



# SARS-CoV-2 Genome Sequence Obtained from Ethiopia

 Molalegne Bitew,<sup>a</sup> Getnet Hailu,<sup>b</sup> Yakob Gebregziabher Tsegay,<sup>c</sup> Keyru Tuki,<sup>a</sup> Kominist Asmamaw,<sup>a</sup> Kassahun Tesfaye,<sup>a</sup> Hailu Dadi,<sup>a</sup> Emanuele Orsini,<sup>e</sup> Simeone Dal Monego,<sup>d</sup> Danilo Licastro,<sup>d</sup>  Alessandro Marcello<sup>e</sup>

<sup>a</sup>Ethiopia Biotechnology Institute (EBTI), Addis Ababa, Ethiopia

<sup>b</sup>Addis Ababa University, Aklilu Lemma Institute of Pathobiology (ALPB), Addis Ababa, Ethiopia

<sup>c</sup>Department of Medical Biotechnology, Institute of Biotechnology, University of Gondar, Gondar, Ethiopia

<sup>d</sup>AREA Science Park Padriciano, Trieste, Italy

<sup>e</sup>Laboratory of Molecular Virology, International Centre for Genetic Engineering and Biotechnology (ICGEB), Trieste, Italy

**ABSTRACT** The coding-complete severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) genome sequences from 15 nasopharyngeal swabs collected in Addis Ababa, Ethiopia, during the period from December 2020 to March 2021 were determined using Illumina MiSeq technology. A sequence analysis identified that the B.1 SARS-CoV-2 lineage was most prevalent with the worrying emergence of B.1.1.7 in June 2021.

The severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) outbreak began in Wuhan, Hubei Province, China, in late December 2019 and spread worldwide as the worst viral pandemic in recent times (1). SARS-CoV-2 is the aetiological agent of COVID-19, which could lead to severe pneumonia (2). SARS-CoV-2 is an enveloped virus belonging to the *Betacoronavirus* genus of the *Coronaviridae* family of RNA viruses with a single-stranded genome of positive polarity. Since the first sequence of the SARS-CoV-2 complete genome was reported in early December 2019, an exponentially increasing number of virus sequences has been reported, with over 2,612,459 complete genome sequences stored in the Global initiative on sharing all influenza data (GISAID) database to date (accessed on 8 April 2021) (3).

Ethiopia in sub-Saharan East Africa has had 306,810 confirmed COVID-19 cases and 4,675 deaths out of the total of 3,239,416 individuals tested by 4 September 2021. There is a strong demand to know whether variants of concern are circulating in this country to inform authorities and to guide public health interventions (4). However, only five sequences from Ethiopia were available on public databases from December 2020 to March 2021, reinforcing the urgent need to increase molecular surveillance (5).

A total of 15 nasopharyngeal swab samples were collected from COVID-19 patients and asymptomatic individuals in Addis Ababa, Ethiopia. Samples were collected from the period of December 2020 to March 2021 at different hospitals, and reverse transcription-quantitative PCR (qRT-PCR) detection was carried out at the Ethiopia Biotechnology Institute (EBTI). Nucleic acids were extracted using an RNA/DNA purification kit (Daan Gene Co., Ltd. of Sun Yat-sen University, Guangzhou, China) according to the manufacturer's protocol and then treated with DNase. RNA was then suspended in RNase-free water, and then the RNA extracts were purified using an RNA clean and concentrator kit (Zymo Research, CA, USA) (6). The diagnostic detection of SARS-CoV-2 RNA was performed with the Beijing Genomic Institute (BGI) real-time fluorescent reverse transcriptase PCR (RT-qPCR) detection kit of SARS-CoV-2 (BGI, China) with primers targeting the ORF 1ab gene (5). A total of 15 samples with cycle threshold ( $C_T$ ) values of  $<30$  were selected randomly for high-throughput genome sequencing. This technique is based mainly on PCR amplification of the genome (7). A Qubit 2.0 fluorimeter (Thermo Fisher Scientific, MA, USA) and Agilent 2100 bioanalyzer (Agilent Technologies,

**Editor** Simon Roux, DOE Joint Genome Institute

**Copyright** © 2022 Bitew 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 Molalegne Bitew, molalegne23@yahoo.com.

The authors declare no conflict of interest.

**Received** 15 December 2021

**Accepted** 3 January 2022

**Published** 3 February 2022

**TABLE 1** Sequencing parameters and phylogenetic traits of sequenced genomes

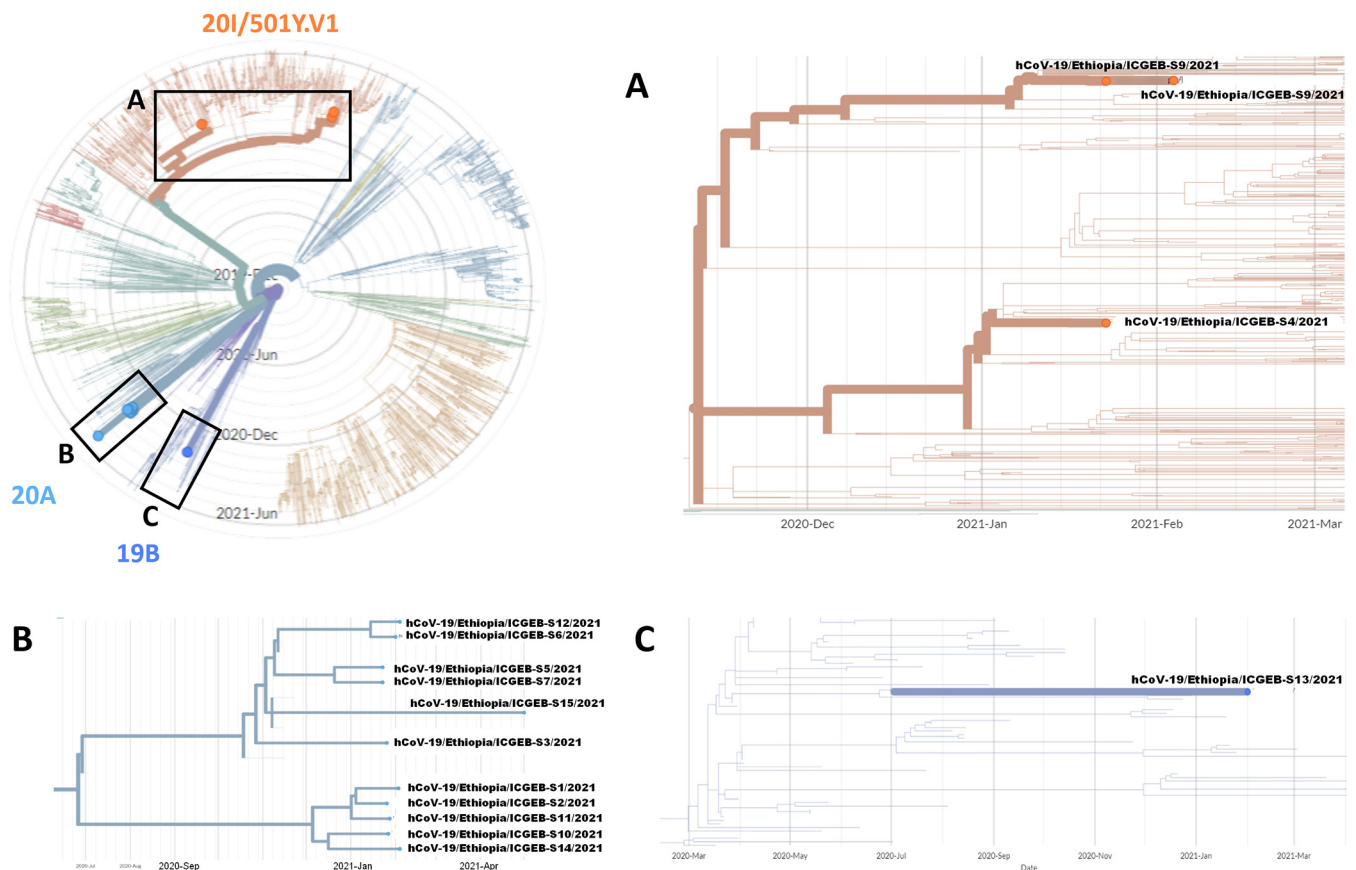
| Sample ID                      | GISAID ID       | Accession no.            | SRA accession no.           | GC content (%) | Complete length (bp) | SARS-CoV-2 coverage (x) | No. of raw reads (millions) | Amino acid substitutions  | Nexstrain clade | PANGO <sup>a</sup> lineage |
|--------------------------------|-----------------|--------------------------|-----------------------------|----------------|----------------------|-------------------------|-----------------------------|---|-----------------|----------------------------|
| hCoV-19/Ethiopia/ICGEB-51/2021 | EPI_ISL_1897644 | <a href="#">OL840802</a> | <a href="#">SRX13137980</a> | 42.6           | 29,878               | 21,793                  | 3.6                         | Spike D614G, Spike E180Q, Spike M1229I, Spike N501Y, N D377Y, N S194L, NS3 Q57H, NSP2 A31V, NSP2 L24F, NSP3 G1001S, NSP3 T350I, NSP3 V521I, NSP5 S284G, NSP12 L648F, NSP12 P323L, NSP13 T588A, NSP13 Y277C, NSP15 K70R  | 20A             | B.1.480                    |
| hCoV-19/Ethiopia/ICGEB-52/2021 | EPI_ISL_1897646 | <a href="#">OL840803</a> | <a href="#">SRX13137981</a> | 43.3           | 29,878               | 21,521                  | 3.8                         | Spike D614G, Spike E180Q, Spike E309Q, Spike M1229I, Spike N501Y, ND377Y, N S194L, NS3 Q57H, NSP2 L24F, NSP3 K1182R, NSP3 S769L, NSP3 V521I, NSP5 S284G, NSP6 A186T, NSP12 L648F, NSP12 P323L, NSP13 P82S, NSP13 Y277C  | 20A             | B.1.480                    |
| hCoV-19/Ethiopia/ICGEB-53/2021 | EPI_ISL_1897648 | <a href="#">OL840804</a> | <a href="#">SRX13137982</a> | 43.3           | 29,878               | 21,621                  | 3.8                         | Spike D614G, N P199L, NS3 Q57H, NSP2 E264K, NSP2 F499L, NSP12 P323L, NSP12 T26I, NSP13 E341D  | 20A             | B.1                        |
| hCoV-19/Ethiopia/ICGEB-54/2021 | EPI_ISL_1897884 | <a href="#">OL840805</a> | <a href="#">SRX13137983</a> | 43.5           | 29,877               | 18,606                  | 3.2                         | Spike A570D, Spike D614G, Spike D1118H, Spike N501Y, Spike P681H, Spike S982A, Spike T716I, ND3L, N G204R, N R203K, N S235F, NS3 V163L, NS8 Q27stop, NS8 R52I, NS8 Y73C, NSP1 E2K, NSP3 A890D, NSP3 G307C, NSP3 I1412T, NSP3 P141S, NSP3 T183I, NSP10 T12I, NSP12 P323L | 20I/501Y.V1     | B.1.1.7                    |
| hCoV-19/Ethiopia/ICGEB-55/2021 | EPI_ISL_1898258 | <a href="#">OL840806</a> | <a href="#">SRX13137984</a> | 42.8           | 29,871               | 21,735                  | 3.8                         | Spike D614G, Spike N440K, Spike S13T, N P199L, NS3 Q57H, NS3 S171L, NSP2 F499L, NSP12 P323L   | 20A             | B.1                        |
| hCoV-19/Ethiopia/ICGEB-56/2021 | EPI_ISL_1898554 | <a href="#">OL840807</a> | <a href="#">SRX13137985</a> | 42             | 29,878               | 22,185                  | 3.6                         | Spike A771S, Spike D614G, Spike S939F, N P199L, NS3 Q57H, NS3 S171L, NSP2 C487S, NSP2 F499L, NSP3 D163G, NSP3 R352H, NSP12 P323L  | 20A             | B.1                        |
| hCoV-19/Ethiopia/ICGEB-57/2021 | EPI_ISL_1898814 | <a href="#">OL840808</a> | <a href="#">SRX13137986</a> | 43.7           | 29,878               | 22,654                  | 4.0                         | Spike A520S, Spike D614G, N P199L, NS3 A110S, NS3 Q57H, NS3 S171L, NSP2 F499L, NSP5 K90R, NSP12 P323L   | 20A             | B.1.404                    |
| hCoV-19/Ethiopia/ICGEB-58/2021 | EPI_ISL_1899047 | <a href="#">OL840809</a> | <a href="#">SRX13137987</a> | 43.7           | 29,874               | 18,443                  | 3.0                         | Spike A570D, Spike D614G, Spike D1118H, Spike N501Y, Spike P330L, Spike P681H, Spike S982A, Spike T716I, Spike Y144del, N D3L, N G204R, N R203K, N S235F, NS3 T89I, NS8 K68stop, NS8 Q27stop,   | 20I/501Y.V1     | B.1.1.7                    |

(Continued on next page)

TABLE 1 (Continued)

| Sample ID                       | GISAID ID       | Accession no. | SRA accession no.           | GC content (%) | Complete length (bp) | SARS-CoV-2 coverage (x) | No. of raw reads (millions) | Amino acid substitutions   | Nexstrain clade | PANGO <sup>a</sup> lineage |
|---------------------------------|-----------------|---------------|-----------------------------|----------------|----------------------|-------------------------|-----------------------------|--|-----------------|----------------------------|
| hCoV-19/Ethiopia/ICGEB-59/2021  | EPI_ISL_1899281 | OL840810      | <a href="#">SRX13137988</a> | 43             | 29,877               | 22,240                  | 3.8                         | NS8 R52I, NS8 Y73C, NSP2 A510V, NSP3 A890D, NSP3 I1412T, NSP3 T183I, NSP12 P227L, NSP12 P323L, NSP14 P451S<br>Spike A570D, Spike D614G, Spike D1118H, Spike N501Y, Spike P681H, Spike S982A, Spike T716I, N D3L, N G204R, N R203K, N S235F, NS3 T89I, NS8 K68stop, NS8 Q27stop, NS8 R52I, NS8 Y73C, NSP3 A890D, NSP3 I1412T, NSP3 T183I, NSP12 P227L, NSP12 P323L, NSP14 P451S | 20I/501Y.V1     | B.1.1.7                    |
| hCoV-19/Ethiopia/ICGEB-S10/2021 | EPI_ISL_1899485 | OL840811      | <a href="#">SRX13137989</a> | 45.7           | 29,878               | 19,306                  | 3.4                         | Spike D614G, Spike G142A, Spike M1229I, Spike T1006I, N S194L, NS3 Q57H, NS8 E59stop, NSP2 L24F, NSP2 M117V, NSP5 S284G, NSP12 E919D, NSP12 P323L, NSP13 L297F   | 20A             | B.1                        |
| hCoV-19/Ethiopia/ICGEB-S11/2021 | EPI_ISL_2241493 | OL840812      | <a href="#">SRX13298198</a> | 43.1           | 29,827               | 5,488                   | 1.2                         | Spike D614G, Spike E180Q, Spike M1229I, Spike N501Y, Spike Q183H, Spike R646P, M V23L, N D377Y, N S194L, NS3 Q57H, NS8 W45C, NSP2 L24F, NSP3 S269A, NSP3 V521I, NSP5 S284G, NSP12 L648F, NSP12 P323L, NSP13 L43F, NSP13 Y277C  | 20A             | B.1.480                    |
| hCoV-19/Ethiopia/ICGEB-S12/2021 | EPI_ISL_2241494 | OL840813      | <a href="#">SRX13298199</a> | 42.1           | 29,827               | 3,936                   | 0.95                        | Spike D614G, Spike R214L, N E367G, N P199L, NS3 Q57H, NS3 S171L, NS7a R25T, NS8 T11I, NSP2 F499L, NSP4 P166S, NSP12 P323L  | 20A             | B.1                        |
| hCoV-19/Ethiopia/ICGEB-S13/2021 | EPI_ISL_2241611 | OL840814      | <a href="#">SRX13298200</a> | 43.4           | 29,817               | 3,407                   | 0.90                        | Spike D614G, Spike R646P, N S202N, NS3 L108F, NS3 T223I, NS8 L84S, NS8 W45S, NSP3 P153L, NSP6 M86I, NSP12 H810Y, NSP12 K103R, NSP14 L157F  | 19B             | A                          |
| hCoV-19/Ethiopia/ICGEB-S14/2021 | EPI_ISL_2241622 | OL840815      | <a href="#">SRX13298201</a> | 44.1           | 29,827               | 2,831                   | 0.69                        | Spike D614G, Spike M1229I, Spike N439K, N G179S, N S194L, NS3 Q57H, NS3 S40L, NSP2 L24F, NSP3 A1872V, NSP5 S284G, NSP8 A74V, NSP12 P323L   | 20A             | B.1.480                    |
| hCoV-19/Ethiopia/ICGEB-S15/2021 | EPI_ISL_2241623 | OL840816      | <a href="#">SRX13298202</a> | 43.4           | 29,827               | 3,129                   | 0.80                        | Spike D614G, Spike T676I, N P199L, NS3 Q57H, NS3 S171L, NSP2 F499L, NSP3 V1201F, NSP5 T198A, NSP12 P323L   | 20A             | B.1                        |

<sup>a</sup> PANGO, Phylogenetic Assignment of Named Global Outbreak Lineages.



**FIG 1** Phylogenetic analysis of sequenced SARS-CoV-2 samples from Ethiopia. Each box highlights phylogenetic relationships of samples within each lineage.

CA, USA) were used to assess the RNA quantities and qualities, respectively. From the total RNA, 100 ng was processed by the Swift Amplicon SARS-CoV-2 research panel for genome amplification and library preparation (Swift Bioscience, USA). High-throughput sequencing was conducted using an Illumina MiSeq sequencer following the standard protocol for paired-end 150-bp reads. The sequencing and data analysis were done at the ARGO Open Lab Platform for Genome Sequencing of the AREA Science Park and at the International Center for Genetic Engineering and Biotechnology (ICGEB), Trieste, Italy. All tools were run with default parameters unless otherwise specified. The raw sequence data were quality controlled using FastQC v0.11.9 (<https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>). Genome assembly was conducted using dedicated Swift dockerized data analysis guidelines (7, 8).

The study was approved by Addis Ababa University, Aklilu Lemma Institute of Pathobiology (ALIP) institutional review board with protocol number ALIPB IRB/52/2013/21. Data were collected after permission was obtained from the institute. All the information obtained from the study participants was kept confidential.

For each sample, sequencing parameters (GC content [%], complete length, genome coverage, and number of raw reads) are listed in Table 1. Briefly, genome length across the samples was 29,817 to 29,878 bp; genome coverage was 18,000× to 22,000× for samples S1 to S10 and 3,000× to 5,000× for samples S11 to S15.

A phylogenetic analysis was conducted on the 15 sequenced samples using the Nextstrain bioinformatics platform (9, 10). The coding complete genome sequences were analyzed in the context of the Nextregions/Africa data set, which is available on GISAID (updated to 7 December 2021). An Ethiopia-focused country-level subsampling strategy was performed, using the reference strain “hCoV-19/Wuhan/WH01/2019” (EPI\_ISL\_402125) as the original root. Results are shown in Fig. 1. A Nextstrain analysis

reveals that SARS-CoV-2 samples from Ethiopia belong to three different Nextstrain clades, as follows: 19B, 20A, and 20I/501Y.V1. A complete description of amino acidic substitutions and lineages for each sample is provided in Table 1.

**Data availability.** The coding complete genome sequences and metadata of all the 15 samples were submitted to the GISAID database ([www.gisaid.org](http://www.gisaid.org)) where they can be accessed through the accession numbers listed in Table 1. The coding complete genome sequences can be also accessed on the NCBI database (identifiers [IDs] OL840802 to OL840816, see Table 1); SRA accessions for each sample are available on the NCBI database under BioProject accession [PRJNA780398](https://www.ncbi.nlm.nih.gov/bioproject/PRJNA780398).

## ACKNOWLEDGMENTS

We are very grateful for the ICGB for collaborating in the sequencing activity and funding the transport of 15 RNA samples from Ethiopia. The funding support of Bioinnovate Africa is also much appreciated. This work was supported, in whole or in part, by the Bill & Melinda Gates Foundation (grant number INV-022,816). Under the grant conditions of the foundation, a Creative Commons Attribution 4.0 Generic License has already been assigned to the author accepted manuscript version that might arise from this submission.

We are also thankful to the staff of the EBTI COVID-19 diagnostic and research lab for their tremendous effort in detecting and characterizing the samples.

## REFERENCES

1. Mo P, Xing Y, Xiao Y, Deng L, Zhao Q, Wang H, Xiong Y, Cheng Z, Gao S, Liang K, Luo M, Chen T, Song S, Ma Z, Chen X, Zheng R, Cao Q, Wang F, Zhang Y. 2021. Clinical characteristics of refractory COVID-19 pneumonia in Wuhan, China. *Clin Infect Dis* 73:e4208–e4213. <https://doi.org/10.1093/cid/ciaa270>.
2. Su S, Wong G, Shi W, Liu J, Lai ACK, Zhou J, Liu W, Bi Y, Gao GF. 2016. Epidemiology, genetic recombination, and pathogenesis of coronaviruses. *Trends Microbiol* 24:490–502. <https://doi.org/10.1016/j.tim.2016.03.003>.
3. Bar-On YM, Flamholz A, Phillips R, Milo R. 2020. SARS-CoV-2 (COVID-19) by the numbers. *Elife* 9:e57309. <https://doi.org/10.7554/eLife.57309>.
4. WHO Health Emergency Dashboard WHO (COVID-19) Homepage. WHO, Geneva, Switzerland. <https://covid19.who.int/region/afro/country/et>.
5. Baba MM, Bitew M, Fokam J, Lelo EA, Ahidjo A, Asmamaw K, Beloumou GA, Bulimo WD, Buratti E, Chenwi C, Dadi H, D'Agaro P, De Conti L, Fainguem N, Gadzama G, Maiuri P, Majanja J, Meshack W, Ndjolo A, Nkenfou C, Oderinde BS, Opanda SM, Segat L, Stuan C, Symekher SL, Takou D, Tesfaye K, Triolo G, Tuki K, Zacchigna S, Marcello A. 2021. Diagnostic performance of colorimetric RT LAMP for the identification of SARS-CoV-2: a multicentre prospective clinical evaluation in sub-Saharan Africa. *EClinicalMedicine* 40: 101101. <https://doi.org/10.1016/j.eclinm.2021.101101>.
6. Licastro D, Rajasekharan S, Dal Monego S, Segat L, D'Agaro P, Marcello A. 2020. Isolation and fulllength genome characterization of SARS-CoV-2 from COVID-19 cases in northern Italy. *J Virol* 94:e00543-20. <https://doi.org/10.1128/JVI.00543-20>.
7. Corman VM, Landt O, Kaiser M, Molenkamp R, Meijer A, Chu DKW, Bleicker T, Brünink S, Schneider J, Schmidt ML, Mulders DGJC, Haagmans BL, van der Veer B, van den Brink S, Wijsman L, Goderski G, Romette J-L, Ellis J, Zambon M, Peiris M, Goossens H, Reusken C, Koopmans MPG, Drosten C. 2020. Detection of 2019 novel coronavirus (2019-nCoV) by real-time RT-PCR. *Euro Surveill* 25:2000045. <https://doi.org/10.2807/1560-7917.ES.2020.25.3.2000045>.
8. Li D, Liu C-M, Luo R, Sadakane K, Lam T-W. 2015. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics* 31:1674–1676. <https://doi.org/10.1093/bioinformatics/btv033>.
9. 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>.
10. Al-Rashedi NAM, Licastro D, Rajasekharan S, Dal Monego S, Marcello A, Munahi MG, Odda BS, Alabdali YAJ, AlObaidi LAH, Jasim A, Abdulzahra IA, Kadhim K, Awad A, Bachay M. 2021. Genome sequencing of a novel coronavirus SARS-CoV-2 isolate from Iraq. *Microbiol Resour Announc* 10: e01316-20. <https://doi.org/10.1128/MRA.01316-20>.