*Clinical Infectious Diseases*

MAJOR ARTICLE

# Genomic Epidemiology and Transmission Dynamics of Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) in Congregate Healthcare Facilities in Santa Clara County, California

**Tara MacCannell,[1,b] Joshua Batson,[2,b] Brandon Bonin,[1] Astha K. C.,[1] Rebecca Quenelle,[1] Betsy Strong,[1] Wen Lin,[1] Sarah L. Rudman,[1] David Dynerman,[2] Patrick Ayscue,[2] George Han,[1] Amy Kistler,[2] and Margarita Elsa Villarino[1]; on behalf of the Chan Zuckerberg Biohub COVIDTracker Consortium and the Santa Clara County Public Health Department's Special Investigations Group[a]**

[1]County of Santa Clara, Public Health Department, San Jose, California, USA; and [2]Chan Zuckerberg Biohub, San Francisco, California, USA

***Background.*** Outbreaks of SARS-CoV-2 in long-term care facilities (LTCFs) cause significant morbidity and mortality. Mapping viral transmission within and between facilities by combining genomic sequencing with epidemiologic investigations enables targeting infection-control interventions.

***Methods.*** We conducted weekly surveillance of residents and staff in LTCFs in Santa Clara County, California, with ≥1 confirmed COVID-19 case between March and July 2020. Positive samples were referred for whole-genome sequencing. Epidemiological investigations and phylogenetic analyses of the largest outbreaks (>30 cases) were carried out in 6 LTCFs (Facilities A through F).

***Results.*** Among the 61 LTCFs in the county, 41 had ≥1 confirmed case during the study period, triggering weekly SARS-CoV-2 testing. The 6 largest outbreaks accounted for 60% of cases and 90% of deaths in LTCFs, although the bed capacity of these facilities represents only 11% of the LTCF beds in the county. Phylogenetic analysis of 196 whole-genome sequences recovered from those facilities showed that each outbreak was monophyletic, with staff and residents sharing a common viral lineage. Outbreak investigations revealed that infected staff members often worked at multiple facilities, and in 1 instance, a staff member infected while working in 1 facility was the likely index case in another.

***Conclusions.*** We detected a pattern of rapid and sustained transmission after a single introduction of SARS-CoV-2 in 6 large LTCF outbreaks, with staff playing a key role in transmission within and between facilities. Infection control, testing, and occupational policies to reduce exposure and transmission risk for staff are essential components to keeping facility residents safe.

**Keywords.** elderly; epidemiology; genomics; healthcare; providers.

Santa Clara County (SCC), an ethnically and economically diverse region of California also known as "Silicon Valley," had an early introduction of the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) novel coronavirus [1]. Its first case of coronavirus disease 2019 (COVID-19) was confirmed in a traveler on 31 January 2020, and subsequent cases were introduced by multiple routes including travel from Wuhan, China; passengers from the Diamond and Grand Princess cruises; and travel from Washington state [2, 3]. The first community-acquired case of SARS-CoV-2 in SCC was announced on 28 February, and within 1 week the virus was found to have spread from the community into long-term care facilities (LTCFs), including skilled-nursing facilities (SNFs), resulting in 4 deaths among 419 cases by 30 April [4]. Residents of these congregate healthcare facilities are particularly prone to poor outcomes including mortality because of advanced age and comorbidities such as hypertension, diabetes, and chronic respiratory conditions. Long-term care facilities also have the potential for rapid and sustained viral transmission due to shared living quarters and close contact between residents and care providers.

To reduce the potential for outbreaks of COVID-19 in LTCFs, the County of Santa Clara Public Health Department (SCCPHD), with assistance from the Centers for Disease Control and Prevention (CDC), provided facilities with infection-control guidance and checklists, increased access to personal protective equipment (PPE), and established policies prohibiting family visitation and group activities. Beginning 1 April, the County introduced response-driven testing in LTCFs,

Received 15 April 2021; editorial decision 8 June 2021; published online 30 July 2021.

[a]The Chan Zuckerberg Biohub COVIDTracker Consortium and the Santa Clara County Public Health Department's Special Investigations Group members are listed in the Acknowledgments section.

[b]Denotes primary co-authorship.

Correspondence: M. E. Villarino, Santa Clara County Department of Public Health, 976 Lenzen Avenue, Suite 1700, San Jose, CA 95126 (elsa.villarino@phd.sccgov.org).

**Clinical Infectious Diseases® 2021;XX(XX):1–7**



https://doi.org/10.1093/cid/ciab553

where a suspected cluster of COVID-19 cases triggered, at a minimum, weekly facility-wide testing of residents and staff. Positive polymerase chain reaction (PCR) specimens were submitted for whole-genome sequencing (WGS).

The use of genomics as a complement to traditional epidemiologic investigations and surveillance has been well established in global health contexts, with successful investigations and responses to outbreaks caused by Ebola, Zika, and seasonal influenza A/B viruses [5–8]. Previous studies have successfully used genomic epidemiology to describe SARS-CoV-2 transmission in hospitals and nursing homes in the United Kingdom and the United States [9, 10]. Viral genome sequencing can provide independent and complementary evidence to contact tracing, which is especially useful in congregate healthcare settings where complex patterns of staff and resident movement within and between facilities create competing hypotheses for transmission routes.

Here, we characterize the outbreak transmission patterns of SARS-CoV-2 between LTCF residents and staff across LTCFs in SCC from 18 March to 31 July 2020. We demonstrate how pairing genomic sequencing of SARS-CoV-2 samples with contact tracing and epidemiologic data revealed relationships between individuals and facilities, and how this additional evidence informed infection-control recommendations and public health decision making.

## METHODS

### Testing and Screening Strategy

Beginning 1 April 2020, facility-wide testing of staff and residents was coordinated by SCCPHD and LTCF management when clusters of 3 or more suspected cases or 1 confirmed case of COVID-19 among staff or residents were detected in any LTCF, which included licensed skilled-nursing, intermediate-care, and assisted-living facilities. As first cases were detected in each facility, the health department engaged in site visits and facility-wide PCR screening (ie, response-driven testing). Nasopharyngeal swabs for all employees and residents were collected weekly for SARS-CoV-2 testing by the SCCPHD laboratory (with a transition to nasal swabs on 18 June). At a minimum, weekly screening was sustained until no new cases were detected. Visitors and nonessential staff were excluded from all LTCFs and were not included in testing. Outbreaks were declared contained in the absence of new suspected or confirmed cases after 2 consecutive weeks. Standardized case line lists contained demographics, clinical data and outcomes, and resident room assignment.

We analyzed all confirmed cases of SARS-CoV-2 detected at 61 LTCFs from 18 March–31 July 2020. Based on the CDC definition, cases of COVID-19 were confirmed by the presence of SARS-CoV-2 RNA as detected by a quantitative molecular amplification detection test (reverse transcription [RT]-PCR) [11]. Whole-genome sequencing on PCR-positive samples from SCC started 1 April 2020. We focused our detailed genomic analysis to 6 LTCF outbreaks (Facilities A to F), each with a minimum of 30 cases of COVID-19.

### Sequencing and Bioinformatics

All respiratory samples from LTCFs were processed for RT-PCR testing at the SCCPHD Laboratory using the CDC protocol for the detection and amplification of SARS-CoV-2 [12]. Positive samples with a cycle threshold (Ct) of 32 or less were forwarded to the Chan Zuckerberg (CZ) Biohub for genomic sequencing via a modified version of the Illumina's Primal-Seq Nextera XT version 2.0 protocol [12, 13], using ARTIC Network V3 primers [14], and paired-end 2 × 150-bp sequencing on Illumina NovaSeq. Consensus sequences were obtained using MN908947.3 as a reference, using minimap2, samtools, and ivar in the CZ Biohub consensus genome pipeline (https://github.com/czbiohub/sc2-illumina-pipeline) [11, 15, 16]. Viral genomes with at least 90% coverage were uploaded to Global Initiative for Sharing All Influenza Data [17].

Phylogenetic assemblies from SCCPHD were constructed using iqtree in the augur pipeline with default settings [18]. Phylogenetic results, paired with demographic and epidemiologic data, were visualized in NextStrain and displayed using baltic (https://github.com/evogytis/baltic) [18]. Nonstudy samples were down-sampled by a factor of 3 after tree construction for ease of display.

### Communication to Facilities

Positive PCR results were communicated to the facility, with an average turnaround time of 48 hours. Efforts to minimize the turnaround time for PCR results and genomic data translated into actionable information. Where applicable, phylogenetic relationships were confidentially communicated to SCCPHD and LTCF management to educate and reinforce infection-control messaging.

### Non-Research Determination

The representative for the Department of Public Health of the Health Services Institutional Review Board (IRB) of the County of Santa Clara reviewed this activity and deemed it to be a public health surveillance activity, granting it a nonresearch activity determination that did not require IRB review.

## RESULTS

### Outbreak Characteristics

The first COVID-19 LTCF outbreaks occurred in mid-March 2020, prompting weekly testing in 41 sites by the end of July 2020. We focused our analysis on the largest outbreaks, which occurred at 6 LTCFs. Facilities A through F comprised 64.5% of cases (491/761) and 90.0% (72/80) of all deaths among LTCFs in the county, despite comprising just 11.3% (761/6722) of the total bed capacity. These facilities included 4 large SNFs and 2

**Table 1. Demographic Characteristics Across 6 Congregate Settings, March 2020–July 2020**

| | Facility A | Facility B | Facility C | Facility D | Facility E | Facility F | Total |
|---|---|---|---|---|---|---|---|
| LTCF capacity, n | 199 | 99 | 201 | 59 | 104 | 104 | 766 |
| LTCF type | SNF | SNF | SNF | SNF | AL/MC | AL/MC | — |
| Confirmed cases, n | 151 | 64 | 126 | 61 | 30 | 59 | 491 |
| Sex, female, n (% of all cases) | 85 (56.3) | 40 (62.5) | 77 (61.1) | 46 (75.4) | 25 (83.3) | 44 (74.6) | 317 (64.6) |
| Resident age, median, y | 75.5 | 79 | 76 | 87 | 84 | 90 | 81 |
| Staff age, median, y | 48 | 53 | 40.5 | 52 | 48 | 47.5 | 47 |
| COVID-19 severity | | | | | | | |
| Hospitalization | 44 | 5 | 11 | 10 | 0 | 1 | 71 |
| ICU admission | 8 | 1 | 23 | 3 | 6 | 5 | 46 |
| Deaths | 27 | 10 | 11 | 11 | 5 | 8 | 72 |

Abbreviations: AL/MC, Assisted Living Memory Care; COVID-19, coronavirus disease 2019; ICU, intensive care unit; LTCF, long-term care facility; SNF, skilled-nursing facility.

assisted-living facilities with memory care services. Table 1 describes the characteristics of their staff (median age, 47 years) and residents (median age, 81 years). Their demographics were comparable to other facilities in the county (data not shown). Moreover, among the Centers for Medicare and Medicaid Services (CMS)–certified SNFs in the county, outbreak size was not statistically associated with facility quality rating or minutes of care per day by nurses or by aides (Spearman correlation coefficients of $\rho = -0.12, 0.05, 0.10$ have $P = 0.51, 0.76, 0.56$ respectively). Likewise, quality ratings among the 4 SNFs with the largest outbreaks included the highest (n = 1), lowest (n = 1), and average (n = 2) facility quality scores.

The large outbreaks came in 2 waves. The initial cases from the 4 SNF outbreaks in March 2020 (Table 1; Facilities A–D) were detected in symptomatic staff. These outbreaks arose early in the pandemic when treatment, diagnostic, and PPE constraints were significant. The next wave included 2 large outbreaks, both detected in June, in assisted-living/memory care centers, with the first cases also identified in staff. Table 2 outlines the timeline and scale of the outbreaks with the recovery of viral genomes from staff and residents. On average, 40.7% of residents as a percentage of facility capacity (range, 12.5–79.6%) tested positive across the 6 facilities, while outbreaks in other county facilities were limited to less than 10% of their capacity (data not shown). All 239 PCR-positive samples for SARS-CoV-2 with a Ct value of no greater than 32 were submitted for WGS, and near-complete (>90%) viral genomes were recovered for 196 samples, representing 50 sequences from staff (27.9% of known staff cases) and 146 from residents (46.8% of known resident cases).

### Phylogenetic Analysis

We constructed a phylogenetic tree consisting of the 196 SARS-CoV-2 genomes from the LTCF facilities investigated together with 509 additional genomes from residents of SCC diagnosed with COVID-19, to understand the relationship between the outbreaks and other circulating genotypes (Figure 1). The additional genomes include COVID-19 cases from hospitals, community clinics, public testing sites, and jails. We found that SCC contained all the major global clades circulating at the time, as defined by NextStrain (19A, 19B, 20A, 20B, 20C) [18]. Each SNF outbreak (Facilities A–D) formed a separate, monophyletic cluster on the tree. Within each outbreak, a plurality of samples (42.3%, on average) shared a single

**Table 2. Skilled-Nursing Facility Outbreak Characteristics and Outcomes, March 2020–July 2020**

| Outbreak Characteristics | Facility A | Facility B | Facility C | Facility D | Facility E | Facility F | Total |
|---|---|---|---|---|---|---|---|
| Outbreak onset (earliest confirmed positive PCR) | March, week 3 | March, week 3 | March, week 1 | March, week 2 | June, week 1 | June, week 3 | N/A |
| Days to outbreak peak | 22 | 19 | 36 | 27 | 13 | 5 | 20.3[a] |
| Days to last case | 54 | 38 | 109 | 47 | 18 | 30 | 49.3[a] |
| COVID-19 positive staff | 49 | 17 | 54 | 18 | 17 | 24 | 179 |
| COVD-19 positive residents | 102 | 47 | 72 | 43 | 13 | 35 | 312 |
| Residents % positive | 51.3 | 47.5 | 79.6 | 35.8 | 12.5 | 33.7 | 40.7 |
| Sequences recovered | 47 | 33 | 32 | 25 | 19 | 40 | 196 |
| Staff sequences (% of known cases) | 9 (18.4) | 8 (47.1) | 8 (14.8) | 1 (5.6) | 10 (58.8) | 14 (58.3) | 50 (27.9) |
| Residents sequences (% of known cases) | 38 (37.3) | 25 (53.2) | 24 (33.3) | 24 (55.8) | 9 (69.2) | 26 (74.3) | 146 (46.8) |
| Sequences with index genotype (% of WGS) | 15 (31.9) | 4 (12.1) | 19 (59.4) | 8 (32.0) | 13 (68.4) | 26 (65.0) | 83 (42.3) |

For SNFs (Facilities A–D), average bed capacity based on Centers for Medicare and Medicaid Services 2019 licensing data (medicare.gov/nursinghomecompare), For assisted-living facilities, total bed capacity was used as the denominator.

Abbreviations: COVID-19, coronavirus disease 2019; PCR, polymerase chain reaction; SNF, skilled-nursing facility; WGS, whole-genome sequencing.

[a]Average time to outbreak peak/resolution in days.

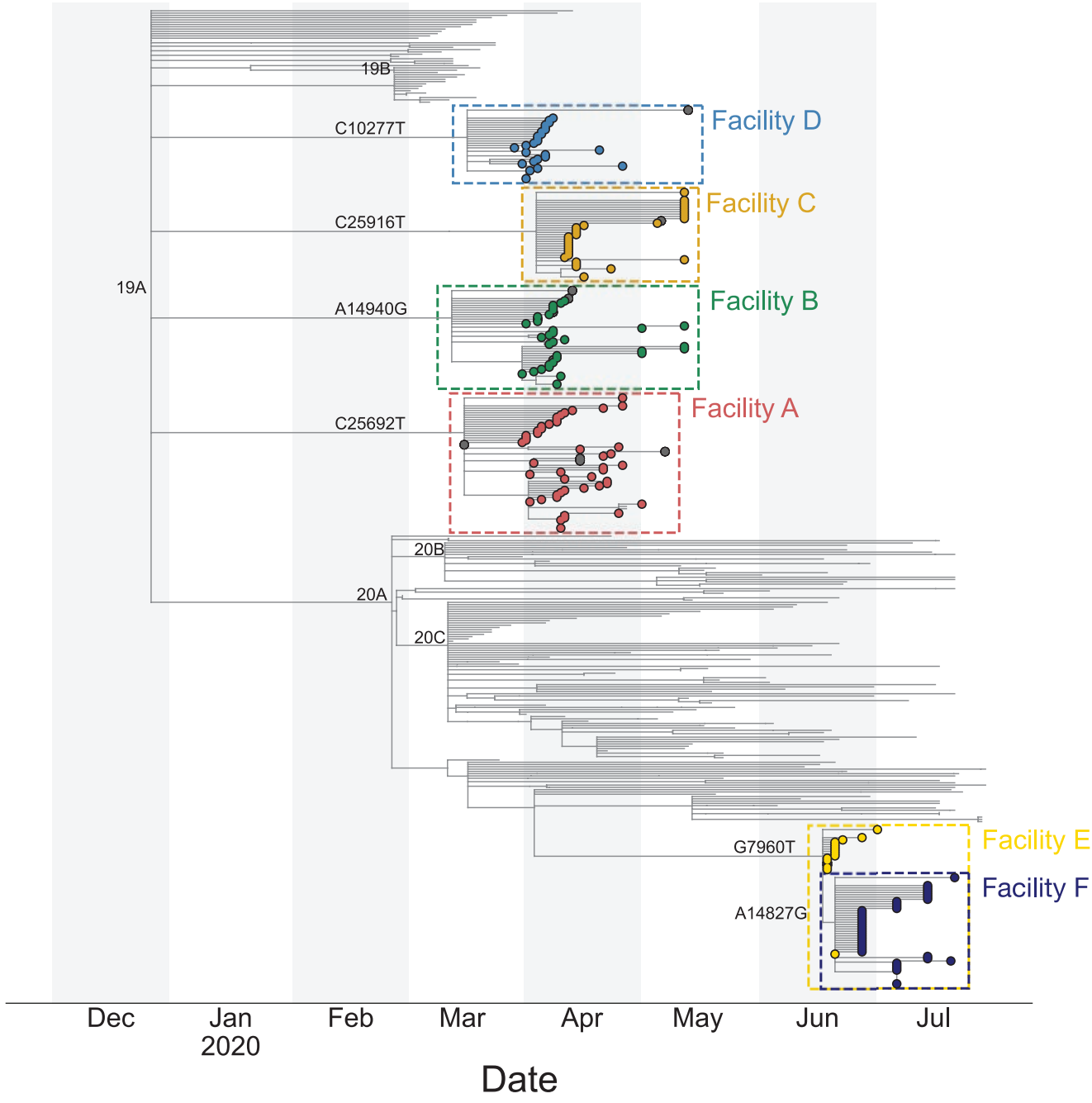


**Figure 1.** Six major SARS-CoV-2 outbreaks in congregate senior healthcare facilities in Santa Clara County are monophyletic. Shown is a maximum-likelihood phylogenetic tree built from 706 viral genomes from Santa Clara County samples collected between February and July 2020. Resident and staff samples from each facility are drawn as dots colored by the facility at which the individual was tested or worked (n = 196). Dashed boxes contain all nodes descended from the index genotype of each facility, defined as the common ancestor of all genotypes sampled from residents at that facility (incoming branches are labeled with the defining SNV). NextStrain clades are labeled 19A–20C. All staff and resident genomes from each facility descend from the index genotype, consistent with a single introduction. Gray nodes (n = 16) indicate contextual samples which are also descended from facility index genotypes, representing potential onward transmission. Abbreviations: SARS-CoV-2, severe acute respiratory syndrome coronavirus 2; SNV, single nucleotide variant.

genotype ancestral to all other genotypes recovered from the outbreak, consistent with an introduction by an index case followed by rapid spread (Table 2). Each outbreak lineage is defined by a small number (1–4) of single nucleotide mutations (denoted as single nucleotide variants [SNVs]) relative to the root Wuhan-Hu-1 genotype, which is their earliest common ancestor (Facility A: C10277T; Facility B: C25692T; Facility C: A14940G; Facility D: C25916T). Interestingly, all sequences belong to the 19A clade, and none of the genomic lineages found at Facilities A through D were of the SCC1 lineage (defined by

SNV G29711T), which defined early community spread within SCCPHD [3], nor the WA1 lineages characterizing the Grand Princess cruises. The outbreaks in both assisted-living facility outbreaks were also monophyletic (Facilities E and F). Notably, the Facility F outbreak lineage was a descendent of the Facility E outbreak lineage: Facility F genomes contained the G7960T mutation associated with the Facility E outbreak as well as an additional A14827G mutation. This genetic connection between the outbreaks was further supported by epidemiological data described below.

Genetic analysis uncovered further instances of staff sharing between facilities. While 37 of the 50 staff genomes (74%) recovered matched the lineage associated with the facility in which the staff member was tested, 13 of the genomes (26%) were found during response-driven testing at a different facility. Further case investigation revealed that each of the 13 staff also worked at the facility where their matching genotype was in circulation, and those cases are colored in Figure 1 according to the latter facility. Even in this small sample, those links suggest transmission opportunities between facilities by infected staff.

#### Transmission Kinetics of SARS-CoV-2

To better understand the kinetics of transmission in LTCFs we studied the genetic and epidemiologic relationships between residents and staff. Figure 2 shows the viral genomes of cases among residents (orange) and staff (blue) from Facilities A and B. Samples from other LTCFs (pink) and the community (dark gray) are also included as a reference. Thirteen genomes, from the earliest cases at Facility A, have an identical genotype. Subsequent mutations define limited onward transmission within the facility. One such transmission chain (Figure 2A, red box) contains a cluster of 4 residents who were roommates or across a corridor from one another. They were transferred to a COVID-19 isolation wing after receiving positive PCR results. Two weeks later, a downstream genotype containing 1 additional SNV was detected in a staff member.

The red box in Figure 2B contains 10 identical genotypes from 9 residents and 1 staff who were tested at Facility B. All of these residents were originally co-located along 1 wing (eg, shared rooms and restrooms, across the hallway). This 1 geographic hotspot within the facility appeared to give rise to genotypically identical cases for 6 weeks from late March to early May.

We also investigated the association between Facilities E and F suggested by the genomic data (Figure 3). One asymptomatic staff member working at both Facilities E and F had a plausible role in introducing COVID-19 to Facility F (red arrow). This individual had a positive PCR result at Facility E in late June and was also the earliest identified case from the outbreak at Facility F. The genome recovered from this case had 1 SNV distinguishing it from the genotype circulating in Facility E and that was subsequently present in all viral genomes recovered from the 59 cases at Facility F. Two asymptomatic certified nursing assistants (CNAs) who shared a household worked in at least 3 LTCFs including Facilities E and F. The viral

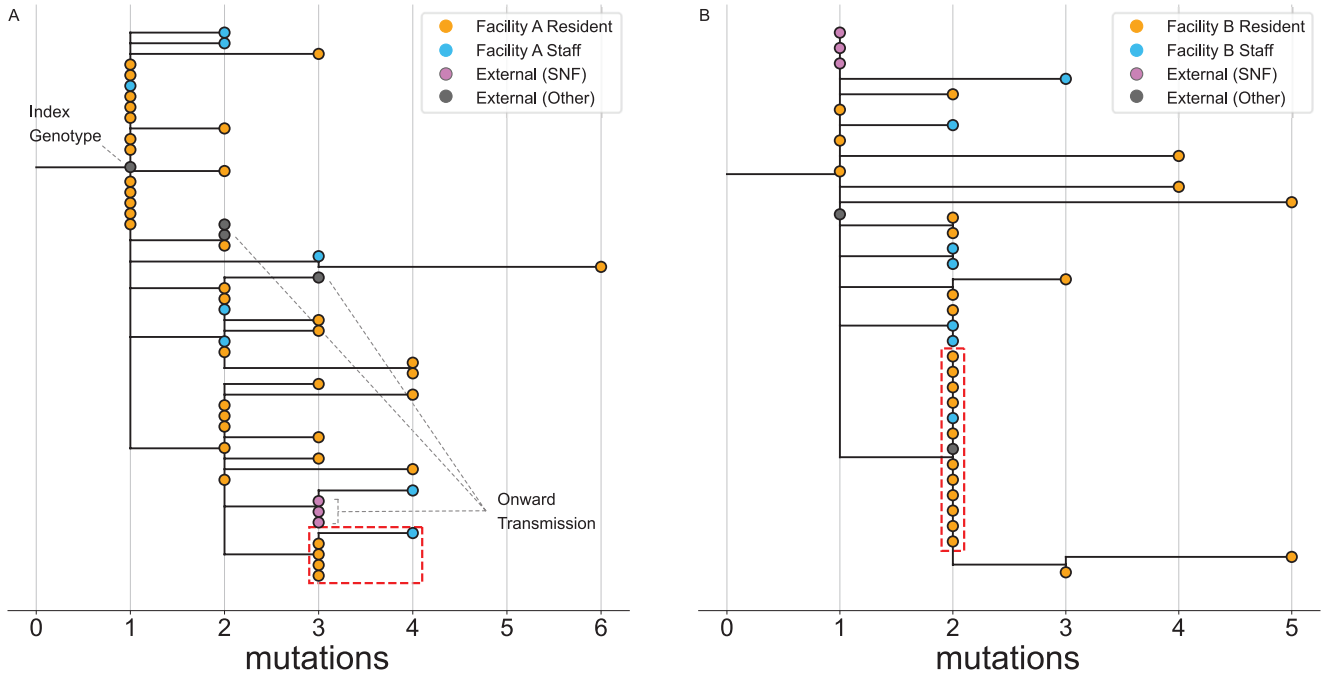


**Figure 2.** Genetic linkages between residents and staff within each facility. Insets of the tree in Figure 1 showing staff and residents from Facilities A (*A*) and B (*B*) together with samples from elsewhere descended from the index genotype. Instances where clusters of staff and residents share additional mutations beyond the index genotype of each outbreak indicate repeated transmission between those groups, and instances where external samples also share those mutations indicate likely onward transmission from the facility back into the community. Red boxes indicate clusters with epidemiological relationships discussed in the text. Abbreviation: SNF, skilled-nursing facility.

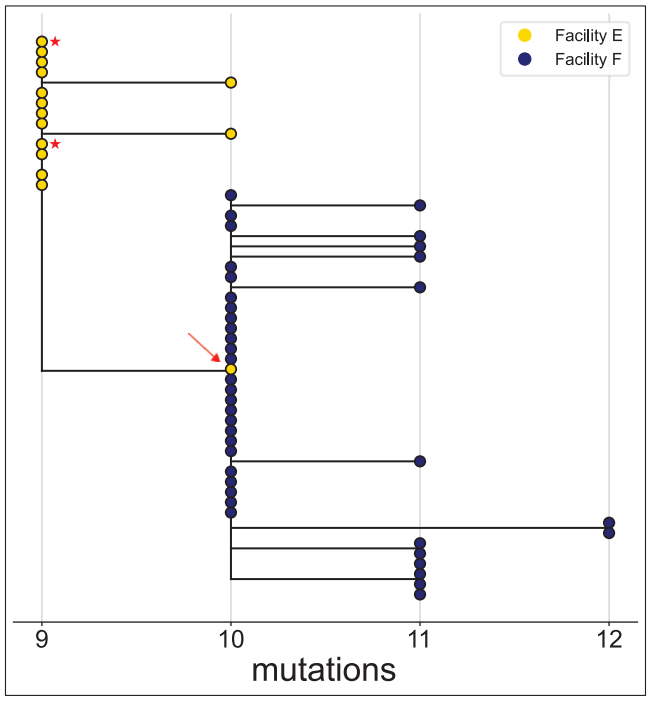


**Figure 3.** Shared staff seed outbreak. The index case of the outbreak at Facility F (red arrow) was a staff member also working at Facility E. The genomic diversity of Facility F is nested entirely inside that of Facility E. Red asterisks indicate 2 additional CNAs who worked at both facilities and share a household. Abbreviation: CNA, certified nursing assistant.

genotypes sequenced from these individuals were identical and matched the dominant genotype associated with Facility E. In this circumstance, transmission was likely associated with the shared household. There was no evidence to indicate that either of the staff members transmitted the virus to others at any of the multiple sites they were employed. Rapid intervention, exclusion from work, and home isolation may have prevented further transmission into the multiple facilities where these providers worked.

#### Staff–Resident Interactions

We also sought to use the genomic data to clarify whether transmission was staff-to-staff, staff-to-resident, or resident-to-resident. Across 146 recovered resident genomes, we identified 10 pairs of resident roommates with identical genomes and 11 pairs of roommates with unmatched genomes differing by at least 1 SNV (consistent with unidentified intermediaries). While roommates share close proximity and an increased likelihood of receiving care from the same staff, other breaches in infection control could present opportunities for transmission. Of the 50 genomes obtained from healthcare workers across the 6 outbreaks, 26 (54.2%) were from CNAs, who perform duties such as feeding, bathing, and toileting activities, involving close or face-to-face contact, increasing the opportunity for person-to-person transmission.

With shelter-in-place orders imposed since 13 March 2020, nonessential care or elective wellness activities were all but discontinued in SNFs, and residents were unable to socialize or circulate as they may have chosen to do normally. While the direction of transmission between residents and staff can be challenging to ascertain, with residents largely confined to their living quarters, transmission was likely amplified by healthcare worker intermediaries.

## DISCUSSION

Our investigation, spanning months of active, response-driven PCR testing for SARS-CoV-2, revealed a consistent pattern of transmission driving large LTCF outbreaks in SCC. In each case, a single viral introduction was followed by rapid proliferation, with the same SARS-CoV-2 lineage persisting for weeks despite infection-prevention efforts. Healthcare providers were infected by the same viral lineage circulating among their patients. While the direction of transmission cannot be established from this analysis, it reinforces the need for setting-specific and real-time infection-control guidance for staff, visitors, and residents [19].

Certified nursing assistants represent a majority of caregiver roles in LTCFs who provide essential face-to-face patient care services. They often care for multiple residents, sometimes exceeding the recommended ratios, and are less likely to receive comprehensive and refresher infection-control training, compared with licensed nursing staff. These factors can elevate transmission risks among CNAs and similar caregiver types. Other investigations have identified comparably credentialed personnel in the United Kingdom and United States to be similarly vulnerable [20, 21]. It is prudent to enhance the infection-control training offered to these provider groups.

In the United States, it is common for staff to work in multiple LTCFs, and a worker exposed in 1 facility may bring the virus to another [20, 22]. In March 2020, the CDC identified staff members working in multiple nursing homes as a likely source of early spread in Washington state [2], and recent work of Chen et al [20] used smartphone data to show a correlation in cross-staff movement and outbreak size across US nursing homes. Genomics can validate such observation and statistical data, and our study provides evidence of staff acting as a transmission link between facilities.

To curb the likelihood of onward transmission within and across healthcare facilities, in-depth contact tracing for healthcare providers may be beneficial. Screening all household and close contacts of positive healthcare providers could provide insight into the intermediaries between community- and healthcare-associated transmission.

Transitioning WGS from research into applied public health requires rapid return of phylogenetic results from clinical

samples, ideally within 3 to 5 days [9]. In our experience, creating an end-to-end process from clinical sample to an assembled phylogenetic tree within 1 business week required multiple stakeholder commitments, including the following: standardized processes moving samples between testing and sequencing laboratories, batched sequencing workflows to control costs, agile analysis methods, and frequent communication. Providing visualizations of phylogenetic data to public health and facility management became an infection-control tool, illustrating, in real-time, targets to break transmission.

While it is challenging to balance the social and behavioral needs of SNF residents under pandemic restrictions, the exclusion of community, family, visitors, and nonessential support services may have prevented multiple introductions into facilities, which could amplify already escalating case counts and resource constraints. However, the necessity of staff entry and interaction with residents creates a baseline level of transmission risk, making appropriate training, testing, cohorting, and support for staff essential components of infection control. Going forward, it is important to continue genomic surveillance in light of SARS-CoV-2 antigenic evolution, waning natural immunity, and shifts in transmission patterns.

## Supplementary Data

Supplementary materials are available at *Clinical Infectious Diseases* online. Consisting of data provided by the authors to benefit the reader, the posted materials are not copyedited and are the sole responsibility of the authors, so questions or comments should be addressed to the corresponding author.

## Notes

***Acknowledgments.*** The authors thank the authors of originating and submitting laboratories of the sequences from GISAID's EpiFlu Database, which are used in this research. In particular, we thank Dr Charles Chiu, Director, UCSF–Abbot Viral Diagnostics Discovery Center; the Ashley Laboratory at Stanford; and the Stanford Clinical Virology Laboratory for sharing other sequences from cases of COVID-19 who were Santa Clara County residents. A full Acknowledgments table is available as supplementary materials.

***Potential conflicts of interest.*** The authors: No reported conflicts of interest. All authors have submitted the ICMJE Form for Disclosure of Potential Conflicts of Interest.

***SCC Public Health Department Special Investigations Unit:*** Rensen Khoshabian, Melanie Diep, Daniel Castillo, Dante Afable, Syeda Iqbal, Maro Boghos, Nagateja Yeddulla, Omar Munoz, Emiko Yamamoto, Doug Medrano, Carrie Ludman, Snizhana Khomych.

***SCC Public Health Nurses:*** Annie Chan, Quynh-Nh Pham, Maite Medina, Stephany Ponce, Yilei Hs, Ann Morales De Aguinaga, Evanthia Phanthavone, Stephany Ponce , Julia Orona.

***Chan Zuckerberg Biohub COVIDTracker Consortium:*** Lab—Karan D. Bhatt, Lienna Chan, Gloria R. Castañeda, Sabrina Mann, Manu Vanaerschot, G. Renuka Kumar, Kalani Ratnasiri, Emily D. Crawford; Genomics—Norma Neff, Michelle Tan, Angela Detweiler, Rene Sit; Data Science—Jack Kamm, Angela Oliveira Pisco, Aaron McGeever, Phoenix Logan, Samantha Hao, James T. Webber, Lucy M Li, Tina Zheng; Rapid Response: Lusajo Mwakibete, Maira Phelps, Vida Ahyong; Chan Zuckerberg Initiative Tech—Shannon Axelrod, Tony Tung, Jonathan Sheu, Mark Zhang; Leadership—Joe DeRisi .

## References

1. US Census Bureau. QuickFacts: Santa Clara County, California. Available at: https://www.census.gov/quickfacts/santaclaracountycalifornia. Accessed 3 October 2020.
2. McMichael TM. COVID-19 in a long-term care facility—King County, Washington, February 27–March 9, 2020. MMWR Morb Mortal Wkly Rep **2020**; 69:339-42. doi: 10.15585/mmwr.mm6912e1.
3. Deng X, Gu W, Federman S, et al. Genomic surveillance reveals multiple introductions of SARS-CoV-2 into Northern California. Science **2020**; 369: 582-7.
4. County of Santa Clara. COVID-19 long-term care facility (LTCF) dashboard—novel coronavirus (COVID-19). Available at: https://www.sccgov.org/sites/covid19/Pages/dashboard-long-term-care-facilities.aspx. Accessed 16 February 2021.
5. Blackburn RM, Frampton D, Smith CM, et al. Nosocomial transmission of influenza: a retrospective cross-sectional study using next generation sequencing at a hospital in England (2012-2014). Influenza Other Respir Viruses **2019**; 13:556-63.
6. Grubaugh ND, Faria NR, Andersen KG, Pybus OG. Genomic insights into Zika virus emergence and spread. Cell **2018**; 172:1160-2.
7. Houlihan CF, Frampton D, Ferns RB, et al. Use of whole-genome sequencing in the investigation of a nosocomial influenza virus outbreak. J Infect Dis **2018**; 218:1485-9.
8. Scarpino SV, Iamarino A, Wells C, et al. Epidemiological and viral genomic sequence analysis of the 2014 Ebola outbreak reveals clustered transmission. Clin Infect Dis **2015**; 60:1079-82.
9. 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. doi: 10.1016/S1473-3099(20)30562-4.
10. Aggarwal D, Myers R, Hamilton WL, et al. The role of genomics in understanding COVID-19 outbreaks in long term care facilities. OSF Preprints **2020**. doi: 10.31219/osf.io/7y9rk.
11. Li H. Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics **2018**; 34:3094-100.
12. Quick J, Grubaugh ND, Pullan ST, et al. Multiplex PCR method for MinION and Illumina sequencing of Zika and other virus genomes directly from clinical samples. Nat Protoc **2017**; 12:1261-76.
13. R&d DP. COVID-19 ARTIC v3 Illumina library construction and sequencing protocol. **2020**. doi: 10.17504/protocols.io.bibtkann.
14. Loman NJ, Quick J. hCoV-2019/nCoV-2019 version 3 amplicon set. Available at: https://artic.network/resources/ncov/ncov-amplicon-v3.pdf. Accessed 22 June 2020.
15. Li H, Handsaker B, Wysoker A, et al; 1000 Genome Project Data Processing Subgroup. The sequence alignment/map format and SAMtools. Bioinformatics **2009**; 25:2078-9.
16. Grubaugh ND, Gangavarapu K, Quick J, et al. An amplicon-based sequencing framework for accurately measuring intrahost virus diversity using PrimalSeq and iVar. Genome Biol **2019**; 20:8.
17. Elbe S, Buckland-Merrett G. Data, disease and diplomacy: GISAID's innovative contribution to global health. Glob Chall **2017**; 1:33-46.
18. Hadfield J, Megill C, Bell SM, et al. Nextstrain: real-time tracking of pathogen evolution. Bioinformatics **2018**; 34:4121-3.
19. Patel MC, Chaisson LH, Borgetti S, et al. Asymptomatic SARS-CoV-2 infection and COVID-19 mortality during an outbreak investigation in a skilled nursing facility. Clin Infect Dis **2020**; 71:2920-6.
20. Chen MK, Chevalier JA, Long EF. Nursing home staff networks and COVID-19. Proc Natl Acad Sci **2021**; 118:1-7. doi: 10.1073/pnas.2015455118.
21. 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.
22. Van Houtven CH, DePasquale N, Coe NB. Essential long-term care workers commonly hold second jobs and double- or triple-duty caregiving roles. J Am Geriatr Soc **2020**; 68:1657-60.