The Journal of Infectious Diseases

BRIEF REPORT

Yearlong COVID-19 Infection Reveals Within-Host Evolution of SARS-CoV-2 in a Patient With B-Cell Depletion

Veronique Nussenblatt,[1,a] Allison E. Roder,[2,a] Sanchita Das,[3] Emmie de Wit,[4] Jung-Ho Youn,[3] Stephanie Banakis[2], Alexandra Mushegian,[2] Christopher Mederos,[2] Wei Wang,[2] Matthew Chung,[2] Lizzette Pérez-Pérez,[4] Tara Palmore,[5] Jennifer N. Brudno,[6] James N. Kochenderfer,[6] and Elodie Ghedin[2]

[1]Laboratory of Clinical Immunology and Microbiology, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, Maryland, USA, [2]Systems Genomics Section, Laboratory of Parasitic Diseases, Division of Intramural Research, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, Maryland, USA, [3]Department of Laboratory Medicine, National Institutes of Health, Bethesda, Maryland, USA, [4]Laboratory of Virology, Division of Intramural Research, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Hamilton, Montana, USA, [5]Clinical Center, National Institutes of Health, Bethesda, Maryland, USA, and [6]Surgery Branch, Center for Cancer Research, National Cancer Institute, National Institutes of Health, Bethesda, Maryland, USA

(See the Editorial Commentary by Lemieux and Luban, on pages 1115–7.)

B-cell–depleting therapies may lead to prolonged disease and viral shedding in individuals infected with severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) and this viral persistence raises concern for viral evolution. We report sequencing of early and late samples from a 335-day infection in an immunocompromised patient. The virus accumulated a unique deletion in the amino-terminal domain of the spike protein, and complete deletion of ORF7b and ORF8, the first report of its kind in an immunocompromised patient. Unique viral mutations found in this study highlight the importance of analyzing viral evolution in protracted SARS-CoV-2 infection, especially in immunosuppressed hosts.

Keywords. COVID-19; SARS-CoV-2; prolonged infection; immunocompromised; viral evolution.

Cell-mediated and humoral immunity are necessary to clear severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) infection [1]. Individuals receiving B-cell–depleting therapies can have protracted disease and prolonged viral shedding [2, 3]. Persistent shedding of viral RNA for weeks to months after onset of symptoms has been reported; however, viable virus is often not detected after 9 days after illness onset [4]. In contrast, viral replication has been detected in immunocompromised patients for several months after initial infection [2, 5, 6]. Persistent viral replication in these patients is likely the result of profound lymphocyte defects due to B- and T-cell–depleting therapies or underlying hematologic disease. Viral persistence in the setting of immunosuppression has raised concern for viral evolution and emergence of variants, especially during treatment with convalescent plasma [3]. In addition to single-nucleotide variants, recent studies have demonstrated that SARS-CoV-2 in immunocompromised hosts is prone to deletion mutations in the spike protein, especially in the S1 region [2, 5, 6]. Deletions across the genome can reflect virus-host interactions and are found in both immunocompetent and immunosuppressed hosts.

Here, we report on a patient with persistent symptomatic viral infection over a period of 335 days. Viral genome sequencing revealed the emergence of 2 unique deletions and showed fixation of early minority variants, displaying viral evolution, a concern in the context of immunosuppression.

METHODS

Approval

Written consent was obtained for human research subjects, as approved by the National Institutes of Health Institutional Review Board (protocol No. NCT02659943).

RNA and Subgenomic RNA qPCR

Detection of the N gene or ORF1a/b was performed on all specimens collected. Amplification of subgenomic transcripts for the E gene (sgE) was done prospectively on samples after day 275, and retrospectively on samples before, as described previously [5].

SARS-CoV-2 Sequencing and Sequence Analysis

Amplification of viral genomes, library preparation, and genome analysis was done according to the protocols available at https://github.com/GhedinSGS/SARS-CoV-2_analysis. Libraries were sequenced on the Illumina NextSeq500 using the 2 × 150 bp paired end protocol. Adapters and primers were trimmed, reads were aligned to the Wuhan/Hu-1 strain (NC_045512.2), and the 2 libraries for each sample were merged. Consensus sequences and variants were identified using the *timo* variant calling pipeline.

Phylogeny and Lineage Identification

Phylogenetic trees containing 266 background sequences from Maryland (obtained from GISAID; Supplementary Table 1) were generated using Nextstrain with default parameters [7]. Lineages were called using Pangolin [8].

Received 23 October 2021; editorial decision 3 December 2021; accepted 21 December 2021; published online 23 December 2021.

aV. N. and A. E. R. contributed equally to this article.

Correspondence: Elodie Ghedin, PhD, NIAID/National Institutes of Health, Bldg 50, Rm 1503, 50 South Drive, Bethesda, MD 20894 (elodie.ghedin@nih.gov).

The Journal of Infectious Diseases® 2022;225:1118–23

https://doi.org/10.1093/infdis/jiab622

RESULTS

Case Presentation

A 48-year-old woman with type 2 diabetes mellitus and in complete remission from past diffuse large B-cell lymphoma presented with fever, headache, nasal congestion, and productive cough on 27 April 2020. The patient's history is relevant for treatment with multiple lines of therapy and anti-CD19 chimeric antigen receptor-modified T-cell therapy [9] in December 2017, resulting in ongoing B-cell aplasia, hypogammaglobulinemia, CD4 lymphopenia, and recurrent upper respiratory infections.

A chest computerized tomography (CT) exam performed on admission showed scattered, bilateral, ground-glass radiodensities and consolidations, and she required 2 L of supplemental oxygen via nasal cannula (NC). Laboratory evaluation revealed a white blood cell count of 4.67×10^9 cells/L (normal range [NR], 3.98×10^9–10.04×10^9 cells/L), absolute lymphocyte count of 0.81×10^9 cells/L (NR, 1.18×10^9–3.74×10^9 cells/L), absolute neutrophil count of 3.41×10^9 cells/L (NR, 1.56×10^9–6.13×10^9 cells/L), IgG of 144 mg/dL (NR, 700–1600 mg/dL), IgM 12 mg/dL (NR, 40–230 g/dL), IgA 31 mg/dL (NR, 70–400 mg/dL), and CD4 count of 202/μL (NR, 354–1565/μL). Nasopharyngeal (NP) swabs were negative for SARS-CoV-2 by polymerase chain reaction (PCR) on 19, 28, and 29 April 2020. Bronchoalveolar lavage (BAL) was performed on 1 May 2020 (day 1) following worsening symptoms and increased oxygen requirement. Broad microbiological testing of the BAL fluid was negative, except for a positive PCR test for SARS-CoV-2. The patient's supplemental oxygen requirement increased, and a vasopressor was initiated, in addition to broad-spectrum antibiotics. On day 2, she received convalescent plasma and 40 g of 10% immune globulin IV for her underlying hypogammaglobulinemia. Remdesivir was not available at the time of initial disease presentation and robust clinical trial data surrounding use of corticosteroids in the acute setting of coronavirus disease 2019 (COVID-19) was not yet available, thus neither were administered at the time.

The patient was discharged a month later but continued to have temperatures of 99–100°F, intermittent episodes of worsening cough and to require 3 L NC supplemental oxygen. Testing for SARS-CoV-2 by PCR on NP swabs was performed monthly for 3 months and every 3 months, thereafter. These were positive intermittently with cycle threshold (Ct) values above 37 (Figure 1). Due to the patient's overall mild to absent symptoms, positive SARS-CoV-2 tests during this period were thought to probably reflect shedding of nonviable virus particles. Chest CTs over the same period showed bilateral increasing multifocal ground-glass opacities with crazy paving pattern and mixed changes and, therefore, organizing pneumonia and superimposed bacterial or fungal infection were considered. The patient preferred conservative management and declined bronchoscopy to rule out a superimposed infection. Induced sputum was negative for bacterial, fungal, or mycobacterial pathogens.

On day 242, prednisone 50 mg daily was initiated for the treatment of COVID-19–related cryptogenic organizing pneumonia and resulted in moderate symptom and slight radiographic improvement. SARS-CoV-2 PCR from a NP sample on day 284 was positive with a Ct value of 27.5, a marked decrease from the previous Ct value, indicating a substantial increase in viral load. This increase in the setting of steroids and only modest decrease in symptoms was concerning for COVID-19 relapse. A Ct value of 32.7 from subgenomic RNA real-time PCR indicated recent virus replication [5] (Figure 1). SARS-CoV-2 antibody testing was negative. Shortly after, the patient reported worsening respiratory symptoms and required increased supplemental oxygen. C-reactive protein (CRP) rose to 144 mg/L (<3.0 mg/L) after prednisone initiation. She was admitted to the hospital in March 2021 (day 313), treated with high-titer convalescent plasma and a 10-day course of remdesivir, and was discharged on day 324 with supplemental oxygen. Prednisone was tapered to physiologic doses of hydrocortisone. Three months later, CRP had normalized, CT chest showed significant decrease in ground glass opacities, and the patient no longer needed supplemental oxygen at rest. She remained SARS-CoV-2 PCR and anti-S and N antibody negative.

Genomic Analyses

Because diagnostic testing for SARS-CoV-2 indicated high viral load 10 months after initial diagnosis, whole-genome sequencing was performed on 5 samples from the original presentation in May 2020 and 2 samples from March 2021 to determine prolonged infection versus reinfection. Assembled consensus sequences were assigned lineage and indicated viral genomes from this patient mapped to the Pango lineage B.1.332. Global surveillance of SARS-CoV-2 genomes reveals that B.1.332 was circulating in March/April of 2020 but was no longer prevalent by March 2021 [8]. Consensus sequences were mapped onto a phylogenetic tree containing 266 background samples from Maryland collected between May 2020 and March 2021 (Supplementary Table 1), using the publicly available Nextstrain software package (Figure 1B) [7]. All samples from this patient clustered on the same branch of the tree, with no intermixed background samples, indicative of a prolonged infection over 335 days rather than a reinfection in March of 2021 (Figure 1B).

The original sample, taken on 1 May 2020, contained 11 consensus changes from the Wuhan/Hu-1 strain (NC_045512.2). To visualize evolution of the virus over time, we compared the consensus nucleotides in the later 6 samples to that of the first sample (day 1). Other samples collected the first month of infection had between 1 and 5 nucleotide level consensus changes, whereas the March 2021 samples had 28 (day 313)

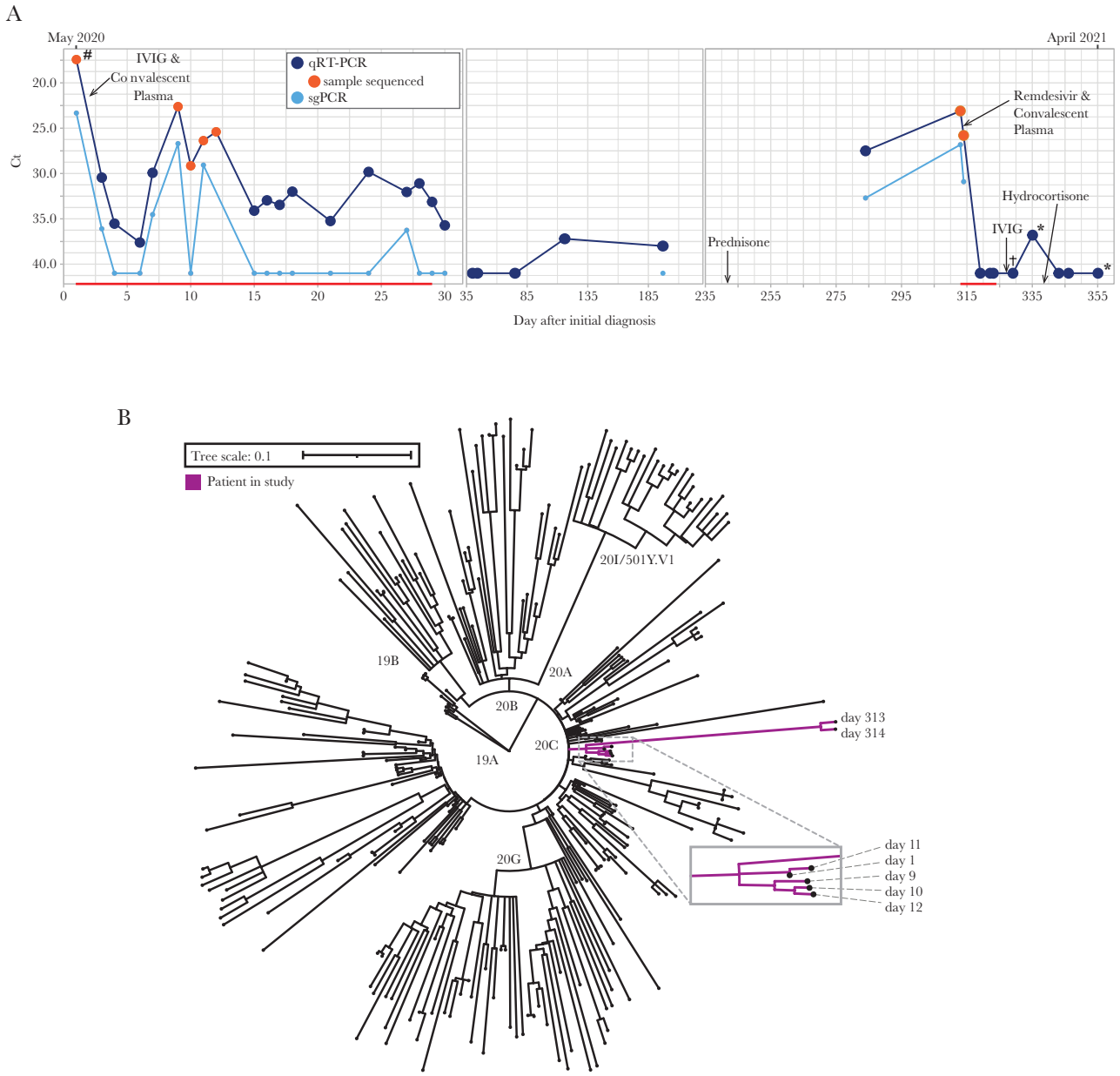

Figure 1. Timeline of diagnostic tests for SARS-CoV-2 and treatment, and phylogeny of sequenced samples. *A*, Nasopharyngeal or oropharyngeal (upper respiratory) specimens were collected for detection of SARS-CoV-2 RNA, except when indicated by the following: # BAL sample, ⋆ days when sputum specimens were collected, and † days when saliva was collected. Specimens with Ct values over 40 were considered negative for SARS-CoV-2 RNA. PCR for subgenomic RNA was performed only on specimens that tested positive for genomic RNA. Samples that were used for next-generation sequencing are indicated with an orange circle. Red lines indicate periods of hospitalization. Treatments administered are indicated with a black arrow and labeled. *B*, Maximum likelihood timed strain tree reconstructed from 266 local sequences from GISAID (Supplementary Table 1). Samples sequenced from the patient in this case study are colored in purple and labeled with day of infection from timeline. Abbreviations: BAL, bronchoalveolar lavage; Ct, cycle threshold; IVIG, intravenous immunoglobulin; qRT-PCR, quantitative reverse transcription polymerase chain reaction; SARS-CoV-2, severe acute respiratory syndrome coronavirus 2; sgPCR, subgenomic polymerase chain reaction.

and 26 (day 314) compared to the initial sample. Of those, 19 and 17, respectively, were nonsynonymous, with 4 substitutions in the spike protein (Figure 2A). More interestingly, the March 2021 samples contained 2 deletions: a gap at nt 22290 to 22298 that led to a unique S:del244-246 and, consequently, a A243G substitution (Figure 2B); and a 497-nt deletion spanning the entire length of the ORF7b coding region and all but 2 amino acids of ORF8 (Figure 2C). Of note, some amino acid changes identified in the March 2021 samples were present as minority variants in the initial samples, suggesting a heterogeneous infection early on (Figure 2D, smaller circles) with eventual fixation, as observed for ORF1a: A3070V, ORF7a: S37F, and N: P365L. Conversely, a consensus change present in the early samples also existed as a minority variant in the last sample in March 2021 (Figure 2D). The observed number of consensus changes in the March 2021 specimens from the

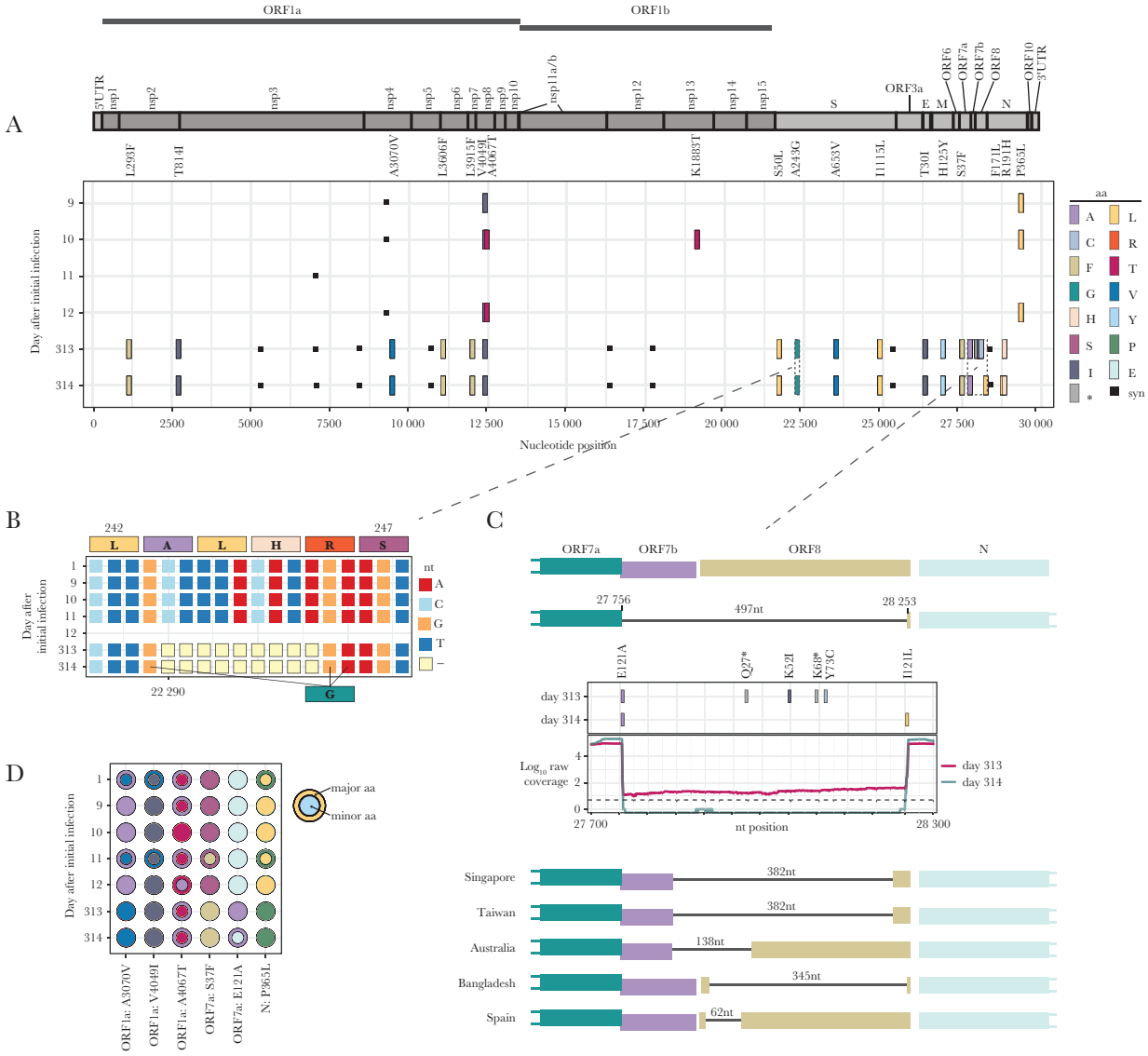

Figure 2. Mutations and deletions in sequenced samples over time. *A*, Tile plot showing consensus changes across the genome as compared to the initial infectious sample (collection date, 1 May 2020, day 1), color-coded by residue. *B*, Schematic showing 9-nt deletion in spike region (nt positions: 22290–22299, spike aa positions shown). *C*, Coverage plot, tile plot, and schematic showing the 497-nt deletion in the ORF7b and ORF8 coding regions. Coverage plot represents $\log_{10}$ raw coverage. Dotted line is at 5 × coverage. Tile plot represents aa changes in this region, as compared to initial infectious sample (collection date: 1 May 2020, day 1), colored by residue as above in (*A*). Deletions identified in previous studies are shown as schematics and labeled by their country of origin. *D*, Circle plot showing major (larger, outer circle) and minor (smaller, inner circle) at locations where a minority variant in one sample exists as a major aa in another sample, colored by residue as in (*A*). ORFs and aa positions within the encoded protein are listed. Abbreviations: aa, amino acid; nt, nucleotide; ORF, open reading frame.

initial sample indicates that the virus acquired the expected number of mutations based on the error rate of the polymerase and evolved within this patient at approximately the same evolutionary rate that has been reported for SARS-CoV-2 in the global population [10], estimated to be around 2 fixed mutations per month.

DISCUSSION

Sequencing data confirmed this patient exhibited a prolonged SARS-CoV-2 infection over 335 days, one of the longest infections reported to date. The existence of ongoing infection is further illustrated by significant clinical improvement and normalization of CRP after viral clearance. The patient displayed COVID-19–like symptoms over the entirety of the infection; however, lower respiratory samples that may have confirmed ongoing viral replication between the patient's 2 admissions were not available.

During the infection, the virus accumulated mutations at approximately the same rate as expected based on the error rate of the polymerase. This was not surprising given the error rate of the polymerase is likely to remain constant regardless of immune status, although which mutations are selected for may differ due to the lack of immune pressure on the virus. Two

important deletions were identified in the later samples, one in the spike protein, and one in the ORF7b and ORF8 regions. The specific spike N-terminal domain (NTD) deletion, del244-246, would impact the supersite and could induce resistance against NTD-directed antibodies [11]. This type of deletion has also been observed in variant B.1.351 (Beta), which contains NTD deletion 242-244 and a R246I mutation [11]. The appearance of this deletion in this patient supports previous observations where chronic SARS-CoV-2 infection in severely immunocompromised hosts receiving convalescent plasma, as was the case for this patient, can lead to variant emergence and reduced sensitivity to neutralizing antibodies [3].

The 497-nt deletion in the ORF7b and ORF8 genes is the longest deletion reported in this region of the genome, and the first in an immunocompromised patient. Other reported ORF7b/ORF8 deletions range from 62 nt to 382 nt in length, with the first instance identified in Singapore in January of 2020 [12, 13]. In vitro analyses of similar deletions indicated mutants replicated to slightly higher levels than wild type (WT) following infection with equal multiplicity of infection (MOI), but showed similar levels of cytopathic effect. This same study further showed that deletion mutant viruses are transmissible, but may be less effective at establishing infection in a new host due to loss of immune evasion features of ORF8 [13]. ORF8 has been established as a key antagonist of innate immunity, eliciting a robust and highly specific antibody response during infection, suggesting that the deletion in competent hosts may be due to immune driven selection [14]. In our case, it is possible that the immunocompromised nature of this patient removes a need for ORF8 during infection. A retrospective cohort study performed on patients in Singapore found that the deletion mutant virus was able to outcompete the WT in some patients that carried a mix of WT and a 382-nt ORF7-ORF8 deletion viruses [15]. We found evidence of a few WT reads in the days 313–314 samples, indicating a possible mixed infection, suggesting that the same competition may have occurred in this patient (Figure 2C).

This case demonstrates that severely immunocompromised patients may experience protracted SARS-CoV-2 infection with mild symptoms and persistent virus replication for a very long period of time. More importantly, it is an example of how prolonged infection can open the door to viral evolution leading to the occurrence of unique mutations, a concern for viral transmission and variant emergence.

Supplementary Data

Supplementary materials are available at *The Journal of Infectious Diseases* online. Supplementary materials consist of data provided by the author that are published to benefit the reader. The posted materials are not copyedited. The contents of all supplementary data are the sole responsibility of the authors. Questions or messages regarding errors should be addressed to the author.

Notes

Acknowledgments. We thank the Department of Laboratory Medicine staff for processing clinical specimens and providing technical assistance. This work utilized the computational resources of the NIH HPC Biowulf cluster (http://hpc.nih.gov).

Financial support. This work was supported by the Division of Intramural Research at the National Institute of Allergy and Infectious Diseases of the National Institutes of Health.

Potential conflicts of interest. J. B. is on the scientific advisory board for Kyverna Therapeutics, Inc; this is an unpaid position. J. K. has received research funding from Kite Pharma, a Gilead Company, and Bristol Myers Squibb; and royalties from Kite Pharma and Kyverna; he has a patent application for an anti-CD19/CD20 CAR construct; and has received drug for a clinical trial from Bellicum Pharmaceuticals. T. P. is an Associate Investigator for a funded study by I-MAB Biopharma, Co. of anti-GM-CSF monoclonal antibody for COVID-19 at a hospital other than where the patient in this report received care; and is an IDWeek 2021 Planning Cochair.

E. G. has received NIH extramural grants. S. D. received payment from the American Society for Microbiology to travel to a conference. All other authors report no potential conflicts.

All authors have submitted the ICMJE Form for Disclosure of Potential Conflicts of Interest. Conflicts that the editors consider relevant to the content of the manuscript have been disclosed.

Presented in part: 2021 Viral Genomics and Evolution (Virtual Conference from Wellcome Genome Campus, UK), 15 September 2021.

Data availability. Data is available in NCBI GenBank under the following accession numbers: MZ385697-MZ385702 and MW990333 and raw sequence reads are available on SRA under bioproject PRJNA784993.

References

1. Sekine T, Perez-Potti A, Rivera-Ballesteros O, et al. Robust T cell immunity in convalescent individuals with asymptomatic or mild COVID-19. Cell **2020**; 183:158–68.e14.
2. Hensley MK, Bain WG, Jacobs J, et al. Intractable coronavirus disease 2019 (COVID-19) and prolonged severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) replication in a chimeric antigen receptor-modified T-cell therapy recipient: a case study. Clin Infect Dis **2021**; 73:e815–21.
3. Kemp SA, Collier DA, Datir RP, et al. SARS-CoV-2 evolution during treatment of chronic infection. Nature **2021**; 592:277–82.
4. Cevik M, Tate M, Lloyd O, Maraolo AE, Schafers J, Ho A. SARS-CoV-2, SARS-CoV, and MERS-CoV viral load dynamics, duration of viral shedding, and infectiousness: a systematic review and meta-analysis. Lancet Microbe **2021**; 2:e13–22.

5. Avanzato VA, Matson MJ, Seifert SN, et al. Case study: prolonged infectious SARS-CoV-2 shedding from an asymptomatic immunocompromised individual with cancer. Cell **2020**; 183:1901–12.e9.
6. Choi B, Choudhary MC, Regan J, et al. Persistence and evolution of SARS-CoV-2 in an immunocompromised host. N Engl J Med **2020**; 383:2291–3.
7. Hadfield J, Megill C, Bell SM, et al. Nextstrain: real-time tracking of pathogen evolution. Bioinformatics **2018**; 34:4121–3.
8. Rambaut A, Holmes EC, O'Toole A, et al. A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology. Nat Microbiol **2020**; 5:1403–7.
9. Brudno JN, Lam N, Vanasse D, et al. Safety and feasibility of anti-CD19 CAR T cells with fully human binding domains in patients with B-cell lymphoma. Nat Med **2020**; 26:270–80.
10. Worobey M, Pekar J, Larsen BB, et al. The emergence of SARS-CoV-2 in Europe and North America. Science **2020**; 370:564–70.
11. Cerutti G, Guo Y, Zhou T, et al. Potent SARS-CoV-2 neutralizing antibodies directed against spike N-terminal domain target a single supersite. Cell Host Microbe **2021**; 29:819–33.e7.
12. Gong YN, Tsao KC, Hsiao MJ, et al. SARS-CoV-2 genomic surveillance in Taiwan revealed novel ORF8-deletion mutant and clade possibly associated with infections in Middle East. Emerg Microbes Infect **2020**; 9:1457–66.
13. Su YCF, Anderson DE, Young BE, et al. Discovery and genomic characterization of a 382-nucleotide deletion in ORF7b and ORF8 during the early evolution of SARS-CoV-2. mBio **2020**; 11:e01610-20.
14. Hachim A, Kavian N, Cohen CA, et al. ORF8 and ORF3b antibodies are accurate serological markers of early and late SARS-CoV-2 infection. Nat Immunol **2020**; 21:1293–301.
15. Young BE, Fong SW, Chan YH, et al. Effects of a major deletion in the SARS-CoV-2 genome on the severity of infection and the inflammatory response: an observational cohort study. Lancet **2020**; 396:603–11.