medRxiv preprint doi: https://doi.org/10.1101/2020.04.17.20069641.this version posted April 22, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

1 **An 81 base-pair deletion in SARS-CoV-2 ORF7a identified from sentinel surveillance in**
2 **Arizona (Jan-Mar 2020)**

3 LaRinda A. Holland[1,*], Emily A. Kaelin[1,2,*], Rabia Maqsood[1], Bereket Estifanos[2,3], Lily I. Wu[1],
4 Arvind Varsani[1,2,4,5], Rolf U. Halden[6,7], Brenda G. Hogue[2,3,8], Matthew Scotch[6,9],
5 Efrem S. Lim[1,2,3,#]

6 [1]Center for Fundamental and Applied Microbiomics, Biodesign Institute, Arizona State
7 University, Tempe, Arizona, USA.

8 [2]School of Life Sciences, Arizona State University, Tempe, Arizona, USA.

9 [3]Center for Immunotherapy, Vaccines and Virotherapy, Biodesign Institute, Arizona State
10 University, Tempe, Arizona, USA.

11 [4]Center of Evolution and Medicine, Arizona State University, Tempe, Arizona, USA.

12 [5]Structural Biology Research Unit, Department of Integrative Biomedical Sciences, University of
13 Cape Town, Observatory, Cape Town 7701, South Africa

14 [6]Center for Environmental Health Engineering, Biodesign Institute, Arizona State University,
15 Tempe, Arizona, USA.

16 [7]One Water One Health Nonprofit Project, Arizona State University Foundation, Tempe,
17 Arizona, USA.

18 [8]Center for Applied Structural Discovery, Biodesign Institute, Arizona State University, Tempe,
19 Arizona, USA.

20 [9]College of Health Solutions, Arizona State University, Phoenix, Arizona, USA.

21 * These authors equally contributed to this work.

22 [#]To whom correspondence should be addressed. Email: Efrem.Lim@asu.edu

medRxiv preprint doi: https://doi.org/10.1101/2020.04.17.20069641.this version posted April 22, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

23 On January 26 2020, the first Coronavirus Disease 2019 (COVID-19) case was reported in
24 Arizona of an individual with travel history (3rd case in the US) (1). Here, we report on early
25 SARS-CoV-2 sentinel surveillance in Tempe, Arizona (USA). Genomic characterization
26 identified an isolate encoding a 27 amino acid in-frame deletion in accessory protein ORF7a,
27 the ortholog of SARS-CoV immune antagonist ORF7a/X4.

28 In anticipation of COVID-19 spreading in the state of Arizona, we initiated a surveillance effort
29 for local emergence of SARS-CoV-2 starting January 24, 2020. We leveraged an ongoing
30 influenza surveillance project at Arizona State University (ASU) Health Services in Tempe,
31 Arizona. Individuals presenting with respiratory symptoms (ILI) were tested for influenza A and
32 B virus (Alere BinaxNOW). Subsequently, we tested influenza-negative nasopharyngeal (NP)
33 swabs for SARS-CoV-2. We extracted total nucleic acid using the bioMérieux eMAG automated
34 platform and performed real-time RT-PCR (qRT-PCR) assays specific for SARS-CoV-2 N and E
35 genes (2, 3). Out of 382 NP swabs collected from January 24, 2020 to March 25, 2020, we
36 detected SARS-CoV-2 in 5 swabs in the week of March 16 to 19 (**Figure 1**). This corresponds
37 to prevalence of 1.31%. Given the estimated 1 – 14-day incubation period for COVID-19, it is
38 possible that the spike in cases might be related to university spring-break holiday travel (March
39 8 – 15) as previously seen in other outbreaks (4, 5).

40 To understand the evolutionary relationships and characterize the SARS-CoV-2 genomes, we
41 performed next-generation sequencing (Illumina NextSeq, 2x76) directly on specimen RNA,
42 thereby avoiding cell culture passage and potentially associated mutations. This generated an
43 NGS dataset of 20.7 to 22.7 million paired-end reads per sample. We mapped quality-filtered
44 reads to a reference SARS-CoV-2 genome (MN908947) using BBMap (version 39.64) to
45 generate three full-length genomes: AZ-ASU2922 (376x coverage), AZ-ASU2923 (50x) and AZ-
46 ASU2936 (879x) (Geneious prime version 2020.0.5). We aligned a total of 222 SARS-CoV-2
47 genome sequences comprising at least 5 representative sequences from phylogenetic lineages

medRxiv preprint doi: https://doi.org/10.1101/2020.04.17.20069641.this version posted April 22, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

48 defined by Rambaut *et al.* (6), ranging from January 5 to March 31, 2020 from 25 different
49 countries. We performed phylogenetic reconstruction with BEAST (version 1.10.4, strict
50 molecular clock, HKY + Γ nucleotide substitution, exponential growth for coalescent model) (7-
51 10). The ASU sequences were phylogenetically distinct indicating that they were independent
52 transmissions (**Figure 2A**).

53 Similar to SARS-CoV, the SARS-CoV-2 genome encodes multiple open reading frames in the 3'
54 region. We found that the SARS-CoV-2 AZ-ASU2923 genome has an 81 base-pair deletion in
55 the ORF7a gene resulting in a 27 amino-acid in-frame deletion (**Figure 2B**). The SARS-CoV
56 ORF7a ortholog is a viral antagonist of host restriction factor BST-2/Tetherin and induces
57 apoptosis (11-14). Based on the SARS-CoV ORF7a structure (15), the 27-aa deletion in SARS-
58 CoV-2 ORF7a maps to the putative signal peptide (partial) and first two beta strands. To
59 validate the deletion, we performed RT-PCR using primers spanning the region and verified by
60 Sanger sequencing the amplicons (**Figure 2C and Supplementary Figure 1**).

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

63

64 **Acknowledgements**

65 We thank the nurses and staff at the ASU Health Services, Arizona Department of Health
66 Services for a SARS-CoV-2 positive sample (AZ_4811) for qRT-PCR assay validation
67 experiments, Nicholas Mellor and the ASU Genomics Facility for technical assistance, the
68 authors, originating and submitting laboratories of the sequences from GISAID's EpiCoV™
69 Database. A complete acknowledgements table is available at
70 https://www.dropbox.com/s/aiybuatgxjunuga/GISAID_CoV2020_Acknowledgements.xlsx?dl=0 .
71 This work was supported by NSF STC Award 1231306 (B.G.H), NIH grants R01 LM013129

medRxiv preprint doi: https://doi.org/10.1101/2020.04.17.20069641.this version posted April 22, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

72 (R.U.H., A.V., M.S.), R00 DK107923 (E.S.L.), J.M. Kaplan Foundation's One Water One Health
73 (Arizona State University Foundation project 30009070) and ASU Core Facilities Seed Funding.

74

75 **Data availability**

76 Sequence data has been deposited to NCBI GenBank and GISAID: SARS-CoV-2 AZ-ASU2922
77 (MT339039, EPI_ISL_424668), SARS-CoV-2 AZ-ASU2923 (MT339040, EPI_ISL_424669) and
78 SARS-CoV-2 AZ-ASU2936 (MT339041, EPI_ISL_424671).

79

80 **References**
81

82 1. Patel A, Jernigan DB. 2020. Initial Public Health Response and Interim Clinical Guidance for
83 the 2019 Novel Coronavirus Outbreak - United States, December 31, 2019-February 4,
84 2020. MMWR Morb Mortal Wkly Rep 69:140-146.
85 2. CDC. 2020. 2019-Novel Coronavirus (2019-nCoV) Real-time rRT-PCR Panel
86 Primers and Probes. https://www.cdc.gov/coronavirus/2019-ncov/downloads/rt-pcr-panel-
87 primer-probes.pdf Accessed 3/4/2020.
88 3. Corman VM, Landt O, Kaiser M, Molenkamp R, Meijer A, Chu DKW, Bleicker T, Brunink S,
89 Schneider J, Schmidt ML, Mulders D, Haagmans BL, van der Veer B, van den Brink S,
90 Wijsman L, Goderski G, Romette JL, Ellis J, Zambon M, Peiris M, Goossens H, Reusken C,
91 Koopmans MPG, Drosten C. 2020. Detection of 2019 novel coronavirus (2019-nCoV) by
92 real-time RT-PCR. Euro Surveill 25.
93 4. Lauer SA, Grantz KH, Bi Q, Jones FK, Zheng Q, Meredith HR, Azman AS, Reich NG,
94 Lessler J. 2020. The Incubation Period of Coronavirus Disease 2019 (COVID-19) From
95 Publicly Reported Confirmed Cases: Estimation and Application. Ann Intern Med
96 doi:10.7326/m20-0504.

medRxiv preprint doi: https://doi.org/10.1101/2020.04.17.20069641.this version posted April 22, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

97 5. Polgreen PM, Bohnett LC, Yang M, Pentella MA, Cavanaugh JE. 2010. A spatial analysis of
98 the spread of mumps: the importance of college students and their spring-break-associated
99 travel. Epidemiol Infect 138:434-41.
100 6. Rambaut A, Holmes EC, Pybus OG. 2020. A dynamic nomenclature for SARS-CoV-2 to
101 assist genomic epidemiology. http://virological.org/t/a-dynamic-nomenclature-for-sars-cov-2-
102 to-assist-genomic-epidemiology/458.
103 7. Suchard MA, Lemey P, Baele G, Ayres DL, Drummond AJ, Rambaut A. 2018. Bayesian
104 phylogenetic and phylodynamic data integration using BEAST 1.10. Virus Evol 4:vey016.
105 8. Hasegawa M, Kishino H, Yano T-a. 1985. Dating of the human-ape splitting by a molecular
106 clock of mitochondrial DNA. Journal of Molecular Evolution 22:160-174.
107 9. Yang Z. 1994. Maximum likelihood phylogenetic estimation from DNA sequences with
108 variable rates over sites: Approximate methods. Journal of Molecular Evolution 39:306-314.
109 10. Pybus OG, Rambaut A. 2002. GENIE: estimating demographic history from molecular
110 phylogenies. Bioinformatics 18:1404-1405.
111 11. Taylor JK, Coleman CM, Postel S, Sisk JM, Bernbaum JG, Venkataraman T, Sundberg EJ,
112 Frieman MB. 2015. Severe Acute Respiratory Syndrome Coronavirus ORF7a Inhibits Bone
113 Marrow Stromal Antigen 2 Virion Tethering through a Novel Mechanism of Glycosylation
114 Interference. J Virol 89:11820-33.
115 12. Yuan X, Wu J, Shan Y, Yao Z, Dong B, Chen B, Zhao Z, Wang S, Chen J, Cong Y. 2006.
116 SARS coronavirus 7a protein blocks cell cycle progression at G0/G1 phase via the cyclin
117 D3/pRb pathway. Virology 346:74-85.
118 13. Tan YJ, Fielding BC, Goh PY, Shen S, Tan TH, Lim SG, Hong W. 2004. Overexpression of
119 7a, a protein specifically encoded by the severe acute respiratory syndrome coronavirus,
120 induces apoptosis via a caspase-dependent pathway. J Virol 78:14043-7.

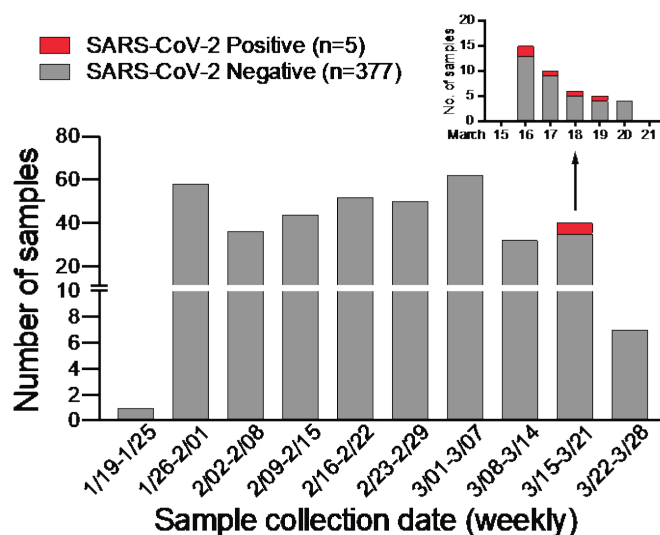
medRxiv preprint doi: https://doi.org/10.1101/2020.04.17.20069641.this version posted April 22, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

121 14. Schaecher SR, Touchette E, Schriewer J, Buller RM, Pekosz A. 2007. Severe acute
122 respiratory syndrome coronavirus gene 7 products contribute to virus-induced apoptosis. J
123 Virol 81:11054-68.
124 15. Nelson CA, Pekosz A, Lee CA, Diamond MS, Fremont DH. 2005. Structure and intracellular
125 targeting of the SARS-coronavirus Orf7a accessory protein. Structure 13:75-85.
126 16. Rambaut A, Drummond AJ, Xie D, Baele G, Suchard MA. 2018. Posterior Summarization in
127 Bayesian Phylogenetics Using Tracer 1.7. Syst Biol 67:901-904.
128 17. Rambaut A. 2018. FigTree v1.4.4, https://github.com/rambaut/figtree.
129 18. O'Toole A, McCrone J. 2020. pangolin, https://github.com/hCoV-2019/pangolin.

130

131

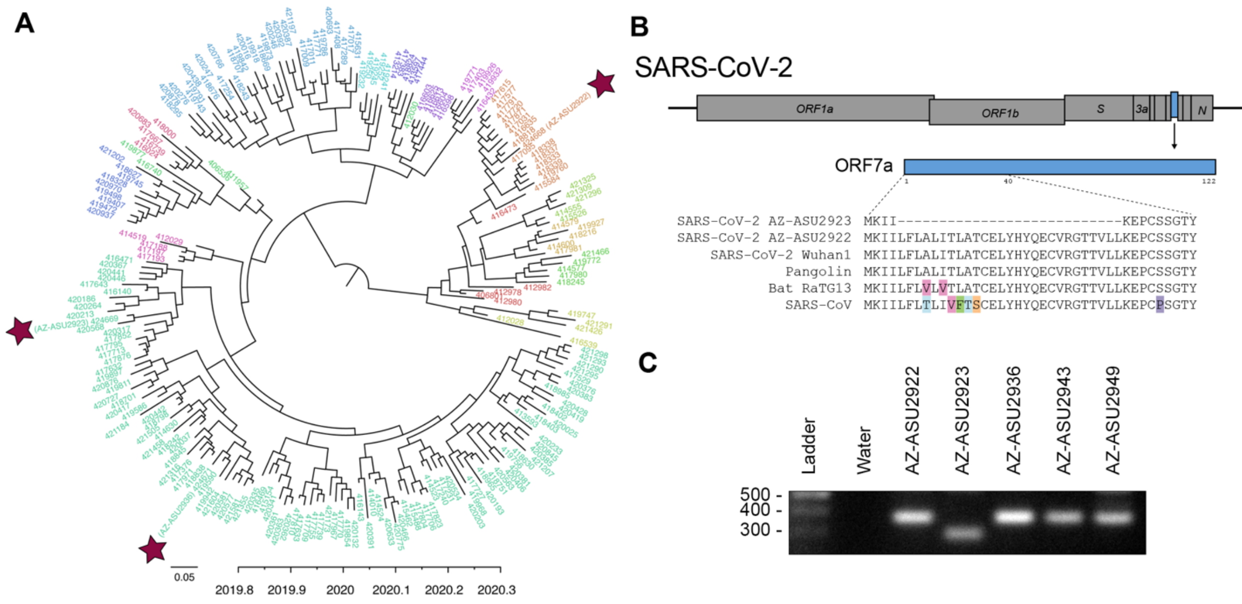
132 **Figure legends**



133

134 **Figure 1: SARS-CoV-2 surveillance in Tempe, AZ from January to March 2020.**
135 Weekly distribution of NP specimens collected by ASU Health Services tested for SARS-CoV-2
136 by qRT-PCR assays. Inset shows SARS-CoV-2 positive NP specimens collected from the week
137 of March 15 – 21, 2020.

medRxiv preprint doi: https://doi.org/10.1101/2020.04.17.20069641.this version posted April 22, 2020. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .



138

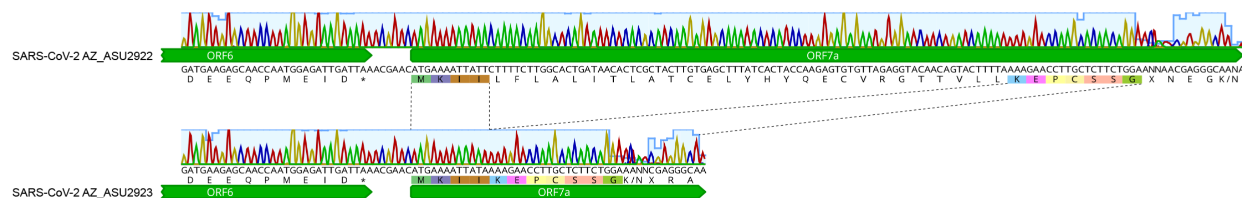
139 **Figure 2: Evolutionary and genomic characterization of SARS-CoV-2 genomes.**

140 (**A**) Bayesian maximum clade credibility (MCC) polar phylogeny of 222 full-length SARS-CoV-2
141 genomes. The 3 new genomes reported in this study are indicated by red stars. Sequenced
142 were aligned in Geneious prime (version 2020.0.3) using the MAFFT v7.450 plugin, and
143 trimmed the 5' and 3' UTR (< 300 nts each). We initiated two independent runs of 500M
144 sampling every 50K steps and used Tracer v1.7.1 (16) to check for convergence and that all
145 ESS values for our statistics were > 200, LogCombiner (7) to combine the models with a 10%
146 burn-in and TreeAnnotator (7) to produce a MCC tree. We used FigTree v1.4.4 (17) to edit the
147 tree and color the tips based on lineages (6), and pangolin (18) to identify the lineages of our 3
148 new sequences based on the established nomenclature (6). The nomenclature consists of two
149 main lineages, A and B, and includes "sub-lineages" (A.1, B.2. *etc.*) up to four levels deep (e.g.
150 A.1.1, B.2.1) (6). For visualization purposes, we grouped all viruses that were not directly
151 assigned to "A" or "B" into their first sub-lineage level and colored tip labels by lineage. B.1
152 lineage: AZ-ASU2923 and AZ-ASU2936; A.1 lineage AZ-ASU2922. (**B**) ORF7a amino acid
153 alignment of SARS-CoV-2 and related genomes. GenBank and GISAID accession numbers:
154 SARS-CoV-2 AZ-ASU2922 (MT339039, EPI_ISL_424668), SARS-CoV-2 AZ-ASU2923

medRxiv preprint doi: https://doi.org/10.1101/2020.04.17.20069641.this version posted April 22, 2020. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

155 (MT339040, EPI_ISL_424669), SARS-CoV-2 AZ-ASU2936 (MT339041, EPI_ISL_424671),
156 SARS-CoV-2 Wuhan1 (MN908947.3), Pangolin (EPI_ISL_410721), Bat-RaTG13
157 (MN996532.1), SARS-CoV (AY278741.1). The 81-bp (27 amino acid) deletion observed in
158 SARS-CoV-2 AZ_ASU2923 ORF7a was not present in the 6,290 SARS-CoV-2 sequences
159 available from GISAID as of April 12, 2020. (**C**) We performed molecular validation by RT-PCR
160 on specimen total nucleic acid extracts with primers flanking the ORF7a N-terminus region. The
161 expected size of amplicons with intact ORF7a region is 377bp, the expected size of an amplicon
162 with the NGS-identified 81bp deletion is 296bp. Primers: SARS2-27144F 5'-
163 ACAGACCATTCCAGTAGCAGTG-3', SARS2-27520r 5'-TGCCCTCGTATGTTCCAGAAG-3'.

164

165 

166 **Supplementary Figure 1.** Sanger sequencing chromatograms verification of 81-bp deletion in
167 SARS-CoV-2 AZ-ASU2923 in the ORF7a region.