




Complete Genome Sequence of a 2019 Novel Coronavirus (SARS-CoV-2) Strain Causing a COVID-19 Case in Morocco

 Sanaâ Lemriss,^a Amal Souiri,^a Narjis Amar,^a Nabil Lemzaoui,^a Omar Mestoui,^a Mohamed Labioui,^a Nabil Ouairiba,^a Ayoub Jibjibe,^a Mahmoud Yartaoui,^a Mohamed Chahmi,^a Marouane El Rhouila,^a Samiha Sellak,^a Nadia Kandoussi,^a Saâd El Kabbaj^a

^aDepartment of Biosafety PCL3, Laboratory of Research and Medical Analysis of Gendarmerie Royale, Rabat, Morocco

ABSTRACT Here, we report a complete genome sequence obtained for a novel severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) strain isolated from a nasopharyngeal swab specimen of a Moroccan patient with coronavirus disease 2019 (COVID-19).

An epidemic of pneumonia, named coronavirus disease 2019 (COVID-19) by the World Health Organization, emerged in Wuhan, China, in December 2019 and has rapidly spread worldwide (1, 2). The virus causing the pneumonia was sequenced, and it was discovered that it is a strain of betacoronavirus (family *Coronaviridae*, genus *Betacoronavirus*); it was named severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) by the Coronavirus Study Group of the International Committee on Taxonomy of Viruses (3). In Morocco, more than 7,800 cases have been reported (4). To control the disease, it is important to characterize the viral genome sequences.

A nasopharyngeal swab specimen was collected on 23 April 2020 from a man who had contact with family clusters that tested positive for COVID-19 in Ouarzazate City, Morocco, and was identified as positive for SARS-CoV-2 (threshold cycle [C_T] value, 16.5 in E-gene and 19.48 in RdRp-gene) by real-time reverse transcriptase PCR (RT-PCR) assay as previously reported (5). All participants provided their written informed consent, and the study protocol was conducted according to ethical requirements for human biological research.

Viral RNA of the same sample was extracted using a QIAamp viral RNA minikit (Qiagen). The full genome was amplified according to the CDC protocol (6), with two sets of 38 primers used in two rounds of nested RT-PCR to produce overlapping PCR products covering the full genome. Amplicons were purified with ExoSAP-IT PCR product cleanup (Applied Biosystems) and quantified with a QFX fluorometer (DeNovix) before being pooled at an equimolar concentration. The library was prepared with the Nextera XT library prep kit (Illumina), purified with Agencourt AMPure XP beads (Beckman Coulter), and quantified with a QFX fluorometer. The resulting DNA library was sequenced with a MiSeq system using 250-bp paired ends (Illumina).

The fastq files (139,949 paired-end reads) were cleaned with Trimmomatic 0.36 (TRAILING:35, SLIDINGWINDOW:20) and then subjected to mapping with the reference SARS-CoV-2 genome (GenBank accession number [MN908947.3](https://www.ncbi.nlm.nih.gov/nuccore/MN908947.3)) using GS Reference Mapper 2.9 with default parameters (7), and a consensus sequence named hCoV-19_Morocco_OUA677_19_2020 was obtained, which has an average coverage of 1,000× and 37.97% GC content.

The open reading frames (ORFs) were predicted using Geneious Prime 2020.1 (8) and annotated using the “CD-Search” tool in the Conserved Domain Database with default parameters (9–11). Sequence alignments with the SARS-CoV-2 genome (GenBank accession number [MN908947.3](https://www.ncbi.nlm.nih.gov/nuccore/MN908947.3)) were done with CLC Genomics Workbench 20, using the “create Alignment 1.02” tools with default parameters (12).

The genome sequence is 29,934 nucleotides (nt) long, including a 5′ untranslated

Citation Lemriss S, Souiri A, Amar N, Lemzaoui N, Mestoui O, Labioui M, Ouairiba N, Jibjibe A, Yartaoui M, Chahmi M, El Rhouila M, Sellak S, Kandoussi N, El Kabbaj S. 2020. Complete genome sequence of a 2019 novel coronavirus (SARS-CoV-2) strain causing a COVID-19 case in Morocco. *Microbiol Resour Announc* 9:e00633-20. <https://doi.org/10.1128/MRA.00633-20>.

Editor Simon Roux, DOE Joint Genome Institute

Copyright © 2020 Lemriss 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 Sanaâ Lemriss, slemriss@lram-fgr.ma.

Received 3 June 2020

Accepted 17 June 2020

Published 2 July 2020

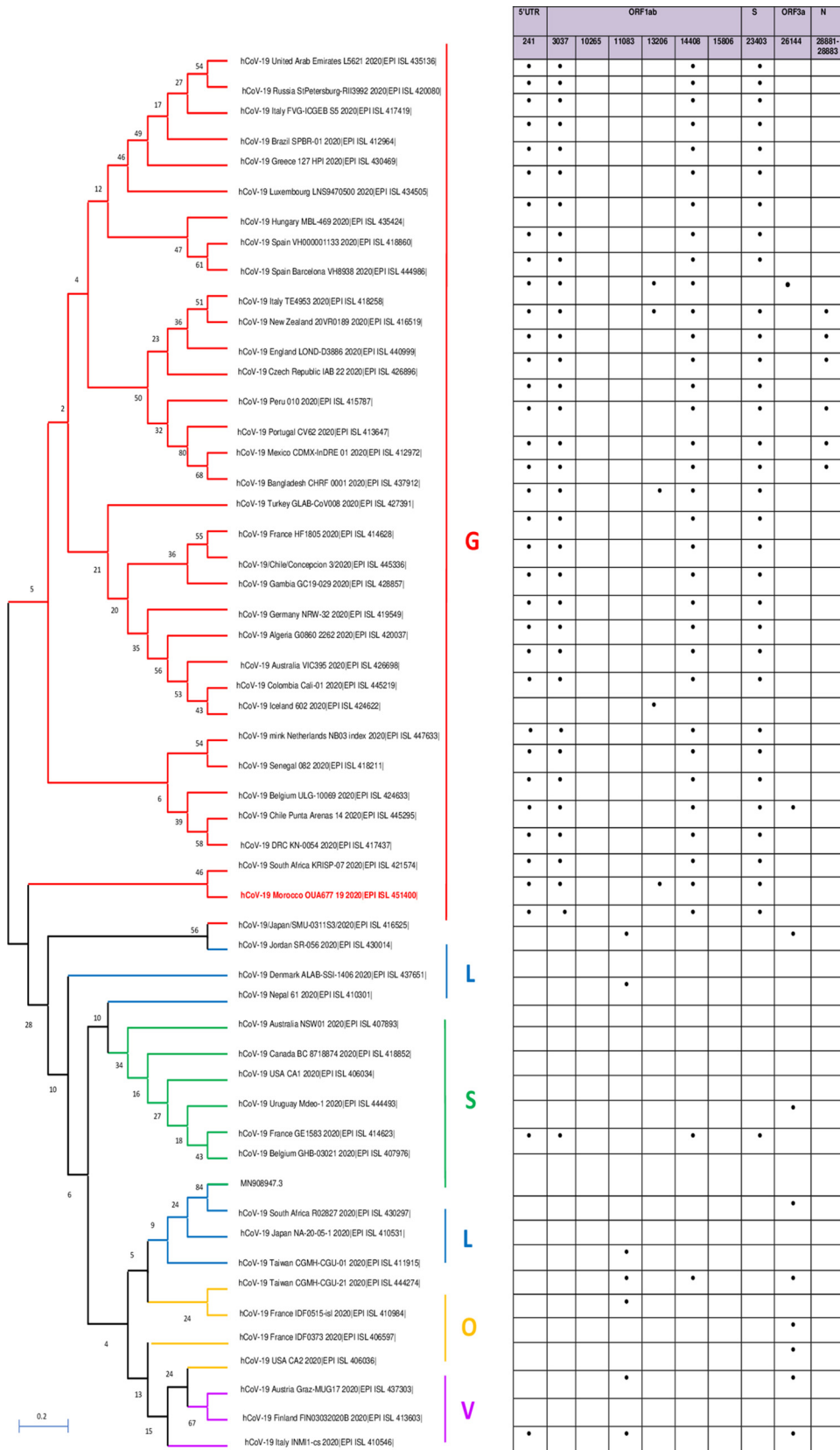


FIG 1 Phylogenetic tree of the complete nucleotide sequence of hCoV-19_Morocco_OUA677_19_2020 and 54 other global strains obtained from the GISAID database, associated with a table representing single-nucleotide polymorphisms (SNPs). The phylogenetic tree was constructed with the neighbor-joining method using MEGAX, and the reliability of each tree branch was estimated by performing 1,000 bootstrap replicates.

region (UTR) (nt 1 to 301), replicase complex ORF1ab (nt 301 to 21590), S gene (nt 21598 to 25419), ORF3a (nt 25428 to 26255), E gene/ORF4 (nt 26280 to 26507), M gene/ORF5 (nt 26558 to 27226), ORF6 (nt 27237 to 27422), ORF7a (nt 27429 to 27794), ORF7b (nt 27791 to 27922), ORF8 gene (nt 27929 to 28294), N gene/ORF9 (nt 28309 to 29568), ORF10 gene (nt 29593 to 29709), and a 3' UTR (nt 29710 to 29921).

Compared with the reference strain (GenBank accession number [MN908947.3](https://doi.org/10.1093/mra/mnaa001)), the hCoV-19_Morocco_OUA677_19_2020 genome has a total of 10 nucleotide variations. We detected mutations in noncoding positions of G to A (position 198), another one of C to T (position 241), two mutations of A to T (position 29881 to 29882), and other mutations in coding regions that generated amino acid changes, such as F924F, P4715L, S5039I, and L6082F in the polyprotein encoded by the ORF1ab gene, D614G in the spike glycoprotein (S), and A155A in the nucleocapsid protein (N) (Fig. 1).

The single-nucleotide polymorphisms (SNPs) of the Moroccan sequence were further defined from 54 sequences of SARS-CoV-2 representing 50 countries all over the world (Fig. 1). Important substitutions were observed among the different ORFs (ORF1ab, 82; S segment, 33; N segment, 18; ORF3a, 10; noncoding region 5' UTR, 35).

Phylogenetic analysis of this virus genome compared with 54 selected sequences showed that it was grouped in SARS-CoV-2 clade G, which includes strains from Asia, Europe, North America, Australia, and Africa (Fig. 1).

We are currently sequencing and analyzing more complete genomes from different regions of Morocco to understand the virus dispersion and to associate this information with epidemiological data.

Data availability. The consensus data for the hCoV-19_Morocco_OUA677_19_2020 genome have been deposited in the GISAID database (accession number EPI_ISL_451400) and GenBank (accession number [MT513758](https://doi.org/10.1093/mra/mnaa001)). The accession numbers for the Illumina MiSeq sequence raw reads in the NCBI Sequence Read Archive (SRA) are [PRJNA637892](https://doi.org/10.1093/mra/mnaa001) (BioProject), [SRR11945456](https://doi.org/10.1093/mra/mnaa001) (SRA), and [SAMN15160097](https://doi.org/10.1093/mra/mnaa001) (BioSample).

ACKNOWLEDGMENT

This study was supported by the Fraternal Gendarmerie Royale, Morocco.

REFERENCES

- Zhou P, Yang XL, Wang XG, Hu B, Zhang L, Zhang W, Si HR, Zhu Y, Li B, Huang CL, Chen HD, Chen J, Luo Y, Guo H, Jiang RD, Liu MQ, Chen Y, Shen XR, Wang X, Zheng XS, Zhao K, Chen QJ, Deng F, Liu LL, Yan B, Zhan FX, Wang YY, Xiao GF, Shi ZL. 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>.
- Munster VJ, Koopmans M, van Doremalen N, van Riel D, de Wit E. 2020. A novel coronavirus emerging in China: key questions for impact assessment. *N Engl J Med* 382:692–694. <https://doi.org/10.1056/NEJMp2000929>.
- 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>.
- World Health Organization. WHO coronavirus disease (COVID-19) dashboard. 2020. <https://covid19.who.int>.
- Corman VM, Landt O, Kaiser M, Molenkamp R, Meijer A, Chu DKW, Bleicker T, Brunink S, Schneider J, Schmidt ML, Mulders D, Haagmans BL, van der Veer B, van den Brink S, Wijsman L, Goderski G, Romette JL, 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://www.eurosurveillance.org/content/10.2807/1560-7917.ES.2020.25.3.2000045>.
- CDC. 2020. CDC comprehensive SARS-CoV-2 sequencing protocols. https://github.com/CDCgov/SARS-CoV-2_Sequencing/tree/master/protocols/CDC-Comprehensive.
- Van Borm S, Rosseel T, Behaeghel I, Saulmont M, Delooz L, Petitjean T, Mathijs E, Vandenbussche F. 2016. Complete genome sequence of bovine polyomavirus type 1 from aborted cattle, isolated in Belgium in 2014. *Genome Announc* 4:e01646-15. <https://doi.org/10.1128/genomeA.01646-15>.
- Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C, Thierer T, Ashton B, Meintjes P, Drummond A. 2012. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* 28:1647–1649. <https://doi.org/10.1093/bioinformatics/bts199>.
- Lu S, Wang J, Chitsaz F, Derbyshire MK, Geer RC, Gonzales NR, Gwadz M, Hurwitz DI, Marchler GH, Song JS, Thanki N, Yamashita RA, Yang M, Zhang D, Zheng C, Lanczycki CJ, Marchler-Bauer A. 2020. CDD/SPARCLE: the conserved domain database in 2020. *Nucleic Acids Res* 48: D265–D268. <https://doi.org/10.1093/nar/gkz991>.
- Kim D, Lee JY, Yang JS, Kim JW, Kim VN, Chang H. 2020. The architecture of SARS-CoV-2 transcriptome. *J Cell* 181:914–921. <https://doi.org/10.1016/j.cell.2020.04.011>.
- Xu J, Hu J, Wang J, Han Y, Hu Y, Wen J, Li Y, Ji J, Ye J, Zhang Z, Wei W, Li S, Wang J, Wang J, Yu J, Yang H. 2003. Genome organization of the SARS-CoV. *Genomics Proteomics Bioinformatics* 1:226–235. [https://doi.org/10.1016/S1672-0229\(03\)01028-3](https://doi.org/10.1016/S1672-0229(03)01028-3).
- Lu R, Zhao X, Li J, Niu P, Yang B, Wu H, Wang W, Song H, Huang B, Zhu N, Bi Y, Ma X, Zhan F, Wang L, Hu T, Zhou H, Hu Z, Zhou W, Zhao L, Chen J, Meng Y, Wang J, Lin Y, Yuan J, Xie Z, Ma J, Liu WJ, Wang D, Xu W, Holmes EC, Gao GF, Wu G, Chen W, Shi W, Tan W. 2020. Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. *Lancet* 395:565–574. [https://doi.org/10.1016/S0140-6736\(20\)30251-8](https://doi.org/10.1016/S0140-6736(20)30251-8).