



Genome Sequence of a SARS-CoV-2 Strain from a COVID-19 Clinical Sample from the Khagrachari District of Bangladesh

M. Imranul Hoq,^a Robiul Hasan Bhuiyan,^b M. Khondakar Raziur Rahman,^c Imam Hossen,^c Sajib Rudra,^c M. Arif Hossain,^c Shanta Paul,^a M. Omer Faruq,^b Mohammad Omar Faruque,^c  H. M. Abdullah Al Masud^a

^aDepartment of Microbiology, Faculty of Biological Sciences, University of Chittagong, Chattogram, Bangladesh

^bDepartment of Biochemistry and Molecular Biology, Faculty of Biological Sciences, University of Chittagong, Chattogram, Bangladesh

^cDepartment of Botany, Faculty of Biological Sciences, University of Chittagong, Chattogram, Bangladesh

M. Imranul Hoq and Robiul Hasan Bhuiyan contributed equally to this work. Author order was determined by alphabetic order.

ABSTRACT This study describes the genome sequence of a severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) strain detected in the nasopharyngeal swab sample of a coronavirus disease 2019 (COVID-19) patient from the southeastern Khagrachari District of Bangladesh.

The ongoing pandemic of coronavirus disease 2019 (COVID-19), which was first reported in Wuhan, China, in December 2019 (1), is caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), a *Betacoronavirus* in the *Coronaviridae* family (2). In Bangladesh, the first COVID-19 case was detected on 8 March 2020 (3). We report here the genome sequence of SARS-CoV-2 from a 43-year-old male from the Khagrachari District of Bangladesh, who was hospitalized with fever, joint ache, diarrhea and difficulty breathing, and tested positive for COVID-19 by reverse transcriptase PCR (RT-PCR) (4). All the protocols were approved by the Ethical Review Board of the University of Chittagong (reference no. CUBIO0001). Informed consent was obtained from the patient, and the sample was collected with the permission of the Directorate General of Health Services of the Government of Bangladesh.

Viral RNA was extracted from the nasopharyngeal swab using a PureLink viral DNA/RNA minikit (catalog no. 12280050; Thermo Fisher Scientific). cDNA synthesis and library preparation were carried out using an Illumina RNA prep with enrichment (L) tagmentation kit (catalog no. 20040537) combined with Illumina respiratory virus oligonucleotide panel v2 (catalog no. 20044311), and the prepared libraries were sequenced on an Illumina MiniSeq system in the paired-end format (read length, 74 bp) according to the manufacturer's instructions. Duplicates and low-quality reads were removed, and coverage plots were created using Illumina DRAGEN RNA pathogen detection version 3.5.15.

The library generated a total of 3,625,976 reads, of which 2,165,542 reads mapped to the reference sequence (GenBank accession no. [MN908947.3](https://doi.org/10.1128/MRA.00189-21)) using human (hg38) and the Illumina respiratory virus panel, with the human control option of the DRAGEN software, and 1,709,268 reads were found unique after exclusion of the duplicates. The FASTQ data files were exported from the Illumina local run manager to the BaseSpace Sequence Hub; a consensus FASTA file was generated using k-mer analysis (GenBank accession no. [NC_045512.2](https://doi.org/10.1128/MRA.00189-21)) of the DRAGEN software; and it was revealed that the genome of this strain has 29,856 bp which starts and ends at positions 7 and 29862, respectively, of the reference sequence (29,903 bp). The whole-genome comparison, using the DRAGEN software, of the strain revealed 99.85% identity, with the reference sequence having a mean coverage depth of

Citation Hoq MI, Bhuiyan RH, Rahman MKR, Hossen I, Rudra S, Hossain MA, Paul S, Faruq MO, Faruque MO, Masud HMAA. 2021. Genome sequence of a SARS-CoV-2 strain from a COVID-19 clinical sample from the Khagrachari District of Bangladesh. *Microbiol Resour Announc* 10:e00189-21. <https://doi.org/10.1128/MRA.00189-21>.

Editor Simon Roux, DOE Joint Genome Institute

Copyright © 2021 Hoq 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 Mohammad Omar Faruque, omf@cu.ac.bd, or H. M. Abdullah Al Masud, masud.mbio@cu.ac.bd.

Received 19 February 2021

Accepted 3 March 2021

Published 1 April 2021

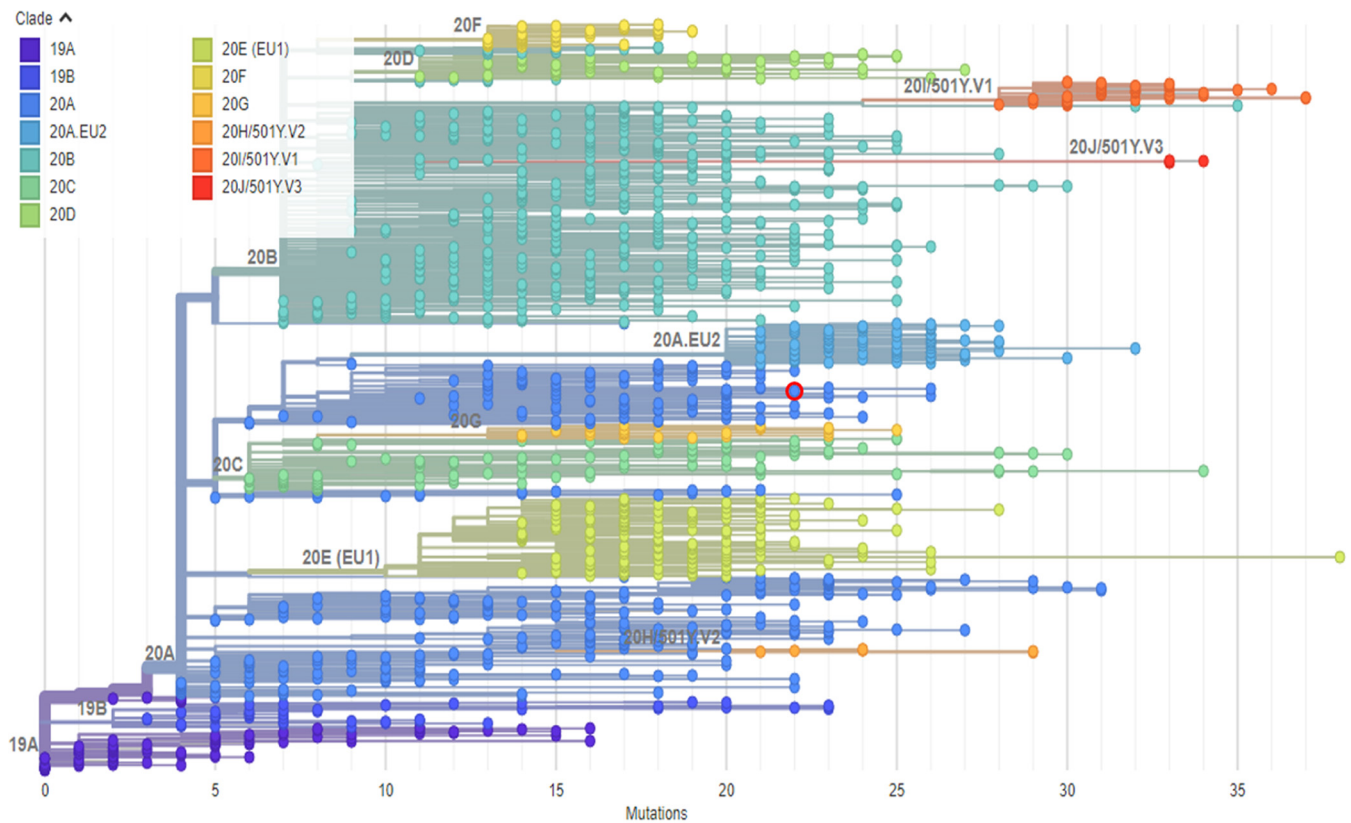


FIG 1 Phylogenetic tree of a SARS-CoV-2 strain from the Khagrachari District of Bangladesh. The tree was constructed on 19 February 2021 using the Nextclade^{beta} version 0.13.0 (clades.nextstrain.org), in which the red circle represents the position of hCoV-19/Bangladesh/CU-CTG-24/2020 (GISAID accession no. EPI_ISL_735496).

303×, whereas no indel was detected. The sequence displays a GC content of 38%. The consensus genome and related sample data were uploaded to the Global Initiative on Sharing All Influenza Data (GISAID) database (accession no. EPI_ISL_735496) on 25 December 2020 (5). Phylogenetic analysis using Nextclade^{beta} version 0.13.0 (clades.nextstrain.org) assigned the new genome to the SARS-CoV-2 clade 20A (Fig. 1) (6). According to the GISAID database basic local alignment search tool (BLAST), the genome shares the highest levels of similarity with sequences uncovered from Saudi Arabia (GISAID accession no. EPI_ISL_678004, EPI_ISL_513151, EPI_ISL_513149, EPI_ISL_437736, and EPI_ISL_437723) and India (GISAID accession no. EPI_ISL_1073009, EPI_ISL_1073011, EPI_ISL_1073010, and EPI_ISL_1073014) (5).

An analysis of the variations, using Genome Detective Virus Tools version 1.132 (7), indicates several changes in the sequence of this strain exhibiting 11 synonymous and 11 nonsynonymous mutations relative to the Wuhan-Hu-1 reference sequence (GenBank accession no. [NC_045512.2](https://www.ncbi.nlm.nih.gov/nuccore/NC_045512.2)) (Table 1). According to the GISAID database, 2 mutations, namely, P681H and V1122L, of the spike glycoprotein of this virus were rare among the SARS-CoV-2 strains recovered in Bangladesh (5). As of 19 February 2021, the mutation P681H was also observed in 5 other strains of SARS-CoV-2 recovered in Bangladesh (GISAID accession no. EPI_ISL_906098, EPI_ISL_906091, EPI_ISL_890237, EPI_ISL_890188, and EPI_ISL_774976), whereas the mutation V1122L is still unique in Bangladesh.

Data availability. The sequence has been deposited in the GISAID database (accession no. EPI_ISL_735496) and GenBank (accession no. [MW599343](https://www.ncbi.nlm.nih.gov/nuccore/MW599343)). The accession number for the raw sequence reads in the NCBI Sequence Read Archive (SRA) is [SRR13718002](https://www.ncbi.nlm.nih.gov/sra/SRR13718002). The BioProject and BioSample accession numbers are [PRJNA701790](https://www.ncbi.nlm.nih.gov/bioproject/PRJNA701790) and [SAMN17911680](https://www.ncbi.nlm.nih.gov/biosample/SAMN17911680), respectively.

TABLE 1 Mutations observed in hCoV-19/Bangladesh/CU-CTG-24/2020 compared with SARS-CoV-2 isolate Wuhan-Hu-1^a

Nucleotide position	Reference nucleotide	Mutated nucleotide	Gene	Amino acid change
241	C	T	5'-UTR ^b	Noncoding
1006	G	T	ORF1ab	K247N
1853	C	T	ORF1ab	None (synonymous mutation)
2836	C	T	ORF1ab	None (synonymous mutation)
3037	C	T	ORF1ab	None (synonymous mutation)
4331	C	T	ORF1ab	None (synonymous mutation)
4755	C	T	ORF1ab	P1497L
6472	C	T	ORF1ab	None (synonymous mutation)
7119	C	T	ORF1ab	S2285F
7247	T	C	ORF1ab	F2328L
14408	C	T	ORF1ab	P4715L
17056	A	G	ORF1ab	M5598V
18877	C	T	ORF1ab	None (synonymous mutation)
22444	C	T	S	None (synonymous mutation)
23403	A	G	S	D614G
23604	C	A	S	P681H
24130	C	T	S	None (synonymous mutation)
24926	G	T	S	V1122L
25563	G	T	ORF3a	Q57H
26735	C	T	M	None (synonymous mutation)
28854	C	T	N	S194L
29260	G	T	N	None (synonymous mutation)

^a GenBank accession no. [NC_045512.2](https://www.ncbi.nlm.nih.gov/nuclot/NC_045512.2).^b UTR, untranslated region.

ACKNOWLEDGMENTS

We thank the Directorate General Health Services of the Government of Bangladesh for the permission for sample collection and Chittagong Veterinary and Animal Sciences University, Bangladesh, for providing the sample along with a test report and patient history. We acknowledge tremendous laboratory support from COVID-19 Testing Laboratory, Central Biological Research Laboratory, and Department of Biochemistry and Molecular Biology of the Faculty of Biological Sciences, University of Chittagong, Bangladesh. We also thank Invent Technologies Ltd., Bangladesh for excellent technical assistance.

This study was financed by the Research and Publication Cell of the University of Chittagong, Bangladesh.

REFERENCES

- Zhou P, Yang X, Wang X, Hu B, Zhang L, Zhang W, Si H, Zhu Y, Li B, Huang C, Chen H, Chen J, Luo Y, Guo H, Jiang R, Liu M, Chen Y, Shen X, Wang X, Zheng X, Zhao K, Chen Q, Deng F, Liu L, Yan B, Zhan F, Wang Y, Xiao G, Shi Z. 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>.
- 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>.
- Hasan K, Shaon AI. 8 March 2020. The first 3 cases of coronavirus confirmed in Bangladesh. *Dhaka Tribune*, Dhaka, Bangladesh. <https://www.dhakatribune.com/health/coronavirus/2020/03/08/iedcr-3-affected-with-coronavirus-in-bangladesh>.
- Sansure Biotech, Inc. 2020. Novel coronavirus (2019-nCoV) nucleic acid diagnostic kit (PCR-fluorescence probing). Sansure Biotech, Inc., Yuelu District, Changsha, Hunan Province, People's Republic of China. <https://www.fda.gov/media/137651/download>.
- Shu Y, McCauley J. 2017. GISAID: Global Initiative on Sharing All Influenza Data: from vision to reality. *Euro Surveill* 22:30494. <https://doi.org/10.2807/15607917.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>.
- 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>.