






Genome Sequences of Two Novel Coronavirus (SARS-CoV-2) Isolates from Dhaka, Bangladesh

M. Abdul Matin,^a  Abdullah Al Nahid,^{c,f}  Bishajit Sarkar,^{a,b} M. Asad Ullah,^{a,b} Yusha Araf,^{b,d} Nihad Adnan,^e Rashedul Islam,^f  Mohammad Shahedur Rahman^{a,b}

^aDepartment of Biotechnology and Genetic Engineering, Faculty of Biological Sciences, Jahangirnagar University, Savar, Dhaka, Bangladesh

^bCOVID Research Cell (CRC), Wazed Miah Science Research Center (WMSRC), Jahangirnagar University, Savar, Dhaka, Bangladesh

^cDepartment of Biochemistry and Molecular Biology, School of Life Sciences, Shahjalal University of Science and Technology, Sylhet, Bangladesh

^dDepartment of Genetic Engineering and Biotechnology, School of Life Sciences, Shahjalal University of Science and Technology, Sylhet, Bangladesh

^eDepartment of Microbiology, Faculty of Biological Sciences, Jahangirnagar University, Savar, Dhaka, Bangladesh

^fX-Genomics Research Laboratory, Dhaka, Bangladesh

M. Abdul Matin and Abdullah Al Nahid contributed equally to this work and these two authors are ordered on the basis of seniority.

ABSTRACT This study reports the genome sequences of two severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) strains detected in the nasopharyngeal swab specimens of two coronavirus disease 2019 (COVID-19) patients from Dhaka, Bangladesh.

The ongoing coronavirus disease 2019 (COVID-19) pandemic is caused by the viral pathogen severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), a member of the *Betacoronavirus* in the *Coronaviridae* family (1). The virus was first detected in late December 2019 in Wuhan, Hubei Province, China (2). On 8 March 2020, the first three cases of COVID-19 in Bangladesh were identified. As of 18 May 2021, a total of 780,857 confirmed cases of the disease with 12,181 deaths had been reported in the country (3). We announce here two genome sequences of SARS-CoV-2 from the collected samples of two female patients, 43 and 28 years old. The samples were collected on 2 and 4 December 2020 during the time period of a relatively reduced number of confirmed COVID-19 cases in Bangladesh (<https://www.worldometers.info/coronavirus>), prior to the second wave of the pandemic in the country from late March to early April 2021 (4). Both patients provided informed consent and were found to be COVID-19 positive by reverse transcription-PCR (RT-PCR) (5). The Ethical Review Board of Jahangirnagar University [reference no. BBEC, JU/M 2020/COVID-19/(8)3] approved all the protocols used to carry out this study.

The viral nucleic acids were extracted from the collected nasopharyngeal swab specimens using the ReliaPrep viral TNA miniprep system (catalog no. AX4820; Promega) according to the manufacturer's protocol. Library preparation was performed using an Illumina RNA prep kit with RNA enrichment Respiratory Virus Oligo Panel following the manufacturer's instructions. Thereafter, an Illumina MiniSeq system was used according to the manufacturer's protocol to sequence the prepared libraries in the paired-end format with a read length of 74 bp.

A total of 7,170,754 and 150,934 reads, respectively, from the SARS-CoV-2 samples were generated and quality assessed with FastQC version 0.11.8 (6). The raw reads were adapter trimmed with Trimmomatic version 0.39 (7), and the surviving 6,716,784 and 142,054 reads were subsequently mapped to the SARS-CoV-2 reference genome (GenBank accession no. [MN908947.3](https://www.ncbi.nlm.nih.gov/nuclseq/MN908947.3)) using the Burrows-Wheeler Aligner (BWA) version 0.7.17 (8). Ultimately, the consensus genomes were generated from the mapped reads using SAMtools version 1.11 (9). The genomes of the two Bangladeshi SARS-CoV-2 samples, i.e., hCoV-19/Bangladesh/JU-WMSRC-Dhaka-5/2020 and hCoV-19/Bangladesh/

Citation Matin MA, Nahid AA, Sarkar B, Ullah MA, Araf Y, Adnan N, Islam R, Rahman MS. 2021. Genome sequences of two novel coronavirus (SARS-CoV-2) isolates from Dhaka, Bangladesh. *Microbiol Resour Annot* 10:e00511-21. <https://doi.org/10.1128/MRA.00511-21>.

Editor Jelle Matthijnsens, KU Leuven

Copyright © 2021 Matin et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

Address correspondence to Rashedul Islam, xgenomics.bd@gmail.com, or Mohammad Shahedur Rahman, rahmanms@juniv.edu.

Received 19 May 2021

Accepted 7 June 2021

Published 8 July 2021

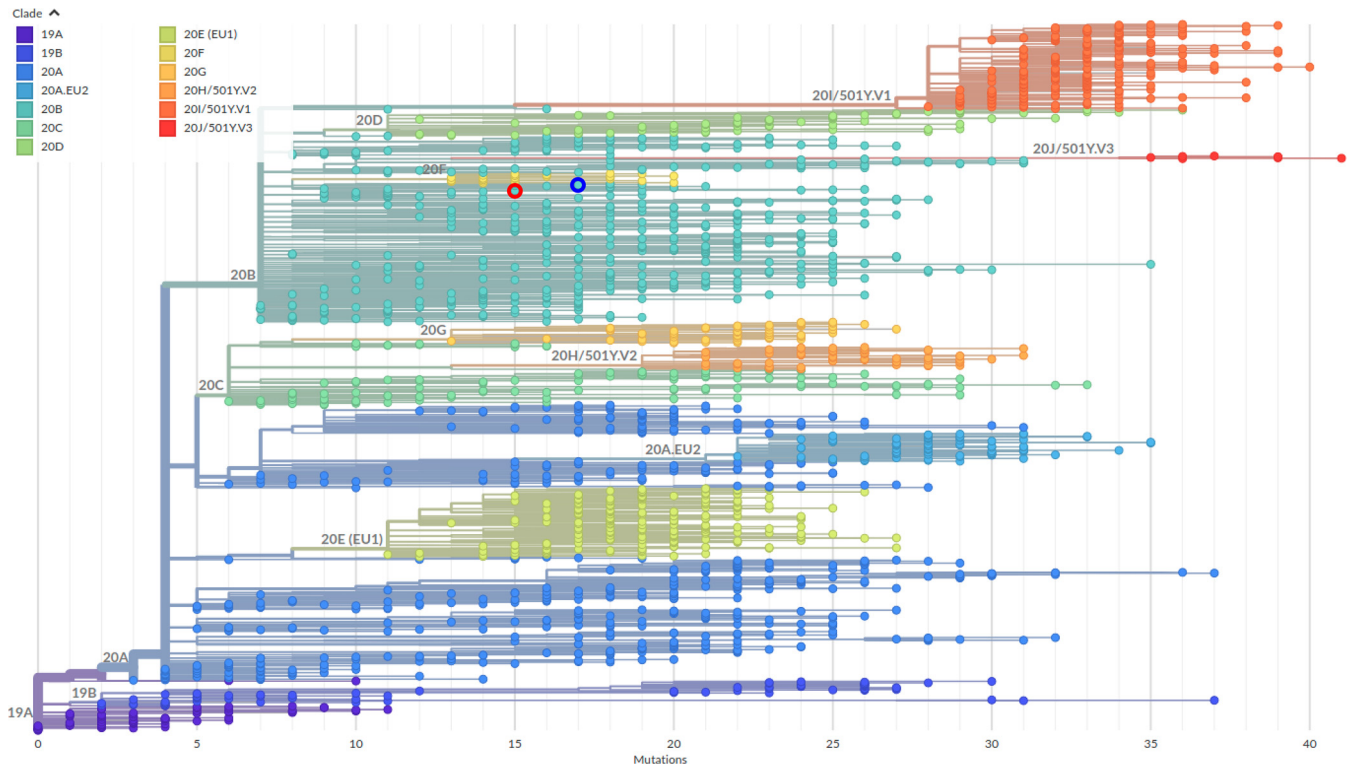


FIG 1 The genomes isolated from the two SARS-CoV-2 samples from the Dhaka District of Bangladesh are presented on a phylogenetic tree. The number of mutations in comparison to the reference genome of the Wuhan-Hu-1 isolate (GenBank accession no. [MN908947.3](https://www.ncbi.nlm.nih.gov/nuccore/MN908947.3)) is represented by the x axis. Nextclade beta version 0.14.2 (clades.nextstrain.org) was used to construct the tree on 4 May 2021, with the red and dark blue circles representing the positions of the hCoV-19/Bangladesh/JU-WMSRC-Dhaka-5/2020 (GISAID accession no. EPI_ISL_884088) and hCoV-19/Bangladesh/JU-WMSRC-Dhaka-8/2020 (GISAID accession no. EPI_ISL_884087) sequences, respectively.

JU-WMSRC-Dhaka-8/2020, were found to have base pair lengths of 29,888 and 29,848, respectively, with an average coverage of over 4,313 \times and 215 \times , respectively. No indels were found, and the GC content was measured to be 38% for both genomes. The consensus genomes along with the related sample metadata were uploaded to the Global Initiative on Sharing All Influenza Data (GISAID) database (10) on 27 January 2021. Nextclade beta version 0.14.2 (clades.nextstrain.org) by Nextstrain (11) was utilized in the phylogenetic analysis, which assigned both sequences to SARS-CoV-2 clade 20B (Fig. 1). The closest ancestral strains of the genomes include sequences from Bangladesh (GISAID accession no. EPI_ISL_959367, EPI_ISL_943554, and EPI_ISL_959386) and New Zealand (GISAID accession no. EPI_ISL_682280) according to Nextclade.

Genome variation analysis of the consensus genomes was performed using the Genome Detective Virus Tool version 1.132 (12), which indicated several changes in the studied genomes relative to the SARS-CoV-2 reference genome sequence of the Wuhan-Hu-1 isolate (GenBank accession no. [NC_045512.2](https://www.ncbi.nlm.nih.gov/nuccore/NC_045512.2)). The sequence of the hCoV-19/Bangladesh/JU-WMSRC-Dhaka-5/2020 strain exhibited 6 synonymous and 9 nonsynonymous mutations, whereas 7 synonymous and 10 nonsynonymous mutations were observed in the sequence of the hCoV-19/Bangladesh/JU-WMSRC-Dhaka-8/2020 strain (Table 1). Except for the Q94L mutation in the ORF7a gene, all nonsynonymous mutations were in both strains. According to the Nextstrain community build of Bangladesh (<https://nextstrain.org/community/CHRF-Genomics/ncovBangladesh@main>), the 9 nonsynonymous mutations common to the two studied genomes are also prevalent in other Bangladeshi SARS-CoV-2 genome sequences. However, the ORF7a Q94L mutation could not be detected in any other Bangladeshi genomes as of 4 May 2021. The most common mutations observed in SARS-CoV-2 genomes from Bangladesh include 5 nonsynonymous

TABLE 1 Mutations detected in both the hCoV-19/Bangladesh/JU-WMSRC-Dhaka-5/2020 (strain A) and hCoV-19/Bangladesh/JU-WMSRC-Dhaka-8/2020 (strain B) sequences compared to the reference genome of the Wuhan-Hu-1 isolate (GenBank accession no. [MN908947.3](https://www.ncbi.nlm.nih.gov/nuccore/MN908947.3))

Strain(s)	Nucleotide position	Reference nucleotide	Mutated nucleotide	Gene	Amino acid change
A, B	241	C	T	5'-UTR	Noncoding
A, B	335	C	T	ORF1ab	R24C
A, B	434	G	A	ORF1ab	E57K
A, B	1163	A	T	ORF1ab	I300F
B	1333	T	G	ORF1ab	None (synonymous mutation)
A, B	3037	C	T	ORF1ab	None (synonymous mutation)
A, B	3652	A	G	ORF1ab	None (synonymous mutation)
A	6229	A	G	ORF1ab	None (synonymous mutation)
A, B	9436	A	G	ORF1ab	None (synonymous mutation)
A, B	14408	C	T	ORF1ab	P4715L
B	18744	C	T	ORF1ab	None (synonymous mutation)
A	21907	G	A	S	None (synonymous mutation)
B	22801	G	T	S	None (synonymous mutation)
A, B	23403	A	G	S	D614G
A, B	23604	C	G	S	P681R
B	27674	A	T	ORF7a	Q94L
A, B	28881	G	A	N	R203K
A, B	28882	G	A	N	R203K
A, B	28883	G	C	N	G204R

mutations that were also found in both studied genomes—I300F and P4715L in the ORF1ab gene, D614G in the S gene, and R203K and G204R in the N gene (13).

As more genomes of COVID-19 patients with various clinical features from various regions of the country are being sequenced on a regular basis, the evolution of the virus must be further investigated and closely monitored in order to track COVID-19 progression in Bangladesh.

Data availability. The SARS-CoV-2 genome sequences from Bangladesh were deposited in the GISAID database (accession no. EPI_ISL_884087 and EPI_ISL_884088), and the raw reads have been submitted to the NCBI Sequence Read Archive (SRA). The BioProject accession no. is [PRJNA701872](https://www.ncbi.nlm.nih.gov/bioproject/PRJNA701872) for both samples. The BioSample and SRA accession no. for the samples are [SAMN17915311](https://www.ncbi.nlm.nih.gov/biosample/SAMN17915311) and [SAMN17915310](https://www.ncbi.nlm.nih.gov/biosample/SAMN17915310) (BioSample) and [SRR13722033](https://www.ncbi.nlm.nih.gov/sra/SRR13722033) and [SRR13721921](https://www.ncbi.nlm.nih.gov/sra/SRR13721921) (SRA).

ACKNOWLEDGMENT

We thank Jahangirnagar University, Dhaka, Bangladesh, for providing the necessary funding and assistance to conduct the study.

REFERENCES

- Gorbalenya AE, Baker SC, Baric RS, de Groot RJ, Drosten C, Gulyaeva AA, Haagmans BL, Lauber C, Leontovich AM, Neuman BW, Penzar D, Perlman S, Poon LLM, Samborskiy DV, Sidorov IA, Sola I, Ziebuhr J, Coronaviridae Study Group of the International Committee on Taxonomy of Viruses. 2020. The species Severe acute respiratory syndrome-related coronavirus: classifying 2019-nCoV and naming it SARS-CoV-2. *Nat Microbiol* 5:536–544. <https://doi.org/10.1038/s41564-020-0695-z>.
- Zhou P, Yang X-L, Wang X-G, Hu B, Zhang L, Zhang W, Si H-R, Zhu Y, Li B, Huang C-L, Chen H-D, Chen J, Luo Y, Guo H, Jiang R-D, Liu M-Q, Chen Y, Shen X-R, Wang X, Zheng X-S, Zhao K, Chen Q-J, Deng F, Liu L-L, Yan B, Zhan F-X, Wang Y-Y, Xiao G-F, Shi Z-L. 2020. A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* 579:270–273. <https://doi.org/10.1038/s41586-020-2012-7>.
- Dong E, Du H, Gardner L. 2020. An interactive Web-based dashboard to track COVID-19 in real time. *Lancet Infect Dis* 20:533–534. [https://doi.org/10.1016/S1473-3099\(20\)30120-1](https://doi.org/10.1016/S1473-3099(20)30120-1).
- Nahid AA, Ghosh A. 2021. Investigating the possible origin and transmission routes of SARS-CoV-2 genomes and variants of concern in Bangladesh. *bioRxiv* 2021.05.24.444482. <https://doi.org/10.1101/2021.05.24.444482>.
- Sansure Biotech, Inc. 2020. Novel coronavirus (2019-nCoV) nucleic acid diagnostic kit (PCR-fluorescence probing). Sansure Biotech, Inc, Changsha, Hunan Province, China. <https://www.fda.gov/media/137651/download>.
- Andrews S. 2010. FastQC: a quality control tool for high throughput sequence data. <https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>.
- Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30:2114–2120. <https://doi.org/10.1093/bioinformatics/btu170>.
- Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25:1754–1760. <https://doi.org/10.1093/bioinformatics/btp324>.
- Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, Whitwham A, Keane T, McCarthy SA, Davies RM, Li H. 2021. Twelve years of SAMtools and BCFtools. *GigaScience* 10:giab008. <https://doi.org/10.1093/gigascience/giab008>.
- Shu Y, McCauley J. 2017. GISAID: Global Initiative on Sharing All Influenza Data: from vision to reality. *Eurosurveillance* 22:30494. <https://doi.org/10.2807/1560-7917.ES.2017.22.13.30494>.
- Hadfield J, Megill C, Bell SM, Huddleston J, Potter B, Callender C, Sagulenko

- P, Bedford T, Neher RA. 2018. Nextstrain: real-time tracking of pathogen evolution. *Bioinformatics* 34:4121–4123. <https://doi.org/10.1093/bioinformatics/bty407>.
12. Vilsker M, Moosa Y, Nooij S, Fonseca V, Ghysens Y, Dumon K, Pauwels R, Alcantara LC, Vanden Eynden E, Vandamme A-M, Deforche K, de Oliveira T. 2019. Genome Detective: an automated system for virus identification from high-throughput sequencing data. *Bioinformatics* 35:871–873. <https://doi.org/10.1093/bioinformatics/bty695>.
13. Hasan MM, Das R, Rasheduzzaman M, Hussain MH, Muzahid NH, Salauddin A, Rumi MH, Mahbubur Rashid SM, Siddiki AZ, 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>.