AMERICAN SOCIETY FOR MICROBIOLOGY | Microbiology Resource Announcements

GENOME SEQUENCES

Genome Sequences of Two GH Clade SARS-CoV-2 Strains Isolated from Patients with COVID-19 in South Korea

Minwoo Kim,[a] Youn-Jung Lee,[b] Jae Sun Yoon,[b] Jin Young Ahn,[b] Jung Ho Kim,[b] Jun Yong Choi,[b] Jong-Won Oh[a]

[a]Department of Biotechnology, Yonsei University, Seoul, South Korea

[b]Division of Infectious Diseases, Department of Internal Medicine, Yonsei University College of Medicine, Seoul, South Korea

Minwoo Kim and Youn-Jung Lee contributed equally to this work. Author order was determined by drawing straws.

ABSTRACT We report the genome sequences of two GH clade severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) strains isolated from nasopharyngeal swabs from patients with coronavirus disease 2019 (COVID-19) in South Korea. These strains had two mutations in the untranslated regions and seven nonsynonymous substitutions in open reading frames, compared with Wuhan/Hu-1/2019, showing 99.96% sequence identity.

A novel coronavirus, severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), belonging to the genus *Betacoronavirus* of the family *Coronaviridae*, emerged in December 2019 in Wuhan, China (1, 2). The ongoing pandemic of coronavirus disease 2019 (COVID-19) caused by SARS-CoV-2 poses an unprecedented threat to global health and health care systems, with over 69 million laboratory-confirmed COVID-19 cases and more than 1.5 million casualties as of December 2020 (3).

Here, we report the genome sequences of two SARS-CoV-2 strains isolated from two patients with COVID-19 who were hospitalized in Severance Hospital, Yonsei University (Seoul, South Korea), in June 2020 during the COVID-19 pandemic. Nasopharyngeal swab specimens were collected on the day on which the patients tested positive for COVID-19 by real-time quantitative PCR using the Allplex 2019-nCoV assay kit (Seegene, Seoul, South Korea), with critical threshold values of 13.92 for patient 6 and 14.34 for patient 8. All of the studies were approved by the institutional review board (IRB) of Severance Hospital, Yonsei University Healthcare System, with written informed consent from the patients (IRB protocol number 4-2020-0076).

Using the QIAamp viral RNA minikit (Qiagen, Hilden, Germany), RNA was extracted from the virus, which had been purified by passaging the swab samples three times on Vero cells (ATCC CCL-81) by the limiting dilution method (4). Viral cDNA synthesized using ProtoScript II reverse transcriptase (New England Biolabs, Ipswich, MA, USA) was amplified as described previously (5, 6), using in-house-designed primer sets and the Illumina platform-based BTSeq SARS-CoV-2 whole-genome sequencing (WGS) kit (Celemics, Seoul, South Korea) for multiplex amplicon sequencing on a MiSeq sequencer (150-bp paired-end mode; Illumina, San Diego, CA, USA). After dual-index filtering and adapter trimming using in-house scripts, reads (69,447 and 66,754 reads for isolates YS006 and YS008, respectively) were mapped to the reference sequence of Wuhan/Hu-1/2019 (GenBank accession number MN988668) (nucleotides 1 to 29870) (7) with BWA v0.7.17-r1188 (8), generating consensus genome sequences of strains SARS-CoV-2/human/KOR/YS006/2020 (29,825 nucleotides) and SARS-CoV-2/human/KOR/YS008/2020 (29,826 nucleotides) isolated from patients 6 and 8, respectively, with average coverage depths of 98.65× and 95.5×, respectively. The consensus sequences for YS006 (nucleotides 16 to 29840) and YS008 (nucleotides 16 to 29841) had no indels. The nearly complete genomes of these isolates, which lack 15 nucleotides and 29 or 30

Citation Kim M, Lee Y-J, Yoon JS, Ahn JY, Kim JH, Choi JY, Oh J-W. 2021. Genome sequences of two GH clade SARS-CoV-2 strains isolated from patients with COVID-19 in South Korea. Microbiol Resour Announc 10:e01384-20. https://doi.org/10.1128/MRA.01384-20.

Editor Simon Roux, DOE Joint Genome Institute

Address correspondence to Jun Yong Choi, seran@yuhs.ac, or Jong-Won Oh, jwoh@yonsei.ac.kr.

Received 4 December 2020
Accepted 14 December 2020
Published 7 January 2021

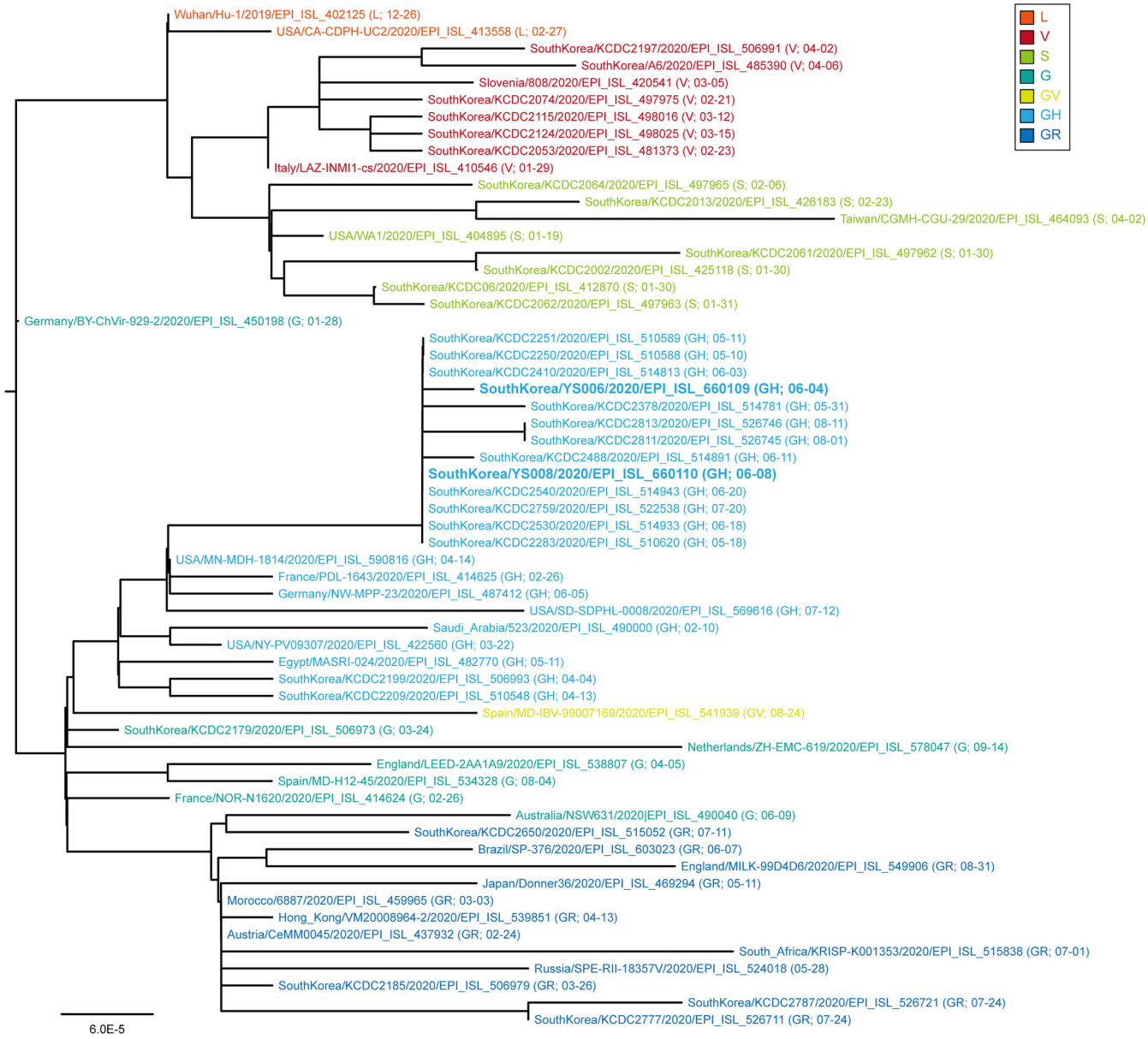

FIG 1 Phylogenetic tree of the genome sequences of SARS-CoV-2/human/KOR/YS006/2020 and SARS-CoV-2/human/KOR/YS008/2020, along with 58 other randomly chosen sequences, including 30 sequences of South Korean isolates of SARS-CoV-2 retrieved from Nextstrain (nextstrain.org). Multiple sequence alignments were conducted with MUSCLE v3.8.31 (12), and the phylogenetic tree was constructed using the Molecular Evolutionary Genetics Analysis (MEGA) v7.0 software by the neighbor-joining method with 1,000 bootstrap replicates (13). Shown in parentheses after the strain name and GISAID accession number are the clade to which it belongs and the collection date (month-day) for each strain. The sequences reported in this announcement are shown in bold. The SARS-CoV-2 isolates are presented in seven different colors for the different clades (shown in the key), classified according to GISAID classification. The scale bar represents the number of nucleotide substitutions per site.

nucleotides from their 5′ and 3′ ends [excluding the poly(A) tail], respectively (99.85% horizontal coverage), were 99.96% identical to the reference sequence. They both have a genomic GC content of 38%. Phylogenetic analysis revealed that these two isolates belong to the GH clade, which is currently most prevalent worldwide (9, 10), according to the GISAID classification (11) (Fig. 1). There were a total of 6 and 7 amino acid substitutions for YS006 and YS008, respectively, in comparison to the reference strain (Table 1). Both strains had C-to-T and G-to-T nucleotide changes in the 5′ untranslated region (UTR) and the 3′ UTR, respectively. A unique S6L substitution in the envelope (E) protein differentiated the YS008 strain from the YS006 and Wuhan/Hu-1/2019 strains.

TABLE 1 Nucleotide and amino acid changes in the YS006 and YS008 strains, in comparison to the reference strain

Nucleotide position	Nucleotide in strain (clade):			Gene name[b]	Amino acid change[c]
	Hu-1 (L)[a]	YS006 (GH)	YS008 (GH)		
241	C	T	T	5′ UTR	
1059	C	T	T	nsp2	T85I
3037	C	T	T	nsp3	
11916	C	T	T	nsp7	S25L
14408	C	T	T	nsp12	P323L
16650	C	T	T	nsp13	
20675	A	T	T	nsp16	Q6L
23403	A	G	G	Spike	D614G
25563	G	T	T	ORF3a	Q57H
26261	C	C	T	E	S6L[d]
29179	G	T	T	N	
29779	G	T	T	3′ UTR	

[a] Reference strain Wuhan/Hu-1/2019 (nucleotides 1 to 29870).
[b] ORF, open reading frame; nsp, nonstructural protein; E, envelope; N, nucleocapsid.
[c] Amino acid changes present in the YS006 and YS008 strains, in comparison to the reference strain.
[d] A nonsynonymous substitution present only in the YS008 strain.

Data availability. The sequences of SARS-CoV-2/human/KOR/YS006/2020 and SARS-CoV-2/human/KOR/YS008/2020 were deposited in the NCBI database (GenBank accession numbers MW345824 and MW345825, respectively) and in the GISAID database (https://www.gisaid.org) (accession numbers EPI_ISL_660109 and EPI_ISL_660110, respectively). The raw reads for the YS006 and YS008 strains were deposited in the NCBI Sequence Read Archive (SRA) database (accession numbers SRR13153716 and SRR13153715, respectively).

ACKNOWLEDGMENTS

This work was supported by National Research Foundation of Korea (NRF) grants (NRF 2019R1H1A2078176 and 2020M3E9A1041759) funded by the Ministry of Science and ICT (MIST), South Korea, and by the Brain Korea 21 (BK21) FOUR program (J.-W.O.). The work conducted by the Department of Internal Medicine, Yonsei University College of Medicine, was supported by the research program funded by the Korea Disease Control and Prevention Agency (2019-ER5408-00), by research grants (2019-ER5101-00, Korea HIV/AIDS Cohort Study, and HI14C1324) funded by the Ministry of Health and Welfare, Republic of Korea, to J.Y.C., and in part by a faculty research grant (2020) from the Department of Internal Medicine, Yonsei University College of Medicine, to J.H.K.

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. Zhu N, Zhang D, Wang W, Li X, Yang B, Song J, Zhao X, Huang B, Shi W, Lu R, Niu P, Zhan F, Ma X, Wang D, Xu W, Wu G, Gao GF, Tan W, China Novel Coronavirus Investigating and Research Team. 2020. A novel coronavirus from patients with pneumonia in China, 2019. N Engl J Med 382:727–733. https://doi.org/10.1056/NEJMoa2001017.
3. World Health Organization. 2020. Coronavirus disease 2019 (COVID-19) situation reports. https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports. Accessed 11 December 2020.
4. Harcourt J, Tamin A, Lu X, Kamili S, Sakthivel S, Murray J, Queen K, Tao Y, Paden C, Zhang J, Li Y, Uehara A, Wang H, Goldsmith C, Bullock H, Wang L, Whitaker B, Lynch B, Gautam R, Schindewolf C, Lokugamage K, Scharton D, Plante J, Mirchandani D, Widen S, Narayanan K, Makino S, Ksiazek T, Plante K, Weaver S, Lindstrom S, Tong S, Menachery V, Thornburg N. 2020. Severe acute respiratory syndrome coronavirus 2 from patient with coronavirus disease, United States. Emerg Infect Dis 26:1266–1273. https://doi.org/10.3201/eid2606.200516.
5. Oude Munnink BB, Nieuwenhuijse DF, Stein M, O'Toole A, Haverkate M, Mollers M, Kamga SK, Schapendonk C, Pronk M, Lexmond P, van der Linden A, Bestebroer T, Chestakova I, Overmars RJ, van Nieuwkoop S, Molenkamp R, van der Eijk AA, GeurtsvanKessel C, Vennema H, Meijer A, Rambaut A, van Dissel J, Sikkema RS, Timen A, Koopmans M, Dutch-Covid-19 Response Team. 2020. Rapid SARS-CoV-2 whole-genome sequencing and analysis for informed public health decision-making in the Netherlands. Nat Med 26:1405–1410. https://doi.org/10.1038/s41591-020-0997-y.
6. Quick J, Grubaugh ND, Pullan ST, Claro IM, Smith AD, Gangavarapu K, Oliveira G, Robles-Sikisaka R, Rogers TF, Beutler NA, Burton DR, Lewis-Ximenez LL, de Jesus JG, Giovanetti M, Hill SC, Black A, Bedford T, Carroll MW, Nunes M, Alcantara LC, Jr, Sabino EC, Baylis SA, Faria NR, Loose M, Simpson JT, Pybus OG, Andersen KG, Loman NJ. 2017. Multiplex PCR method for MinION and Illumina sequencing of Zika and other virus genomes directly from clinical samples. Nat Protoc 12:1261–1276. https://doi.org/10.1038/nprot.2017.066.
7. Chen L, Liu W, Zhang Q, Xu K, Ye G, Wu W, Sun Z, Liu F, Wu K, Zhong B, Mei Y, Zhang W, Chen Y, Li Y, Shi M, Lan K, Liu Y. 2020. RNA based mNGS

approach identifies a novel human coronavirus from two individual pneumonia cases in 2019 Wuhan outbreak. Emerg Microbes Infect 9:313–319. https://doi.org/10.1080/22221751.2020.1725399.

8. 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.
9. Alm E, Broberg EK, Connor T, Hodcroft EB, Komissarov AB, Maurer-Stroh S, Melidou A, Neher RA, O'Toole Á, Pereyaslov D. 2020. Geographical and temporal distribution of SARS-CoV-2 clades in the WHO European Region, January to June 2020. Eurosurveillance 25:2001410. https://doi.org/10.2807/1560-7917.ES.2020.25.32.2001410.
10. Mercatelli D, Giorgi FM. 2020. Geographic and genomic distribution of SARS-CoV-2 mutations. Front Microbiol 11:1800. https://doi.org/10.3389/fmicb.2020.01800.
11. GISAID. 2020. Clade and lineage nomenclature aids in genomic epidemiology studies of active hCoV-19 viruses. https://www.gisaid.org/references/statements-clarifications/clade-and-lineage-nomenclature-aids-in-genomic-epidemiology-of-active-hcov-19-viruses. Accessed 25 November 2020.
12. Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res 32:1792–1797. https://doi.org/10.1093/nar/gkh340.
13. 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.