 2021)
9502C>T	Synonymous	ORF1ab	None	7	-	-	-	-	-
10323A>G	Missense	ORF1ab	K3353R	8	Neutral	-0.13	Tolerated	CD4+ T Cell epitope	-
11083G>T	Missense	ORF1ab	L3606F	11	Decrease	-1.00	Affect function	Altered Ordered interface, Altered Transmembrane protein, Gain of Sulfation at Y3607, CD4+ and CD8+ T Cell epitope	More prevalent in asymptomatic cases (Aiewsakun et al., 2020; Wang et al., 2020)
14408C>T	Missense	ORF1ab	P4715L	317	Decrease	-0.83	Tolerated	CD8+ T Cell epitope, RdRp zone,	Frequently observed as co-mutation with 3037C>T (98%), 14408C>T (98%) and 23403A>G (97%) linked with European clade and high infection rate (Rahimi et al., 2021). Associated with higher mutation rate (Pachetti et al., 2020)
15324C>T	Synonymous	ORF1ab	None	8	-	-	-	-	-
15738C>T	Synonymous	ORF1ab	None	7	-	-	-	-	-
18877C>T	Synonymous	ORF1ab	None	24	-	-	-	-	Associated with mutation density in M and E gene (Eskier et al., 2020a)
19723G>T	Missense	ORF1ab	V6487F	17	Decrease	-1.55	Tolerated	-	-
21575C>T	Missense	S	L5F	8	Decrease	-0.98	Tolerated	-	May increase hydrophobicity of the signal peptide thus facilitate in viral secretion from cell (Zhan et al., 2020) hence increase infectivity (Li et al., 2020). It also increases epitope binding affinity (Guo and Guo, 2020)
21855C>T	Missense	S	S98F	6	Neutral	0.00	Tolerated	-	-
22444C>T	Synonymous	S	None	24	-	-	-	-	Found to co-evolve with 28854C>T and unique to Indian isolates (Banerjee et al., 2020)
23403A>G	Missense	S	D614G	315	Decrease	-0.93	Tolerated	B cell epitope	Discussed in detail in the discussion part
25494G>T	Synonymous	ORF3a	None	19	-	-	-	-	-
25504C>G	Missense	ORF3a	Q38E	8	Decrease	-1.02	Affect function	CD4+ and CD8+ T Cell epitope, transmembrane protein	-
25563G>T	Missense	ORF3a	Q57H	30	Decrease	-2.03	Affect function	CD4+ T Cell epitope	Decrease ion permeability of ORF3a channel pore and possibly decrease viral release and immune response (Alam et al., 2020)
26735C>T	Synonymous	M	None	29	-	-	-	-	-
26895C>T	Missense	M	H125Y	6	Neutral	0.06	Tolerated	CD8+ T Cell epitope, Altered Transmembrane protein, Altered Ordered interface	-
28854C>T	Missense	N	S194L	25	Neutral	-0.38	Tolerated	B cell epitope	Enhance N-E and decrease N-M interactions thus may promote viral release and attenuate viral assembly (Wu et al., 2021)
28881G>A	Missense	N	R203K	279	Decrease	-0.93	Tolerated	B cell epitope, PCR target area, transmembrane protein	Enhance N-E interaction thus may promote viral release and infectivity (Wu et al., 2021).
28882G>A	Missense	N	R203K	279	Decrease	-0.93	Tolerated	-	-

(continued on next page)

Table 2 (continued)

Genomic change	Type of mutation	Gene	Amino acid change	No. of samples	Structural Prediction Effect (SVM3)	DDG Value (kcal/mol)	Functional prediction effect	Predicted molecular mechanism change	Finding from other studies
28883G>C	Missense	N	G204R	279	Decrease	-0.52	Affect function	B cell epitope, PCR target area, transmembrane protein	B cell epitope, PCR target area, transmembrane protein Enhance N-E interaction thus may promote viral release and infectivity (Wu et al., 2021).
29403A>G	Missense	N	D377G	6	Neutral	-0.44	Tolerated	-	-
29742G>A	Non-coding	3'-UTR	-	7	-	-	-	-	Create miR-3664-5p binding site but degradation of viral RNA by host protein is unaltered (Mukherjee and Goswami, 2020).
29848T>A	Non-coding	3'-UTR	-	8	-	-	-	-	-
29850A>T	Non-coding	3'-UTR	-	8	-	-	-	-	-
29852A>T	Non-coding	3'-UTR	-	7	-	-	-	-	-
29853G>A	Non-coding	3'-UTR	-	8	-	-	-	-	-

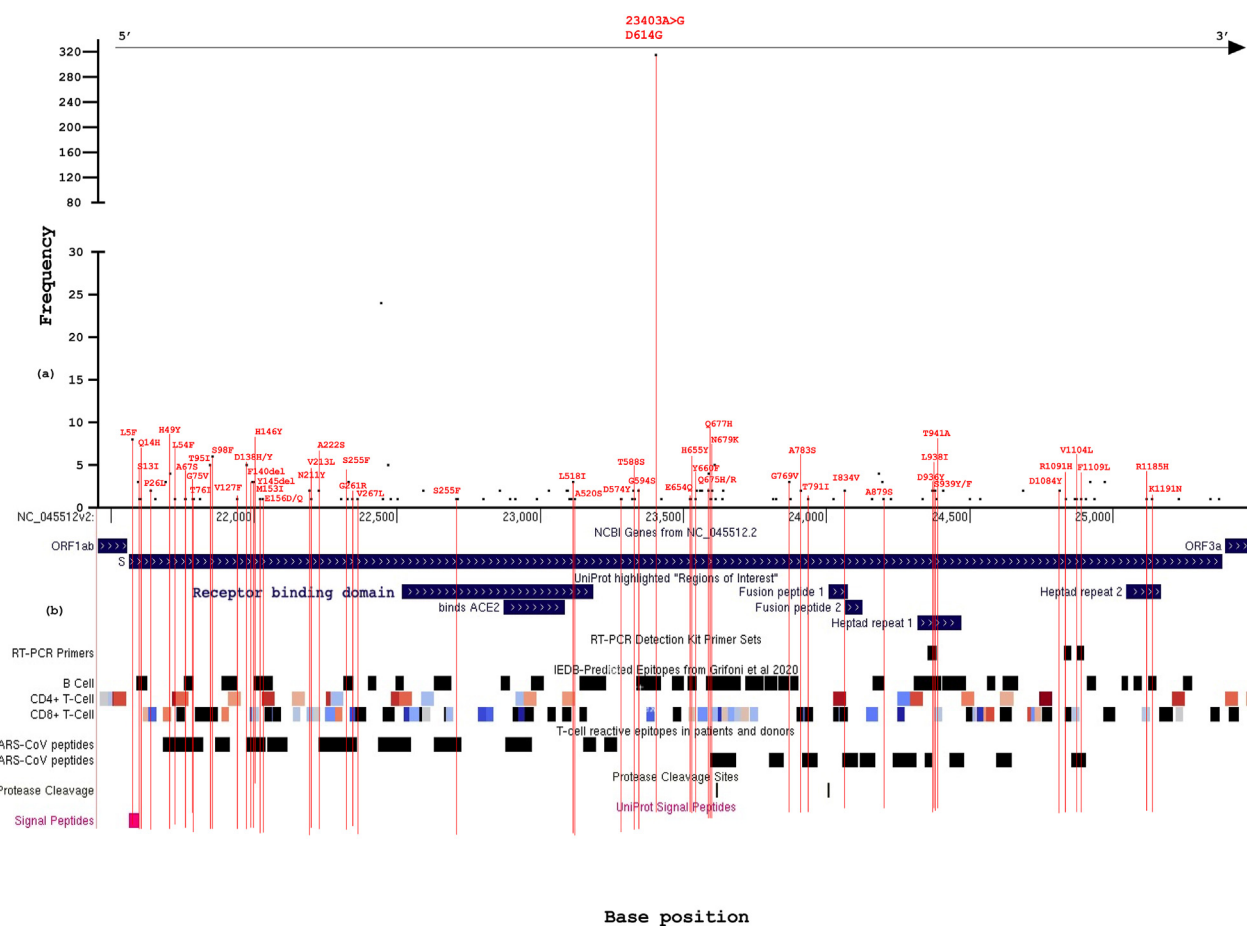


Figure 2. (a) Frequency of SNVs along the base positions of SARS-CoV-2 S gene, missense mutations are annotated in red text (b) The missense SNVs aligned (red lines) to different regions of SARS-CoV-2 S gene, uniprot regions of interest, RT-PCR diagnostic primer set, B-Cell and T-Cell predicted epitope regions, SARS-CoV T Cell epitope regions (M1 and M2 peptides), human protein interaction, protease cleavage and signal peptide regions derived from UCSC genome browser.

reported D614G mutation increases infectivity of the virus as it was found to be associated with higher viral load and higher infectious titre (Korber et al., 2020a, 2020b; Zhang et al., 2020a).

The D614 amino acid is located between S1 and S2 junction of spike protein. The cleavage of S1-S2 junction by host protease is crucial for entry into the host cell and multiple cleavage sites enhances the fusion of SARS-CoV with host cell membrane (Belouzard et al., 2009). One study predicted D614G mutation introduces a novel protease (elastase)

cleavage site that may enhance the fusion of viral envelop to the host cell membrane hence further facilitate viral RNA entry into the host cell (Bhattacharyya et al., 2020). Studies identified the 614G variant of the virus can get more functional advantage in population with delC variant (rs35074065 site) of TMPRSS2 gene (Bhattacharyya et al., 2020; Russo et al., 2020). This variant is common in Europe, America and South Asia but extremely rare in East Asia according to 1000 Genome project data (NCBI, 2020). This may explain the spread of D614G or G clade in mostly

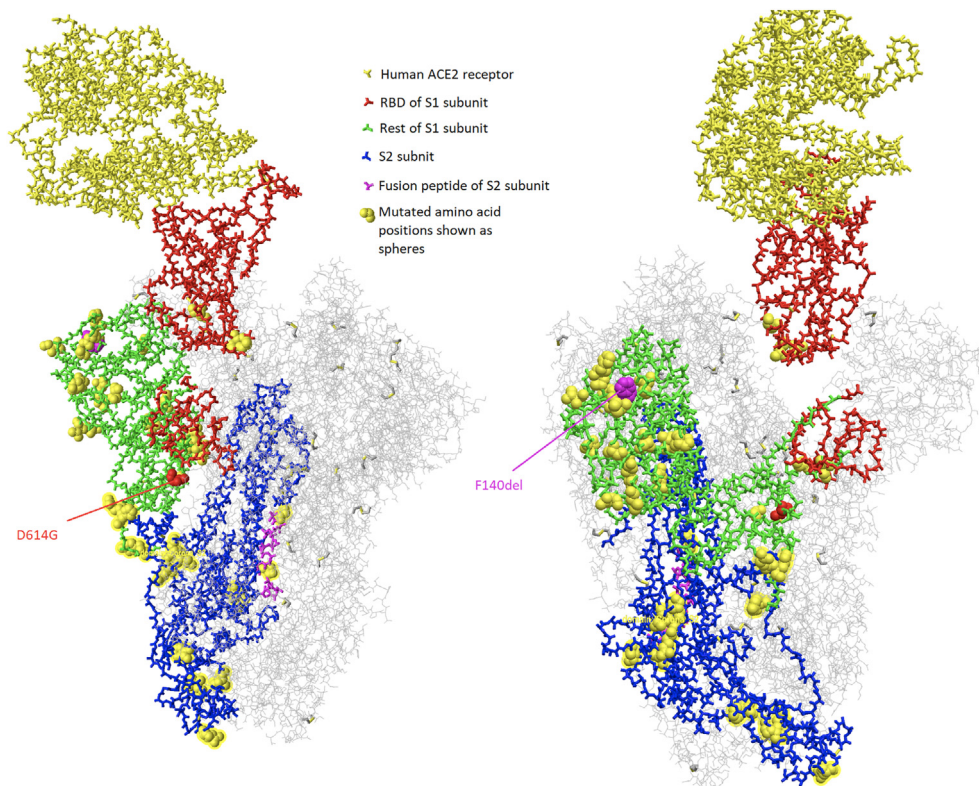


Figure 3. Mutated amino acid positions shown in the human ACE2 receptor bound structure (PDB ID: 6acj) of SARS-CoV-2 spike protein (except L5F, S13I, Q14H, G75V, T76I, Y145del, H146Y, Q675 H/R, Q677H, N679K, I834V, R1185H and K1195N as the regions were not covered in the structure). Mutations are shown in only one chain of the structure.

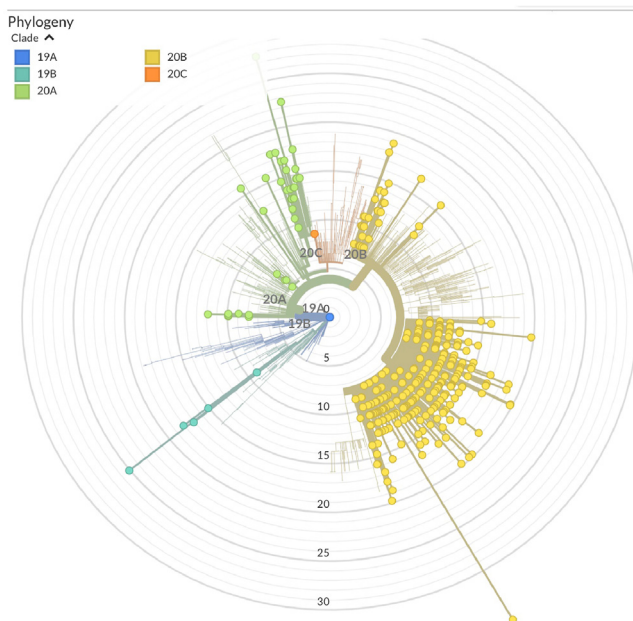


Figure 4. Radial presentation of phylogeny tree formed from 324 SARS-CoV-2 genomes from Bangladesh compared to the reference SARS-CoV-2 sequence shows the isolates belonged to all of the nextstrain clades and the 20B clade is the most prominent.

Europe, America and recently in South Asia. While some studies suggested that D614G mutation is associated with higher fatality rate (Becerra-Flores and Cardozo, 2020; Toyoshima et al., 2020), several other studies reported no significant association (Korber et al., 2020b;

Cassia Wagner et al., 2020). There is no significant association between viral load and clinical outcome or survival was found in another recent study (Argyropoulos et al., 2020). Considering these analyses it can be said that the current evidence is not clear about the impact of D614G mutation alone on the disease severity and mortality as multiple other stronger factors play role especially age and comorbidity (Grubaugh et al., 2020).

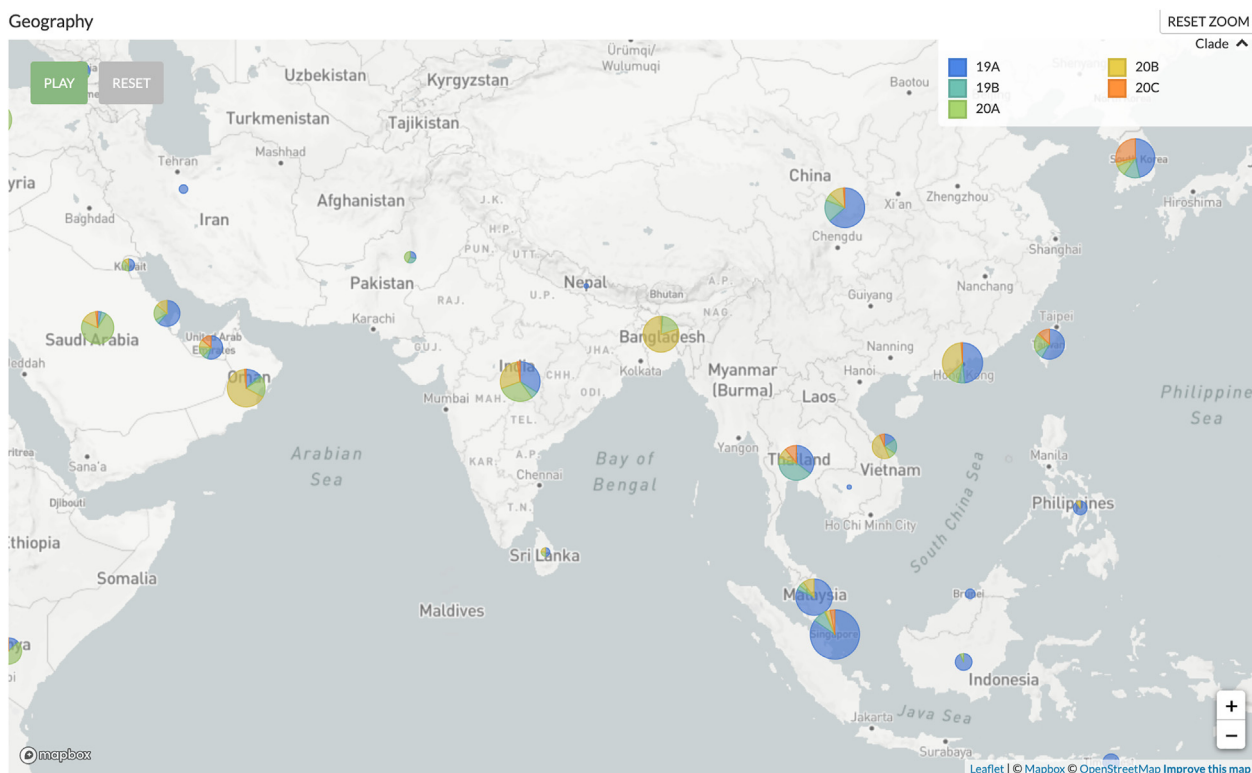
There have been concerns about the impact of D614G mutation on vaccine development but it is clear that the mutation does not take place in the receptor binding region of the spike protein which is the primary target of the neutralizing antibodies. Also, studies suggested that in natural infection, antibodies generated from D614 variant can cross neutralize G614 variant viruses. Hence this is unlikely that the mutation will have a drastic effect on the immunogenicity of the virus and less likely to have any impact on vaccine efficacy (Grubaugh et al., 2020; Hu et al., 2020; Ozono et al., 2020).

The second most frequent mutation in our analysis was a trinucleotide change resulting in two amino acid changes which are R203K and G204R in N protein. Our analysis revealed these mutations occurred in a PCR target area and B-cell epitope region. Though functionally tolerated, this change was predicted to decrease the protein stability. This finding is similar with other studies those reported R203K and G204R destabilizes N protein structure but may enhance interaction with SARS-CoV-2 Envelop protein that may promote viral release (Rahman et al., 2020; Wu et al., 2021).

The I300F mutation (occurred in 78% isolates) was predicted to reduce the stability of NSP2 protein the function of which is not yet confirmed. One study suggested that the amino acid is positioned within the internal groove of the protein and less likely to interact with host factors (Ul Alam et al., 2020). There two other functionally significant mutations which are Q57H in ORF3a and S194L in N protein. Our analysis suggested the Q57H decreases protein stability with altered protein function may result in loss of a CD4+ T Cell epitope similar to

Table 3. Location-wise of distribution of isolates from different clades.

Location (Division)	Nextstrain/GISAID Clade	No. of Isolates	Primary Countries
Dhaka	19A/L	1/1	Asia: China/Thailand
	20A/G,GH	6/5,1	N America/Europe/Asia: USA, Belgium, India
	20B/GR,O	77/75,2	Europe: UK, Belgium, Sweden
Chittagong	19B/S	5/5	Asia: China
	20A/G,GH,O	14/2,11,1	N America/Europe/Asia: USA, Belgium, India
	20B/GR	55/55	Europe: UK, Belgium, Sweden
Rajshahi	20A/GH	2/2	N America/Europe/Asia: USA, Belgium, India
	20B/GR	30/30	Europe: UK, Belgium, Sweden
Rangpur	20A/GH	1/1	N America/Europe/Asia: USA, Belgium, India
	20B/GR,G	24/23,1	Europe: UK, Belgium, Sweden
Khulna	20A/O,GH	5/3,2	N America/Europe/Asia: USA, Belgium, India
	20B/GR	22/22	Europe: UK, Belgium, Sweden
Sylhet	20A/GH	2/2	N America/Europe/Asia: USA, Belgium, India
	20B/GR	19/19	Europe: UK, Belgium, Sweden
Barishal	20A/GH	2/2	N America/Europe/Asia: USA, Belgium, India
	20B/G,GR	21/1,20	Europe: UK, Belgium, Sweden
Unknown	20A/G,GH	6/4,2	N America/Europe/Asia: USA, Belgium, India
	20B/GR	16/16	Europe: UK, Belgium, Sweden
	20C/GH	1/1	N America: USA

**Figure 5.** Comparison of clade distribution in different regions of Asia from subsampled global analysis in nextstrain.org.

other studies (Kim et al., 2020; Wu et al., 2021). One study mentioned that the Q57H mutation in ORF3a protein may decrease ion permeability by creating a tighter constriction in channel pore and possibly decrease viral release and immune response (Ul Alam et al.). The S194L was predicted to have neutral effect in our study with a reduced DDG value and may attenuate viral assembly as reported in an earlier study (Wu et al., 2021).

We have observed several mutations in some of the RT-PCR target regions. Though it is yet unknown that if mismatch in primer template changes the accuracy and precision of RT-PCR based COVID-19

diagnosis, we recommend avoidance of using primers containing mutation prone regions for better diagnosis. Several other low frequency mutations we found to be associated with higher infectivity and manifestation of specific symptoms which are mentioned in Table 2.

As a limitation of our study, we couldn't derive any clinical information of the patients from whom the samples were collected. The functional significance described in this paper are only computational prediction based and may not always reflect clinical scenario. Also, the genomic sequences were derived using different sequencing platforms (i.e. Illumina, Ion Torrent etc.) and methods (Sanger and Next-generation

sequencing) by different laboratories which may have impacted the quality of the sequences hence impacted our analysis. We have found one sequenced that has no mutation compared to the reference sequence which is very unlikely and may possibly be a submission error as the sample was collected long after the original Wuhan outbreak. We hope our findings will create scopes for further research specially including clinical data and also help identifying changes in pathogenicity and infectivity pattern of the virus.

Declarations

Author contribution statement

Dr. Md. Marufur Rahman: Conceived and designed the experiments; Performed the experiments; Analyzed and interpreted the data; Contributed reagents, materials, analysis tools or data; Wrote the paper.

Dr. Shirmin Bintay Kader: Performed the experiments; Analyzed and interpreted the data; Contributed reagents, materials, analysis tools or data; Wrote the paper.

Dr. S M Shahriar Rizvi: Conceived and designed the experiments; Analyzed and interpreted the data; Contributed reagents, materials, analysis tools or data; Wrote the paper.

Funding statement

This work was supported by the Centre for Medical Biotechnology (CMBT), Management Information System, Directorate General of Health Services, Bangladesh.

Data availability statement

Data included in article/supp. material/referenced in article.

Competing interest statement

The authors declare no conflict of interest.

Additional information

Supplementary content related to this article has been published online at <https://doi.org/10.1016/j.heliyon.2021.e07866>.

Acknowledgements

We acknowledge who were involved in the process of sample collection, genome sequencing and sequence data submission in the GISAID database. We also acknowledge Dr. Senjuti Saha, Scientist, Child Health Research Foundation Bangladesh, who inspired and guided us throughout the writing of this article.

References

- A*Star-Bioinformatics-Institute, 2020. CoVsurver - CoronaVirus Surveillance Server [Online]. Available: <https://corona.bii.a-star.edu.sg/>. (Accessed 19 September 2020).
- Addetia, A., Xie, H., Roychoudhury, P., Shrestha, L., Loprieno, M., Huang, M.-L., Jerome, K., Greninger, A., 2020. Identification of multiple large deletions in ORF7a resulting in in-frame gene fusions in clinical SARS-CoV-2 isolates. medRxiv.
- Aiewsakun, P., Wongtrakongate, P., Thawornwattana, Y., Hongeng, S., Thitithanyanont, A., 2020. SARS-CoV-2 genetic variations associated with COVID-19 severity. MedRxiv.
- Alam, A.R.U., Islam, O.K., Hasan, M.S., Al-Emran, H.M., Jahid, M.I.K., Hossain, M.A., 2020. Evolving infection paradox of SARS-CoV-2: fitness costs virulence? MedRxiv.
- Argyropoulos, K.V., Serrano, A., Hu, J., Black, M., Feng, X., Shen, G., Call, M., Kim, M.J., Lytle, A., Belovarac, B., 2020. Association of initial viral load in severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) patients with outcome and symptoms. Am. J. Pathol. 190, 1881–1887.
- Armengaud, J., Delaunay-Moisan, A., Thuret, J.Y., Van Anken, E., Acosta-Alvear, D., Aragón, T., Arias, C., Blondel, M., Braakman, I., Collet, J.F., 2020. The importance of naturally attenuated Sars-Cov-2 in the fight against Covid-19. Environ. Microbiol. 22, 1997–2000.
- Arpaia, N., Barton, G.M., 2011. Toll-like receptors: key players in antiviral immunity. Curr. Opin. Virol. 1, 447–454.
- Atkinson, N.J., Witteveldt, J., Evans, D.J., Simmonds, P., 2014. The influence of CpG and UpA dinucleotide frequencies on RNA virus replication and characterization of the innate cellular pathways underlying virus attenuation and enhanced replication. Nucleic Acids Res. 42, 4527–4545.
- Bai, Y., Jiang, D., Lon, J.R., Chen, X., Hu, M., Lin, S., Chen, Z., Meng, Y., Du, H., 2020. Evolution and molecular characteristics of SARS-CoV-2 genome. bioRxiv.
- Banerjee, A., Sarkar, R., Mitra, S., Lo, M., Dutta, S., Chawla-Sarkar, M., 2020. The novel coronavirus enigma: phylogeny and analyses of coevolving mutations among the sars-cov-2 viruses circulating in India. JMIR Bioinform. Biotechnol. 1, e20735.
- Becerra-Flores, M., Cardozo, T., 2020. SARS-CoV-2 viral spike G614 mutation exhibits higher case fatality rate. Int. J. Clin. Pract.
- Begum, F., Mukherjee, D., Das, S., Thagrikri, D., Tripathi, P.P., Banerjee, A.K., Ray, U., 2020. Specific mutations in SARS-CoV2 RNA dependent RNA polymerase and helicase alter protein structure, dynamics and thus function: effect on viral RNA replication. bioRxiv.
- Belouzard, S., Chu, V.C., Whittaker, G.R., 2009. Activation of the SARS coronavirus spike protein via sequential proteolytic cleavage at two distinct sites. Proc. Natl. Acad. Sci. Unit. States Am. 106, 5871–5876.
- Bhattacharyya, C., Das, C., Ghosh, A., Singh, A.K., Mukherjee, S., Majumder, P.P., Basu, A., Biswas, N.K., 2020. Global spread of SARS-CoV-2 subtype with spike protein mutation D614G is shaped by human genomic variations that regulate expression of TMPRSS2 and MX1 genes. bioRxiv.
- 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, W306–W310.
- Cassia Wagner, P.R., Frazar, Chris D., Lee, Jover, Müller, Nicola F., Moncla, Louise H., et al., 2020. Comparing Viral Load and Clinical Outcomes in Washington State across D614G substitution in spike protein of SARS-CoV-2 [Online]. Available: <https://github.com/blas/ncov-wa-d614g>. (Accessed 21 August 2020).
- Cleemput, S., Dumon, W., Fonseca, V., Abdool Karim, W., Giovanetti, M., Alcantara, L.C., Deforche, K., De Oliveira, T., 2020. Genome Detective Coronavirus Typing Tool for rapid identification and characterization of novel coronavirus genomes. Bioinformatics 36, 3552–3555.
- Costela-Ruiz, V.J., Illescas-Montes, R., Puerta-Puerta, J.M., Ruiz, C., Melguizo-Rodríguez, L., 2020. SARS-CoV-2 infection: the role of cytokines in COVID-19 disease. Cytokine Growth Factor Rev.
- Demir, A.B., Benvenuto, D., Abacıoğlu, Y.H., Angeletti, S., Ciccozzi, M., 2020. Identification of the nucleotide substitutions in 62 SARS-CoV-2 sequences from Turkey. Turk. J. Biol. 44, 178–184.
- Elbe, S., Buckland-Merrett, G., 2017. Data, disease and diplomacy: GISAID's innovative contribution to global health. Glob. Challng. 1, 33–46.
- Eskier, D., Suner, A., Oktay, Y., Karakülah, G., 2020. Mutations of SARS-CoV-2 nsp14 exhibit strong association with increased genome-wide mutation load. PeerJ 8, e10181.
- Eskier, D., Karakülah, G., Suner, A., Oktay, Y., 2020. RdRp mutations are associated with SARS-CoV-2 genome evolution. bioRxiv.
- Fernandes, J.D., Hinrichs, A.S., Clawson, H., Gonzales, J.N., Lee, B.T., Nassar, L.R., Raney, B.J., Rosenbloom, K.R., Nerli, S., Rao, A.A., 2020. The UCSC SARS-CoV-2 genome browser. bioRxiv.
- Flower, T.G., Buffalo, C.Z., Hooy, R.M., Allaire, M., Ren, X., Hurley, J.H., 2021. Structure of SARS-CoV-2 ORF8, a rapidly evolving immune evasion protein. Proc. Natl. Acad. Sci. Unit. States Am. 118.
- Forster, P., Forster, L., Renfrew, C., Forster, M., 2020. Phylogenetic network analysis of SARS-CoV-2 genomes. Proc. Natl. Acad. Sci. Unit. States Am. 117, 9241–9243.
- Gisaaid-Clade, 2020. GISAID - Clade and Lineage Nomenclature Aids in Genomic Epidemiology of Active hCoV-19 Viruses [Online]. Available: <https://www.gisaid.org/references/statements-clarifications/clade-and-lineage-nomenclature-aids-in-genomic-epidemiology-of-active-hcov-19-viruses/>. (Accessed 19 September 2020).
- GISAID-INITIATIVE, 2020. Available: <https://www.gisaid.org/>. (Accessed 19 September 2020).
- Gordon, D.E., Jang, G.M., Bouhaddou, M., Xu, J., Obernier, K., White, K.M., O'meara, M.J., Rezelj, V.V., Guo, J.Z., Swaney, D.L., 2020. A SARS-CoV-2 protein interaction map reveals targets for drug repurposing. Nature 1–13.
- 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 182, 794–795.
- Guo, E., Guo, H., 2020. CD8 T cell epitope generation toward the continually mutating SARS-CoV-2 spike protein in genetically diverse human population: implications for disease control and prevention. PloS One 15, e0239566.
- Hadfield, J., Megill, C., Bell, S.M., Huddleston, J., Potter, B., Callender, C., Sagulenko, P., Bedford, T., Neher, R.A., 2018. Nextstrain: real-time tracking of pathogen evolution. Bioinformatics 34, 4121–4123.
- Holmes, E.C., 2009. The Evolution and Emergence of RNA Viruses. Oxford University Press.
- Hu, J., He, C.L., Gao, Q., Zhang, G.J., Cao, X.X., Long, Q.X., Deng, H.J., Huang, L.Y., Chen, J., Wang, K., 2020. The D614G mutation of SARS-CoV-2 spike protein enhances viral infectivity. BioRxiv.
- IEDCR, 2020. Available: <https://www.iedcr.gov.bd/>. (Accessed 12 October 2020).
- IOM, 2020. IOM Assists Vulnerable Returning Migrants Impacted by the COVID-19 Pandemic. International Organization for Migration [Internet]. [cited 2020 Aug 8]. Available from: <https://bangladesh.iom.int/news/iom-assists-vulnerable-returning-g-migrants-impacted-covid-19-pandemic>.

- Islam, M.R., Hoque, M.N., Rahman, M.S., Alam, A.R.U., Akther, M., Puspo, J.A., Akter, S., Sultana, M., Crandall, K.A., Hossain, M.A., 2020. Genome-wide analysis of SARS-CoV-2 virus strains circulating worldwide implicates heterogeneity. *Sci. Rep.* 10, 1–9.
- Khailany Ra, S.M., Ozaslan, M., 2020. Genomic characterization of a novel SARS-CoV-2. *Gene Rep.* 19, 100682.
- Kim, J.-S., Jang, J.-H., Kim, J.-M., Chung, Y.-S., Yoo, C.-K., Han, M.-G., 2020. Genome-Wide identification and characterization of point mutations in the SARS-CoV-2 genome. *Osong Publ. Health Res. Perspect.* 11, 101.
- Kopecky-Bromberg, S.A., Martinez-Sobrido, L., Palese, P., 2006. 7a protein of severe acute respiratory syndrome coronavirus inhibits cellular protein synthesis and activates p38 mitogen-activated protein kinase. *J. Virol.* 80, 785–793.
- Korber, B., Fischer, W., Gnanakaran, S.G., Yoon, H., Theiler, J., Abfalterer, W., Foley, B., Giorgi, E.E., Bhattacharya, T., Parker, M.D., 2020a. Spike mutation pipeline reveals the emergence of a more transmissible form of SARS-CoV-2. *bioRxiv*.
- Korber, B., Fischer, W.M., Gnanakaran, S., Yoon, H., Theiler, J., Abfalterer, W., Hengartner, N., Giorgi, E.E., Bhattacharya, T., Foley, B., 2020b. Tracking changes in SARS-CoV-2 Spike: evidence that D614G increases infectivity of the COVID-19 virus. *Cell* 182, 812–827 e19.
- Li, Q., Wu, J., Nie, J., Zhang, L., Hao, H., Liu, S., Zhao, C., Zhang, Q., Liu, H., Nie, L., 2020. The impact of mutations in SARS-CoV-2 spike on viral infectivity and antigenicity. *Cell* 182, 1284–1294 e9.
- Mercatelli, D., Giorgi, F.M., 2020 Jul 22. Geographic and genomic distribution of SARS-CoV-2 mutations [Internet]. *Front. Microbiol.* [cited 2020 Aug 19];11:1800. Available from: <https://www.frontiersin.org/article/10.3389/fmicb.2020.01800/full>.
- Mukherjee, M., Goswami, S., 2020. Global cataloguing of variations in untranslated regions of viral genome and prediction of key host RNA binding protein-microRNA interactions modulating genome stability in SARS-CoV-2. *PLoS One* 15, e0237559.
- NCBI, 2020. rs35074065 RefSNP Report - dbSNP - NCBI [Internet]. [cited 2020 Aug 20]. Available from: https://www.ncbi.nlm.nih.gov/snp/rs35074065#frequency_tab.
- NEXTCLADE, 2020. Available: <https://clades.nextstrain.org/>. (Accessed 19 September 2020).
- Nextstrain-Github, 2020. *ncov/naming_clades.md* at Master · Nextstrain/ncov · GitHub [Online]. Available: https://github.com/nextstrain/ncov/blob/master/docs/naming_clades.md. (Accessed 19 September 2020).
- Ng, P.C., Henikoff, S., 2003. SIFT: predicting amino acid changes that affect protein function. *Nucleic Acids Res.* 31, 3812–3814.
- Ozono, S., Zhang, Y., Ode, H., Seng, T.T., Imai, K., Miyoshi, K., Kishigami, S., Ueno, T., Iwatani, Y., Suzuki, T., 2020. Naturally mutated spike proteins of SARS-CoV-2 variants show differential levels of cell entry. *bioRxiv*.
- Pachetti, M., Marini, B., Benedetti, F., Giudici, F., Mauro, E., Storicci, P., Masciovecchio, C., Angeletti, S., Ciccozzi, M., Gallo, R.C., 2020. Emerging SARS-CoV-2 mutation hot spots include a novel RNA-dependent-RNA polymerase variant. *J. Transl. Med.* 18, 1–9.
- Pejaver, V., Urresti, J., Lugo-Martinez, J., Pagel, K.A., Lin, G.N., Nam, H.-J., Mort, M., Cooper, D.N., Sebat, J., Iakoucheva, L.M., 2017. MutPred2: inferring the molecular and phenotypic impact of amino acid variants. *bioRxiv*. Preprint.
- Rabbi, M.F.A., Khan, M.I., Hasan, S., Chalita, M., Hasan, K.N., Sufian, A., Hosen, M.B., Polol, M.N.I., Naima, J., Lee, K., 2021. Large scale genomic and evolutionary study reveals SARS-CoV-2 virus isolates from Bangladesh strongly correlate with European origin and not with China. *bioRxiv*.
- Rahimi, A., Mirzazadeh, A., Tavakolpour, S., 2021. Genetics and genomics of SARS-CoV-2: a review of the literature with the special focus on genetic diversity and SARS-CoV-2 genome detection. *Genomics* 113, 1221–1232.
- Rahman, M.S., Islam, M.R., Alam, A.R.U., Islam, I., Hoque, M.N., Akter, S., Rahaman, M.M., Sultana, M., Hossain, M.A., 2020. Evolutionary dynamics of SARS-CoV-2 nucleocapsid protein (N protein) and its consequences. *bioRxiv*.
- Rambaut, A., Holmes, E.C., O’Toole, A., Hill, V., McCrone, J.T., Ruis, C., Plessis, L. du, Pybus, O.G., 2020. A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology. *Nat. Microbiol.* 5, 1403–1407.
- Russo, R., Andolfo, I., Lasorsa, V.A., Iolascon, A., Capasso, M., 2020. Genetic analysis of the novel SARS-CoV-2 host receptor TMPRSS2 in different populations. *bioRxiv*.
- Shishir, T.A., Naser, I.B., Faruque, S.M., 2021. In silico comparative genomics of SARS-CoV-2 to determine the source and diversity of the pathogen in Bangladesh. *PLoS One* 16, e0245584.
- Siddiqui, T., Sultana, M., Sultana, R., Akhter, S., 2020. Labour Migration from Bangladesh 2018 Achievements and Challenges [Internet]. [cited 2020 Aug 8]. Available from: <https://www.forum-asia.org/uploads/wp/2019/05/Migration-Trend-Analysis-2018-RMMRU.pdf>.
- Tan, W., Zhao, X., Ma, X., Wang, W., Niu, P., Xu, W., Gao, G.F., Wu, G., 2020. A novel coronavirus genome identified in a cluster of pneumonia cases—Wuhan, China 2019–2020. *China CDC Weekly* 2, 61–62.
- Toyoshima, Y., Nemoto, K., Matsumoto, S., Nakamura, Y., Kiyotani, K., 2020. SARS-CoV-2 genomic variations associated with mortality rate of COVID-19. *J. Hum. Genet.* 1–8.
- Ul Alam, A.R., Rafiul Islam, M., Shaminur Rahman, M., Islam, O.K., Anwar Hossain, M., 2020. Understanding the possible origin and genotyping of first Bangladeshi SARS-CoV-2 strain. *J. Med. Virol.*
- Wang, R., Chen, J., Hozumi, Y., Yin, C., Wei, G.-W., 2020. Decoding asymptomatic COVID-19 infection and transmission. *J. Phys. Chem. Lett.* 11, 10007–10015.
- Wei, Y., Silke, J.R., Aris, P., Xia, X., 2020. Coronavirus genomes carry the signatures of their habitats. *PLoS One* 15, e0244025.
- WORLDOMETER, 2020. Available: <https://www.worldometers.info/coronavirus/country/bangladesh/>. (Accessed 18 October 2020).
- Wu, S., Tian, C., Liu, P., et al., 2021. Effects of SARS-CoV-2 mutations on protein structures and intraviral protein–protein interactions. *J. Med. Virol.* 93, 2132–2140.
- Yao, H.-P., Lu, X., Chen, Q., Xu, K., Chen, Y., Cheng, L., Liu, F., Wu, Z., Wu, H., Jin, C., 2020. Patient-derived Mutations Impact Pathogenicity of SARS-CoV-2. *CELL-D-20-01124*.
- Yew, N.S., Cheng, S.H., 2004. Reducing the immunostimulatory activity of CpG-containing plasmid DNA vectors for non-viral gene therapy. *Expert Opin. Drug Deliv.* 1, 115–125.
- Zhan, X.-Y., Zhang, Y., Zhou, X., Huang, K., Qian, Y., Leng, Y., Yan, L., Huang, B., He, Y., 2020. Molecular evolution of SARS-CoV-2 structural genes: evidence of positive selection in spike glycoprotein. *bioRxiv*.
- Zhang, L., Jackson, C.B., Mou, H., Ojha, A., Rangarajan, E.S., Izard, T., Farzan, M., Choe, H., 2020a. The D614G mutation in the SARS-CoV-2 spike protein reduces S1 shedding and increases infectivity. *bioRxiv*.
- Zhang, Y., Zhang, J., Chen, Y., Luo, B., Yuan, Y., Huang, F., Yang, T., Yu, F., Liu, J., Liu, B., 2020b. The ORF8 protein of SARS-CoV-2 mediates immune evasion through potentially downregulating MHC-I. *bioRxiv*.
- Zhou, P., Yang, X.-L., Wang, X.-G., Hu, B., Zhang, L., Zhang, W., Si, H.-R., Zhu, Y., Li, B., Huang, C.-L., 2020. A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* 579, 270–273.