Whole-Genome Sequences of SARS-CoV-2 Isolates from the Dominican Republic

R. Paulino-Ramirez,[a] E. Riego,[b] A. Vallejo-Degaudenzi,[a] V. V. Calderon,[b] L. Tapia,[a] Patricia León,[b] Danilo Licastro,[c] Simeone Dal Monego,[c] Sreejith Rajasekharan,[d] Emanuele Orsini,[d] Alessandro Marcello[d]

[a]Instituto de Medicina Tropical & Salud Global, Universidad Iberoamericana, Santo Domingo, Dominican Republic
[b]Molecular Biology Department, Referencia Clinical Laboratory, Santo Domingo, Dominican Republic
[c]AREA Science Park, Trieste, Italy
[d]Laboratory of Molecular Virology, International Centre for Genetic Engineering and Biotechnology, Trieste, Italy

ABSTRACT Here, we report the genome sequences of five severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) strains that were obtained from symptomatic individuals with travel histories during community surveillance in the Dominican Republic in 2020. These sequences provide a starting point for further genomic studies of gene flow and molecular diversity in the Caribbean nation. Phylogenetic analysis suggests that all genomes correspond to the B.1 variant.

Coronaviruses have emerged in the past century, causing epidemics and pandemics of zoonotic strains (1). Coronavirus disease 2019 (COVID-19), the disease caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), a member of the family *Coronaviridae*, genus *Betacoronavirus*, has caused a global pandemic with unprecedented impact on humans and has highlighted the weaknesses of the response systems for health emergencies in developing countries (2).

After the first case was identified in Wuhan, China, in late December 2020, the virus was first reported in La Hispaniola on 28 February 2020, with an expanding wave of transmission all over the Dominican Republic (3–5). The first cases in the country were detected in Greater Santo Domingo (which includes the capital city and the neighboring province) and San Francisco de Macoris, a north-central town in the Duarte Province (Fig. 1A) (6–8).

A total of five samples were collected from symptomatic individuals with travel histories during community surveillance in February to May 2020. Cases were associated with a history of travel to Italy and were geographically from a north-central community (in the Duarte Province) and Greater Santo Domingo (Fig. 1A), representing the first confirmed SARS-CoV-2 cases in the country. Samples were collected using nasopharyngeal swabs. RNA extraction was performed using the MagMax viral/pathogen nucleic acid isolation kit in the KingFisher Flex automated extraction system (Thermo Fisher Scientific) following the manufacturer's protocols. For each sample, 100 ng of total RNA was processed using the Zymo-Seq RiboFree ribosomal depletion library preparation kit (Zymo Research) (9). A Qubit 2.0 fluorometer (Thermo Fisher Scientific, MA, USA) and Agilent 2100 Bioanalyzer (Agilent Technologies, CA, USA) were used to assess RNA quantity and quality. Total RNA was processed using library construction based on the Swift Amplicon SARS-CoV-2 research panel (Swift Biosciences, USA), which provides optimal coverage and data quality. High-throughput sequencing was conducted using an Illumina MiSeq sequencer following the standard procedure. The raw sequence data were quality controlled using FastQC v0.11.9 37 (https://www.bioinformatics.babraham.ac.uk/projects/fastqc). Genome assembly was conducted using dedicated Swift guidelines (10); the genome size, coverage depth, and overall GC content of each genome are indicated in Table 1 (11). All tools were run with default parameters unless otherwise specified.
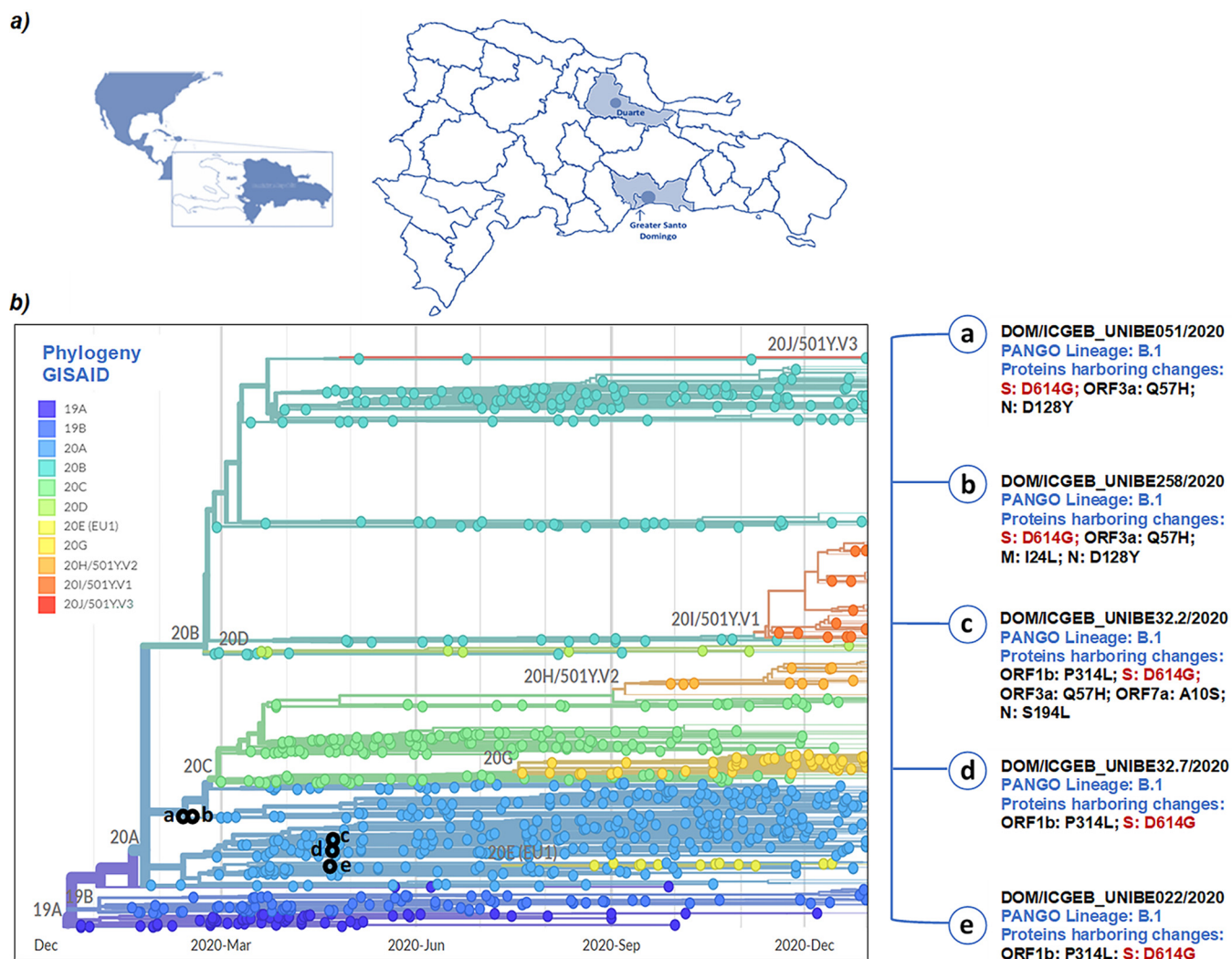
FIG 1 (A) Map showing the island of La Hispaniola in the Caribbean basin (right), the Dominican Republic (dark blue), and localities (dark blue circles) where samples were collected in provinces in blue, i.e., Duarte (located in the north-central plateau) and Greater Santo Domingo (on the southern coast), which includes the capital city (Distrito Nacional) (arrow) and the nearby municipalities. Maps were generated under the ArcGIS software v10.8.1 license. (B) Phylogenetic analysis of SARS-CoV-2 including five genome sequences collected in the Dominican Republic. Available genomes were retrieved from GISAID (https://www.gisaid.org) on 1 September 2021 but were limited to December 2020 to better allocate the early sequences detected. Colors depict clades based on mutation marks using the GISAID standardized nomenclature. All analyzed sequences were classified as B or B.1 (PANGO Lineage), harboring S protein changes in the 614 codon (D614G) (in red); unique mutations in S protein are highlighted in blue.

Phylogenetic tree analysis was conducted using the Nextstrain bioinformatics platform (http://nextstrain.org/ncov) with the maximum likelihood option and the JTT matrix (12–15). The complete genome sequences were analyzed in the context of the Nextregions/North American data set, which is available at the GISAID site (updated to 1 September 2021). A Dominican Republic-focused country-level subsampling strategy was performed, using the

TABLE 1 Bioinformatic details of analyzed sequences

Parameter	Finding for strain:				
	ICGEB_UNIBE32.7	ICGEB_UNIBE051	ICGEB_UNIBE258	ICGEB_UNIBE022	ICGEB_UNIBE32.2
SRA accession no.	SRX12761935	SRX9816443	SRX9816442	SRX9816441	SRX9816440
GISAID clade	O	GH	GH	G	GH
No. of raw reads	929,390	893,092	76,794	1,203,034	1,070,364
Genome size (bp)	29,585	29,901	29,580	29,902	29,748
SARS-CoV-2 coverage depth ($\times$)	2,761	5,000	83	5,447	567
GC content (%)	48.7	44.8	45.6	43.5	46.2

reference strain hCoV-19/Wuhan/WH01/2019 (GISAID accession number EPI_ISL_402125) as the original root. Figure 1B shows the genetic relationship between Dominican samples and other strains in the GISAID database. This study represents the starting point to further explore the genomic diversity of SARS-CoV-2 in the Dominican Republic, since its introduction, human mobilization across the terrestrial border with Haiti, and the tourism industry.

The institutional review board at Universidad Iberoamericana (UNIBE) (CEI-2020-16) and the National Bioethical Committee (020-2021) approved this study.

Data availability. Sequences were deposited in the NCBI database (BioSample accession numbers SAMN17274443, SAMN17274444, SAMN17274445, SAMN17274446, and SAMN22555599). The raw reads were deposited in the NCBI Sequence Read Archive (SRA) database (SRA accession numbers SRX12761935, SRX9816443, SRX9816442, SRX9816441, and SRX9816440) under BioProject number PRJNA691021. Genome sequences were deposited in the NCBI GenBank database (GenBank accession number OK523387, OK523388, OK523389, OK523390, and OK542388) and in the GISAID database (GISAID accession numbers EPI_ISL_523811, EPI_ISL_523812, EPI_ISL_525467, EPI_ISL_525468, and EPI_ISL_525469).

REFERENCES

1. Chinese SARS Molecular Epidemiology Consortium. 2004. Molecular evolution of the SARS coronavirus during the course of the SARS epidemic in China. Science 303:1666–1669. https://doi.org/10.1126/science.1092002.
2. Walker PGT, Whittaker C, Watson OJ, Baguelin M, Winskill P, Hamlet A, Djafaara BA, Cucunubá Z, Olivera Mesa D, Green W, Thompson H, Nayagam S, Ainslie KEC, Bhatia S, Bhatt S, Boonyasiri A, Boyd O, Brazeau NF, Cattarino L, Cuomo-Dannenburg G, Dighe A, Donnelly CA, Dorigatti I, van Elsland SL, FitzJohn R, Fu H, Gaythorpe KAM, Geidelberg L, Grassly N, Haw D, Hayes S, Hinsley W, Imai N, Jorgensen D, Knock E, Laydon D, Mishra S, Nedjati-Gilani G, Okell LC, Unwin HJ, Verity R, Vollmer M, Walters CE, Wang H, Wang Y, Xi X, Lalloo DG, Ferguson NM, Ghani AC. 2020. The impact of COVID-19 and strategies for mitigation and suppression in low- and middle-income countries. Science 369:413–422. https://doi.org/10.1126/science.abc0035.
3. Paulino-Ramirez R, Tapia L. 2020. Learning from pandemics in the Americas: the Dominican Republic Programmatic response against a novel coronavirus (COVID-19). InterAm J Med Health 3:e202003024. https://doi.org/10.31005/iajmh.v3i0.104.
4. Petrone ME, Earnest R, Lourenço J, Kraemer MU, Paulino-Ramirez R, Grubaugh ND, Tapia L. 2021. Asynchronicity of endemic and emerging mosquito-borne disease outbreaks in the Dominican Republic. Nat Commun 12:151. https://doi.org/10.1038/s41467-020-20391-x.
5. Paulino-Ramirez R, Báez AA, Vallejo Degaudenzi A, Tapia L. 2020. Seroprevalence of specific antibodies against SARS-CoV-2 from hotspot communities in the Dominican Republic. Am J Trop Med Hyg 103:2343–2346. https://doi.org/10.4269/ajtmh.20-0907.
6. Direccion General de Epidemiologia, Ministry of Health. 2020. COVID-19 special bulletins. http://www.digepisalud.gob.do/documentos/?drawer=Boletines%20epidemiologicos*Boletines%20semanales*2020. Accessed 28 December 2020.
7. Tapia L. 2020. COVID-19 and fake news in the Dominican Republic. Am J Trop Med Hyg 102:1172–1174. https://doi.org/10.4269/ajtmh.20-0234.
8. Du P, Ding N, Li J, Zhang F, Wang Q, Chen Z, Song C, Han K, Xie W, Liu J, Wang L, Wei L, Ma S, Hua M, Yu F, Wang L, Wang W, An K, Chen J, Liu H, Gao G, Wang S, Huang Y, Wu AR, Wang J, Liu D, Zeng H, Chen C. 2020. Genomic surveillance of COVID-19 cases in Beijing. Nat Commun 11:5503. https://doi.org/10.1038/s41467-020-19345-0.
9. Licastro D, Rajasekharan S, Dal Monego S, Segat L, D'Agaro P, Marcello A, Regione FVG Laboratory Group on COVID-19. 2020. Isolation and full-length 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.
10. Swift Biosciences. 2021. Swift Normalase amplicon SARS-CoV-2 panel (SNAP) dockerized data analysis guidelines. Swift Biosciences, Ann Arbor, MI. https://swiftbiosci.com/wp-content/uploads/2021/02/TEC-006_SARS-CoV-2_TechNote_Rev3.pdf.
11. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup. 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics 25:2078–2079. https://doi.org/10.1093/bioinformatics/btp352.
12. Jones DT, Taylor WR, Thornton JM. 1992. The rapid generation of mutation data matrices from protein sequences. Comput Appl Biosci 8:275–282. https://doi.org/10.1093/bioinformatics/8.3.275.
13. Edgar RC. 2004. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. BMC Bioinformatics 5:113. https://doi.org/10.1186/1471-2105-5-113.
14. Hatcher EL, Zhdanov SA, Bao Y, Blinkova O, Nawrocki EP, Ostapchuck Y, Schäffer AA, Brister JR. 2017. Virus Variation Resource: improved response to emergent viral outbreaks. Nucleic Acids Res 45:D482–D490. https://doi.org/10.1093/nar/gkw1065.
15. 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.