Complete Genome Sequence of a SARS-CoV-2 Strain Sampled in Morocco in May 2020, Obtained Using Sanger Sequencing

Abderrazak Rfaki,[a,b] Nadia Touil,[b,c] Mouhssine Hemlali,[a,b] Sanaâ Alaoui Amine,[a,b] Marouane Melloul,[a,b] Moulay Abdelaziz El Alaoui,[a,d] Hicham Elannaz,[c] Amine Idriss Lahlou,[c] Mostafa Elouennass,[c] Khalid Ennibi,[c] Elmostafa El Fahime[a,b]

[a]Molecular Biology and Functional Genomics Platform, National Center for Scientific and Technical Research (CNRST), Rabat, Morocco
[b]Genomic Center for Human Pathologies (GENOPATH), Faculty of Medicine and Pharmacy, University Mohammed V, Rabat, Morocco
[c]Unité de Culture Cellulaire, CVMIT, HMI Med V, Rabat, Morocco
[d]Laboratoire de Botanique et de Protection des Plantes, UFR de Mycologie, Département de Biologie, Faculté des Sciences, Université Ibn Tofail, Kenitra, Morocco

Abderrazak Rfaki and Nadia Touil contributed equally to this work. Author order was determined on the basis of seniority.

ABSTRACT The complete genome sequence of a severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) strain was obtained. The strain was isolated from a nasopharyngeal swab specimen from a female patient in Rabat, Morocco, with coronavirus disease 2019 (COVID-19). This strain belongs to clade 20A and has 12 mutations and 8 amino acid substitutions compared to the reference strain Wuhan/Hu-1/2019.

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), a member of the *Coronaviridae* family and *Betacoronavirus* genus, is the causative agent of the coronavirus disease 2019 (COVID-19) pandemic. As of April 2020, positive cases rose continually, reached a peak in November 2020, and then decreased gradually until now (http://www.covidmaroc.ma/pages/Accueilfr.aspx). Analysis of the SARS-CoV-2 genome is crucial for fully understanding its spread and evolution as well as for vaccine development (1). Currently, SARS-CoV-2 whole-genome sequences are often generated by next-generation sequencing (NGS) (2), because of the fast speed and parallelism of the standard NGS methods; however, Sanger sequencing can also provide high-quality genome sequences (3).

Here, we report the complete genome sequence of SARS-CoV-2 from a Moroccan patient, using Sanger sequencing technology, based on multiple amplified nucleic acid fragments, according to CDC (4) protocol.

Strain SARS-CoV-2/human/MAR/Morocco_HMIMV_279CC/2020 was recovered from a nasopharyngeal swab specimen from a 48-year-old female patient in Rabat, Morocco, with COVID-19 on 10 May 2020.

Viral RNA was extracted using a MagPurix instrument (Zinexts Life Science, UK) and quality controlled using a NanoDrop spectrophotometer (Thermo Fisher Scientific, Germany). cDNA was generated using a Tetro cDNA synthesis kit (Bioline, UK), and real-time PCR (RT-PCR) was performed using MyFi mix (Bioline) according to the manufacturer's protocol.

The whole-genome sequence of strain SARS-CoV-2/human/MAR/Morocco_HMIMV_279CC/2020 was generated using 38 pairs of specific primers covering the whole virus genome using a previously reported method (4). The PCR amplification yielded 38 amplicons, which were confirmed by gel electrophoresis analysis (1% agarose) and purified using an ExoSAP-IT PCR cleanup kit (Applied Biosystems, USA). The sequence reactions were performed using the BigDye terminator cycle sequencing kit v3.1 (Applied Biosystems), purified using the BigDye XTerminator purification kit (Applied Biosystems), and then loaded onto a SeqStudio genetic analyzer capillary sequencer (Applied Biosystems) according to the manufacturer's instructions.
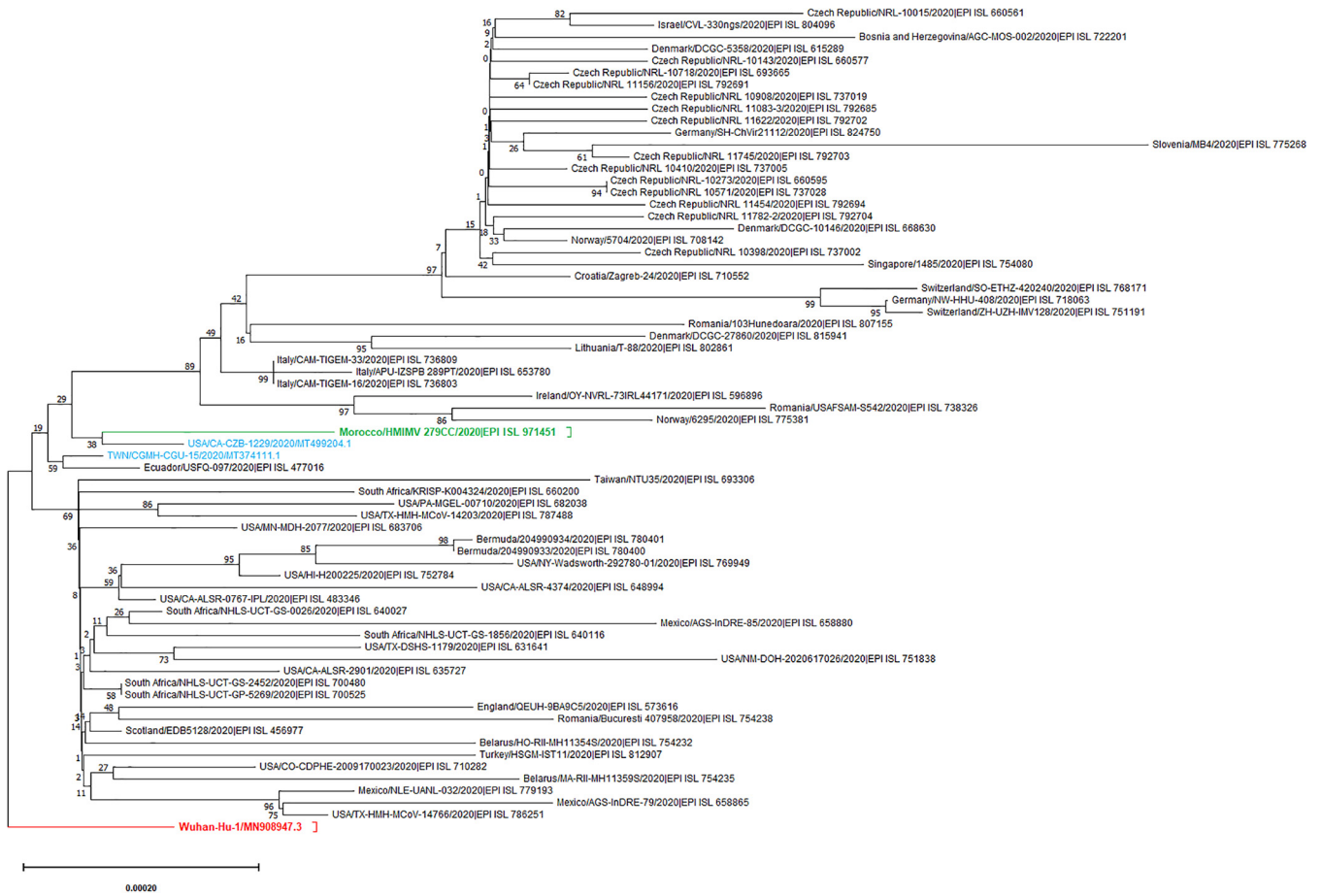
FIG 1 Phylogenetic tree of strain SARS-CoV-2/human/MAR/Morocco_HMIMV_279CC/2020. A total of 69 viral genome sequences are displayed, including (i) 66 genomes provided by Nextstrain (https://nextstrain.org/sars-cov-2/) on 10 February 2021, counting the sequence reported in this announcement (in green); (ii) 2 mostly similar genome sequences from NCBI, USA/CA-CZB-1229/2020 (GenBank accession number MT499204.1) and TWN/CGMH-CGU-15/2020 (MT374111.1) (in blue); and (iii) the reference genome of the SARS-CoV-2 Wuhan-Hu-1 isolate (MN908947.3) (in red). Multiple sequence alignments were conducted using MAFFT v7.221 (6) through the Galaxy platform (7), and the phylogenetic tree was constructed using MEGA X v10.2.0 (8) with the neighbor-joining algorithm with 1,000 bootstrap replicates. The tree structure was rooted in the position of reference strain Wuhan-Hu-1. Bar, number of nucleotide substitutions per site.

The obtained nucleotide sequences were evaluated, edited, and assembled, based on mapping to the reference sequence of the SARS-CoV-2 Wuhan-Hu-1 isolate (GenBank accession number MN908947.3) using Unipro UGENE software v37.0. The length of the final consensus sequence is 29,903 nucleotides with 37.96% GC content. NCBI BLASTN v2.11.0+ (5) showed that the genome was mostly similar (99.97%) to SARS-CoV-2/human/USA/CA-CZB-1229/2020 (MT499204.1) and to SARS-CoV-2/human/TWN/CGMH-CGU-15/2020 (MT374111.1), reported in the USA and Taiwan, respectively (Fig. 1).

Sequence quality control, clade assignment, and mutation determination were done using the Nextclade tool v0.12.0 (https://clades.nextstrain.org/). All tools were run using default parameters unless otherwise specified. The results showed that the SARS-CoV-2/human/MAR/Morocco_HMIMV_279CC/2020 genome belongs to clade 20A, with 12 mutations (A2568T, C3037T, C5884T, C8169T, C9907G, C14408T, C17104T, A20268G, G21795T, A23403G, G25563T, and G29734C) and 8 amino acid substitutions (ORF1a, Q768L; ORF1a, S2635L; ORF1a, Y3214; ORF1b, P314L; ORF1b, H1213Y; ORF3a, Q57H; S, R78M; and S, D614G), without amino acid deletions or PCR primer changes.

Data availability. The complete nucleotide sequence has been deposited in NCBI GenBank under accession number MW582699 and in the GISAID database under identifier EPI_ISL_971451.

ACKNOWLEDGMENTS

REFERENCES

1. 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.
2. 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.
3. Quiñones-Mateu ME, Avila S, Reyes-Teran G, Martinez MA. 2014. Deep sequencing: becoming a critical tool in clinical virology. J Clin Virol 61:9–19. https://doi.org/10.1016/j.jcv.2014.06.013.
4. Lu R, Niu P, Zhao L, Wang H, Wang W, Tan W. 2020. Sequencing the complete genome of COVID-19 virus from clinical samples using the Sanger method. China CDC Wkly 2:447–452. https://doi.org/10.46234/ccdcw2020.088.
5. Zhang Z, Schwartz S, Wagner L, Miller W. 2000. A greedy algorithm for aligning DNA sequences. J Comput Biol 7:203–214. https://doi.org/10.1089/10665270050081478.
6. 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.
7. Afgan E, Baker D, Batut B, van den Beek M, Bouvier D, Cech M, Chilton J, Clements D, Coraor N, Grüning BA, Guerler A, Hillman-Jackson J, Hiltemann S, Jalili V, Rasche H, Soranzo N, Goecks J, Taylor J, Nekrutenko A, Blankenberg D. 2018. The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update. Nucleic Acids Res 46:W537–W544. https://doi.org/10.1093/nar/gky379.
8. Kumar S, Stecher G, Li M, Knyaz C, Tamura K. 2018. MEGA X: molecular evolutionary genetics analysis across computing platforms. Mol Biol Evol 35:1547–1549. https://doi.org/10.1093/molbev/msy096.