



Coding-Complete Genome Sequence of a SARS-CoV-2 Variant Obtained from Raw Sewage at the University of Tennessee—Knoxville Campus

K. T. Ash,^{a,b} I. Alamillo,^b Y. Li,^b D. C. Joyner,^{a,b} D. E. Williams,^c P. J. McKay,^b B. M. Green,^d C. Iler,^e S. E. DeBlander,^f F. Kara-Murdoch,^{a,c,g} C. M. Swift,^{a,c}  T. C. Hazen^{a,b,d,h,i}

^aBiosciences Division, Oak Ridge National Laboratory, Oak Ridge, Tennessee, USA

^bDepartment of Civil and Environmental Sciences, University of Tennessee, Knoxville, Tennessee, USA

^cCenter for Environmental Biotechnology, University of Tennessee, Knoxville, Tennessee, USA

^dDepartment of Microbiology, University of Tennessee, Knoxville, Tennessee, USA

^eFacilities Services Department, University of Tennessee, Knoxville, Tennessee, USA

^fCollege of Natural Science, Michigan State University, East Lansing, Michigan, USA

^gBattelle Memorial Institute, Columbus, Ohio, USA

^hDepartment of Earth and Planetary Sciences, University of Tennessee, Knoxville, Tennessee, USA

ⁱInstitute for a Secure and Sustainable Environment, University of Tennessee, Knoxville, Tennessee, USA

ABSTRACT Reported here is a coding-complete genome sequence of a SARS-CoV-2 variant obtained from raw wastewater samples at the University of Tennessee—Knoxville campus. This sequence provides insight into SARS-CoV-2 variants that circulate on large college campuses but remain mostly undetected.

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) (family *Coronaviridae*, genus *Betacoronavirus*) first appeared in Wuhan, China, causing the novel infectious disease coronavirus 2019 (COVID-19) (1). As the disease began to spread across the globe, many cities turned to wastewater-based epidemiology (WBE) as a method to survey their populations for the presence of the SARS-CoV-2 virus (2). Universities, and other institutions, also began to use WBE methods to closely monitor their campuses for the presence of the SARS-CoV-2 virus (3).

The wastewater surveillance of student housing at University of Tennessee (UT)—Knoxville's campus began in September 2020. The purpose of the WBE surveillance project was to monitor on-campus student housing, in collaboration with saliva testing, to detect and isolate individuals who were infected or exposed to COVID-19. Wastewater was collected directly from 47 buildings via a dispense valve or manhole and tested once a week until June 2021.

Wastewater samples were pasteurized for 2 h at 60°C, followed by filtration through a 0.45- μ M nitrocellulose filter and then a 0.22- μ M nitrocellulose filter. Fifteen milliliters of the filtered samples was concentrated using the 10 kDa Amicon Ultra centrifugal filter units, and genomic RNA was extracted using a Qiagen QIAamp viral RNA minikit (4). The samples were tested for the presence of SARS-CoV-2 by reverse transcriptase quantitative PCR (RT-qPCR), which utilized TaqPath 1-step RT-qPCR master mix, CG, along with the CDC-developed primer/probe assays 2019-nCoV_N1 and 2019-nCoV_N3. Additionally, the primer/probe assay for the *Pepper mild mottle virus* (PMMoV) was used to serve as an internal positive control (4). All assays were performed at recommended conditions. Once testing was completed, the samples were stored at -80°C .

Sample UTK-840 was obtained from a sewer manhole in front of a student residential building on the UT-Knoxville campus, on 12 October 2020. This sample was selected for

Editor Simon Roux, DOE Joint Genome Institute

Copyright © 2021 Ash 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 T. C. Hazen, tchazen@utk.edu.

Received 26 October 2021

Accepted 3 November 2021

Published 24 November 2021

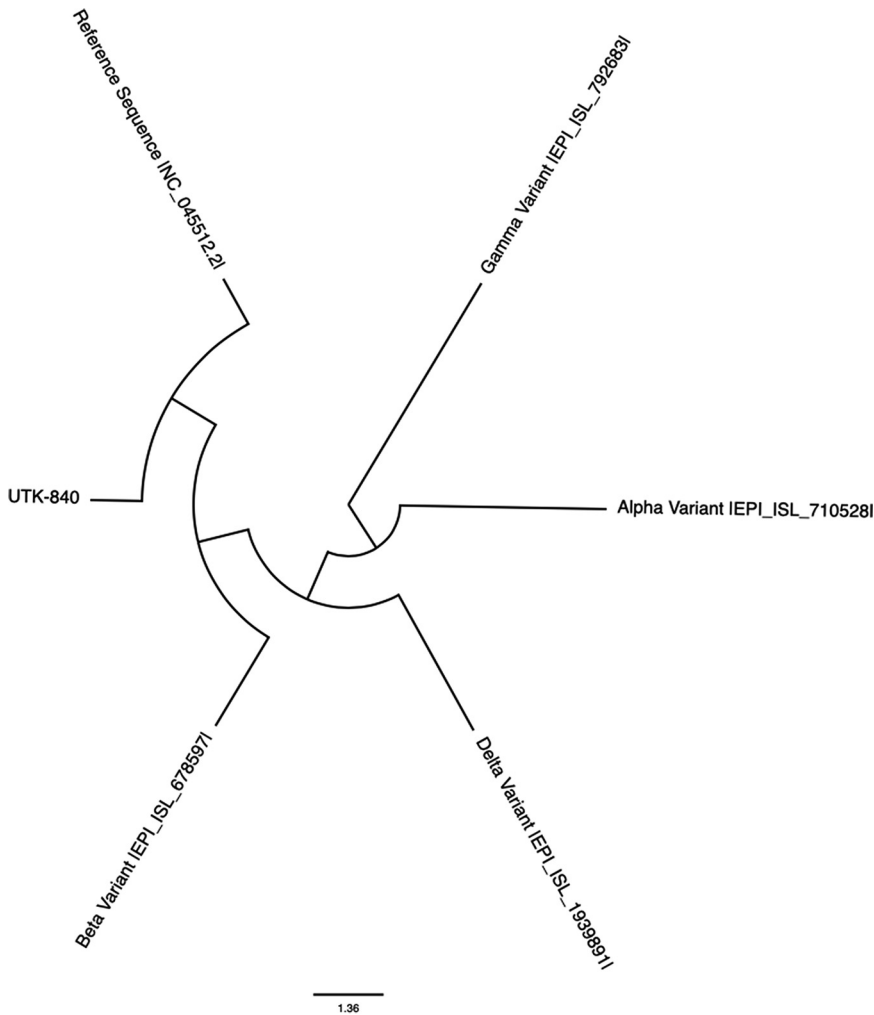


FIG 1 Phylogenetic tree of the UTK-840 variant in relation to the reference variant and known variants of concern. The evolutionary distance was inferred using the UPGMA method (9). The optimal tree is shown. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree (10). The evolutionary distances were computed using the Jukes-Cantor method (11) and are given in the number of base substitutions per site. This analysis involved 6 whole-genomic sequences. The codon positions included were first, second, third, and noncoding. All ambiguous positions were removed for each sequence pair (pairwise deletion option). There were a total of 29,907 positions in the final data set. Evolutionary analyses were conducted using MEGA11 (12, 13).

sequencing due to its high average quantification cycle (C_q) value ($C_q = 24.31$) from the RT-qPCR analysis. The sample was amplified using the PCR tiling ARTIC primer pools and prepared for sequencing using the NEBNext Ultra II DNA library prep kit for Illumina (NEB catalog number E7645L) (5). Sequencing was performed on a NovaSeq 6000 instrument at the Institute for Genome Sciences, University of Maryland. The genome assembly was produced using viral-ngs v1.25.0 utilizing viral-pipelines v2.1.19.0 from the Broad Institute (6). The Illumina paired-end reads were aligned to the SARS-CoV-2 reference strain (GenBank accession number [NC_045512.2](https://www.ncbi.nlm.nih.gov/nuccore/NC_045512.2)). The total number of reads aligned to the reference was 15,317,116, with a mean length of 249.7 bp. The mean coverage depth was 132,599.4 reads, with a standard deviation of 125,391.3 reads.

This SARS-CoV-2 isolate is 29,874 bp long, with a GC content of 39.3%. A BLAST comparison, under default parameters, revealed that the UTK-840 sequence shares 99.93% nucleotide similarity with the Wuhan reference strain (GenBank accession number [NC_045512.2](https://www.ncbi.nlm.nih.gov/nuccore/NC_045512.2)) (7). The differences between the sequences are represented by 20 single nucleotide mutations, of which 15 are transition and 5 are transversion mutations. Half of these

mutations occurred in the Orf1ab gene. Three mutations are in the nucleocapsid gene and only one in the spike gene, which is the D614G mutation that has been observed in most of the variants of concern. A sequence comparison made using the Audacity Instant program on GISAID, under default parameters, revealed the UTK-840 sequence to be closely related to lineage B.1.2 and the GH neighbor clade (8) (Fig. 1).

Data availability. The whole-genome sequence has been deposited at GenBank under the accession number [MZ265375](https://www.ncbi.nlm.nih.gov/nuccore/MZ265375). The raw sequencing reads have been deposited at the NCBI SRA database under the accession number [SRR16202037](https://www.ncbi.nlm.nih.gov/sra/SRR16202037).

ACKNOWLEDGMENTS

We acknowledge the University of Tennessee Office of Research and Engagement for providing funding for this project and the Maryland Institute for Genome Sciences (IGS) Microbiome Service Laboratory (MSL) (University of Maryland) for sequencing of the sample.

REFERENCES

- 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>.
- World Health Organization. 2020. Status of environmental surveillance for SARS-CoV-2 virus: scientific brief, 5 August 2020. World Health Organization, Geneva, Switzerland.
- Betancourt WQ, Schmitz BW, Innes GK, Prasek SM, Pogreba Brown KM, Stark ER, Foster AR, Sprissler RS, Harris DT, Sherchan SP, Gerba CP, Pepper IL. 2021. COVID-19 containment on a college campus via wastewater-based epidemiology, targeted clinical testing and an intervention. *Sci Total Environ* 779: 146408. <https://doi.org/10.1016/j.scitotenv.2021.146408>.
- Wu F, Xiao A, Zhang J, Moniz K, Endo N, Armas F, Bonneau R, Brown MA, Bushman M, Chai PR, Duvallet C, Erickson TB, Foppe K, Ghaeli N, Gu X, Hanage WP, Huang KH, Lee WL, Matus M, McElroy KA, Nagler J, Rhode SF, Santillana M, Tucker JA, Wuertz S, Zhao S, Thompson J, Alm EJ. 2020. SARS-CoV-2 RNA concentrations in wastewater foreshadow dynamics and clinical presentation of new COVID-19 cases. *Sci Total Environ* 805: 150121. <https://doi.org/10.1016/j.scitotenv.2021.150121>.
- Itokawa K, Sekizuka T, Hashino M, Tanaka R, Kuroda M. 2020. Disentangling primer interactions improves SARS-CoV-2 genome sequencing by multiplex tiling PCR. *PLoS One* 15:e0239403. <https://doi.org/10.1371/journal.pone.0239403>.
- Park D, Tomkins-Tinch C, Ye S, Jungreis I, Shlyakhter I, Metsky H, Hanna, Lin M, Le V, Lin A, pvanheus, Krasilnikova LA, Leinweber K. 2019. broadinstitute/viralngs: v1.25.0. Zenodo. <https://doi.org/10.5281/zenodo.3509008>.
- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *J Mol Biol* 215:403–410. [https://doi.org/10.1016/S0022-2836\(05\)80360-2](https://doi.org/10.1016/S0022-2836(05)80360-2).
- Shu Y, McCauley J. 2017. GISAID: global initiative on sharing all influenza data—from vision to reality. *Euro Surveill* 22:30494. <https://doi.org/10.2807/1560-7917.ES.2017.22.13.30494>.
- Sneath PHA, Sokal RR. 1973. Numerical taxonomy: the principles and practice of numerical classification. W. H. Freeman and Co, San Francisco, CA.
- Felsenstein J. 1985. Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* 39:783–791. <https://doi.org/10.1111/j.1558-5646.1985.tb00420.x>.
- Jukes TH, Cantor CR. 1969. Evolution of protein molecules, p 21–132. *In* Munro HN (ed), *Mammalian protein metabolism*. Academic Press, Cambridge, MA. <https://doi.org/10.1016/B978-1-4832-3211-9.50009-7>.
- Tamura K, Stecher G, Kumar S. 2021. MEGA11: Molecular Evolutionary Genetics Analysis version 11. *Mol Biol Evol* 38:3022–3027. <https://doi.org/10.1093/molbev/msab120>.
- Stecher G, Tamura K, Kumar S. 2020. Molecular Evolutionary Genetics Analysis (MEGA) for macOS. *Mol Biol Evol* 37:1237–1239. <https://doi.org/10.1093/molbev/msz312>.