



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.



## The status and analysis of common mutations found in the SARS-CoV-2 whole genome sequences from Bangladesh

Sadniman Rahman<sup>a,1</sup>, Md. Asaduzzaman Shishir<sup>b</sup>, Md Ismail Hosen<sup>c</sup>, Miftahul Jannat Khan<sup>d</sup>, Ashiqul Arefin<sup>e</sup>, Ashfaqu Muid Khandaker<sup>a,\*,1</sup>

<sup>a</sup> Branch of Genetics and Molecular Biology, Department of Zoology, University of Dhaka, Bangladesh

<sup>b</sup> Department of Microbiology, Primeasia University, Dhaka, Bangladesh

<sup>c</sup> Department of Biochemistry and Molecular Biology, University of Dhaka, Bangladesh

<sup>d</sup> Department of Anesthesiology, Bangabandhu Sheikh Mujib Medical University (BSMMU), Dhaka, Bangladesh

<sup>e</sup> Combined Military Hospital, Dhaka, Bangladesh

### ARTICLE INFO

Edited by Dominic Voon

#### Keywords:

SARS-CoV-2

Common mutations

Block mutation

SR rich motif

### ABSTRACT

Rapid emergence of covid-19 variants by continuous mutation made the world experience continuous waves of infections and as a result, a huge number of death-toll recorded so far. It is, therefore, very important to investigate the diversity and nature of the mutations in the SARS-CoV-2 genomes. In this study, the common mutations occurred in the whole genome sequences of SARS-CoV-2 variants of Bangladesh in a certain timeline were analyzed to better understand its status. Hence, a total of 78 complete genome sequences available in the NCBI database were obtained, aligned and further analyzed. Scattered Single Nucleotide Polymorphisms (SNPs) were identified throughout the genome of variants and common SNPs such as: 241:C>T in the 5'UTR of Open Reading Frame 1A (ORF1A), 3037: C>T in Non-structural Protein 3 (NSP3), 14,408: C>T in ORF6 and 23,402: A>G, 23,403: A>G in Spike Protein (S) were observed, but all of them were synonymous mutations. About 97% of the studied genomes showed a block of tri-nucleotide alteration (GGG>AAC), the most common non-synonymous mutation in the 28,881–28,883 location of the genome. This block results in two amino acid changes (203–204: RG>KR) in the SR rich motif of the nucleocapsid (N) protein of SARS-CoV-2, introducing a lysine in between serine and arginine. The N protein structure of the mutant was predicted through protein modeling. However, no observable difference was found between the mutant and the reference (Wuhan) protein. Further, the protein stability changes upon mutations were analyzed using the I-Mutant2.0 tool. The alteration of the arginine to lysine at the amino acid position 203, showed reduction of entropy, suggesting a possible impact on the overall stability of the N protein. The estimation of the non-synonymous to synonymous substitution ratio (dN/dS) were analyzed for the common mutations and the results showed that the overall mean distance among the N-protein variants were statistically significant, supporting the non-synonymous nature of the mutations. The phylogenetic analysis of the selected 78 genomes, compared with the most common genomic variants of this virus across the globe showed a distinct cluster for the analyzed Bangladeshi sequences. Further studies are warranted for conferring any plausible association of these mutations with the clinical manifestation.

**Abbreviations:** CoVs, Coronaviruses; +ssRNA, positive single-stranded RNA; RdRp, RNA-dependent RNA polymerase; NSP, Nonstructural Protein; PLP, Papain-like Protease; ORFs, Open Reading Frames; RTC, Replication–Transcription Complex; RBD, Receptor-Binding Domain; ACE2, Angiotensin-Converting Enzyme 2; TMPRSS2, Transmembrane Protease Serine 2; sgRNAs, Sub-genomic RNAs; ERGIC, ER-Golgi intermediate compartment; NTD, N-terminal Domain; CTD, C-terminal Domain; SNP, Single Nucleotide Polymorphism; IRF3, Interferon Regulatory Factor 3; NFkB, Nuclear Factor kappa B; GSK3, Glycogen Synthase Kinase 3; CDK, Cyclin Dependent Kinases; ECM, Extracellular Matrix Protein; COX2, Cyclooxygenase 2; DGHS, General of Health Services.

\* Corresponding author.

E-mail address: [muid.zoo@du.ac.bd](mailto:muid.zoo@du.ac.bd) (A.M. Khandaker).

<sup>1</sup> Equal contribution.

<https://doi.org/10.1016/j.genrep.2022.101608>

Received 21 May 2021; Received in revised form 30 March 2022; Accepted 31 March 2022

Available online 4 April 2022

2452-0144/© 2022 Elsevier Inc. All rights reserved.

## 1. Introduction

SARS-CoV-2, the causative agent of COVID-19 infection, is a faster spreading pathogenic virus than the earlier SARS and MERS coronaviruses and belongs to the  $\beta$ -coronavirus genera (Naqvi et al., 2020). SARS-CoV-2 pathogenesis involves both the innate as well as the adaptive immune system (Morse et al., 2020) leading to the activation of signaling cascades, culminating in the release of cytokines, and chemokines and causes the recruitment of immune cells to the site of infection (Fung and Liu, 2014). And the dysregulation of the host's immune response leads to excessive inflammation, altered adaptive immune response, and sometimes even to death (Moens and Meys, 2020). Furthermore, emergence of new variants due to the mutation in the viral genome is facilitating newer clinical manifestations (Bakhshandeh et al., 2021). Although most mutations in the SARS-CoV-2 genome are predicted to be very insignificant, a small proportion might affect the functional properties, modify the infectivity, severity of disease or interactions with host immunity (Harvey et al., 2021).

The complete genome of SARS-CoV-2 is about 29.9 kb (Wuhan variant) with a GC content of 38% and composed of 12 functional open reading frames (ORFs) (Khailany et al., 2020; Naqvi et al., 2020). The ORF1a and ORF1b (5'-3') encode 16 non-structural proteins (NSP1–NSP16), i.e. polyproteins (Alanagreh et al., 2020) among which NSP3 (4955–5900 bp) and NSP5 (10,055–10,977 bp) encode for proteases (Fig. 1) that engage in the cleaving of polypeptides and blocking of the host's innate immune response (Rastogi et al., 2020).

Worldwide, multiple genomic variants, harboring different mutations in the spike protein of the SARS-CoV-2 have been detected, such as the B.1.1.7 (first detected in UK, September 2020), B.1.351 (South Africa, December 2020), P.1 (Detected in Japan from Brazilian travelers, January 2021), B.1.427/B.1.429 (USA, February 2021) and B.1.617.2 (India, 2021) variants (Davies and Jarvis, 2021; Zhou et al., 2021; Sabino et al., 2021; McCallum et al., 2021; Adam, 2021).

The first positive case of SARS-CoV-2 infection in Bangladesh was detected through RT-PCR assays in three Bangladeshi individuals on 07th March 2020 (Anwar et al., 2020). Since then, there are certain reports on the genome analysis of the SARS-CoV-2 from Bangladesh, the physiological conditions of the patients, association of the comorbidities to the severity as well as the comparison of global and local mutations (Hasan et al., 2021; Mannan et al., 2021; Rahman et al., 2021).

In this present study, we specifically monitored and analyzed 78 curated whole-genome sequences of SARS-CoV-2 submitted at the NCBI genome databases from Bangladesh to understand the commonly found mutations and the nature of those mutations. Thus, understanding the nature of common mutations in a timeline will help in analyzing the diverse SARS-CoV-2 genomes in the country.

## 2. Methodology

This study has been conducted based on the analysis of the whole genome sequences of the SARS-CoV-2 submitted from Bangladesh to the NCBI database (<https://www.ncbi.nlm.nih.gov/sars-cov-2/>) from June to October 2020. Sequences used in this study have been given in the Table 1 with their accession numbers.

Alignment of multiple sequences was performed with the submitted whole genome sequences of SARS-CoV-2 from Bangladesh using MEGA X software. The whole genome sequence of Wuhan variant was used as the reference (Accession: NC\_045512.2) sequence. At first, polymorphism along with conserved regions was counted from the aligned sequences. Then the mutation position and the specific mutations subtypes were recorded. The frequency of the mutations occurring at the analyzed genomes were calculated and presented in Table 1.

We further predicted the structure of the N-protein with or without the mutations, using the SWISS-MODEL tool. At first, the Template for building the model was searched with BLAST and HHblits against the SWISS-MODEL template library (SMTL, last update: 2020-12-30, last included PDB release: 2020-12-25). The target sequence was searched with BLAST against the primary amino acid sequence contained in the SMTL. A total of 67 templates were found. For each identified template, the template's quality was predicted from features of the target-template alignment. The templates with the highest quality were selected for model building. Models were built based on the target-template alignment using ProMod3 of the SWISS-MODEL. The global and per-residue model quality was assessed using the QMEAN scoring function.

Further, the I-Mutant2.0: tool was applied for predicting the stability changes of the N-protein upon mutations (Capriotti et al., 2005). The phylogenetic analysis and the difference between the nonsynonymous and synonymous distances (dN-dS) per site from averaging over all sequence pairs of each gene were calculated using the MEGA X. The dN-dS analyses were conducted using the Nei-Gojobori model.

The genetic relatedness of the Corona virus strains of Bangladesh was estimated with other variants by using the Neighbor-Joining method (Saitou and Nei, 1987). The bootstrap consensus tree inferred from 1000 replicates (Felsenstein, 1985) is taken to represent the evolutionary history of the taxa analyzed (Felsenstein, 1985). The evolutionary distances were computed using the maximum composite likelihood method (Felsenstein, 1985). This analysis involved 71 and 45 genome sequences among which majority were from Bangladesh and the rest of the sequences were of different variants. Branches corresponding to partitions reproduced in less than 50% bootstrap replicates are collapsed. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) is shown next to the branches (Felsenstein, 1985). The evolutionary distances were computed using the Maximum Composite Likelihood method (Tamura et al., 2004) and are in the units of the number of base substitutions per

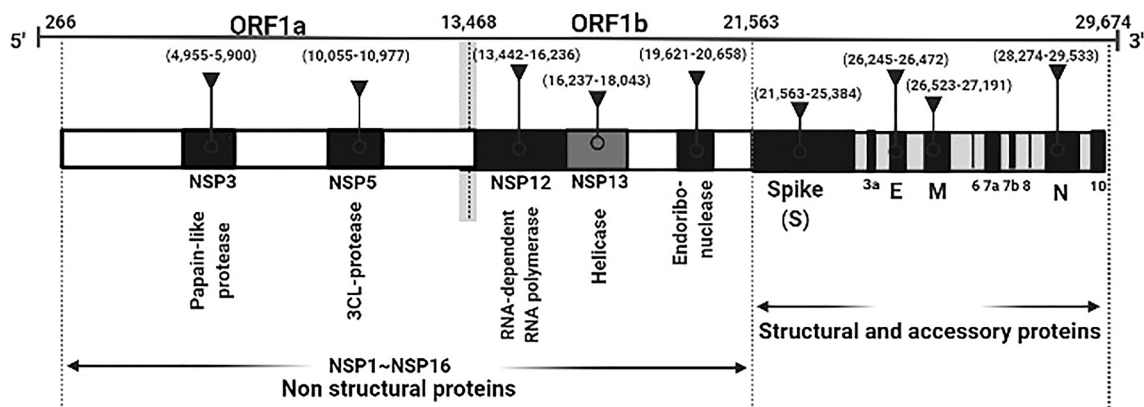


Fig. 1. SARS-CoV-2 whole genome structure and organization.

**Table 1**

Details of mutations identified in the SARS-CoV-2 whole genome sequences submitted from Bangladesh from June 2020 to December 2021.

Position	From	To	Type of mutation	Accession no.	Submitted on	Mutation frequency (%)
205	T	C	SM	MT876547.1 MT876554.1 MT876599.1 MT876606.1	12 August 2020 12 August 2020 12 August 2020 12 August 2020	6.55
*241	C	T	SM	All from BD		100
683	C	T	SM	MT876433.1	12 August 2020	1.63
751	C	T	SM	MT607252.1	15 June 2020	1.63
1036	T	C	SM	MT655744.1	23 June 2020	1.63
1148	G	T	SM	MT876572.1	12 August 2020	1.63
1457	C	T	SM	MT876526.1	12 August 2020	1.63
1734	C	T	SM	MT876547.1	12 August 2020	1.63
1820	G	A	SM	MT657958.1 MT876547.1	23 June 2020 12 August 2020	3.27
2057	A	G	SM	MT607252.1	15 June 2020	1.63
2110	C	A	SM	MT876556.1	12 August 2020	1.63
2113	C	T	SM	MT731937.1	08 July 2020	1.63
2210	G	T	SM	MT581414.1 MW531680.1	31 October 2020 27 January 2021	3.27
2288	G	A	SM	MT876599.1	12 August 2020	1.63
2388	C	T	SM	MT731932.1 MT731933.1 MT731935.1	08 July 2020 08 July 2020 08 July 2020	4.91
2805	A	C	SM	MT876556.1	12 August 2020	1.63
2910	C	T	SM	MT876555.1 MW624725.1	12 August 2020 18 February 2021	3.27
*3037	C	T	SM	All from BD		100
3077	G	A	SM	MT581419.1 MT581416.1	31 October 2020	3.27
3163	T	C	SM	MT876556.1	12 August 2020	1.63
3234	A	G	SM	MT657958.1	23 June 2020	1.63
3533	T	C	SM	MT666099.1	25 June 2020	1.63
3688	C	T	SM	MT581419.1 MT581418.1 MT581417.1 MT581416.1 MT581415.1	31 October 2020	8.19
3754	A	G	SM	MT666068.1	25 June 2020	1.63
3871	G	T	SM	MT581411.1	31 October 2020	1.63
3961	C	T	SM	MT601275.1 MT648676.1 MT655744.1 MT655746.1 MT876432.1 MT876555.1 MT876525.1 MT876572.1 MT581413 MT581410.1	12 June 2020 22 June 2020 23 June 2020 23 June 2020 12 August 2020 12 August 2020 12 August 2020 12 August 2020 31 October 2020 31 October 2020	16.39
4024	G	A	SM	MT876572.1	12 August 2020	1.63
4113	C	T	SM	MT876607.1	12 August 2020	1.63
4298	G	T	SM	MT581415.1	31 October 2020	1.63
4300	G	T	SM	MT876525.1	12 August 2020	1.63
4444	G	T	SM	MT581423.1 MT581422.1 MT581420.1	31 October 2020	4.91
4503	A	T	SM	MT664171.1	25 June 2020	1.63
4522	G	T	SM	MT601281.1	12 June 2020	1.63
4579	T	A	SM	MT667351.1	25 June 2020	1.63
4778	A	G	SM	MT876598.1	12 August 2020	1.63
5037	G	C	SM	MT876599.1	12 August 2020	1.63
5366	A	G	SM	MT655744.1	23 June 2020	1.63
5621	C	T	SM	MT607252.1	15 June 2020	1.63
5832	C	T	SM	MT655744.1	23 June 2020	1.63
5950	G	T	SM	MT876555.1	12 August 2020	1.63
6120	C	T	SM	MT731937.1	08 July 2020	1.63
6359	G	A	SM	MT664106.1 MT664109.1	25 June 2020 25 June 2020	3.27
6578	C	T	SM	MT581416.1	31 October 2020	1.63
6807	C	T	SM	MT876547.1	12 August 2020	1.63
7113	C	T	SM	MT876527.1	12 August 2020	1.63
7528	C	T	SM	MT581419.1	31 October 2020	1.63
8026	A	G	SM	MT731937.1	08 July 2020	1.63
8127	C	T	SM	MT664171.1	25 June 2020	1.63
8156	T	G	SM	MT876599.1	12 August 2020	1.63

(continued on next page)

Table 1 (continued)

Position	From	To	Type of mutation	Accession no.	Submitted on	Mutation frequency (%)
8327	C	T	SM	MT666099.1	25 June 2020	1.63
8371	G	T	SM	MT581423.1 MT581422.1 MT581420.1	31 October 2020	4.91
9050	C	T	SM	MT731936.1	08 July 2020	1.63
9223	C	T	SM	MT657271.1	23 June 2020	1.63
9246	C	T	SM	MT655744.1	23 June 2020	1.63
9502	C	T	SM	MT731937.1	08 July 2020	1.63
9532	C	T	SM	MT876556.1	12 August 2020	1.63
9565	C	T	SM	MT666099.1	25 June 2020	1.63
9828	A	G	SM	MT731936.1	08 July 2020	1.63
10,198	C	T	SM	MT731935.1	08 July 2020	1.63
10,252	C	T	SM	MT731932.1 MT581413	08 July 2020 31 October 2020	3.27
10,323	A	G	SM	MT876606.1	12 August 2020	1.63
10,834	C	T	SM	MT876525.1	12 August 2020	1.63
10,870	G	T	SM	MT876571.1	12 August 2020	1.63
11,036	C	A	SM	MT876555.1	12 August 2020	1.63
11,042	G	T	SM	MT876554.1	12 August 2020	1.63
11,083	G	T	SM	MT731937.1	08 July 2020	1.63
11,719	G	A	SM	MT876526.1	12 August 2020	1.63
11,761	G	T	SM	MT667351.1	25 June 2020	1.63
11,824	C	T	SM	MT876599.1	12 August 2020	1.63
12,061	A	G	SM	MT876527.1	12 August 2020	1.63
12,070	G	T	SM	MT876607.1	12 August 2020	1.63
12,085	C	T	SM	MT731937.1	08 July 2020	1.63
12,357	C	T	SM	MT664107.1	25 June 2020	1.63
12,672	C	T	SM	MT648676.1	22 June 2020	1.63
12,936	A	C	SM	MT876547.1	12 August 2020	1.63
13,201	G	T	SM	MT664107.1	25 June 2020	1.63
13,348	G	T	SM	MT601281.1	12 June 2020	1.63
13,812	G	T	SM	MT731932.1	08 July 2020	1.63
13,920	G	A	SM	MT876525.1	12 August 2020	1.63
14,110	C	A	SM	MT655744.1	23 June 2020	1.63
*14408	C	T	SM	All from BD		100
14,645	C	T	SM	MT581410.1 MW624725.1	31 October 2020 18 February 2021	3.27
15,324	C	T	SM	MT581419.1 MT581418.1 MT581417.1 MT581416.1 MT581415.1	31 October 2020	8.19
15,540	C	T	SM	MT581413	31 October 2020	1.63
15,543	G	T	SM	MT876554.1	12 August 2020	1.63
15,714	C	T	SM	MT601281.1	12 June 2020	1.63
15,738	C	T	SM	MT876546.1	12 August 2020	1.63
15,960	C	T	SM	MT731937.1	08 July 2020	1.63
15,982	G	T	SM	MT876525.1	12 August 2020	1.63
16,596	C	T	SM	MT876607.1	12 August 2020	1.63
16,830	C	T	SM	MT731934.1	08 July 2020	1.63
16,939	T	C	SM	MT731936.1	08 July 2020	1.63
17,259	G	T	SM	MT667351.1	25 June 2020	1.63
17,427	G	T	SM	MT664171.1	25 June 2020	1.63
17,678	C	T	SM	MT876554.1	12 August 2020	1.63
18,105	G	T	SM	MT664106.1	25 June 2020	1.63
18,131	C	T	SM	MT876607.1	12 August 2020	1.63
18,457	C	T	SM	MT876607.1	12 August 2020	1.63
18,735	A	G	SM	MT601275.1	12 June 2020	1.63
18,877	C	T	SM	MT601281.1 MT664107.1	12 June 2020 25 June 2020	3.27
19,162	G	T	SM	MT581419.1	31 October 2020	1.63
19,273	C	T	SM	MT655948.1	23 June 2020	1.63
19,398	G	A	SM	MT655750.1	23 June 2020	1.63
19,723	G	T	SM	MT731932.1	08 July 2020	1.63
20,436	C	T	SM	MT655744.1	23 June 2020	1.63
20,480	C	T	SM	MT655746.1	23 June 2020	1.63
20,628	C	T	SM	MT731936.1	08 July 2020	1.63
20,679	G	T	SM	MT876548.1	12 August 2020	1.63
20,774	G	A	SM	MT581412	31 October 2020	1.63
20,893	G	T	SM	MT664107.1	25 June 2020	1.63
20,955	T	C	SM	MT876571.1	12 August 2020	1.63
21,204	G	T	SM	MT655746.1	23 June 2020	1.63
21,216	C	T	SM	MT876554.1	12 August 2020	1.63
21,306	C	T	SM	MT655746.1	23 June 2020	1.63
21,595	C	T	SM	MT876431.1	12 August 2020	1.63
21,639	C	T	SM	MT655746.1	23 June 2020	1.63

(continued on next page)

Table 1 (continued)

Position	From	To	Type of mutation	Accession no.	Submitted on	Mutation frequency (%)
21,707	C	T	SM	MT876546.1	12 August 2020	1.63
21,855	C	T	SM	MT876555.1	12 August 2020	1.63
21,941	G	T	SM	MT655745.1	23 June 2020	1.63
21,998	C	T	SM	MT876606.1	12 August 2020	3.27
				MW624725.1	18 February 2020	
22,199	G	T	SM	MT581413	31 October 2020	1.63
22,343	G	C	SM	MT655742.1	23 June 2020	1.63
22,444	C	T	SM	MT655742.1	23 June 2020	3.27
				MT664107.1	25 June 2020	
22,501	T	C	SM	MT876598.1	12 August 2020	1.63
23,029	C	T	SM	MT664106.1	25 June 2020	3.27
				MT664109.1	25 June 2020	
23,095	A	G	SM	MT664105.1	25 June 2020	3.27
				MT664175.1	25 June 2020	
23,101	T	G	SM	MT648676.1	22 June 2020	1.63
23,202	C	T	SM	MT581410.1	31 October 2020	1.63
23,230	C	T	SM	MT581418.1	31 October 2020	1.63
23,268	T	G	SM	MT581418.1	31 October 2020	1.63
*23402	A	G	SM	All from BD except MT664107.1	October 2020	98.36
*23403	A	G	SM	All sequences	October 2020	100
23,952	T	G	SM	MT581419.1	31 October 2020	1.63
23,586	A	G	SM	MT731933.1	08 July 2020	1.63
23,587	G	T	SM	MT648676.1	22 June 2020	1.63
23,599	T	G	SM	MT876571.1	12 August 2020	1.63
23,608	G	T	SM	MT657271.1	23 June 2020	1.63
23,934	C	T	SM	MT655745.1	23 June 2020	1.63
23,957	G	A	SM	MT581419.1	31 October 2020	1.63
24,181	C	T	SM	MT876555.1	12 August 2020	1.63
24,383	A	G	SM	MT876607.1	12 August 2020	1.63
25,494	G	T	SM	MT601281.1	12 June 2020	1.63
25,563	G	T	SM	MT601281.1	12 June 2020	1.63
25,597	T	C	SM	MT655750.1	23 June 2020	1.63
25,615	A	G	SM	MT657271.1	23 June 2020	1.63
25,644	G	T	SM	MT731935.1	8 July 2020	1.63
25,713	T	C	SM	MT876548.1	12 August 2020	1.63
25,883	G	T	SM	MT876554.1	12 August 2020	1.63
25,906	G	T	SM	MT581423.1	31 October 2020	3.27
				MT581422.1		
26,058	C	T	SM	MT876556.1	12 August 2020	1.63
26,302	T	C	SM	MT666099.1	25 June 2020	1.63
26,735	C	T	SM	MT601281.1	12 June 2020	1.63
				MT664107.1	25 June 2020	
26,895	C	T	SM	MT731937.1	08 July 2020	1.63
27,199	C	T	SM	MT876606.1	12 August 2020	1.63
27,316	A	T	SM	MT601281.1	12 June 2020	1.63
27,389	C	T	SM	MT664109.1	25 June 2020	1.63
27,518–19	GC	TT	NSM	MT666099.1	25 June 2020	1.63
27,675	A	C	SM	MT876554.1	12 August 2020	1.63
27,944	C	T	SM	MT607252.1	15 June 2020	1.63
27,945	C	G	SM	MT581414.1	31 October 2020	1.63
28,008	A	G	SM	MT876599.1	12 August 2020	1.63
28,085	G	A	SM	MT655742.1	23 June 2020	1.63
28,115	C	T	SM	MT664107.1	25 June 2020	1.63
28,304	A	G	SM	MT876556.1	12 August 2020	1.63
28,321	G	T	SM	MT601281.1	12 June 2020	1.63
28,435	C	T	SM	MT731937.1	08 July 2020	1.63
28,521	A	G	SM	MT667351.1	25 June 2020	1.63
28,687	C	T	SM	MT581411.1	31 October 2020	1.63
28,690	G	A	SM	MT601287.1	12 June 2020	1.63
28,831	C	T	SM	MT876571.1	12 August 2020	1.63
28,854	C	T	SM	MT601281.1	12 June 2020	3.27
				MT664107.1	25 June 2020	
**28,881–83	GGG	AAG	NSM	All except for Ref seq, MT601281.1 and MT664107.1 Sequences submitted in 2021		96.72
28,888	T	C	SM	MT876546.1	12 August 2020	1.63
28,893	C	T	SM	MT581415.1	31 October 2020	1.63
28,903	G	T	SM	MT581415.1	31 October 2020	1.63
28,960	G	T	SM	MT731935.1	8 July 2020	1.63
29,081	G	T	SM	MT664107.1	25 June 2020	1.63
29,085	C	T	SM	MT664107.1	25 June 2020	1.63
29,218	C	T	SM	MT655744.1	23 June 2020	1.63
29,260	G	T	SM	MT657958.1	23 June 2020	1.63
29,296	C	T	SM	MT876554.1	12 August 2020	1.63
29,403	A	G	SM		31 October 2020	4.91

(continued on next page)

Table 1 (continued)

Position	From	To	Type of mutation	Accession no.	Submitted on	Mutation frequency (%)
				MT581423.1		
				MT581422.1		
				MT581420.1		
29,431	G	T	SM	MT601281.1	12 June 2020	1.63
29,614	C	T	SM	MT664107.1	25 June 2020	1.63
29,661	T	C	SM	MT876555.1	12 August 2020	1.63
29,666	C	T	SM	MT876527.1	12 August 2020	1.63
29,688	G	T	SM	MT876571.1	12 August 2020	1.63
29,736	G	T	SM	MT657958.1	23 June 2020	1.63
29,741	C	T	SM	MT876547.1	12 August 2020	3.27
				MT581414.1	31 October 2020	
29,447	G	T	SM	MT655742.1	23 June 2020	1.63
29,753	T	C	SM	MT876546.1	12 August 2020	1.63
29,785	T	C	SM	MT648676.1	22 June 2020	1.63

Here, SM = Synonymous Mutation; NSM = Non-synonymous Mutation. \* showing the nucleotide positions of commonly found SNPs whereas \*\* showing commonly found triple base (block) mutations.

site. This analysis involved 45 nucleotide sequences. All ambiguous positions were removed for each sequence pair (pairwise deletion option). There were a total of 29,677 positions in the final dataset. Evolutionary analyses were conducted in MEGA X (Kumar et al., 2018).

### 3. Result and discussion

After analyzing all 78 complete genome sequences of SARS-CoV-2 submitted from Bangladesh, a bloc of tri-nucleotide of GGG>AAC (triple base mutation) was most commonly observed in the 28,881–28,883 location of the genome as missense in nature (non-synonymous) (Table 1). However, other mutations in the genome were found as single nucleotide polymorphism (SNPs), among them some were also common but synonymous mutations, such as: 241:C>T in the 5'UTR of ORF1A, 3037: C>T in 'NSP3' and 14,408: C>T in ORF6 (Table 1). The A>G mutations located in the Spike Glycoprotein of the virus at positions 23,402 and 23,403 were also very frequent (98.36% and 100% respectively, Table 1). However, these were synonymous mutations with no structural implications.

Phylogenetic analysis of the whole genome sequences of the 71 SARS-CoV-2 sequences (61 Bangladeshi, 1 Wuhan and 9 most common variants) showed that the strains isolated from Bangladesh were more closely related to Wuhan variant. The tree generated two main clusters, cluster 1 and cluster 2. Cluster 1 was comprised of the variants emerged later viz. Gamma, Iota, Mu, Kappa, Beta, Delta, Alpha, Eta, Lambda variants. All the analyzed sequences from Bangladesh were in cluster 2 along with the variant from Wuhan. Cluster 2 formed five sub-clusters: sub-cluster 1 (SC1), SC2, SC3, SC4 and SC5. Since the strains were also labeled with their times, it could be speculated that the minimum mutations occurred during the month of October 2020 compared to the month of June, July and August (Fig. 2A).

The tree generated from the estimation of genetic relatedness among the initially prevailing strains revealed that the causative strains of infection in Bangladesh were very similar to the Wuhan variant during the period of June 2020 to October 2020 (Fig. 2A). While considering the sequences of both 2020 and 2021 (representative 2 sequences from each month), two distinct clusters were found: cluster 1 (C1) and cluster 2 (C2) (Fig. 2B). One local strain (MT655746.1) of June 2020 belonging to the cluster 1, seemed to prevail up to January 2021 and its further circulation was not observed whereas another local strain (MT664175.1) of June 2020 belonging to the cluster 2a, seemed to circulate almost for one year, up to May 2021. Again, a strain similar to MT581415.1 (October 2020) was observed in September 2021.

Cluster 2 produced two separate sub-clusters C2a and C2b. Sub-cluster C2a contains the sequences of June 2020 to October 2020 and a single sequence from the month of May 2021 which indicates that this particular strain of 2020 continued to circulate till mid of 2021.

Eight (8) months' representative strains of 2021 were observed in the

Sub-cluster C2b where different variants that occurred worldwide were also found. Sequence similarity was observed among three (3) strains from June 2020, one (1) strain from July 2021, and the variants that occurred worldwide. It could be possible that the strain prevailing in the month of June 2020 continued to circulate till July 2021.

Another interesting observation was the prevalence of strains similar to the omicron variant during the period of Jan 2021 to April 2021 in Bangladesh although the variant was announced in the month of November 2021. This strain could be assumed to be introduced in the month of January 2021 but was no more detected after April 2021.

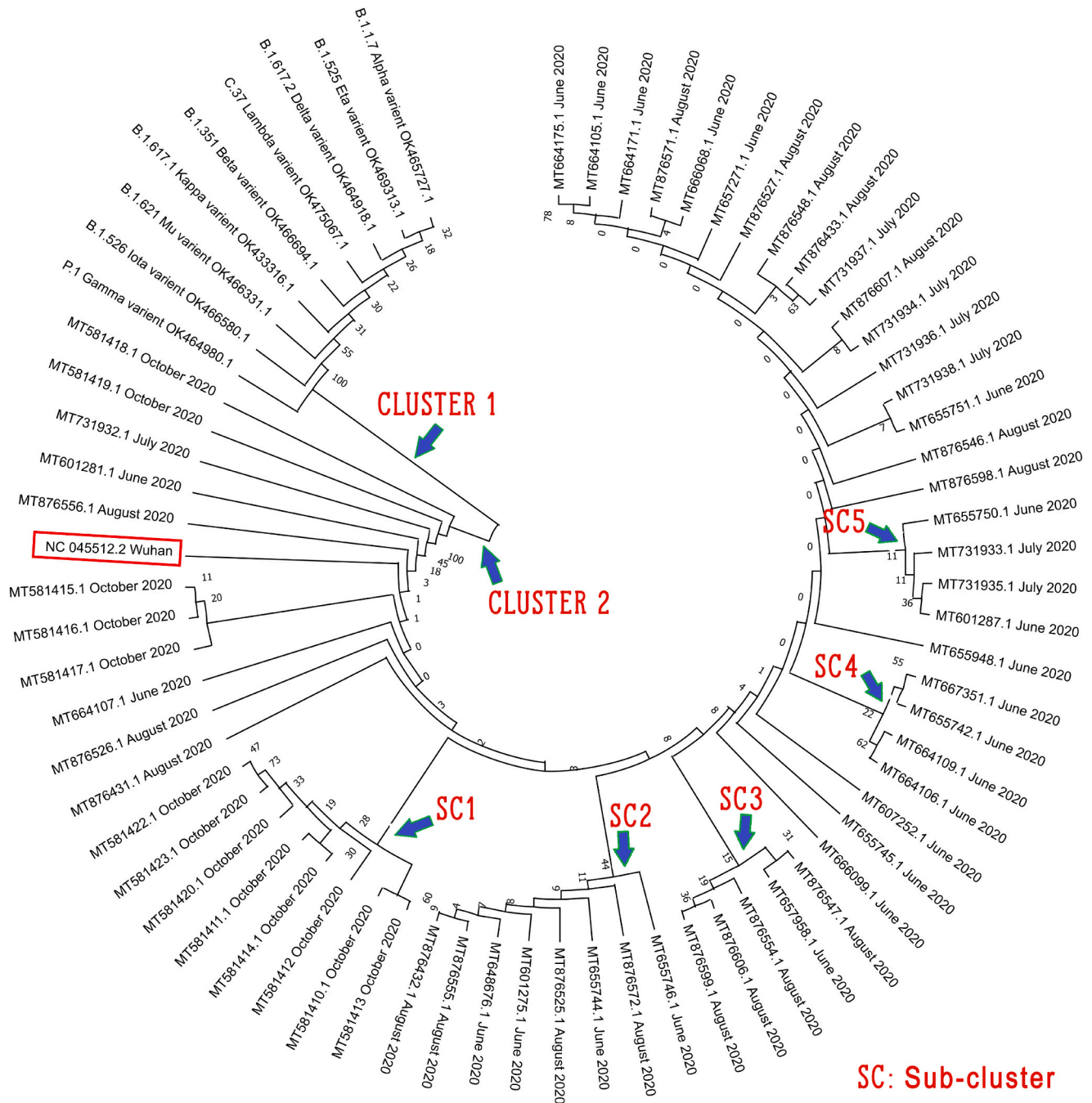
Moreover, the strains of June 2020 could be classified into three types which continued to circulate till January, May and July 2021.

In this current study we especially focused on the most prevalent non-synonymous mutation (28,881–28,883: GGG>AAC) in order to analyze the impact on the pathogenicity of SARS-CoV-2. According to the NCBI reference genome (Wuhan), 28,881–28,883: GGG>AAC bloc results in two amino acids (203–204: RG>KR) changes in the nucleocapsid (N) protein of the SARS-CoV-2 (Fig. 3A, B). Garvin et al. (2020) and Dey et al. (2021) also noticed similar block mutations in N protein of SARS-CoV-2 around the globe.

We performed dN-dS analysis for estimating the non-synonymous to synonymous substitution ratio (dN/dS) for the N-, S- and the NSP 3 genes. Our results for dN-dS analysis of the N-protein of the 61 analyzed sequences showed that the overall dN-dS *p*-value for all the three genes to be <1.00, indicating a constraint selection (amino acid changes disfavored, Table 2).

While looking at the surrounding sequence of these amino acids (Fig. 3B), it appears that the mutation discontinues a serine-arginine (S-R) dipeptide by introducing a lysine in-between them which is a basic and polar hydrophilic charged (+) amino acid. Basically, arginine provides the protein structure with more stability than lysine (Sokalingam et al., 2012). So, the incorporation of lysine in the motif could have impacts on the overall distinctive properties of the protein as reported before (Tylor et al., 2009). Especially, the serine-arginine dipeptide disordering may hamper the phosphorylation of the SR-rich domain. This phosphorylation event is critical for cellular localization and regulation of the N protein synthesis (Maitra et al., 2020). Notably the GSK3 (glycogen synthase kinase 3) phosphorylation site at Ser202 and a CDK (cyclin dependent kinase) phosphorylation site at Ser206 are in the vicinity of our identified block mutation. We thought that this interaction would contribute to reduction of conformational entropy and might affect protein structure. In this study, the change of N protein stability upon mutations at the amino acid positions 203–204 (RG>KR) was predicted using I-Mutant 2.0 tool and found that the incorporation of Lysine in 203 amino acid position predicted a reduction of entropy ( $\Delta\Delta G = -2.26$ ) and thus affecting its stability (Table 3). The structure of the protein with the block mutation was predicted and compared with the reference sequence (Wuhan variant) by SWISS-MODEL, a protein

A



**Fig. 2.** Phylogenetic analysis: the genetic relatedness of corona virus strains of Bangladesh with other variants was estimated by using the Neighbor-Joining method using MEGA X.

The phylogenetic tree revealed that during the period of June 2020 to October 2020, the causative strains of infection in Bangladesh were very similar to the Wuhan variant (Fig. 2A). The sequences from the year 2020 and 2021 resulted in two distinct clusters- C1 and C2 (Fig. 2B). Cluster 2 produced two separate sub-clusters C2a and C2b.

modeling tool (Fig. 3C). However, no observable difference was found in the block mutation area of the predicted N protein (GGG>AAC). On the other hand, Maitra et al., 2020 found that three miRNAs binding in the mutation site 28,881–3 can regulate the mutant pathogenicity. Taken together with these data, we suggest that the block mutation may regulate the stability and function of N protein rather than the structure of the protein.

Kang et al. (2020) reported that SARS-CoV-2 N protein which is 419 amino acid (aa) long, consists of three highly conserved parts: an N-terminal domain (NTD) that binds RNA, a C-terminal domain (CTD) for dimerization of the protein, and a linker region called SR-rich (serine-arginine) motif. The SR-rich motif is located in the middle region covering amino acids 177–207 in between NTD and CTD (Surjit and Lal, 2010). Our observed common triple base mutation results in amino acid



B

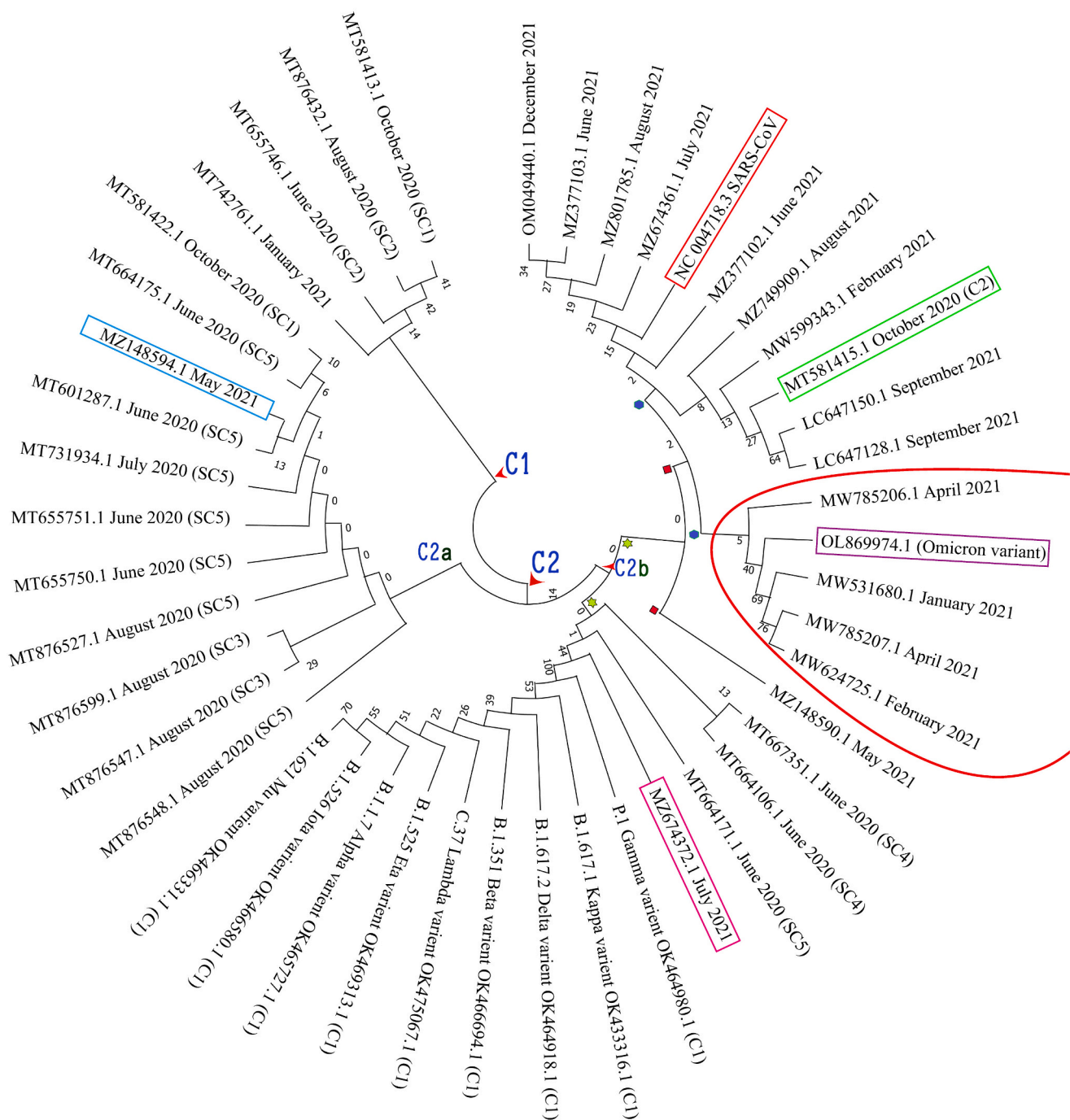
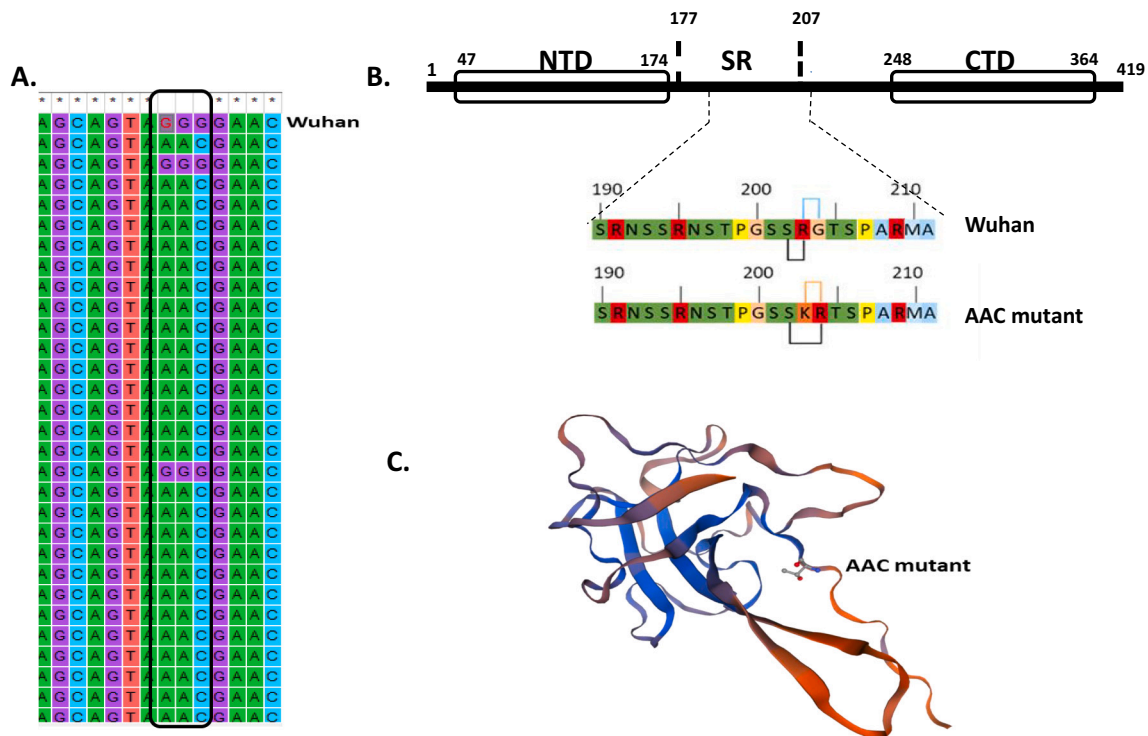


Fig. 2. (continued).

changes of R203K and G204R which occurred in the SR-rich motif (Fig. 3B). Tylor et al. (2009) revealed that SR rich motif is important for viral replication, N protein multimerization and RNA splicing. So, any disruption in the motif could affect the overall structure and function of the N protein. Tylor et al. (2009) also found that the mutated SR rich motif resulted in an extreme reduction of viral infectivity and brought about a remarkable deficiency in the viral replication capacity.

Mutational analysis also showed that the N-protein of the SARS-CoVs suppresses the activity of the cyclin-CDK complex function leading to hypophosphorylation followed by cell cycle arrest (Surjit and Lal, 2010). Again, the interaction between hnRNPA1 and N-protein through the

middle region (aa 161–210) of N-protein was considered to regulate the viral RNA synthesis (Luo et al., 2004). Mutation in this region might hamper the viral RNA synthesis. N-protein also interacts with B23, a phosphoprotein in the nucleus, through aa 175–210 i.e. the SR-rich motif which may contribute a significant role in centrosome duplication (Zeng et al., 2008). In addition, N-protein governs the upregulation of Cyclooxygenase 2 (COX2), the inflammatory agent. It was further revealed that the N-protein binds to the NFkB response element present in the COX2 promoter region through a 68 aa residue binding domain (aa 136–204) and activates its transcription (Yan et al., 2006). Interestingly the common block mutation (R203K and G204R) in some of the



**Fig. 3.** Analysis of the common block mutation: A. Representative image showing GGG>AAC (28881–28,883) mutation in the SARS-CoV-2 Bangladesh variants compared with Wuhan variant (first row) as revealed upon the alignment analyzed by clustalW. The block mark indicates the site of the triple base mutation. B. The mutation confers two amino acid changes (203–204: RG>KR) in the SR rich motif of N protein. NTD and CTD represent N terminal domain and C terminal domain respectively. The wild type (Wuhan) N protein with intact S-R dipeptide and the mutated (AAC mutant) N protein were intercalated with the insertion of lysine between S and R amino acids. Bars on the top indicate the wild type and mutated amino acids respectively. The bottom bars indicate the insertion of lysine between S and R amino acids. C. Predicted structure of N protein from the GGG>AAC mutated sequence (GenBank accession: MT876546.1) using SWISS-MODEL tool.

**Table 2**  
Predicting Non-synonymous to synonymous substitution rate of the common mutations.

Target gene harboring common mutations	dN-dS analysis for overall mean distance	
	p-Value	Std error
N	0.00	0.00
S	0.00	0.00
NSP3	0.00	0.00

**Table 3**  
Prediction of protein stability changes upon mutations using I-Mutant v2.0 (Capriotti et al., 2005).

Position of the amino acid in N-protein	WT	Mutant	DDG	pH	T
203	R	K	-2.26	7.0	25
204	G	R	0.00	7.0	25

WT: amino acid in wild-type protein, mutant: new amino acid after mutation, DDG: DG(NewProtein)-DG(WildType) in Kcal/mol, DDG < 0: decrease stability, DDG > 0: increase stability, T: temperature in Celsius degrees, pH: -log[H+].

genomes from Bangladesh occurred in the above-mentioned binding domain. So, it could be presumed that the triple base block mutation might have affected the COX2 transcriptional upregulation leading to reduced inflammation. However, COX2 expression was not investigated in this study, further experiments are required for the prediction.

**4. Conclusion**

The GGG>AAC non-synonymous mutation remained most frequent

in the Bangladeshi population during the study period. We predicted that the mutation is responsible for the reduced stability of the N protein due to the intercalation of the amino acid. However, due to the lack of experimental evidence, many questions regarding the influence of these mutations still remain elusive.

**Funding**

This research work was funded by the Centennial research grant, University of Dhaka, Bangladesh.

**CRedit authorship contribution statement**

SH was involved in Conceptualization; Data curation; Formal analysis; original draft writing.

MAS was involved in the Data curation; Formal analysis, Methodology, original draft writing, review and editing.

MIH was involved with Data Curation, Formal analysis, Methodology, Validation; Visualization and revised manuscript review and editing.

MJK was involved in Data curation; Formal analysis; original draft writing, review & editing.

AA was involved in Data curation; Formal analysis; original draft writing, review & editing.

AMK was involved in Conceptualization; Methodology, Data curation; Formal analysis; original draft writing, revised manuscript review and editing, overall Supervision; Validation; Visualization.

**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.

## References

- Adam, D., 2021. What scientists know about new, fast-spreading coronavirus variants. *Nature* 594, 7861.
- Alanagreh, L., Alzoughool, F., Atoum, M., 2020. The human coronavirus disease COVID-19: its origin, characteristics, and insights into potential drugs and its mechanisms. *Pathogens* 9, 331. <https://doi.org/10.3390/pathogens9050331>.
- Anwar, S., Nasrullah, M., Hosen, M.J., 2020. COVID-19 and Bangladesh: challenges and how to address them. *Front Public Heal.* 8, 154.
- Bakhshandeh, B., Jahanafrooz, Z., Abbasi, A., Goli, M.B., Sadeghi, M., Mottaqi, M.S., Zamani, M., 2021. Mutations in SARS-CoV-2; consequences in structure, function, and pathogenicity of the virus. *Microb. Pathog.* 154, 104831 <https://doi.org/10.1016/j.micpath.2021.104831>.
- Capriotti, E., Fariselli, P., Casadio, R., 2005. I-Mutant2.0: predicting stability changes upon mutation from the protein sequence or structure. *Nucl. Acids Res.* 33, 306–310.
- Davies, N.G., Jarvis, C.I., 2021. CMMID COVID-19 Working Group, Edmunds WJ, Jewell NP, Diaz-Ordaz K, Keogh RH. Increased mortality in community-tested cases of SARS-CoV-2 lineage B.1.1.7. *Nature* 593, 7858. <https://doi.org/10.1038/s41586-021-03426-1>. Epub 2021 Mar 15.
- Dey, T., Chatterjee, S., Manna, S., Nandy, A., Basak, S.C., 2021. Identification and computational analysis of mutations in SARS-CoV-2. *Comput. Biol. Med.* 129, 104166 <https://doi.org/10.1016/j.combiomed.2020.104166>.
- Felsenstein, J., 1985. Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* 39, 783–791.
- Fung, T.S., Liu, D.X., 2014. Coronavirus infection, ER stress, apoptosis and innate immunity. *Front. Microbiol.* <https://doi.org/10.3389/fmicb.2014.00296>.
- Garvin, M.R., Prates, T.E., Pavicic, M., et al., 2020. Potentially adaptive SARS-CoV-2 mutations discovered with novel spatiotemporal and explainable AI models. *Genome Biol.* 21, 304. <https://doi.org/10.1186/s13059-020-02191-0>.
- Harvey, W.T., Carabelli, A.M., Jackson, B., Gupta, R.K., Thomson, E.C., Harrison, E.M., Ludden, C., Reeve, R., Rambaut, A., Peacock, S.J., Robertson, D.L., 2021. SARS-CoV-2 variants, spike mutations and immune escape. *Nat. Rev. Microbiol.* 19, 409–424. <https://doi.org/10.1038/s41579-021-00573-0>.
- Hasan, M.M., Das, R., Rasheduzzaman, M., Hussain, M.H., Muzahid, N.H., Salauddin, A., Rumi, M.H., Mahbubur Rashid, S.M., Siddiki, A.Z., Mannan, A., 2021. Global and local mutations in bangladeshi SARS-CoV-2 genomes. *Virus Res.* 297, 198390 <https://doi.org/10.1016/j.virusres.2021.198390>.
- Kang, S., Yang, M., Hong, Z., Zhang, L., Huang, Z., Chen, X., He, S., Zhou, Ziliang, Zhou, Zhechong, Chen, Q., Yan, Y., Zhang, C., Shan, H., Chen, S., 2020. Crystal structure of SARS-CoV-2 nucleocapsid protein RNA binding domain reveals potential unique drug targeting sites. *Acta Pharm. Sin. B* 10, 1228–1238. <https://doi.org/10.1016/j.apsb.2020.04.009>.
- Khailany, R.A., Safdar, M., Ozaslan, M., 2020. Genomic characterization of a novel SARS-CoV-2. *Gene Rep.* 19 <https://doi.org/10.1016/j.genrep.2020.100682>.
- Kumar, S., Stecher, G., Li, M., Knyaz, C., Tamura, K., 2018. MEGA X: molecular evolutionary genetics analysis across computing platforms. *Mol. Biol. Evol.* 35, 1547–1549. <https://doi.org/10.1093/molbev/msy096>.
- Luo, H., Ye, F., Sun, T., Yue, L., Peng, S., Chen, J., Li, G., Du, Y., Xie, Y., Yang, Y., Shen, J., Wang, Y., Shen, X., Jiang, H., 2004. In vitro biochemical and thermodynamic characterization of nucleocapsid protein of SARS. *Biophys. Chem.* 112, 15–25. <https://doi.org/10.1016/j.bpc.2004.06.008>.
- Maitra, A., Sarkar, M.C., Raheja, H., Biswas, N.K., Chakraborti, S., Singh, A.K., Ghosh, S., Sarkar, S., Patra, S., Mondal, R.K., Ghosh, T., Chatterjee, A., Banu, H., Majumdar, A., Chinnaswamy, S., Srinivasan, N., Dutta, S., Das, S., 2020. Mutations in SARS-CoV-2 viral RNA identified in eastern India: possible implications for the ongoing outbreak in India and impact on viral structure and host susceptibility. *J. Biosci.* 45, 1–18. <https://doi.org/10.1007/s12038-020-00046-1>.
- Mannan, A., Mehedi, H.M.H., Chy, N.U.H.A., Qayum, M.O., Akter, F., Rob, M.A., Biswas, P., Hossain, S., Ayub, M.I., 2021. A multi-Centre, cross-sectional study on coronavirus disease 2019 in Bangladesh: clinical epidemiology and short-term outcomes in recovered individuals. *New Microbes New Infect.* 40, 100838 <https://doi.org/10.1016/j.nmni.2021.100838>.
- McCallum, M., Bassi, J., Marco, A., De, Chen, A., Walls, A.C., Iulio, Di, J., et al., 2021. SARS-CoV-2 immune evasion by the B.1.427/B.1.429 variant of concern. *Science* 373 (6555), 648–654. <https://doi.org/10.1126/science.abi7994>.
- Moens, L., Meyts, I., 2020. Recent human genetic errors of innate immunity leading to increased susceptibility to infection. *Curr. Opin. Immunol.* <https://doi.org/10.1016/j.coi.2019.12.002>.
- Morse, J.S., Lalonde, T., Xu, S., Liu, W.R., 2020. Learning from the past: possible urgent prevention and treatment options for severe acute respiratory infections caused by 2019-nCoV. *ChemBioChem* 21, 730–738. <https://doi.org/10.1002/cbic.202000047>.
- Naqvi, A.A.T., Fatima, K., Mohammad, T., Fatima, U., Singh, I.K., Singh, A., Atif, S.M., Hariprasad, G., Hasan, G.M., Hassan, M.I., 2020. Insights into SARS-CoV-2 genome, structure, evolution, pathogenesis and therapies: structural genomics approach. <https://doi.org/10.1016/j.bbadis.2020.165878>.
- Rahman, M.A., Shanjana, Y., Tushar, M.I., Mahmud, T., Rahman, G.M.S., Milan, Z.H., Sultana, T., Chowdhury, A.M.L.H., Bhuiyan, M.A., Islam, M.R., Reza, H.M., 2021. Hematological abnormalities and comorbidities are associated with COVID-19 severity among hospitalized patients: experience from Bangladesh. *PLoS One* 16, e0255379. <https://doi.org/10.1371/journal.pone.0255379>.
- Rastogi, M., Pandey, N., Shukla, A., Singh, S.K., 2020. SARS coronavirus 2: from genome to infectome. *Respir. Res.* <https://doi.org/10.1186/s12931-020-01581-z>.
- Sabino, E.C., Buss, L.F., Carvalho, M.P.S., Prete, C.A., Crispim, M.A.E., Fraijoli, N.A., et al., 2021. Resurgence of COVID-19 in Manaus, Brazil, despite high seroprevalence. *Lancet* 397. <https://doi.org/10.1016/S0140-6736>.
- Saitou, N., Nei, M., 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* 4.
- Sokalingam, S., Raghunathan, G., Soundararajan, N., Lee, S.G., 2012. A study on the effect of surface lysine to arginine mutagenesis on protein stability and structure using green fluorescent protein. *PLoS One* 7. <https://doi.org/10.1371/journal.pone.0040410>.
- Surjit, M., Lal, S.K., 2010. The nucleocapsid protein of the SARS coronavirus: structure, function and therapeutic potential. In: *Molecular Biology of the SARS-Coronavirus*. Springer, Berlin Heidelberg, pp. 129–151. [https://doi.org/10.1007/978-3-642-03683-5\\_9](https://doi.org/10.1007/978-3-642-03683-5_9).
- Tamura, K., Nei, M., Kumar, S., 2004. Prospects for inferring very large phylogenies by using the neighbor-joining method. *Proc. Natl. Acad. Sci. U. S. A.* 101, 11030–11035.
- Taylor, S., Andonov, A., Cutts, T., Cao, J., Grudsky, E., Van Domselaar, G., Li, X., He, R., 2009. The SR-rich motif in SARS-CoV nucleocapsid protein is important for virus replication. *Can. J. Microbiol.* 55, 254–260. <https://doi.org/10.1139/W08-139>.
- Yan, X., Hao, Q., Mu, Y., Timani, K.A., Ye, L., Zhu, Y., Wu, J., 2006. Nucleocapsid protein of SARS-CoV activates the expression of cyclooxygenase-2 by binding directly to regulatory elements for nuclear factor-kappa B and CCAAT/enhancer binding protein. *Int. J. Biochem. Cell Biol.* 38, 1417–1428. <https://doi.org/10.1016/j.biocel.2006.02.003>.
- Zeng, Y., Ye, L., Zhu, S., Zheng, H., Zhao, P., Cai, W., Su, L., She, Y., Wu, Z., 2008. The nucleocapsid protein of SARS-associated coronavirus inhibits B23 phosphorylation. *Biochem. Biophys. Res. Commun.* 369, 287–291. <https://doi.org/10.1016/j.bbrc.2008.01.096>.
- Zhou, D., Dejnirattisai, W., Ren, J., Stuart, D.I., Screaton, Correspondence, G.R., Supasa, P., et al., 2021. Evidence of escape of SARS-CoV-2 variant B.1.351 from natural and vaccine-induced sera. *Cell* 15, 184.