

Unraveling a Nosocomial Outbreak of COVID-19: The Role of Whole-Genome Sequence Analysis

Suzy E. Meijer,^{1,7,a} Noam Harel,^{4,5,a} Ronen Ben-Ami,^{1,7} Meital Nahari,^{1,7} Michal Yakubovsky,^{1,7} Howard S. Oster,^{3,7} Albert Kolomansky,^{3,7} Ora Halutz,^{2,7} Orly Laskar,⁶ Oryan Henig,^{1,7} Adi Stern,^{4,5,©} and Yael Paran^{1,7,©}

¹Department of Infectious Diseases and Epidemiology, Tel Aviv Sourasky Medical Center, Tel Aviv, Israel, ²Laboratory of Microbiology, Tel Aviv Sourasky Medical Center, Tel Aviv, Israel, ³Department of Internal Medicine, Tel Aviv Sourasky Medical Center, Tel Aviv, Israel, ⁴The Shmunis School of Biomedicine and Cancer Research, Tel Aviv University, Tel Aviv, Israel, ⁵Edmond J. Safra Center for Bioinformatics at Tel Aviv University, Tel Aviv, Israel, ⁶The Department of Infectious Diseases, Israel Institute for Biological Research, Ness-Ziona, Israel, ⁷Sackler Faculty of Medicine, Tel Aviv University, Tel Aviv, Israel

Background. The coronavirus disease 2019 (COVID-19) pandemic poses many epidemiological challenges. The investigation of nosocomial transmission is usually performed via thorough investigation of an index case and subsequent contact tracing. Notably, this approach has a subjective component, and there is accumulating evidence that whole-genome sequencing of the virus may provide more objective insight.

Methods. We report a large nosocomial outbreak in 1 of the medicine departments in our institution. Following intensive epidemiological investigation, we discovered that 1 of the patients involved was suffering from persistent COVID-19 while initially thought to be a recovering patient. She was therefore deemed to be the most likely source of the outbreak. We then performed whole-genome sequencing of the virus of 14 infected individuals involved in the outbreak.

Results. Surprisingly, the results of whole-genome sequencing refuted our initial hypothesis. A phylogenetic tree of the samples showed multiple introductions of the virus into the ward, 1 of which led to a cluster of 10 of the infected individuals. Importantly, the results pointed in the direction of a specific index patient that was different from the 1 that arose from our initial investigation.

Conclusions. These results underscore the important added value of using whole-genome sequencing in epidemiological investigations as it may reveal unexpected connections between cases and aid in understanding transmission dynamics, especially in the setting of a pandemic where multiple possible index cases exist simultaneously.

Keywords. COVID-19; epidemiology; nosocomial; outbreak; sequencing.

Since the beginning of the coronavirus disease 2019 (COVID-19) pandemic, epidemiological investigations have proven key in both curbing viral transmission chains and furthering our understanding of how the virus spreads in order to improve infection prevention measures. These investigations rely on the thorough questioning of an index case and evaluation of other relevant information in order to establish the source of infection and identify possible contacts who may have been infected by the index patient. Notably, this type of investigation has a subjective component, and despite efforts to control for this, it is often difficult to reach definite conclusions.

Whole-genome sequencing has been used in investigations of hospital outbreaks of drug-resistant bacteria [1, 2]. Since the

beginning of the COVID-19 pandemic, sequencing has played an important role in tracking the spread of the disease within countries and across borders [3–10]. However, there are limited data on the use of whole-genome sequencing to determine the dynamics of nosocomial transmission of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) [11–13]. Here, we performed sequencing of viral isolates coupled with phylogenetic analysis in order to determine the origins and direction of transmission of a nosocomial outbreak in a medicine department.

Setting

We report a large-scale outbreak of COVID-19 that occurred in a medicine department at the Tel Aviv Sourasky Medical Center, a 1500-bed tertiary care hospital in Tel Aviv, Israel.

Considering the role of asymptotically infected individuals in SARS-CoV-2 transmission [14–18], periodic screening of patients and health care personnel is performed in our institution using nasopharyngeal swabs for real-time polymerase chain reaction (RT-PCR) of SARS-CoV-2. Any case of COVID-19 that is detected among patients outside the designated COVID-19 departments or among health care personnel undergoes a structured investigation in an effort to minimize further transmission.

Received 15 January 2021; editorial decision 8 March 2021; accepted 9 March 2021.

^aEqual contribution

Correspondence: Suzy Meijer, Dalia Rabikovitch 3/22, Tel Aviv 4647000, Israel (suzy.meijer@gmail.com).

Open Forum Infectious Diseases® 2021

© The Author(s) 2021. Published by Oxford University Press on behalf of Infectious Diseases Society of America. This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs licence (<http://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial reproduction and distribution of the work, in any medium, provided the original work is not altered or transformed in any way, and that the work is properly cited. For commercial re-use, please contact journals.permissions@oup.com
DOI: 10.1093/ofid/ofab120

Outbreak Description

On September 22, 2020, routine screening revealed 2 positive results for COVID-19. One subject was a patient admitted to the medium-care room (Patient 9), and the other was a member of the nursing team (Staff member 1). Patient 9 had a cycle threshold (CT) value of 36. He tested negative 5 days earlier upon admission. This patient had no signs or symptoms suggestive of COVID-19, so his positive result was interpreted as reflecting remnants of RNA at the end of an unnoticed infection or a false-positive test. Staff member 1 was asymptomatic at the time of screening but upon questioning appeared to have had minor symptoms that could be attributable to COVID-19 the week before. Coincidentally, another staff member on the ward (Staff member 2) had developed symptoms on that same day and had subsequently tested positive for SARS-CoV-2.

As a rule, when 2 or more infections are found in 1 department, we perform screening of all patients and staff members on that ward. We thus discovered that 15 out of 30 patients were infected with SARS-CoV-2, as well as 2 additional staff members. The infected patients were located in rooms throughout the entire ward (Figure 1). Only 6 out of 15 infected patients had symptoms attributable to COVID-19. Five of the infected patients had walked around the ward, while the other 10 were either bedridden or admitted in isolation (for reasons other than COVID-19) and had not left their rooms. Thirteen of the 15 patients had tested negative upon admission (Figure 2). Only 1 patient (Patient 1) had tested positive upon admission, but as she was known to be recovering from a prior SARS-CoV-2 infection diagnosed 8 weeks earlier, she was interpreted as being

no longer infectious and was therefore not transferred to a designated COVID-19 ward. Another patient (Patient 2) had not been screened upon admission.

On further assessment, Patient 1 was noted to suffer from follicular lymphoma and was receiving maintenance therapy with obinutuzumab (an anti-CD20 monoclonal antibody). The patient had ongoing respiratory symptoms and infiltrates on chest imaging since testing positive for SARS-CoV-2 in July, suggesting persistent COVID-19. Recent evidence has highlighted the occurrence of protracted SARS-CoV-2 infection in patients receiving B-cell-depleting therapy [19–21]. When admitted, this patient was evaluated using guidelines from the Israeli Ministry of Health, based on international guidelines [22], to determine active vs past infection. As she had 4 negative tests before her positive test upon admission, according to these criteria she was considered a recovering patient and was therefore initially overlooked in our investigation. During the department-wide screening, she tested positive again, but this time with decreasing CT values. Also, SARS-CoV-2 was recovered in cell culture performed on her nasopharyngeal swab (methods in the Appendix). In the absence of other leads as to how the outbreak started, we hypothesized that this patient was the index case and examined several possible scenarios explaining means by which SARS-CoV-2 could have been transmitted from her to other patients and staff. As she was bedridden and had not left her room, these theories included aerosol transmission originating from oxygen therapy through high-flow nasal cannula (which is considered an aerosol-generating procedure) combined with potential defects in local

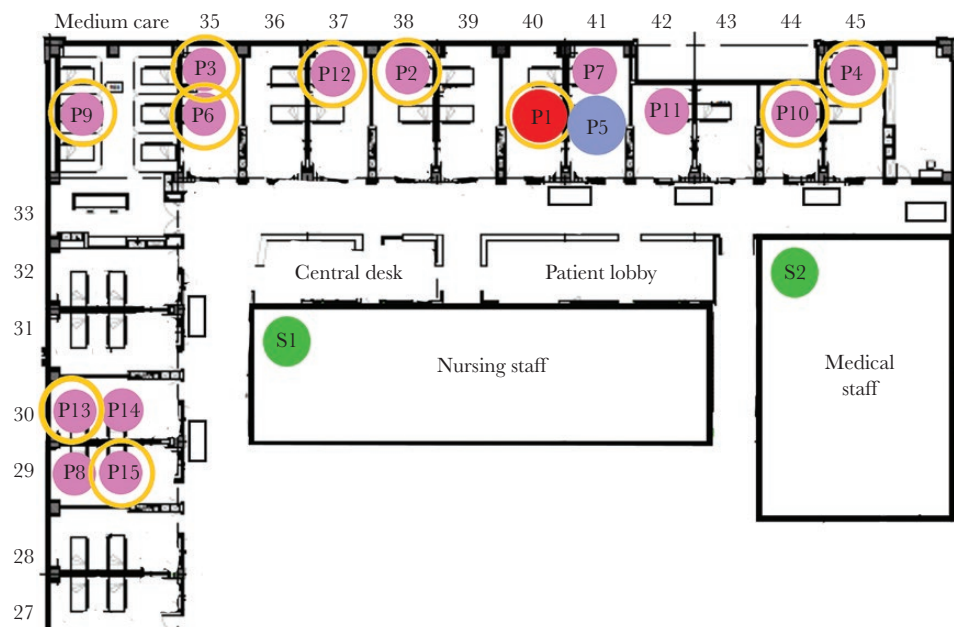


Figure 1. Floor scheme of the medicine department involved in the outbreak, with a map of the patients who tested positive for SARS-CoV-2. Yellow circle, immobile patient. Abbreviations: P, infected patient; P1 (red), originally assumed index patient; P5 (blue), index patient as suggested by sequencing; S, infected staff; SARS-CoV-2, severe acute respiratory syndrome coronavirus 2.

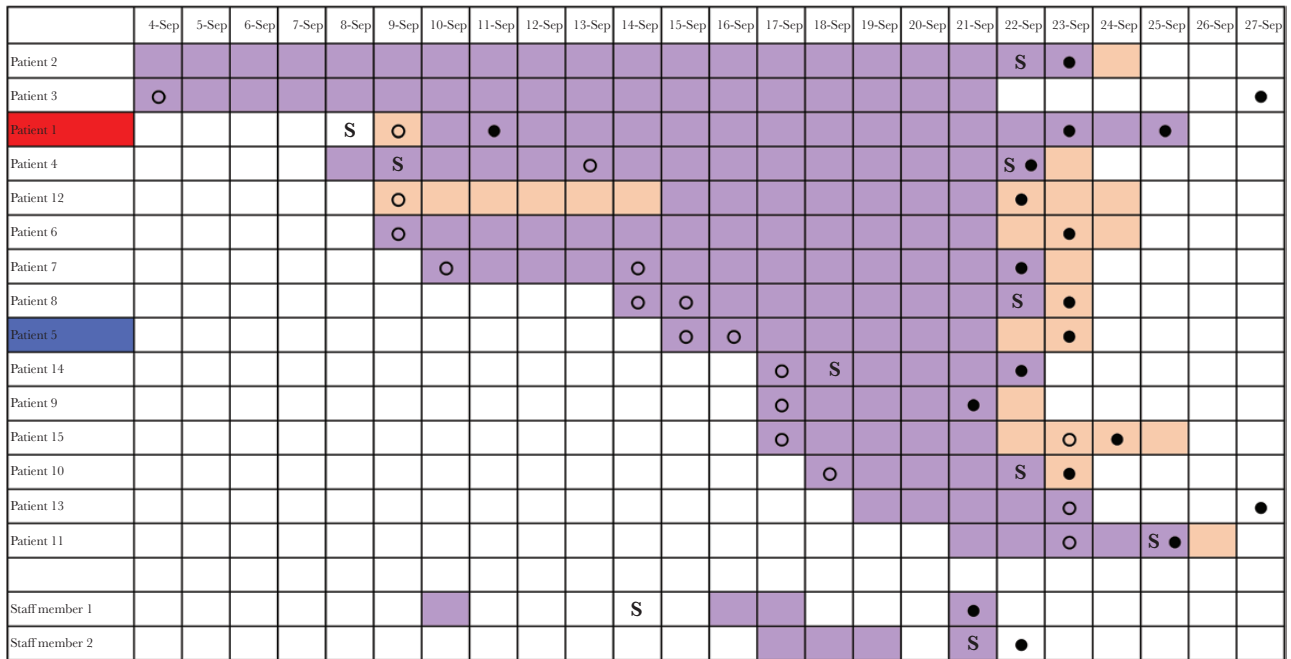


Figure 2. Schematic overview of hospital stay, onset of symptoms, and diagnosis of COVID-19 of patients involved in the outbreak. Top row, dates of stay in the month of September 2020. Purple, hospital stay in the described medicine department. Blue, hospital stay in another department in the hospital. Patient 1 (red), initially assumed index case. Patient 5 (blue), index case as suggested by sequencing. ○, SARS-CoV-2 PCR negative. ●, SARS-CoV-2 PCR positive. Abbreviations: COVID-19, coronavirus disease 2019; S, onset of symptoms; SARS-CoV-2, severe acute respiratory syndrome coronavirus 2.

air filtering. Finally, in order to obtain a better understanding of the dynamics of this department-wide outbreak, we performed whole-genome sequencing of all available SARS-CoV-2 amplicons.

METHODS

Virus Genome Sequencing

Samples were obtained from 15 of the 17 infected individuals, including 3 samples at different time points from Patient 1. SARS-CoV-2 RNA that was extracted from nasopharyngeal swabs underwent whole-genome sequencing using the V3 ARTIC protocol (<https://artic.network/ncov-2019>). Briefly, reverse transcription, multiplex PCR, and adaptor ligation were performed, and samples were run on an Illumina MiSeq using 250-cycle V2 kits in the Technion Genome Center (Israel).

Determining Genome Consensus Sequences

Raw reads were trimmed using pTrimmer (<https://bmcbioinformatics.biomedcentral.com/articles/10.1186/s12859-019-2854-x>) and mapped to the typical reference genome of SARS-CoV-2 (GenBank ID MN908947) using our AccuNGS pipeline. We performed quality control of all sequences, requiring that $\geq 95\%$ of the genome was sequenced, at coverage ≥ 10 . This led to the exclusion of 1 sample and to the retainment of 16 samples. Consensus sequences were determined for each sample using 2 different approaches. The first and main approach required a substitution (compared with

the reference) to be present in 80% of the reads to be added to the consensus. Positions with coverage < 10 or a substitution appearing at between 50% and 80% of the reads were deemed unknown and assigned Ns. Additionally, short stretches (< 20) of “known” bases appearing between unknown areas (Ns) were masked with Ns as well in order to not interfere with alignments. The second, more lenient approach, used to deal with a sequence that was not optimally sequenced, was to use the majority rule, meaning every position was assigned the base that appeared most frequently (coverage < 10 still receiving Ns).

Phylogenetic Analysis

We further gathered an additional 85 sequences from Israel from similar time points (June to September 2020). We used MAFFT to align the sequences and reconstructed a phylogenetic tree using PhyML, which relies on a maximum likelihood approach. Bootstrap values were assigned to each split (node) of the tree as a measure of confidence ($n = 1000$). The more lenient approach for consensus creation was used to validate the clustering of the sequence of Patient 5 with the rest of the department cluster, which raised the bootstrap of this cluster from 486/1000 to 840/1000.

RESULTS

We set out to explore the relationship between the 16 samples successfully sequenced from the described outbreak, with a larger set of sequences from Israel serving as a reference.

We first noted that the 16 samples were spread across 5 independent clades (Figure 3). The branches separating these clades were equivalent to at least 9 substitutions, strongly suggesting that there were 5 independent introductions of the virus into the ward. One of these introductions led to a tightly linked cluster, composed of 8 sequences that were 100% identical to each other, and 2 additional sequences (of Patient 2 and Patient 5) that differed by 1–2 substitutions from this cluster. The sequences obtained from the hypothesized index case, Patient 1, were very distant from the cluster and from all other sequences from the outbreak: At least 11 substitutions separated the purported index case and the cluster. Given that there is on average 1 substitution every second transmission, this roughly translates to 22 transmission events, making it highly

unlikely that the suspected index case is indeed the source of the outbreak. When focusing on the cluster, we noted that 1 sequence (that of Patient 5) appeared to be slightly ancestral to the rest of the cluster. This was based on 2 substitutions absent in this sequence and present in the rest. We thus cautiously suggest that Patient 5 may be the source of the outbreak, because of the assumption that it is more likely to gain mutations (substitutions) over time rather than to lose them.

Cross-referencing this information with our epidemiological investigation, Patient 5 appeared to have developed mild symptoms that could be attributed to COVID-19 2 days after his admission. During his hospital stay, he interacted with multiple other patients and caregivers while refusing to don a face mask despite repeated reprimands from the staff. Additionally, before this patient

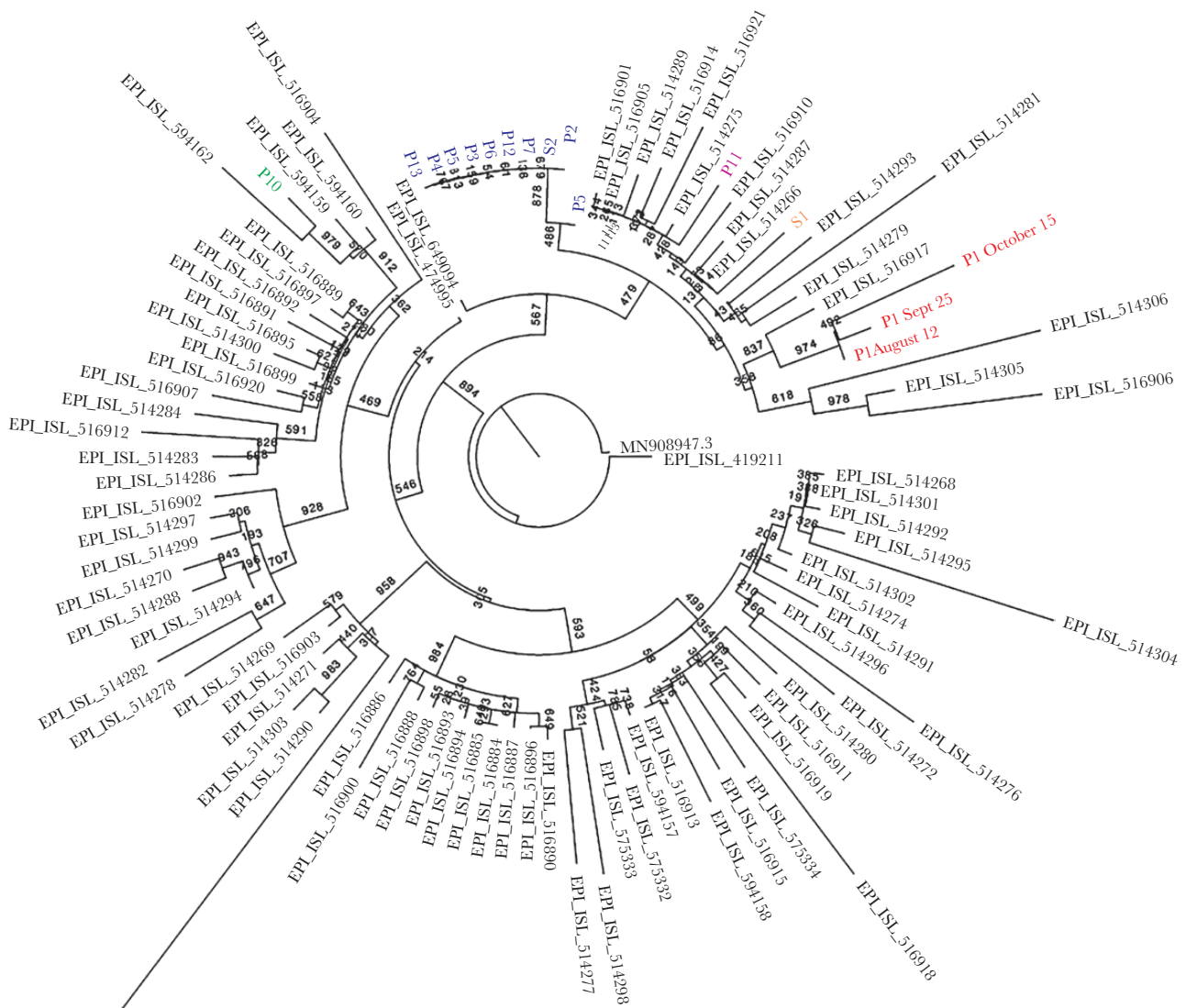


Figure 3. Phylogenetic tree of SARS-CoV-2 sequences obtained in the epidemiological investigation. Samples from the investigation are in color; each of the 5 clades colored differently indicates a separate introduction into the ward. Sequences in gray are unrelated samples from Israel from a similar time point. Bootstrap values are shown on branches ($n = 1000$). Abbreviations: P, patient; P1, originally assumed index patient; P5, index patient as suggested by sequencing; S, staff member; SARS-CoV-2, severe acute respiratory syndrome coronavirus 2.

was diagnosed with COVID-19, he was transferred to another department, where 5 staff members were subsequently infected. Therefore, this patient was identified as a likely superspreader.

CONCLUSIONS

In this epidemiological investigation, whole-genome sequencing provided us with a number of new and unforeseen insights. First, the results showed that the patient we had assumed to be the most likely index case apparently had infected no one on the ward. Our assumption was based on laboratory data and clinical circumstances. Specifically, the patient had a positive SARS-CoV-2 PCR on admission with decreasing CT values on subsequent sampling and a positive viral culture, all the while being treated with an aerosol-generating procedure. If not for the results of the sequencing, the investigation may have likely concluded that she was the source of the outbreak via aerosol transmission, a conclusion with dramatic implications for future infection control measures.

Remarkably, the phylogenetic tree suggested a different and surprising potential index case, namely Patient 5. This patient had a negative test upon his admission a week earlier but developed mild symptoms 2 days later. With the results of the sequencing in hand, we hypothesized that he was infected at home and introduced SARS-CoV-2 into the ward. His negative test upon admission could either reflect him still being in the incubation period of the infection, or the test could have been falsely negative. The fact that so many people were infected by him could plausibly be explained by the fact that he was known to interact with many people while consistently refusing to wear a mask. Furthermore, the patients in the cluster could have been infected directly by the index patient or through secondary infection by others in the cluster, including Staff member 2.

Our epidemiological investigation has a number of limitations. First, for technical reasons, we were unable to perform whole-genome sequencing of 3 out of 17 specimens of the involved cases. Even though we consider it unlikely, the results of sequencing of these missing specimens may have led to different conclusions. Second, assessing the direction of spread by the results of whole-genome sequencing is based on the assumed evolution of mutations and therefore cannot be determined with absolute certainty. However, whole-genome sequencing can be used more definitively to rule out hypotheses on a source of infection, as we show here.

We conclude that whole-genome sequencing is an important and powerful tool in epidemiological investigations, as it may reveal unexpected connections between cases and aid in understanding the dynamics of spread of a disease. Especially during the COVID-19 pandemic, with high prevalence of disease both outside and inside the hospital, it can be challenging to determine the origins of an outbreak when many possible index cases exist simultaneously. Furthermore, establishing infectivity

alone is not enough to identify the source of nosocomial transmission, as not every patient who is capable of infecting in fact does so.

APPENDIX. METHODS FOR CELL CULTURE

Viral swabs were inserted into refrigerated transfer buffer-containing tubes (Copan). Tubes containing the swabs were vortexed for 1 minute. Two mL of the buffer was transferred to a new 15-mL tube and centrifuged (5000 G, 5 minutes, 4°C), and the supernatant was transferred through a 0.22- μ m filter. Vero E6 (ATCC CRL-1586) was cultured in DMEM supplemented with 10% fetal bovine serum (FBS), MEM nonessential amino acids, 2 mM of L-Glutamine, 100 units/mL of penicillin, 0.1 mg/mL of streptomycin, and 12.5 units/mL of nystatin (Biological Industries, Israel). Each supernatant sample was added in duplicate to cells' monolayers in 12-well plates (Costar; 0.2 mL/well) for 1 hour, followed by the addition of 2 mL of MEM containing 2% FBS, MEM nonessential amino acids, 2 mM of L-Glutamine, 100 units/mL of penicillin, 0.1 mg/mL of streptomycin, 12.5 units/mL of nystatin, and 0.15% sodium bicarbonate (Biological Industries, Israel). Plates were further incubated at 37°C, 5% CO₂, for 5 days. SARS-CoV-2 (GISAID accession EPI_ISL_406862), kindly provided by Bundeswehr Institute of Microbiology, Munich, Germany, was used as the positive control at a concentration of 60 pfu/mL. Cytopathic effect (CPE) was microscopically determined. To confirm that the cytopathic effect was due to SARS-Cov-2, a real-time RT-PCR for the virus was performed on cell supernatant.

Acknowledgments

Potential conflicts of interest. All authors: no reported conflicts of interest. 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.

Patient consent. Our investigation was approved by the local Helsinki committee with approval No. 1042-20-TLV. The Helsinki committee has exempted us from obtaining written informed consent of the patients involved in our manuscript.

References

1. García-Fernández S, Frentrup M, Steglich M, et al. Whole-genome sequencing reveals nosocomial *Clostridioides difficile* transmission and a previously unsuspected epidemic scenario. *Sci Rep* 2019; 9:6959.
2. Sui W, Zhou H, Du P, et al. Whole genome sequence revealed the fine transmission map of carbapenem-resistant *Klebsiella pneumoniae* isolates within a nosocomial outbreak. *Antimicrob Resist Infect Control* 2018; 7:70.
3. Oude Munnink BB, Nieuwenhuijse DE, Stein M, et al; Dutch-Covid-19 response team. Rapid SARS-CoV-2 whole-genome sequencing and analysis for informed public health decision-making in the Netherlands. *Nat Med* 2020; 26:1405–10.
4. Stefanelli P, Faggioni G, Lo Presti A, et al. Whole genome and phylogenetic analysis of two SARS-CoV-2 strains isolated in Italy in January and February 2020: additional clues on multiple introductions and further circulation in Europe. *Euro Surveill* 2020; 25:2000305.
5. Böhmer MM, Buchholz U, Corman VM, et al. Investigation of a COVID-19 outbreak in Germany resulting from a single travel-associated primary case: a case series. *Lancet Infect Dis* 2020; 20:920–8.
6. Xavier J, Giovanetti M, Adelino T, et al. The ongoing COVID-19 epidemic in Minas Gerais, Brazil: insights from epidemiological data and SARS-CoV-2 whole genome sequencing. *Emerg Microbes Infect* 2020; 9:1824–34.

7. Choi EM, Chu DKW, Cheng PKC, et al. In-flight transmission of SARS-CoV-2. *Emerg Infect Dis* **2020**; 26:2713–6.
8. Chan JF, Yuan S, Kok KH, et al. A familial cluster of pneumonia associated with the 2019 novel coronavirus indicating person-to-person transmission: a study of a family cluster. *Lancet* **2020**; 395:514–23.
9. Wang X, Zhou Q, He Y, et al. Nosocomial outbreak of COVID-19 pneumonia in Wuhan, China. *Eur Respir J* **2020**; 55:2000544.
10. Taylor J, Carter RJ, Lehnertz N, et al; Minnesota Long-Term Care COVID-19 Response Group. Serial testing for SARS-CoV-2 and virus whole genome sequencing inform infection risk at two skilled nursing facilities with COVID-19 outbreaks - Minnesota, April-June 2020. *MMWR Morb Mortal Wkly Rep* **2020**; 69:1288–95.
11. Lucey M, Macori G, Mullane N, et al. Whole-genome sequencing to track SARS-CoV-2 transmission in nosocomial outbreaks. *Clin Infect Dis*. **In press**.
12. Takenouchi T, Iwasaki YW, Harada S, et al. Clinical utility of SARS-CoV-2 whole genome sequencing in deciphering source of infection. *J Hosp Infect* **2021**; 107:40–4.
13. Meredith LW, Hamilton WL, Warne B, et al. Rapid implementation of SARS-CoV-2 sequencing to investigate cases of health-care associated COVID-19: a prospective genomic surveillance study. *Lancet Infect Dis* **2020**; 20:1263–71.
14. Rivett L, Sridhar S, Sparkes D, et al. Screening of healthcare workers for SARS-CoV-2 highlights the role of asymptomatic carriage in COVID-19 transmission. *Elife* **2020**; 9:e58728.
15. Huff HV, Singh A. Asymptomatic transmission during the COVID-19 pandemic and implications for public health strategies. *Clin Infect Dis*. **In press**.
16. Yanes-Lane M, Winters N, Fregonese F, et al. Proportion of asymptomatic infection among COVID-19 positive persons and their transmission potential: a systematic review and meta-analysis. *PLoS One* **2020**; 15:e0241536.
17. Furukawa et al. www.cdc.gov. Accessed 1 January 2021.
18. He D, Zhao S, Lin Q, et al. The relative transmissibility of asymptomatic COVID-19 infections among close contacts. *Int J Infect Dis* **2020**; 94:145–7.
19. Baang JH, Smith C, Mirabelli C, et al. Prolonged SARS-CoV-2 replication in an immunocompromised patient. *J Infect Dis*. **In press**.
20. 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.
21. Aydillo T, Gonzalez-Reiche AS, Aslam S, et al. Shedding of viable SARS-CoV-2 after immunosuppressive therapy for cancer. *N Engl J Med* **2020**; 383:2586–8.
22. CDC. www.cdc.gov. Accessed 1 January 2021.