



Genome Sequence of SARS-CoV-2 Isolate Cali-01, from Colombia, Obtained Using Oxford Nanopore MinION Sequencing

 Diana Lopez-Alvarez,^{a,b}  Beatriz Parra,^c  Wilmer J. Cuellar^a

^aVirology Laboratory, International Center for Tropical Agriculture (CIAT), Cali, Colombia

^bFaculty of Agricultural Sciences, National University of Colombia, Palmira, Colombia

^cVirology Laboratory, COVID-19 UniValle Team, Department of Microbiology, Universidad del Valle, Cali, Colombia

ABSTRACT We report the genome sequence of a severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) isolate obtained from a patient with symptoms of coronavirus disease 2019 (COVID-19) who was infected in Cali, Colombia. The patient had no recent travel record and did not require hospitalization. The virus genome was obtained using Oxford Nanopore MinION sequencing.

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) (family *Coronaviridae*, genus *Betacoronavirus*) is spreading in the Americas and causing coronavirus disease 2019 (COVID-19) (1). Up to 14 May 2020, more than 13,000 cases of SARS-CoV-2 infection had been reported in Colombia, with 1,237 of those in the city of Cali. Genomic information and, above all, the development of local capacity for rapid molecular characterization of pathogens are crucial for understanding the genomic composition of the virus population and responding in a timely way to this and other emerging pathogens. Portable, next-generation sequencing technologies offer a quick approach to responding to this challenge. We report the coding-complete genome sequence of SARS-CoV-2 isolate Cali-01, obtained from a Colombian patient with no recent record of international travel.

SARS-CoV-2 isolate Cali-01 was obtained from a nasopharyngeal swab from a 45-year-old patient who presented with symptoms of COVID-19, including fever, cough, and throat congestion (2). The sample was collected in viral transport medium (VTM) following WHO protocols (3) and submitted to the Virology Laboratory of the Universidad del Valle for molecular diagnostics. Viral RNA was purified from 200 μ l of VTM using a QIAamp viral RNA minikit (Qiagen), and quantitative reverse transcription-PCR was carried out by following the CDC 2019-nCoV protocol (4). For genome sequencing, the RNA sample (cycle threshold value, 16) was diluted 1:10 for cDNA synthesis using random hexamer primers and was PCR amplified using a total of 218 primer pairs covering the whole virus genome (ARTIC Network amplicon sequencing protocol v2, with primers v3) (5). Libraries were prepared from 126 ng of DNA using a genomic DNA by ligation sequencing kit (SQK-LSK109; Oxford Nanopore Technologies), and sequencing was conducted using a FLO-MIN106D (R9.5) flow cell. Base calling was performed in real time using MinKNOW v2.0 (raw reads, 103,428 bases; N_{50} , 408 bp). The assembly was performed in two steps (using default parameters). First, for quality control and filtering of reads (fragments of 400 to 700 bp), we used the guppyplex script of the ARTIC Network bioinformatics protocol (<https://artic.network/ncov-2019/ncov2019-bioinformatics-sop.html>), followed by a reference assembly with minimap2 (6) and Pilon (7), using the sequence of the Wuhan-Hu-1 isolate (GenBank accession number [MN908947.3](https://genbank.ncbi.nlm.nih.gov/GenBank/FASTA/NC_023020.1)). The consensus sequence was quality checked using Qualimap v2.2.1 (8) and covered

Citation Lopez-Alvarez D, Parra B, Cuellar WJ. 2020. Genome sequence of SARS-CoV-2 isolate Cali-01, from Colombia, obtained using Oxford Nanopore MinION sequencing. *Microbiol Resour Announc* 9:e00573-20. <https://doi.org/10.1128/MRA.00573-20>.

Editor John J. Dennehy, Queens College

Copyright © 2020 Lopez-Alvarez 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 Beatriz Parra, beatriz.parra@correounivalle.edu.co, or Wilmer J. Cuellar, w.cuellar@cgiar.org.

Received 18 May 2020

Accepted 20 May 2020

Published 25 June 2020

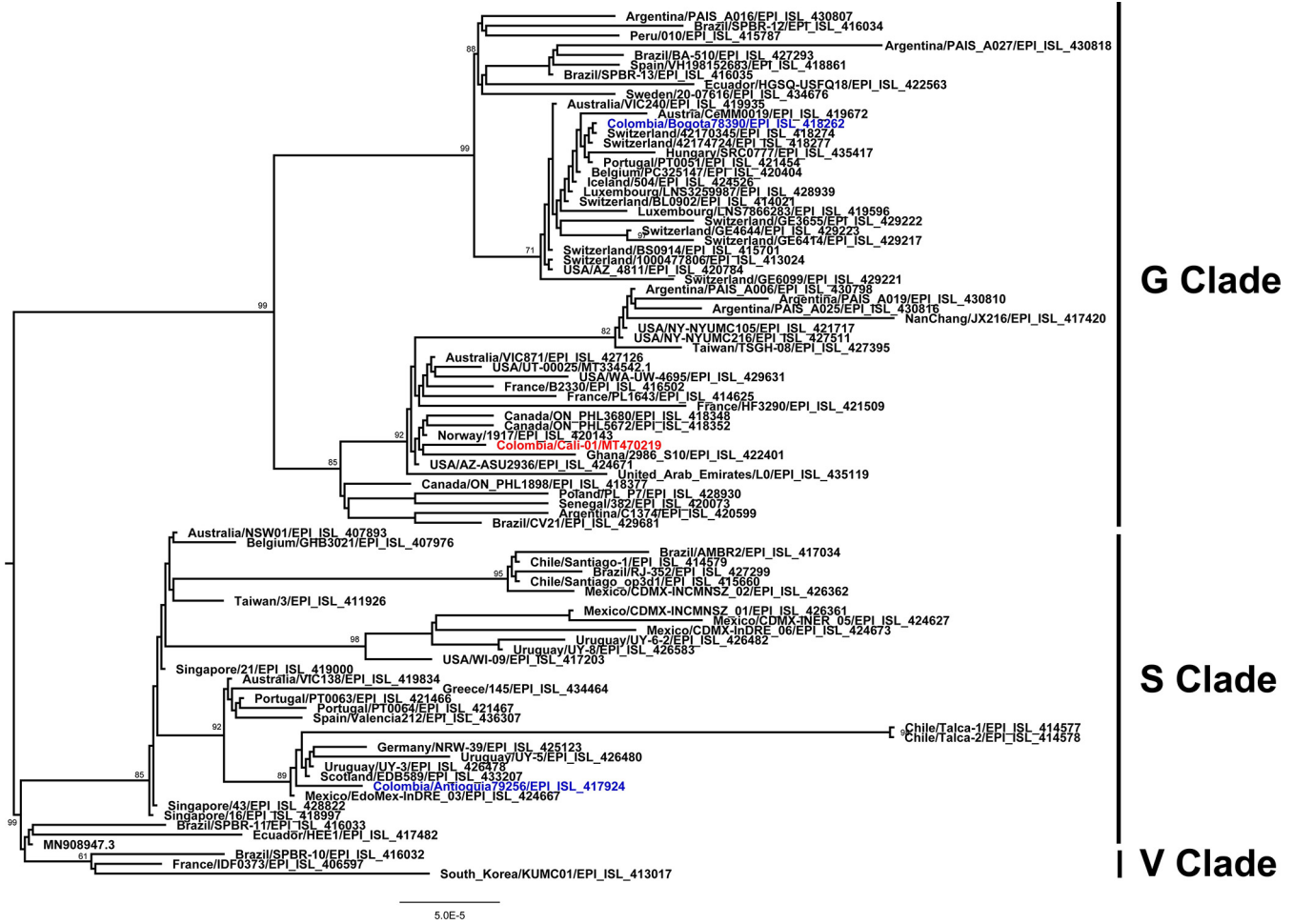


FIG 1 Maximum likelihood phylogeny of 89 genome sequences reported for SARS-CoV-2 isolates belonging to the G, S, and V clades, obtained from the GISAID database (<https://www.gisaid.org>). Values on the branches indicate bootstrap support. Isolate Cali-01 is shown in red, while other isolates from Colombia (GISAID accession numbers EPI_ISL_417924 [Antioquia] and EPI_ISL_418262 [Bogotá]) are shown in blue.

the complete reference genome (29,903 nucleotides; G+C content, 40.68%) in a contig of 102,381 reads, with an average coverage depth of 1,111×.

For phylogenetic analysis, a data set with 89 additional genome sequences, representing different GISAID-proposed clades (Fig. 1), was retrieved from GISAID (<https://www.gisaid.org>). The alignment was performed using MAFFT (9), implemented via the rapid phylodynamic alignment pipeline provided by Augur (github.com/nextstrain/augur) (10). A maximum likelihood phylogeny was reconstructed using IQ-TREE 2.0.3 (11) with a GTR+G model, which was selected using jModelTest 2 (12). We detected seven mutations, compared to the reference genome, of which three (C241T, C3037T, and A23122T) are in a noncoding position. Mutations T265I and P4715L are located in the *orf1ab* gene, Q57H in the *orf3a* gene, and D614G in the S gene; this last change clusters isolates Cali-01 and Bogotá78390 in clade G, while isolate Antioquia79256 remains in clade S (due to the L84S change in the *orf8* gene) (Fig. 1). The Cali-01 sequence reported here contributes to the ongoing international effort to track and understand the SARS-CoV-2 pandemic.

This work obtained institutional review board/ethics committee exemption from the Facultad de Salud of the Universidad del Valle because it relies on laboratory results that did not involve experimentation with humans. The required documentation is available upon request.

Data availability. The genome sequence of SARS-CoV-2 isolate Cali-01 was deposited in GISAID and GenBank under accession numbers [EPI_ISL_445219](https://gisaid.org/record/EPI_ISL_445219) and [MT470219](https://genbank.org/record/MT470219),

respectively. Raw reads were deposited in the GenBank Sequence Read Archive (SRA) under accession number [SRX8341156](#) (BioProject number [PRJNA632678](#) and BioSample number [SAMN14917563](#)).

ACKNOWLEDGMENTS

We are grateful to Ana Maria Leiva from the Virology Laboratory of the International Center for Tropical Agriculture (CIAT) for bioinformatics support. The sequencing work was carried out at the Virology Laboratory, Department of Microbiology, of the Universidad del Valle, in Cali. We thank the tireless group of volunteers, including faculty members, laboratory technicians, professionals, and students, from the Universidad del Valle (COVID-19 UniValle team) who are supporting the diagnostics of SARS-CoV-2 in Colombia.

This work was funded by the Universidad del Valle and the International Center for Tropical Agriculture (CIAT).

REFERENCES

1. Wu F, Zhao S, Yu B, Chen Y-M, Wang W, Song Z-G, Hu Y, Tao Z-W, Tian J-H, Pei Y-Y, Yuan M-L, Zhang Y-L, Dai F-H, Liu Y, Wang Q-M, Zheng J-J, Xu L, Holmes EC, Zhang Y-Z. 2020. A new coronavirus associated with human respiratory disease in China. *Nature* 579:265–269. <https://doi.org/10.1038/s41586-020-2008-3>.
2. Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y, Zhang L, Fan G, Xu J, Gu X, Cheng Z, Yu T, Xia J, Wei Y, Wu W, Xie X, Yin W, Li H, Liu M, Xiao Y, Gao H, Guo L, Xie J, Wang G, Jiang R, Gao Z, Jin Q, Wang J, Cao B. 2020. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet* 395:497–506. [https://doi.org/10.1016/S0140-6736\(20\)30183-5](https://doi.org/10.1016/S0140-6736(20)30183-5).
3. World Health Organization. 2020. Laboratory testing for coronavirus disease (COVID-19) in suspected human cases: interim guidance. WHO/COVID-19/laboratory/2020.5. World Health Organization, Geneva, Switzerland. <https://apps.who.int/iris/handle/10665/331501>.
4. Centers for Disease Control and Prevention. 2020. CDC 2019-novel coronavirus (2019-nCoV) real-time RT-PCR diagnostic panel. Centers for Disease Control and Prevention, Atlanta, GA. <https://www.fda.gov/media/134922/download>.
5. Quick J. 2020. nCoV-2019 sequencing protocol v2. protocols.io, Berkeley, CA. <https://doi.org/10.17504/protocols.io.bdp715rn>.
6. Li H. 2018. minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* 34:3094–3100. <https://doi.org/10.1093/bioinformatics/bty191>.
7. Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, Cuomo CA, Zeng Q, Wortman J, Young SK, Earl AM. 2014. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One* 9:e112963. <https://doi.org/10.1371/journal.pone.0112963>.
8. Okonechnikov K, Conesa A, García-Alcalde F. 2016. Qualimap 2: advanced multi-sample quality control for high-throughput sequencing data. *Bioinformatics* 32:292–294. <https://doi.org/10.1093/bioinformatics/btv566>.
9. Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol* 30:772–780. <https://doi.org/10.1093/molbev/mst010>.
10. 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>.
11. Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. 2015. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum likelihood phylogenies. *Mol Biol Evol* 32:268–274. <https://doi.org/10.1093/molbev/msu300>.
12. Darriba D, Taboada GL, Doallo R, Posada D. 2012. jModelTest 2: more models, new heuristics and parallel computing. *Nat Methods* 9:772. <https://doi.org/10.1038/nmeth.2109>.