Contents lists available at ScienceDirect

Journal of Clinical Virology

CORRESPONDENCE

A 336-nucleotide in-frame deletion in ORF7a gene of SARS-CoV-2 identified in genomic surveillance by next-generation sequencing

Public health laboratories (PHLs) across the US have been taking a central role in the response to the coronavirus disease 2019 (COVID-19) pandemic [1]. Currently, PHLs have increased their capacity to sequence whole genomes of the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), the causative agent of COVID-19, which enhances surveillance. SARS-CoV-2 genomes are known to use non-canonical translation mechanisms such as leaky scanning, ribosomal frameshifting, and alternative initiation, as is commonly observed in the coronavirus family [2–6]. Here, we report the longest in-frame deletion in ORF7a gene identified in a patient using Next-Generation Sequencing (NGS) during baseline surveillance.

A nasal sample was collected on July 7th, 2021, from a previously tested SARS-CoV-2 positive pediatric asymptomatic female in San Mateo County Public Health Laboratory, California, USA. The presence of SARS-CoV-2 in a specimen was confirmed by the CDC 2019-Novel Coronavirus Real-Time RT-PCR Diagnostic Panel, that targets two unique regions in the SARS-CoV-2 nucleocapsid gene (N1 and N2) and the human RNase P gene (RP) [7], with the cycle threshold values of 15.4, 15.6, and 26.8, respectively. Subsequently, the total nucleic acids from the SARS-CoV-2 positive specimen were used for NGS library preparation utilizing the Respiratory Pathogen ID/AMR Enrichment Panel (RPIP) Kit (Illumina). The concentration and distribution size of the prepared library was checked on a Qubit 4 fluorometer (Life Technologies) and 4200 TapeStation (Agilent Technologies), respectively. Paired-end NGS was performed on an Illumina MiniSeq instrument using the 150-cycle MiniSeq high output reagent kit (Illumina). Sequence data were assembled using the Explify RPIP Analysis App v.3.2.8 (IDbyDNA) and validated by the viral-ngs v.2.1.8 workflows (Broad Viral Genomics) on the Terra platform (app.terra.bio) [8]. Briefly, the SARS-CoV-2 genome was assembled using a reference-based assembly approach of quality-filtered raw reads with the reference genome Wuhan-Hu-1 (NC_045512.2) [9]. A total of 8363,544 raw paired-end reads were generated. The genome consensus of the isolate, designated CA-SMCPHL-072321.3, was called from 2,071,958 SARS-CoV-2 reads using the minimap2 aligner [10]. This resulted in the mean read depth coverage of 5,066 reads. The final genome consensus was 29,548 nucleotides (nt) long and was deposited in GISAID under the accession number EPI_ISL_6159211. The genome was classified as

Nextstrain clade 20I [11] from Pango lineage B.1.1.7 [12]. During the annotation of this genome, we noticed a 336-nt deletion in the gene-encoding accessory protein ORF7a (27,418–27,753) (Fig. 1). Although numerous studies have been performed for some other SARS-CoV-2 viral proteins [13–17], studies on putative activity and role of ORF7a are just starting to arise. It has been demonstrated that SARS-CoV-2 ORF7a protein inhibits type I interferon (IFN-1) signaling [18,19], interacts with CD14+ monocytes [20], induces the nuclear factor kappa B (NF-κB) pathway [21], and thus triggers expression of proinflammatory cytokines, including interleukin-6 (IL-6) and tumor necrosis factor alpha (TNF-α). This forms the basis of a likely mechanism
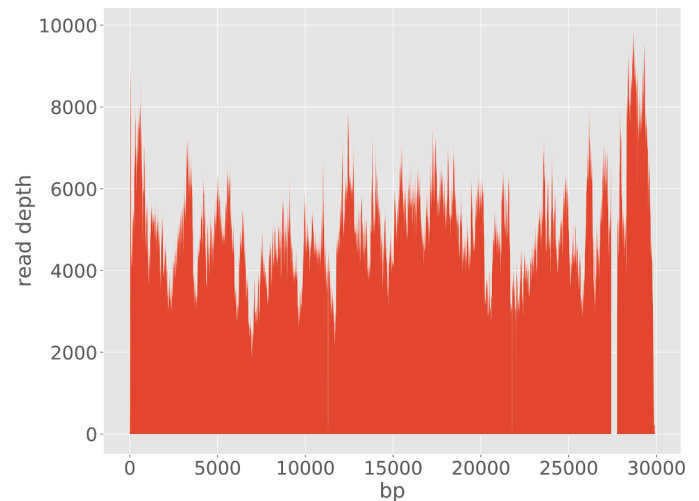

Fig. 1. Coverage plot across hCoV-19/USA/CA-SMCPHL-072321.3/2021 (EPI_ISL_6159211) whole-genome assembly. Coverage depth values were obtained by mapping Illumina reads to the reference Wuhan-Hu-1 genome NC_045512.2 and expressed as the number of sequence reads at each nucleotide position along the length of the CA-SMCPHL-072321.3 genome, which shows a deletion in the ORF7a gene (around 27.5 Kbp). Mean sequencing depth was 5,066 reads.
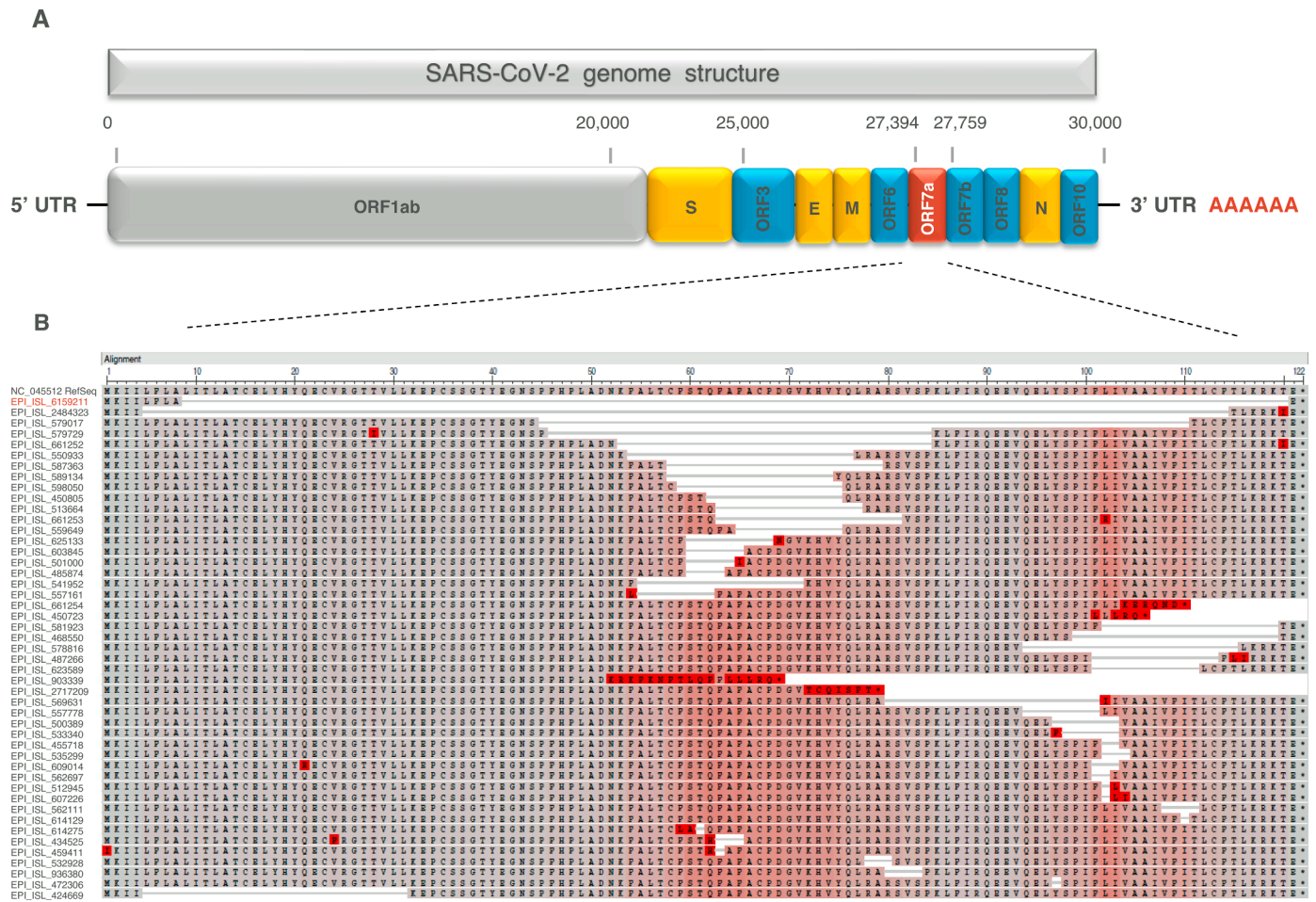
Fig. 2. Complete SARS-CoV-2 genome organization diagram (A) and alignment of ORF7a gene sequences (B). To avoid sequences with gaps due to poor sequencing coverage, only sequences containing truncations and deletions in the ORF7a gene, that had high sequencing coverage, were downloaded from GISAID (Table S1). The multiple amino acid alignments of the ORF7a gene sequences ($n = 47$) with reference genome Wuhan-Hu-1 NC_045512.2 were performed by MUSCLE algorithm in MEGA software v. 7.0.26. Frequency-based differences coloring of aligned ORF7a proteins were visualized in NCBI MSA Viewer v. 1.20.1. The GISAID accession number of the hCoV-19/USA/CA-SMCPHL-072321.3/2021 (EPI_ISL_6159211) from this study is shown in red. The 336-nt (112-amino acid) deletion in CA-SMCPHL-072321.3 ORF7a was the longest observed deletion within the ORF7a gene sequences available from GISAID and GenBank as of October 2021.

through which ORF7a mediates the potentially fatal cytokine storm progression, indicating that ORF7a may be a key viral factor for clinical severity of COVID-19 [22].

Previous studies reported complete ORF7a gene loss [23,24], multiple length in-frame and frame-shift deletions in ORF7a [25–31], as well as large ORF7a$^{\Delta370}$, ORF7a$^{\Delta227}$, and ORF7a$^{\Delta392}$ that resulted in the fusion of ORF7a with ORF6, ORF7b, and ORF8 genes, respectively [32, 33]. However, the ORF7a$^{\Delta336}$ reported here is currently the longest detected in-frame deletion within the ORF7a gene, which makes the gene only 30 nt long (Fig. 2). ORF7a deletions are found to impact SARS-CoV-2 pathogenicity [31]. *In vitro* viral challenge experiments demonstrated that the C-terminal truncation of ORF7a results in a replication defect, which was found to be associated with elevated IFN response to SARS-CoV-2 [31]. It was also shown that complete deletion of ORF7a reduces viral replication [34]. This suggests that strains with deletions in ORF7a are more likely to emerge in immunocompromised patients. Hence, further experiments are needed to determine the functional outcomes of different deletions.

Unfortunately, a variety of deletions in the ORF7a region can be under-investigated similarly to ORF8 deletions that have been shown to often not be reported by the standard data analysis pipelines, which are frequently simply represented by stretch ambiguous bases or as gaps in the consensus sequence due to poor NGS coverage [35]. Thus, non-canonical genes are generally excluded from genomic and clinical

analyses despite their importance for understanding SARS-CoV-2 evolution and replication dynamics, which have vital implications in vaccine development and control strategies for COVID-19 [6,36,37]. These findings highlight the necessity of submission of the raw sequencing reads in public databases in order to assess the spread of deletion strains.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper

A table of the contributors is available in Supplementary Table S1.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.jcv.2022.105105.

References

[1] C. Huang, Y. Wang, X. Li, L. Ren, J. Zhao, Y. Hu, L. Zhang, G. Fan, J. Xu, X. Gu, Z. Cheng, T. Yu, J. Xia, Y. Wei, W. Wu, X. Xie, W. Yin, H. Li, M. Liu, Y. Xiao, H. Gao, L. Guo, J. Xie, G. Wang, R. Jiang, Z. Gao, Q. Jin, J. Wang, B Cao, Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China, Lancet 395 (2020) 497–506.

[2] Ardern Z., Wei X., Nelson W.C. 2011. SARS-CoV-2: don't ignore non-canonical genes. https://virological.org. Accessed October 02, 2018.

[3] A.E. Firth, A putative new SARS-CoV protein, 3c, encoded in an ORF overlapping ORF3a, J. Gen. Virol. 101 (2020) 1085–1089.

[4] I. Jungreis, C.W. Nelson, Z. Ardern, Y. Finkel, N.J. Krogan, K. Sato, J. Ziebuhr, N. Stern-Ginossar, A. Pavesi, A.E. Firth, A.E. Gorbalenya, M. Kellis, Conflicting and ambiguous names of overlapping ORFs in the SARS-CoV-2 genome: a homology-based resolution, Virology 558 (2021) 145–151.

[5] B. Malone, N. Urakova, E.J. Snijder, E.A. Campbell, Structures and functions of coronavirus replication-transcription complexes and their relevance for SARS-CoV-2 drug design, Nat. Rev. Mol. Cell Biol. 23 (2022) 21–39.

[6] A.E. Firth, I. Brierley, Non-canonical translation in RNA viruses, J. Gen. Virol. 93 (2012) 1385–1409.

[7] X. Lu, L. Wang, S.K. Sakthivel, B. Whitaker, J. Murray, S. Kamili, B. Lynch, L. Malapati, S.A. Burke, J. Harcourt, A. Tamin, N.J. Thornburg, J.M. Villanueva, S. Lindstrom, US CDC real-time reverse transcription PCR panel for detection of severe acute respiratory syndrome Coronavirus 2, Emerg. Infect. Dis. 26 (2020).

[8] J.E. Lemieux, K.J. Siddle, B.M. Shaw, C. Loreth, S.F. Schaffner, A. Gladden-Young, G. Adams, T. Fink, C.H. Tomkins-Tinch, L.A. Krasilnikova, K.C. DeRuff, M. Rudy, M.R. Bauer, K.A. Lagerborg, E. Normandin, S.B. Chapman, S.K. Reilly, M. N. Anahtar, A.E. Lin, A. Carter, C. Myhrvold, M.E. Kemball, S. Chaluvadi, C. Cusick, K. Flowers, A. Neumann, F. Cerrato, M. Farhat, D. Slater, J.B. Harris, J.A. Branda, D. Hooper, J.M. Gaeta, T.P. Baggett, J. O'Connell, A. Gnirke, T.D. Lieberman, A. Philippakis, M. Burns, C.M. Brown, J. Luban, E.T. Ryan, S.E. Turbett, R. C. LaRocque, W.P. Hanage, G.R. Gallagher, L.C. Madoff, S. Smole, V.M. Pierce, E. Rosenberg, et al., Phylogenetic analysis of SARS-CoV-2 in Boston highlights the impact of superspreading events, Science 371 (6529) (2021) eabe3261.

[9] F. Wu, S. Zhao, B. Yu, Y.M. Chen, W. Wang, Z.G. Song, Y. Hu, Z.W. Tao, J.H. Tian, Y.Y. Pei, M.L. Yuan, Y.L. Zhang, F.H. Dai, Y. Liu, Q.M. Wang, J.J. Zheng, L. Xu, E. C. Holmes, Y.Z. Zhang, A new coronavirus associated with human respiratory disease in China, Nature 579 (2020) 265–269.

[10] H. Li, Minimap2: pairwise alignment for nucleotide sequences, Bioinformatics 34 (2018) 3094–3100.

[11] J. Hadfield, C. Megill, S.M. Bell, J. Huddleston, B. Potter, C. Callender, P. Sagulenko, T. Bedford, R.A. Neher, Nextstrain: real-time tracking of pathogen evolution, Bioinformatics 34 (2018) 4121–4123.

[12] A. Rambaut, E.C. Holmes, Á. O'Toole, V. Hill, J.T. McCrone, C. Ruis, Plessis L du, O.G. Pybus, A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology, Nat. Microbiol. 5 (2020) 1403–1407.

[13] S. Kang, M. Yang, Z. Hong, L. Zhang, Z. Huang, X. Chen, S. He, Z. Zhou, Q. Chen, Y. Yan, C. Zhang, H. Shan, S Chen, Crystal structure of SARS-CoV-2 nucleocapsid protein RNA binding domain reveals potential unique drug targeting sites, Acta Pharm. Sin. B 10 (2020) 1228–1238.

[14] J. Lan, J. Ge, J. Yu, S. Shan, H. Zhou, S. Fan, Q. Zhang, X. Shi, Q. Wang, L. Zhang, X Wang, Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor, Nature 581 (2020) 215–220.

[15] A.C. Walls, Y.J. Park, M.A. Tortorici, A. Wall, A.T. McGuire, D. Veesler, Structure, function, and antigenicity of the SARS-CoV-2 spike glycoprotein, Cell 183 (2020) 1735.

[16] D. Wrapp, N. Wang, K.S. Corbett, J.A. Goldsmith, C.L. Hsieh, O. Abiona, B. S. Graham, J.S McLellan, Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation, Science 367 (2020) 1260–1263.

[17] R. Yan, Y. Zhang, Y. Li, L. Xia, Y. Guo, Q. Zhou, Structural basis for the recognition of SARS-CoV-2 by full-length human ACE2, Science 367 (2020) 1444–1448.

[18] Z. Cao, H. Xia, R. Rajsbaum, X. Xia, H. Wang, P.Y Shi, Ubiquitination of SARS-CoV-2 ORF7a promotes antagonism of interferon response, Cell. Mol. Immunol. 18 (2021) 746–748.

[19] H. Xia, Z. Cao, X. Xie, X. Zhang, J.Y. Chen, H. Wang, V.D. Menachery, R. Rajsbaum, P.Y Shi, Evasion of type I interferon by SARS-CoV-2, Cell Rep. 33 (2020) 108234.

[20] Z. Zhou, C. Huang, Z. Huang, L. Su, S. Kang, X. Chen, Q. Chen, S. He, X. Rong, F. Xiao, J. Chen, S. Chen, Structural insight reveals SARS-CoV-2 ORF7a as an immunomodulating factor for human CD14, iScience 24 (2021), 102187.

[21] C.M. Su, L. Wang, D Yoo, Activation of NF-κB and induction of proinflammatory cytokine expressions mediated by ORF7a protein of SARS-CoV-2, Sci. Rep. 11 (2021) 13464.

[22] A. Lobiuc, D. Șterbuleac, O. Sturdza, M. Dimian, M Covasa, A conservative replacement in the transmembrane domain of SARS-CoV-2 ORF7a as a putative risk factor in COVID-19, Biology (Basel) 10 (2021).

[23] H. Tse, D.C. Lung, S.C. Wong, K.F. Ip, T.C. Wu, K.K. To, K.H. Kok, K.Y. Yuen, G. K. Choi, Emergence of a severe acute respiratory syndrome Coronavirus 2 virus variant with novel genomic architecture in Hong Kong, Clin. Infect. Dis. 73 (2021) 1696–1699.

[24] N. Mazur-Panasiuk, L. Rabalski, T. Gromowski, G. Nowicki, M. Kowalski, W. Wydmanski, P. Szulc, M. Kosinski, K. Gackowska, N. Drweska-Matelska, J. Grabowski, A. Piotrowska-Mietelska, B. Szewczyk, K. Bienkowska-Szewczyk, J. Swadzba, P. Labaj, M. Grzybek, K Pyrc, Expansion of a SARS-CoV-2 Delta variant with an 872nt deletion encompassing, Euro Surveill. 26 (2021).

[25] K. Joonlasak, E.M. Batty, T. Kochakarn, B. Panthan, K. Kümpornsin, P. Jiaranai, A. Wangwiwatsin, A. Huang, N. Kotanan, P. Jaru-Ampornpan, W. Manasatienkij, T. Watthanachockchai, K. Rakmanee, A.R. Jones, S. Fernandez, I. Sensorn, S. Sungkanuparph, E. Pasomsub, C. Klungthong, T. Chookajorn, W Chantratita, Genomic surveillance of SARS-CoV-2 in Thailand reveals mixed imported populations, a local lineage expansion and a virus with truncated ORF7a, Virus Res. 292 (2021), 198233.

[26] Rosenthal S.H., Kagan R.M., Gerasimova A., Anderson B., Grover D., Hua M., Liu Y., Owen R., Lacbawan F. 2020. Identification of eight SARS-CoV-2 ORF7a deletion variants in 2,726 clinical specimens. bioRxiv 2020.12.10.418855.

[27] A.T. Pyke, N. Nair, A.F. van den Hurk, P. Burtonclay, S. Nguyen, J. Barcelon, C. Kistler, S. Schlebusch, J. McMahon, F. Moore, Replication kinetics of B.1.351 and B.1.1.7 SARS-CoV-2 variants of concern including assessment of a B.1.1.7 mutant carrying a defective ORF7a gene, Viruses 13 (2021).

[28] Panzera Y., Ramos N., Frabasile S., Calleros L., Marandino A., Tomás G., Techera C., Grecco S., Fuques E., Goñi N., Ramas V., Coppola L., Chiparelli H., Sorhouet C., Mogdasy C., Arbiza J., Delfraro A., Pérez R. 2021. A deletion in SARS-CoV-2 ORF7 identified in COVID-19 outbreak in Uruguay. Transbound Emerg. Dis.

[29] Foster C.S.P., Rawlinson W.D. 2021. Rapid spread of a SARS-CoV-2 Delta variant with a frameshift deletion in ORF7a. medRxiv 2021.08.18.21262089.

[30] L.A. Holland, E.A. Kaelin, R. Maqsood, B. Estifanos, L.I. Wu, A. Varsani, R. U. Halden, B.G. Hogue, M. Scotch, E.S. Lim, An 81-nucleotide deletion in SARS-CoV-2 ORF7a identified from sentinel surveillance in Arizona (January to March 2020), J. Virol. 94 (2020).

[31] A. Nemudryi, A. Nemudraia, T. Wiegand, J. Nichols, D.T. Snyder, J.F. Hedges, C. Cicha, H. Lee, K.K. Vanderwood, D. Bimczok, M.A. Jutila, B. Wiedenheft, SARS-CoV-2 genomic surveillance identifies naturally occurring truncation of ORF7a that limits immune suppression, Cell Rep. 35 (2021), 109197.

[32] A. Addetia, H. Xie, P. Roychoudhury, L. Shrestha, M. Loprieno, M.L. Huang, K. R. Jerome, A.L Greninger, Identification of multiple large deletions in ORF7a resulting in in-frame gene fusions in clinical SARS-CoV-2 isolates, J. Clin. Virol. 129 (2020), 104523.

[33] H. Tse, S.C. Wong, K.F. Ip, V.C. Cheng, K.K. To, D.C. Lung, G.K. Choi, Genome sequences of three SARS-CoV-2 ORF7a deletion variants obtained from patients in Hong Kong, Microbiol. Resour. Announc. 10 (2021).

[34] T. Thi Nhu Thao, F. Labroussaa, N. Ebert, P. V'kovski, H. Stalder, J. Portmann, J. Kelly, S. Steiner, M. Holwerda, A. Kratzel, M. Gultom, K. Schmied, L. Laloli, L. Hüsser, M. Wider, S. Pfaender, D. Hirt, V. Cippà, S. Crespo-Pomar, S. Schröder, D. Muth, D. Niemeyer, V.M. Corman, M.A. Müller, C. Drosten, R. Dijkman, J. Jores, V Thiel, Rapid reconstruction of SARS-CoV-2 using a synthetic genomics platform, Nature 582 (2020) 561–565.

[35] D. Brandt, M. Simunovic, T. Busche, M. Haak, P. Belmann, S. Jünemann, T. Schulz, L.J. Klages, S. Vinke, M. Beckstette, E. Pohl, C. Scherer, A. Sczyrba, J. Kalinowski, Multiple occurrences of a 168-nucleotide deletion in SARS-CoV-2 ORF8, unnoticed by standard amplicon sequencing and variant calling pipelines, Viruses 13 (2021).

[36] J.S.Y. Ho, Z. Zhu, I. Marazzi, Unconventional viral gene expression mechanisms as therapeutic targets, Nature 593 (2021) 362–371.

[37] A. Pavesi, Origin, evolution and stability of overlapping genes in viruses: a systematic review, Genes (Basel) 12 (2021).

Shantelle Lucas, Morris Saffold Jones II, Sangita Kothari, Adrian Madlambayan, Christopher Ngo, Carmen Chan, Iryna V. Goraichuk[*]

San Mateo County Public Health Laboratory, 225 37th Avenue, San Mateo, CA 94403, United States of America

[*] Corresponding author.
E-mail address: goraichuk@outlook.com (I.V. Goraichuk).