AMERICAN SOCIETY FOR MICROBIOLOGY | Microbiology Resource Announcements

GENOME SEQUENCES

Genome Sequence of a SARS-CoV-2 P.1 Variant of Concern (20J/501Y.V3) from Bangladesh

M. Murshed Hasan Sarkar,[a] Mohammad Fazle Alam Rabbi,[b,c] Shahina Akter,[a] Tanjina Akhtar Banu,[a] Barna Goswami,[a] Iffat Jahan,[a] M. Saddam Hossain,[a] Eshrar Osman,[a] Mohammad Samir Uzzaman,[d] M. Ahashan Habib,[a] Abu Sayeed Mohammad Mahmud,[a] Firoz Kabir,[b] Kazi Nadim Hasan,[b,e] M. Mizanur Rahman,[b,f] M. Abdul Khaleque,[b,e] Sharif Akhteruzzaman,[b,g] M. Salim Khan[a]

[a]Bangladesh Council of Scientific & Industrial Research (BCSIR), Dhaka, Bangladesh
[b]DNA Solution Ltd., Dhaka, Bangladesh
[c]Department of Soil, Water and Environment, University of Dhaka, Dhaka, Bangladesh
[d]SciTech Consulting and Solutions, Dhaka, Bangladesh
[e]Department of Biochemistry and Microbiology, School of Health and Life Sciences, North South University, Dhaka, Bangladesh
[f]NIPRO JMI Pharma, Dhaka, Bangladesh
[g]Department of Genetic Engineering and Biotechnology, University of Dhaka, Dhaka, Bangladesh

M. Murshed Hasan Sarkar and Mohammad Fazle Alam Rabbi contributed equally to this article. Author order was determined by the alphabetical order of affiliated institutes.

ABSTRACT This study reports the coding-complete genome sequence, with variant identifications and phylogenetic analysis, of a severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) P.1 variant (20J/501Y.V3), obtained from an oropharyngeal swab specimen from a female Bangladeshi patient diagnosed with coronavirus disease 2019 (COVID-19) with no travel history.

Since March 2021, the confirmed number of positive COVID-19 cases and the associated death toll have been increasing in Bangladesh compared to late 2020 and early 2021 national surveillance data (1). Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is the etiological agent of COVID-19 and belongs to the family *Coronaviridae* and the genus *Betacoronavirus*. Variants of the virus have been reported already, among them variants of great concern (B.1.1.7, 20I/501Y.V1, and B.1.351 [20H/501Y.V2]) due to their possible increased transmissibility and demonstrated resistance to neutralizing antibody therapy targeted at the N-terminal domain (NTD) or the receptor-binding domain (RBD) (2). Variant P.1 (20J/501Y.V3) is also a variant of concern and shares three substitution mutations (K417T, E484K, and N501Y) at the RBD site similar to the B.1.351 (20H/501Y.V2) variant. Here, we report the coding-complete genome sequence of a SARS-CoV-2 P.1 variant (20J/501Y.V3) from Bangladesh. Ethical approval was received from the National Institute of Laboratory Medicine and Referral Center (NILMRC) of the Bangladesh government to sequence strains from human subjects at the Genomic Research Laboratory of BCSIR, where this study was carried out.

A patient with a suspected COVID-19 case with a fever and respirational distress visited the outdoor patient unit of DNA Solution Ltd. on 18 February 2021. The 37-year-old female patient had a history of asthma with no other known comorbidities. The patient had no travel history but had come into contact with a traveler from South Africa. An oropharyngeal swab specimen was collected and was tested for SARS-CoV-2 RNA using a commercial one-step COVID-19 real-time PCR kit (Sansure Biotech Inc., Changsha, China) following the manufacturer's instructions. The sample tested positive for SARS-CoV-2 RNA with a cycle threshold (C_T) value of less than 30. The sample was then sent out for sequencing at the Genomic Research Laboratory of the Bangladesh Council of Scientific and Industrial Research (BCSIR). There, viral RNA was extracted using the ReliaPrep viral total nucleic acid purification kit (Promega, USA). Viral cDNA libraries were prepared using the RNA prep

Citation Sarkar MMH, Rabbi MFA, Akter S, Banu TA, Goswami B, Jahan I, Hossain MS, Osman E, Uzzaman MS, Habib MA, Mahmud ASM, Kabir F, Hasan KN, Rahman MM, Khaleque MA, Akhteruzzaman S, Khan MS. 2021. Genome sequence of a SARS-CoV-2 P.1 variant of concern (20J/501Y.V3) from Bangladesh. Microbiol Resour Announc 10:e00524-21. https://doi.org/10.1128/MRA.00524-21.

Editor Simon Roux, DOE Joint Genome Institute

Copyright © 2021 Sarkar et al. This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International license.

Address correspondence to M. Salim Khan, k2salim@yahoo.com.

Received 20 May 2021
Accepted 8 June 2021
Published 8 July 2021

Downloaded from https://journals.asm.org/journal/mra on 09 July 2023 by 220.247.165.130.

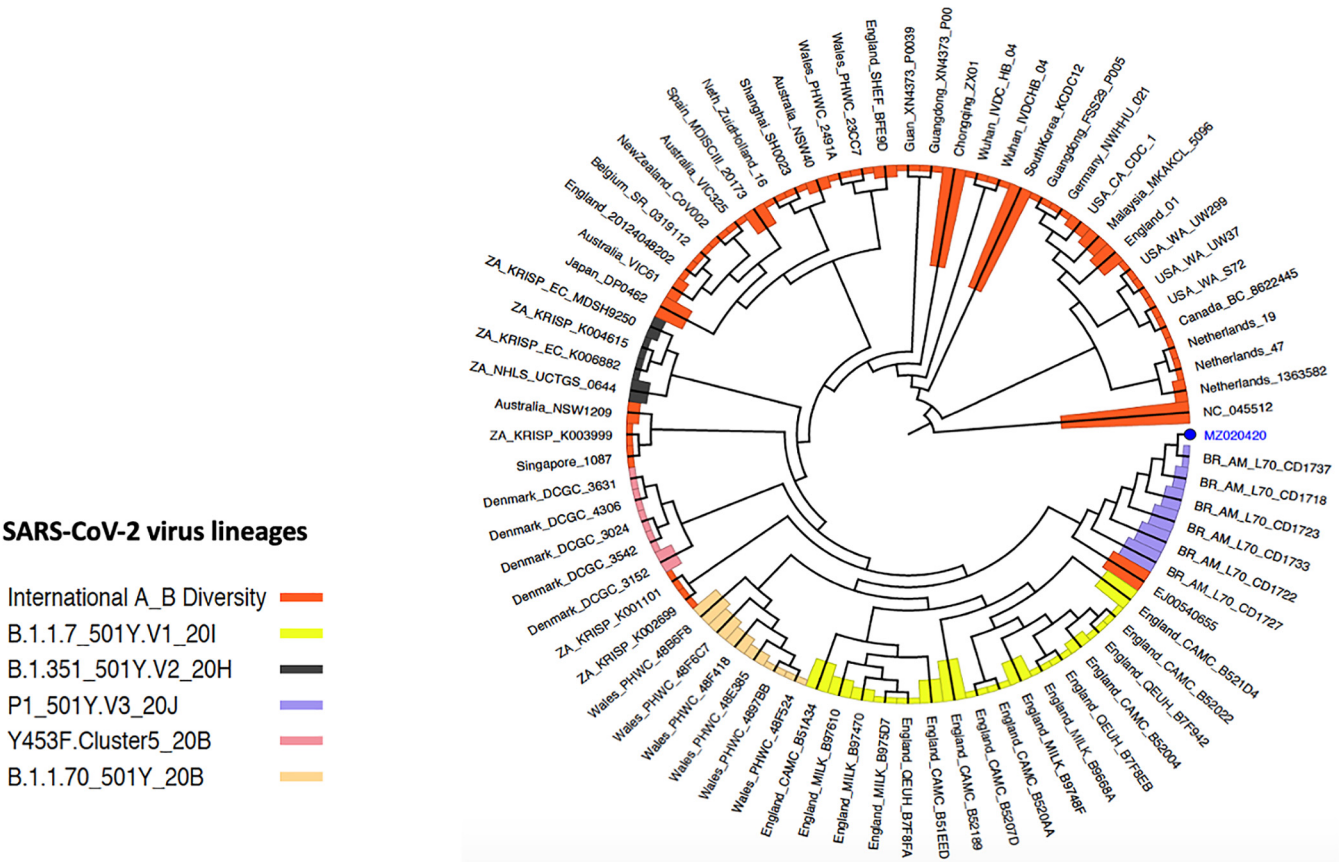

FIG 1 Schematic polar diagram displaying the phylogenetic tree of the strain found under the GenBank accession number MZ020420. The FASTA file was uploaded to the Genome Detective Coronavirus Typing Tool version 1.17. The toolkit uses BLAST and phylogenetic methods to assign a lineage to the virus sequence according to a method described by Cleemput et al. (7). The genome of interest is of the P.1 (20J/501Y.V3) lineage.

with enrichment (L) tagmentation kit according to the manufacturer's protocol (Illumina, CA, USA). After tagmentation and amplification, the sample was enriched using the Illumina respiratory virus oligonucleotide panel version 2, and a library was generated. The library was sequenced using the MiniSeq sequencing system with an output of 2 × 74-bp paired-end reads. A total of 4,831,986 reads were generated, of which 4,753,260 (98.37%) belonged to the SARS-CoV-2 genome. The sequenced reads of SARS-CoV-2 genomes were assembled using Illumina DRAGEN RNA Pathogen Detection software version 3.5.14 with default settings. The sequencing data were deposited in the NCBI GenBank database (accession number MZ020420).

Further, genome assembly of the raw data was performed using the assembly toolkit in the EzCOVID19 cloud service (3), provided on the EzBioCloud website (4). The assembly was performed by aligning reads to the Wuhan predefined reference genome (GenBank accession number NC_045512.2). The genome sequence consisted of 29,789 nucleotides and covered 99.62% of the SARS-CoV-2 Wuhan reference genome (NC_045512.2), with a GC content of 38%. A consensus genome was created to retain the unique variations of the raw data using the EzCOVID19 cloud service (3, 5). The consensus genome was then compared against the same reference genome to calculate single nucleotide variations (SNVs) and positions. The SNVs were compared against GISAID clade variation markers (6). The genome presented here belonged to the P.1 (20J/501Y.V3) lineage of clade GR. This was further verified by analyzing the raw data using the default parameters of the Coronavirus Typing Tool offered by Genome Detective (7) (Fig. 1).

Data availability. The SARS-CoV-2 sequence was deposited in GenBank under the accession number MZ020420. The raw sequence reads have been deposited in the NCBI Sequence Read Archive (SRA) under the accession number SRX11071271 and the BioProject accession number PRJNA733209.

Downloaded from https://journals.asm.org/journal/mra on 09 July 2023 by 220.247.165.130.

ACKNOWLEDGMENTS

We acknowledge the brave phlebotomists and technicians of DNA Solution Ltd. and BCSIR, who are serving humankind during this challenging pandemic situation. We also thank the patient who allowed us to sequence the viral genome from her clinical sample.

REFERENCES

1. Bangladesh Directorate General of Health Services, Ministry of Health and Family Welfare. Coronavirus COVID-19 dashboard, 2021. http://103.247.238.92/webportal/pages/covid19.php.
2. Wang P, Nair MS, Liu L, Iketani S, Luo Y, Guo Y, Wang M, Yu J, Zhang B, Kwong PD, Graham BS, Mascola JR, Chang JY, Yin MT, Sobieszczyk M, Kyratsous CA, Shapiro L, Sheng Z, Huang Y, Ho DD. 2021. Antibody resistance of SARS-CoV-2 variants B.1.351 and B.1.1.7. Nature 593:130–135. https://doi.org/10.1038/s41586-021-03398-2.
3. EzCOVID19. https://eztaxon-e.ezbiocloud.net/tools/sc2?id=5bbb155a-dd3e-4e9b-9b10-54b4e67926bd.
4. Yoon S-H, Ha S-M, Kwon S, Lim J, Kim Y, Seo H, Chun J. 2017. Introducing EzBioCloud: a taxonomically united database of 16S rRNA gene sequences and whole-genome assemblies. Int J Syst Evol Microbiol 67:1613–1617. https://doi.org/10.1099/ijsem.0.001755.
5. Mercatelli D, Giorgi FM. 2020. Geographic and genomic distribution of SARS-CoV-2 mutations. Front Microbiol 11:1800. https://doi.org/10.3389/fmicb.2020.01800.
6. GISAID. Clade and lineage nomenclature aids in genomic epidemiology of active hCoV-19 viruses. https://www.gisaid.org/references/statements-clarifications/clade-and-lineage-nomenclature-aids-in-genomic-epidemiology-of-active-hcov-19-viruses/.
7. Cleemput S, Dumon W, Fonseca V, Abdool Karim W, Giovanetti M, Alcantara LC, Deforche K, de Oliveira T. 2020. Genome detective coronavirus typing tool for rapid identification and characterization of novel coronavirus genomes. Bioinformatics 36:3552–3555. https://doi.org/10.1093/bioinformatics/btaa145.

Downloaded from https://journals.asm.org/journal/mra on 09 July 2023 by 220.247.165.130.