

An 81-Nucleotide Deletion in SARS-CoV-2 ORF7a Identified from Sentinel Surveillance in Arizona (January to March 2020)

LaRinda A. Holland,^a Emily A. Kaelin,^{a,b} Rabia Maqsood,^a Bereket Estifanos,^{b,c} Lily I. Wu,^a  Arvind Varsani,^{a,b,d,e} Rolf U. Halden,^{f,g} Brenda G. Hogue,^{b,c,h}  Matthew Scotch,^{f,i}  Efrem S. Lim^{a,b,c}

^aCenter for Fundamental and Applied Microbiomics, Biodesign Institute, Arizona State University, Tempe, Arizona, USA

^bSchool of Life Sciences, Arizona State University, Tempe, Arizona, USA

^cCenter for Immunotherapy, Vaccines and Virotherapy, Biodesign Institute, Arizona State University, Tempe, Arizona, USA

^dCenter of Evolution and Medicine, Arizona State University, Tempe, Arizona, USA

^eStructural Biology Research Unit, Department of Integrative Biomedical Sciences, University of Cape Town, Observatory, Cape Town, South Africa

^fCenter for Environmental Health Engineering, Biodesign Institute, Arizona State University, Tempe, Arizona, USA

^gOne Water One Health Nonprofit Project, Arizona State University Foundation, Tempe, Arizona, USA

^hCenter for Applied Structural Discovery, Biodesign Institute, Arizona State University, Tempe, Arizona, USA

ⁱCollege of Health Solutions, Arizona State University, Phoenix, Arizona, USA

LaRinda A. Holland and Emily A. Kaelin contributed equally to this work. Author order was determined alphabetically.

KEYWORDS COVID-19, ORF7a, SARS-CoV-2

On 26 January 2020, the first coronavirus disease 2019 (COVID-19) case was reported in Arizona (third case in the United States) (1). Here, we report on early severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) sentinel surveillance in Tempe, Arizona. Genomic characterization identified an isolate encoding a 27-amino-acid in-frame deletion in accessory protein ORF7a, the ortholog of SARS-CoV immune antagonist ORF7a/X4.

In anticipation of COVID-19 spreading in Arizona, we initiated a surveillance effort for the local emergence of SARS-CoV-2 starting 24 January 2020. We leveraged an ongoing influenza surveillance project at Arizona State University (ASU) Health Services in Tempe, Arizona. Individuals presenting with respiratory symptoms (ILI) were tested for influenza A and B viruses (Alere BinaxNOW). Subsequently, we tested influenza-negative nasopharyngeal (NP) swabs for SARS-CoV-2. We extracted total nucleic acid using the bioMérieux eMAG automated platform and performed real-time quantitative reverse transcription-PCR (qRT-PCR) assays specific for SARS-CoV-2 N and E genes (2, 3). Out of 382 NP swabs collected from 24 January to 25 March 2020, we detected SARS-CoV-2 in 5 swabs from 16 to 19 March (Fig. 1). This corresponds to a prevalence of 1.31%. Given the estimated 1- to 14-day incubation period for COVID-19, the spike in cases might be related to university spring break holiday travel (8 to 15 March), as previously seen in other outbreaks (4, 5).

To understand the evolutionary relationships and characterize the SARS-CoV-2 genomes, we performed next-generation sequencing (NGS; Illumina NextSeq, 2 \times 76) directly on specimen RNA, thereby avoiding cell culture passage and potentially associated mutations. This generated an NGS data set of 20.7 to 22.7 million paired-end reads per sample. We mapped quality-filtered reads to a reference SARS-CoV-2 genome (MN908947) using BBMap (version 39.64) to generate three full-length genomes: AZ-ASU2922 (376 \times coverage), AZ-ASU2923 (50 \times), and AZ-ASU2936 (879 \times) (Geneious prime, version 2020.0.5). We aligned a total of 222 SARS-CoV-2 genome sequences comprising at least 5 representative sequences from phylogenetic lineages defined by Rambaut et al. (6). We performed phylogenetic reconstruction with BEAST (version 1.10.4; strict molecular clock, HKY + Γ nucleotide substitution, exponential growth for coalescent model) (7–10). The ASU sequences were phylogenetically distinct, indicating independent transmissions (Fig. 2A).

Citation Holland LA, Kaelin EA, Maqsood R, Estifanos B, Wu LI, Varsani A, Halden RU, Hogue BG, Scotch M, Lim ES. 2020. An 81-nucleotide deletion in SARS-CoV-2 ORF7a identified from sentinel surveillance in Arizona (January to March 2020). *J Virol* 94:e00711-20. <https://doi.org/10.1128/JVI.00711-20>.

Editor Tom Gallagher, Loyola University Chicago

Copyright © 2020 Holland 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 Efrem S. Lim, Efrem.Lim@asu.edu.

Accepted manuscript posted online 1 May 2020

Published 1 July 2020

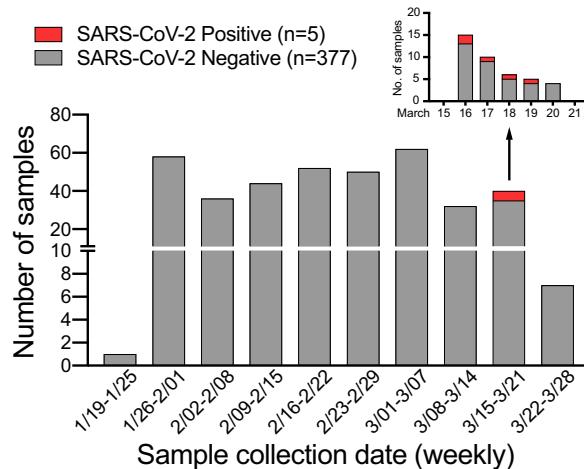


FIG 1 SARS-CoV-2 surveillance in Tempe, AZ, from January to March 2020. Shown is the weekly distribution of NP specimens collected by ASU Health Services and tested for SARS-CoV-2 by qRT-PCR assays. The inset shows SARS-CoV-2-positive NP specimens collected from the week of 15 to 21 March 2020.

Like SARS-CoV, the SARS-CoV-2 genome carries multiple open reading frames in the 3' region. We found that the SARS-CoV-2 AZ-ASU2923 genome has an 81-nucleotide (nt) deletion in the ORF7a gene, resulting in a 27-amino-acid in-frame deletion (Fig. 2B). The SARS-CoV ORF7a ortholog is a viral antagonist of host restriction factor BST-2/

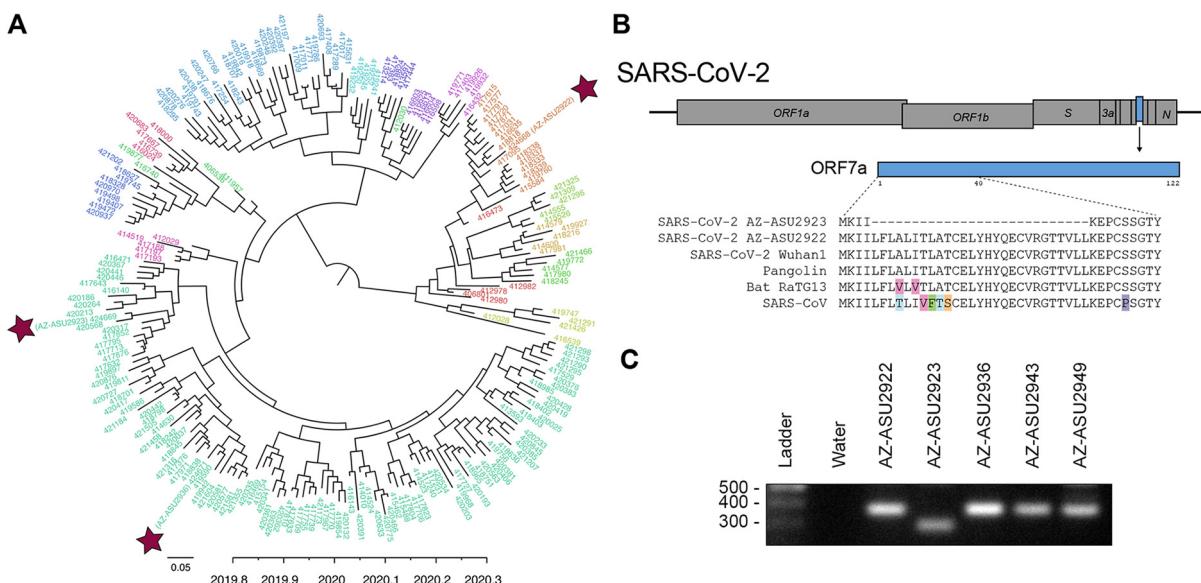


FIG 2 Evolutionary and genomic characterization of SARS-CoV-2 genomes. (A) Bayesian maximum clade credibility (MCC) polar phylogeny of 222 full-length SARS-CoV-2 genomes. The 3 new genomes reported in this study are indicated by red stars. Sequences were aligned in Geneious prime (version 2020.0.3) using the MAFFT v7.450 plugin, and the 5' and 3' untranslated regions (<300 nt each) were trimmed. We initiated two independent runs of 500M sampling every 50K steps and used Tracer v1.7.1 (18) to check for convergence and that all effective sample size (ESS) values for our statistics were >200, LogCombiner (7) to combine the models with a 10% burn-in, and TreeAnnotator (7) to produce an MCC tree. We used FigTree, v1.4.4 (19), to edit the tree and color the tips based on lineages (6) and pangolin (20) to identify the lineages of our 3 new sequences based on the established nomenclature (6). The nomenclature consists of two main lineages, A and B, and includes sublineages (A.1, B.2, etc.) up to four levels deep (e.g., A.1.1, B.2.1, etc.) (6). For visualization purposes, we grouped all viruses that were not directly assigned to A or B into their first sublineage level and colored tip labels by lineage. B.1 lineage, AZ-ASU2923 and AZ-ASU2936; A.1 lineage, AZ-ASU2922. (B) ORF7a amino acid alignment of SARS-CoV-2 and related genomes. GenBank and GISAID accession numbers were the following: SARS-CoV-2 AZ-ASU2922, MT39039 and EPI_ISL_424668; SARS-CoV-2 AZ-ASU2923, MT39040 and EPI_ISL_424669; SARS-CoV-2 AZ-ASU2936, MT39041 and EPI_ISL_424671; SARS-CoV-2 Wuhan1, MN908947.3; Pangolin, EPI_ISL_410721; Bat-RaTG13, MN996532.1; and SARS-CoV, AY278741.1. The 81-nt (27-amino-acid) deletion observed in SARS-CoV-2 AZ-ASU2923 ORF7a was not present in the 6,290 SARS-CoV-2 sequences available from GISAID as of 12 April 2020. (C) We performed molecular validation by RT-PCR on specimen total nucleic acid extracts with primers flanking the ORF7a N terminus region. The expected size of amplicons with an intact ORF7a region is 377 bp, and the expected size of an amplicon with the NGS-identified 81-nt deletion is 296 bp. Primers were SARS2-27144F, 5'-ACAGACCATTCCAGTAGCAGTG-3', and SARS2-27520r, 5'-TGCCCTCGATGTTCCAGAAG-3'.

Tetherin and induces apoptosis (11–14). Based on the SARS-CoV ORF7a structure (15), the 27-amino-acid deletion in SARS-CoV-2 ORF7a maps to the putative signal peptide (partial) and first two beta strands. To validate the deletion, we performed RT-PCR using primers spanning the region and verified the amplicons by Sanger sequencing (Fig. 2C). Similar deletions in SARS-CoV-2 genomes are emerging, notably in the ORF8 gene (16), that may reduce virus fitness (17). Hence, further experiments are needed to determine the functional consequences of the ORF7a deletion.

Collectively, although global NGS efforts indicate that SARS-CoV-2 genomes are relatively stable, dynamic mutations can be selected in symptomatic individuals.

ACKNOWLEDGMENTS

We thank the nurses and staff at ASU Health Services, Arizona Department of Health Services, for a SARS-CoV-2-positive sample (AZ_4811) for qRT-PCR assay validation experiments, Nicholas Mellor and the ASU Genomics Facility for technical assistance, and the authors and originating and submitting laboratories of the sequences from the GISAID's EpiCoV database. A complete acknowledgments table for sequences obtained from this database is available at https://www.dropbox.com/s/aiybuatgxjunuga/GISAID_CoV2020_Acknowledgements.xlsx?dl=0.

This work was supported by NSF STC award 1231306 (B.G.H.), NIH grants R01 LM013129 (R.U.H., A.V., and M.S.) and R00 DK107923 (E.S.L.), J. M. Kaplan Foundation's One Water One Health (Arizona State University Foundation project 30009070), and ASU Core Facilities Seed Funding.

REFERENCES

- Patel A, 2019-nCoV CDC Response Team, Jernigan DB. 2020. Initial public health response and interim clinical guidance for the 2019 novel coronavirus outbreak—United States, December 31, 2019–February 4, 2020. *MMWR Morb Mortal Wkly Rep* 69:140–146. <https://doi.org/10.15585/mmwr.mm6905e1>.
- CDC. 2020. 2019-Novel coronavirus (2019-nCoV) real-time rRT-PCR panel primers and probes. Centers for Disease Control and Prevention, Atlanta, GA. <https://www.cdc.gov/coronavirus/2019-ncov/downloads/rt-pcr-panel-primer-probes.pdf>. Accessed 4 March 2020.
- 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, Wijnsman 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://doi.org/10.2807/1560-7917.ES.2020.25.3.2000045>.
- Lauer SA, Grantz KH, Bi Q, Jones FK, Zheng Q, Meredith HR, Azman AS, Reich NG, Lessler J. 10 March 2020. The incubation period of coronavirus disease 2019 (COVID-19) from publicly reported confirmed cases: estimation and application. *Ann Intern Med* <https://doi.org/10.7326/m20-0504>.
- Polgreen PM, Bohnett LC, Yang M, Pentella MA, Cavanaugh JE. 2010. A spatial analysis of the spread of mumps: the importance of college students and their spring-break-associated travel. *Epidemiol Infect* 138: 434–441. <https://doi.org/10.1017/S0950268809990719>.
- Rambaut A, Holmes EC, Hill V, O'Toole A, McCrone JT, Ruis C, du Plessis L, Pybus OG. 2020. A dynamic nomenclature for SARS-CoV-2 to assist genomic epidemiology. *bioRxiv* <https://doi.org/10.1101/2020.04.17.046086>.
- Suchard MA, Lemey P, Baele G, Ayres DL, Drummond AJ, Rambaut A. 2018. Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10. *Virus Evol* 4:vey016. <https://doi.org/10.1093/ve/vey016>.
- Hasegawa M, Kishino H, Yano T-A. 1985. Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. *J Mol Evol* 22: 160–174. <https://doi.org/10.1007/bf02101694>.
- Yang Z. 1994. Maximum likelihood phylogenetic estimation from DNA sequences with variable rates over sites: approximate methods. *J Mol Evol* 39:306–314. <https://doi.org/10.1007/bf00160154>.
- Pybus OG, Rambaut A. 2002. GENIE: estimating demographic history from molecular phylogenies. *Bioinformatics* 18:1404–1405. <https://doi.org/10.1093/bioinformatics/18.10.1404>.
- Taylor JK, Coleman CM, Postel S, Sisk JM, Bernbaum JG, Venkataraman T, Sundberg EJ, Frieman MB. 2015. Severe acute respiratory syndrome coronavirus ORF7a inhibits bone marrow stromal antigen 2 virion tethering through a novel mechanism of glycosylation interference. *J Virol* 89:11820–11833. <https://doi.org/10.1128/JVI.02274-15>.
- Yuan X, Wu J, Shan Y, Yao Z, Dong B, Chen B, Zhao Z, Wang S, Chen J, Cong Y. 2006. SARS coronavirus 7a protein blocks cell cycle progression at G0/G1 phase via the cyclin D3/pRb pathway. *Virology* 346:74–85. <https://doi.org/10.1016/j.virol.2005.10.015>.
- Tan YJ, Fielding BC, Goh PY, Shen S, Tan TH, Lim SG, Hong W. 2004. Overexpression of 7a, a protein specifically encoded by the severe acute respiratory syndrome coronavirus, induces apoptosis via a caspase-dependent pathway. *J Virol* 78:14043–14047. <https://doi.org/10.1128/JVI.78.24.14043-14047.2004>.
- Schaecher SR, Touchette E, Schriewer J, Buller RM, Pekosz A. 2007. Severe acute respiratory syndrome coronavirus gene 7 products contribute to virus-induced apoptosis. *J Virol* 81:11054–11068. <https://doi.org/10.1128/JVI.01266-07>.
- Nelson CA, Pekosz A, Lee CA, Diamond MS, Fremont DH. 2005. Structure and intracellular targeting of the SARS-coronavirus Orf7a accessory protein. *Structure* 13:75–85. <https://doi.org/10.1016/j.str.2004.10.010>.
- Su YC, Anderson DE, Young BE, Zhu F, Linster M, Kalimuddin S, Low JG, Yan Z, Jayakumar J, Sun L, Yan GZ, Mendenhall IH, Leo Y-S, Lye DC, Wang L-F, Smith GJ. 2020. Discovery of a 382-nt deletion during the early evolution of SARS-CoV-2. *bioRxiv* <https://doi.org/10.1101/2020.03.11.987222>.
- Muth D, Corman VM, Roth H, Binger T, Dijkman R, Gottula LT, Gloza-Rausch F, Balboni A, Battilani M, Rihtarić D, Toplak I, Ameneiros RS, Pfeifer A, Thiel V, Drexler JF, Müller MA, Drosten C. 2018. Attenuation of replication by a 29 nucleotide deletion in SARS-coronavirus acquired during the early stages of human-to-human transmission. *Sci Rep* 8:15177. <https://doi.org/10.1038/s41598-018-33487-8>.
- Rambaut A, Drummond AJ, Xie D, Baele G, Suchard MA. 2018. Posterior summarization in bayesian phylogenetics using Tracer 1.7. *Syst Biol* 67:901–904. <https://doi.org/10.1093/sysbio/syy032>.
- Rambaut A. 2018. FigTree v1.4.4. <https://github.com/rambaut/figtree>.
- O'Toole A, McCrone J. 2020. pangolin. <https://github.com/hCoV-2019-pangolin>.