Title:

Utilization of Whole Genome Sequencing to Understand SARS-CoV-2 Transmission Dynamics in Long-Term Care Facilities, Correctional Facilities and Meat Processing Plants in Minnesota, March – June 2020

Authors:

Nicholas B Lehnertz, Xiong Wang, Jacob Garfin, Joanne Taylor, Jennifer Zipprich, Brittany VonBank, Karen Martin, Dana Eikmeier, Carlota Medus, Brooke Wiedinmyer, Carmen Bernu, Matthew Plumb, Kelly Pung, Margaret A. Honein, Rosalind Carter, Duncan MacCannell, Kirk E. Smith, Kathryn Como-Sabetti, Kris Ehresmann, Richard Danila, Ruth Lynfield

Affiliations:

Minnesota Department of Health, St. Paul, Minnesota, USA (N. Lehnertz, X. Wang, J. Garfin, J. Zipprich, B. VonBank, K. Martin, D. Eikmeier, C. Medus, B. Wiedinmyer, C. Medus, M. Plumb, K. Pung, K. Smith, K. Como-Sabetti, K. Ehresmann, R. Danila, R. Lynfield)
Center for Disease Control and Prevention, Atlanta, Georgia, USA (J. Taylor, M. Honein, R. Carter, D. MacCannell)

Abstract:

Congregate settings and high-density workplaces have endured a disproportionate impact from COVID-19. In order to provide further understanding of the transmission patterns of SARS-CoV-2 in these settings, whole genome sequencing (WGS) was performed on samples obtained from 8 selected outbreaks in Minnesota from March – June, 2020. WGS and phylogenetic analysis was conducted on 319 samples, constituting 14.4% of the 2,222 total SARS-CoV-2-positive individuals associated with these outbreaks. Among the sequenced specimens, three LTCFs and both correctional facilities had spread associated with a single genetic sequence. A

fourth LTCF had outbreak cases associated with two distinct sequences. In contrast, cases associated with outbreaks in the two meat processing plants represented multiple SARS-CoV-2 sequences. These results suggest that a single introduction of SARS-CoV-2 into a facility can result in a widespread outbreak, and early identification and cohorting of cases, along with continued vigilance with infection prevention and control measures is imperative.

Text

Introduction:

COVID-19 has disproportionately impacted adults residing in long-term care facilities (LTCFs) in the United States (1-5). LTCF outbreaks have caused high numbers of hospitalizations and death. Similar findings have been reported in correctional facilities, with SARS-CoV-2 infection incidence among inmates and staff over five times greater and age-adjusted mortality over three times greater than that of the general population (6-8). Workers in high-density workplaces (e.g., meat processing plants) have similarly been heavily affected, with a disproportionate impact on minority populations (9-11).

The first identified COVID-19 case in Minnesota was detected on March 6, 2020. Shortly thereafter, COVID-19 outbreaks occurred across the state, including in LTCFs (March 12, 2020) and meat processing plants (March 15, 2020), followed shortly thereafter by correctional facilities (March 25, 2020). From March 6 through June 30, 2020, the Minnesota Department of Health (MDH) identified and responded to 1,060 distinct outbreaks of COVID-19 in LTCFs, comprising 4,421 resident cases and 3,002 staff cases. In addition, there were 4 discrete outbreaks in correctional facilities that resulted in a total of 382 cases, and 68 outbreaks in meat

processing plants resulting in approximately 2,616 cases among employees¹; outbreaks in these 3 settings accounted for 31.3% of all identified cases in Minnesota.

For outbreaks in both congregate settings and high-density workplaces, it was difficult to confirm the temporal and relational aspect of SARS-CoV-2 transmission; the role of intra-facility spread versus multiple introductions was difficult to disentangle. Whole-genome sequencing (WGS) of outbreak case specimens can be used to understand transmission dynamics and relatedness of viral pathogens in infectious disease outbreaks (12-15). Unprecedented efforts to sequence SARS-CoV-2 genomes have occurred at the local, regional, national and international level in order to investigate potential reinfections (16-19), nosocomial transmission (20), patterns of community spread (21-23), and sources of introduction of SARS-CoV-2 without known epidemiologic links (24).

In Minnesota, as part of the CDC SARS-CoV-2 Sequencing for Public Health Emergency Response, Epidemiology and Surveillance (SPHERES) consortium, the Minnesota Molecular Surveillance of SARS-CoV-2 (MN-SOS) initiative solicited outbreak specimens for sequencing and genetic variation analysis to understand viral transmission patterns in congregate settings and meat processing plants. WGS of a convenience sample of SARS-CoV-2 positive specimens associated with outbreaks was performed to supplement epidemiological information and assess whether single or multiple introductions likely occurred during a facility outbreak, and to evaluate molecular relatedness.

Methods:

Three types of outbreak settings (LTCF, correctional facilities and meat processing plants) were chosen for WGS, with selection of specific facilities based in part on the impact and severity of the outbreak, the need for further clarity regarding transmission patterns, and the

availability of samples. Selected outbreaks occurred between March 6 and June 30, 2020 at four unique LTCFs (LTCF A-D), two correctional facilities (Correctional Facilities A and B), and two meat processing plants (Processing Plant A and B), along with identified cases in individuals residing in the same county as Processing Plant A (Community Samples A). An outbreak at a LTCF was defined as ≥ 1 confirmed COVID-19 case in a resident or staff member. In correctional facilities, an outbreak was defined as 1) ≥ 2 cases in the inmate population > 7 days after intake to a new facility, with an epidemiologic link², or 2) ≥ 2 cases in correctional staff with an epidemiologic link³ at the facility, or 3) ≥ 1 facility-acquired⁴ COVID-19 cases in an inmate. An outbreak at meat processing facilities was defined as ≥ 3 laboratory-confirmed COVID-19 cases among workers at a facility who resided in separate households. On June 1, a time component was added to the outbreak definition in meat processing plants, adding that case onset dates are within 14 days of each other. Cases at all outbreak locations were defined as an individual testing SARS-CoV-2 positive by reverse transcription polymerase chain reaction (RT-PCR) using the original CDC assay protocol (25).

Epidemiologic data, including gender, age, symptom status, symptom onset date, residence, occupation, and potential source of exposure were collected via interview with persons having lab-confirmed SARS-CoV-2.

WGS was performed by the MDH Public Health Laboratory (PHL) on available SARS-CoV-2 RT-PCR positive specimens, collected from March 6 through June 30, 2020. Specimens were obtained from the nasopharynx, the anterior nares or oropharynx. SARS-CoV-2 RNA extracts were acquired either as residual from clinical testing at MDH-PHL or other clinical laboratories serving Minnesota residents. cDNA and tiled amplicons were created as described in the ARTIC Network nCoV-2019 sequencing protocol (26). Sequencing Libraries for Illumina

next-generation sequencing were prepared according to the Nextera DNA Flex protocol created by the State Public Health Bioinformatics Group (StaPH-B) (27) and sequenced using 2x250 bp Illumina V2 chemistry on MiSeq instruments. Consensus SARS-CoV-2 genome sequences for each specimen were generated with the StaPH-B Toolkit Monroe pipeline (28). Assembled SARS-CoV-2 genomes were individually reviewed in Geneious Prime 2019.2.1 (29) and genomes with gaps greater than 125 nucleotides were discarded.⁵

The Augur toolkit (30) was used for alignment of SARS-CoV-2 genome consensus sequences, phylogenetic tree generation, and incorporation of epidemiological sequence metadata. Genomes were aligned with MAFFT version 7.310 with options “--keeplength --reorder --anysymbol --nomemesave --adjustdirection” (31). Variation in sequences identified in the first 54 and last 67 bases of the Wuhan-Hu-1 (GenBank: MN908947.3) reference sequence was masked during tree generation due to the inability of the tiled-amplicon sequencing approach to reliably generate sequence in those regions. IQ-TREE version 1.6.1 was used for phylogenetic tree creation with parameters “-ninit 2 -n 2 -me 0.05” (32). Output from Augur was visualized using Auspice as hosted by the nextstrain team at <http://auspice.us/> (30). The resulting trees were visualized with the Interactive Tree of Life (iTOL) (33) with branch lengths rounded and scaled represent mutations from the reference. Pangolin lineages for all samples were retrieved following submission of assemblies to GISAID (30, 34-35).

Genetically closely related sequences (i.e., clusters) were defined as cases which were associated with an outbreak and formed a monophyletic clade on the state-wide phylogenetic tree.

In accordance with federal human subjects protection regulations at 45 CFR §46.101c and §46.102d and with the Guidelines for Defining Public Health Research and Public Health

Non-Research, the project was reviewed by a human subjects protection coordinator at the CDC and in Minnesota DOH and determined to be a non-research, public health response exempt from institutional review board evaluation.

Results:

WGS and subsequent phylogenetic analysis was successfully conducted on 319 samples associated with the eight outbreaks, constituting 14.4% of the 2,222 total positive cases identified from outbreaks in Minnesota through June 2020. Specimens were obtained from staff and residents from four different LTCFs (180 specimens sequenced out of 505 identified cases, 35.6%), staff and inmates from 2 different correctional facilities (110 specimens sequenced out of 544 identified cases, 20.2%) and among employees at 2 different meat processing plants, along with identified community cases (29 samples sequenced out of 1,173 identified cases, 2.5%) (Table 1). Among most of the specimens that were sequenced, three LTCFs and both correctional facilities had spread associated with a single genetic sequence unique to each outbreak facility. A fourth LTCF had outbreak cases associated with two distinct sequences. In contrast, cases associated with outbreaks in the two meat processing plants represented multiple SARS-CoV-2 sequences. (Figure 1)

Single Cluster in Long-Term Care Facilities:

During the COVID-19 outbreak at LTCF A⁷ between April 15 and June 11 (Figure 2A), 51 of 77 residents, and 38 of 108 health care personnel (HCP) tested were confirmed to be infected, following identification of positive HCP. Specimens from 17 residents (33.3% of cases) and 10 staff (26.3% of cases) were available for WGS. SARS-CoV-2 viral sequences from these 27 persons were genetically closely related (Pangolin lineage B.1.2). Viral genomes from two

HCP sampled on April 30 and one resident sampled on May 18 at facility A did not cluster with each other or the primary outbreak cluster, though all were a part of the broad Pangolin lineage B.1.

In LTCF B from April 29 - June 11 (Figure 2B), 114 of 182 tested residents and 76 of 233 tested HCP were confirmed to be SARS-CoV-2 positive, following identification of a SARS-CoV-2 positive resident on April 29. All 82 sequenced specimens from this facility, including 77 residents (67.5% of cases) and 5 HCP (6.6% of cases), were closely related (Pangolin lineage B.1.116).

The first identified COVID-19 case at LTCF C (Figure 2C) was on April 24. Four positive HCP and 3 symptomatic residents were identified by April 30. Throughout May and June, facility-wide testing was implemented, with approximately 941 residents and staff tests performed, with 80 positive residents and 52 positive staff identified. Phylogenetic analysis of the 32 successfully sequenced genomes, including 9 residents (10.8% of cases) and 23 staff (41% of cases) showed 29 of the 32 cases were closely related (Pangolin lineage B.1.2). The remaining three cases (Pangolin lineages B.1 and B.4) were not closely related to each other nor identified with further transmission.

Single Cluster in Correctional Facilities:

In late March, 2020, an outbreak of SARS-CoV-2 was identified in Correctional Facility A (Figure 3A). The first identified case was an inmate who became symptomatic and tested positive for SARS-CoV-2 on March 25. By March 30th, a total of 7 confirmed cases and 6 suspected cases among the inmate population were identified. Between March 30 and April 7, 15 staff tested positive for SARS-CoV-2. Analysis of the genetic relatedness of the virus from 34

inmates (73.9% of cases) and 15 staff (18.3% of cases) from Correctional Facility A were all closely related (Pangolin lineage A.1).

In early June, 2020, an outbreak was identified in Correctional Facility B (Figure 3B). The investigation revealed that an employee had developed symptoms consistent with COVID-19 on May 13, tested positive for SARS-CoV-2 on May 14, and was subsequently excluded from work and isolated at home. Approximately two weeks later, three additional cases, one staff member and two inmates from the same unit as the index patient, tested positive for SARS-CoV-2. A point-prevalence survey (PPS) on June 1 revealed 63 COVID-19 positive inmates among the 87 tested. Subsequent facility-wide testing of both staff and inmates identified cases in other units, 83 new cases in inmates and one new staff case identified among the approximately 2,200 persons tested. Phylogenetic analysis of this outbreak among the 1 staff member (0.5% of cases) and 60 inmates (29.1% of cases) at Correctional Facility B shows that all cases were closely related (Pangolin lineage B.1.2) and genetically identical to, or plausibly descended from, the sequence of SARS-CoV-2 from the index case.

Linking LTCF C with Correctional Facility B:

During the epidemiologic investigation at LTCF C, it was learned that a HCP at LTCF C was a household contact of a Correctional Facility B employee. Both individuals became symptomatic at the same time, and both subsequently tested positive in mid-May. SARS-CoV-2 genome sequences recovered from these two household contacts were identical to each other, and also to the genomic sequences recovered from 32 inmates at Correctional Facility B (Figure 4). In addition, this genomic sequence differs by only a single mutation (G5617T) from 13 cases sequenced from LTCF C.

Two Distinct Clusters in a Long-Term Care Facility:

LTCF D (Figure 2D) is a 100 bed facility with approximately 78 residents and 100 staff that experienced an outbreak beginning on April 17, 2020 with a symptomatic HCP. Identification of the first resident and staff case occurred on April 20, 2020, with subsequent testing identifying a total of 53 positive residents and 21 positive staff. While epidemiologically similar to other LTCF outbreaks, an analysis of the genetic relatedness among 39 sequenced cases demonstrated two distinct genetic clusters present in the facility during approximately the same time period. In contrast to the outbreaks in LTCFs A, B and C, viruses from both clusters appeared to circulate simultaneously throughout the facility, each contributing to the outbreak. All sequenced cases from LTCF D belonged to the broad Pangolin lineage B.1.

Multiple Clusters in Meat Processing Plants:

In early April 2020, an outbreak was detected at Processing Plant A (Figure 5A), a large primary and secondary meat processor. The outbreak at Processing Plant A continued for several weeks until mid-May when the number of cases among workers began to increase rapidly. From March 15 to July 1, there were 446 confirmed cases who reported working at Processing Plant A, including 4 (1%) cases who tested positive in March (management and office staff), 5 (1%) in April, 211 (47%) in May, and 226 (51%) in June. Analysis demonstrated that of the 16 samples (3.7% of cases) sequenced between March 15 and June 3, there were at least 6 unrelated clusters or singleton cases. While the majority of genomes sequenced from Processing Plant B belong to Pangolin lineages B.1, B1.2, B.1.26, one early case is genetically quite different (Pangolin lineage A.1). An interview confirmed that this early case had traveled out of the state during their exposure period (14 days before symptom onset).

Between May 15 and June 1, we sequenced samples obtained from eight cases in the county where Processing Plant A is located was performed (Community Samples A). From these eight samples, a total of 5 clusters were identified. Five of the eight samples were closely related with three of the clusters from Processing Plant A, while the remaining three samples formed two distinct clusters. Four of the five sequences from Community Samples A that clustered with sequences from Processing Plant A had sequences that were identical to sequences from Processing Plant A, and all four had no known contact with a verified case.

In mid-April 2020, an outbreak was identified among employees at Processing Plant B (Figure 5B), another large meat processing plant. By May 1, there were 649 confirmed cases among workers at Processing Plant B. Sequencing of the 5 available samples from Processing Plant B (0.7% of cases) identified one cluster and two singleton genomes, all belonging to Pangolin lineage B.1.

Discussion:

WGS identified three primary patterns of genetic relatedness among cases in various outbreak settings: 1) outbreaks where cases were part of a single, genetically related cluster, 2) an outbreak with two unique clusters of cases, each contributing to the outbreak during the same time period, and 3) outbreaks where there were multiple genetically distinct sequences present.

Phylogenetic analyses of the viral sequences from available specimens associated with outbreaks in LTCF A, LTCF B and LTCF C were all consistent with at least one primary cluster impacting each facility, suggesting that a single introduction of SARS-CoV-2 into a facility can result in a widespread outbreak. This is similar to previous reports, where WGS has evidenced rapid spread in high density settings as opposed to multiple introductions contributing to the

outbreak (20). Cases from LTCF D, in contrast, formed 2 distinct genetic clusters, one of 17 related samples and the other consisting of 22 samples. This is consistent with a potential scenario where there were two separate, independent introductions into the facility, and subsequent parallel intra-facility spread of each individually distinct sequence.

Phylogenetic analysis conducted in both LTCF A and LTCF C also demonstrated the presence of “outlier” SARS-CoV-2 viral sequences that were not genetically closely related to the primary cluster in each facility. This suggests community-acquired infection, and subsequent introduction of SARS-CoV-2 into the facility (3). Two of the three outlier cases at LTCF C tested positive over a month after the first identified case. Similarly, two of the three outlier cases identified at LTCF A were identified 10 days after the first identified case, while the third tested positive 28 days later. It is not possible to determine whether these introductions of distinct genetic sequences resulted in additional spread given that WGS characterization was not performed on all positive samples in each facility, and not all HCP or residents were tested. However, the timing of the identification of these outlier cases after the date of the first identified primary case suggest that mitigation strategies implemented after the initial identification of the outbreak, including cohorting strategies, infection prevention and control measures (IPC), and correct use of personal protective equipment (PPE) may have been effective in preventing intra-facility transmission of these late outlier cases, as has been previously reported (3, 22-23).

WGS identified a different genetic landscape in processing plants, one in which several distinct sequences contributed to the facility outbreak. This finding is despite only 2.5% of positive cases at the processing plants having been sequenced, suggesting that increased sequencing may have identify even greater diversity. Additionally, several genomes identified at Processing Plant A were either identical or closely related to genomes present in the surrounding

community (Community Samples A). Six of the eight sequenced Community Samples A were from persons with no known epidemiologic link to a case at Processing Plant A strongly suggesting an unrecognized connection. The benefit of WGS to identify previously unrecognized transmission patterns has previously been established (20, 36). While no definitive conclusions can be made regarding the direction of transmission, WGS provided strong evidence of a worker-community member spread, with communal housing, multigenerational families and group transportation as hypothesized factors potentially contributing to this transmission pattern.

WGS has shown to improve understanding of an outbreak upon retrospective analysis (3, 20-22), to provide justification for specific public health measures (22-23), and to add insight to transmission patterns in high-risk settings. Our work provides further support for the use of WGS in these situations, while identifying several additional public health implications. WGS has demonstrated that outbreaks in LTCFs and correctional facilities can be the result of a single introduction. Continued vigilance with facility-wide staff screening, and subsequent exclusion of symptomatic HCPs or staff and those with known or suspected contacts is imperative. WGS has demonstrated extensive intra-facility spread, with closely related sequences comprising all or the vast majority of cases contributing to the outbreak. Measures, such as IPC, consistent and correct use of PPE, cohorting of known positive residents and exclusion of positive HCPs must be maintained. WGS has also illuminated the transmission patterns in meat processing plants, including the multiple introductions identified through the multiple closely related sequences identified, and the related community strains. WGS has illustrated the need for community-level mitigation to prevent introductions in high-density worksites, including accessible community-wide testing, housing and transportation strategies, in addition to facility-level measures to prevent unintended introduction into the workplace.

There are several important limitations to this study. First, only a subset of specimens were available for sequencing due to different laboratory specimen retention policies. For example, at LTCF B, samples from only 5 staff were available for sequencing. Similarly, in Processing Plant B, only 5 samples were available due to a clinical testing laboratory protocol that resulted in the discarding of samples after approximately 7 days. In addition, not all available patient samples were able to be successfully sequenced, owing primarily to degraded or low concentrations of viral RNA in the specimen.

Another limitation is that not all staff and employees at the LTCFs, correctional facilities and processing plants agreed to be tested. Because of the incomplete genomic picture at each setting, definitive conclusions about single introductions in LTCF A and LTCF D are speculative, and these individual introductions may have actually resulted in some viral transmission that was not identified in the study.

Finally, we were not able to present sociodemographic data such as race or ethnicity pertaining to cases, or the facility population, associated with these eight outbreaks in Minnesota, due to limitations in the case investigation process and incomplete case data. This limitation is of particular importance, due to the disproportionate impact of COVID-19 on Black, Indigenous, and People of Color (BIPOC). Since BIPOC disproportionately experience incarceration, and a high proportion of meat processing plant employees are persons from immigrant communities, these settings can serve to amplify racial and ethnic health disparities related to COVID-19.

Long-term care facilities, correctional facilities, and high-density workplace settings have many factors that are hypothesized to contribute to rapid transmission of SARS-CoV-2. These include insufficient resources and training in IPC, difficulties implementing social distancing due

to close habitation and/or work environment, and delayed case detection and access to care (8, 11, 37). WGS has demonstrated that many outbreaks in Minnesota were caused by single introductions of SARS-CoV-2, highlighting the importance of consistent and correct usage of PPE, rigorous and systematic IPC, environmental control measures, and systematic testing of residents and staff to identify asymptomatic cases. As this pandemic continues, community mitigation strategies and strong enforcement of policies to reduce the risk of introducing SARS-CoV-2 virus into congregate settings are more important than ever. Similarly, IPC and aggressive containment practices are vital to mitigate the spread of SARS-CoV-2 upon introduction into a facility. WGS can be a useful tool to supplement epidemiological information and examine the role of facility and community factors contributing to SARS-COV-2 outbreaks in high risk settings.

Acknowledgments:

We would like to gratefully acknowledge the contributions from the following: MDH MEDSS Team, MDH Public Health Lab, MDH Case Investigation Team, MDH Outbreak Investigation Team, MDH Long-Term Care Facility Team, Matthew Binnicker and Joseph Yao at Mayo Clinic Laboratories, and Dr. Andrew C. Nelson and Sophia Yohe at M Health Fairview Molecular Diagnostics Laboratory.

Disclaimers:

The findings and conclusions in this report are those of the authors and do not necessarily represent the official position of the US Centers for Disease Control and Prevention.

Author Bio:

Dr. Nicholas Lehnertz is a Medical Specialist in Infectious Disease Epidemiology, Prevention and Control at the Minnesota State Department of Health. His current research involves the epidemiology of COVID-19 transmission patterns, clinical characteristics of pre-symptomatic COVID-19 infection in residents of long-term care facilities, and the understanding of human belief systems surrounding COVID-19.

Dr. Nicholas Lehnertz contact info: Minnesota Department of Health, 625 Robert Street North, St. Paul, MN 55164-0975 nick.lehnertz@state.mn.us 651-201-5270

Footnotes:

¹ Only including individuals interviewed and where workplace information was provided.

² An epidemiologic link among inmates is defined as residing on the same unit or ward within a 14-day time period of each other.

³ An epidemiologic link among correctional staff is defined as having the potential to have been within 6 feet for 15 minutes or longer while working in the facility during the 14 days prior to prior to the onset of symptoms; for example, worked on the same unit during the same shift. An epidemiologic link also requires that cases among correctional staff neither share a household together, nor were identified as close contacts with each other outside of the facility during the standard case investigation.

⁴ Facility-acquired COVID-19 infection in an inmate/resident is defined as a confirmed diagnosis 14 days or more after entry to the facility, without an exposure during the previous 14 days to another setting where an outbreak was known or suspected to be occurring.

⁵ Review consisted of discarding samples where read alignments to the Wuhan-Hu-1 reference had internal gaps greater than 125 nucleotides and removing trailing A nucleotides from consensus genomes in cases where genome alignments had reads mapped only to the reference sequence poly(A) tail and none of the preceding reference sequence.

⁶ No cases or samples sequenced after June 30 included in study. An outbreak is defined as closed if there are no new COVID-19 cases for 28 days after the onset date of the last case. The outbreak at Correctional Facility A was considered closed as of July 20, the outbreak at Correctional Facility B was considered closed as of August 5. The outbreaks at both Processing Plant A and Processing Plant B are considered ongoing as of November 6, 2020.

⁷ A full accounting of the outbreaks at LTCF A and LTCF B were previously published. *See* Taylor J, Carter RJ, Lehnertz N, et al. 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;69:1288–1295. DOI: <http://dx.doi.org/10.15585/mmwr.mm6937a3>

References

1. Arons MM, Hatfield KM, Reddy SC, Kimball A, James A, Jacobs JR, et al. Presymptomatic SARS-CoV-2 Infections and Transmission in a Skilled Nursing Facility.

N Engl J Med. 2020 May 28;382(22):2081-2090. doi: 10.1056/NEJMoa2008457. Epub
2020 Apr 24. PMID: 32329971; PMCID: PMC7200056.

2. Kimball A, Hatfield KM, Arons M, James A, Taylor J, Spicer K, et al. Asymptomatic and
Presymptomatic SARS-CoV-2 Infections in Residents of a Long-Term Care Skilled
Nursing Facility - King County, Washington, March 2020. MMWR Morb Mortal Wkly
Rep. 2020 Apr 3;69(13):377-381. doi: 10.15585/mmwr.mm6913e1. PMID: 32240128;
PMCID: PMC7119514.

3. Taylor J, Carter RJ, Lehnertz N, Kazazian L, Sullivan M, Wang X, et al. 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 Sep 18;69(37):1288-1295. doi:
10.15585/mmwr.mm6937a3. PMID: 32966272; PMCID: PMC7498172.

4. Dora AV, Winnett A, Jatt LP, Davar K, Watanabe M, Sohn L, et al. Universal and Serial
Laboratory Testing for SARS-CoV-2 at a Long-Term Care Skilled Nursing Facility for
Veterans - Los Angeles, California, 2020. MMWR Morb Mortal Wkly Rep. 2020 May
29;69(21):651-655. doi: 10.15585/mmwr.mm6921e1. PMID: 32463809; PMCID:
PMC7269604.

5. Ladhani SN, Chow JY, Janarthanan R, Fok J, Crawley-Boevey E, Vusirikala A, et al.
Increased risk of SARS-CoV-2 infection in staff working across different care homes:
enhanced CoVID-19 outbreak investigations in London care Homes. J Infect. 2020
Oct;81(4):621-624. doi: 10.1016/j.jinf.2020.07.027. Epub 2020 Jul 29. PMID: 32735893;
PMCID: PMC7387283.

6. Wallace M, Hagan L, Curran KG, Williams SP, Handanagic S, Bjork A, et al. COVID-19 in Correctional and Detention Facilities - United States, February-April 2020. MMWR Morb Mortal Wkly Rep. 2020 May 15;69(19):587-590. doi: 10.15585/mmwr.mm6919e1. PMID: 32407300.
7. Montoya-Barthelemy AG, Lee CD, Cundiff DR, Smith EB. COVID-19 and the Correctional Environment: The American Prison as a Focal Point for Public Health. Am J Prev Med. 2020 Jun;58(6):888-891. doi: 10.1016/j.amepre.2020.04.001. Epub 2020 Apr 17. PMID: 32387174; PMCID: PMC7164863.
8. Watson C, Warmbrod L, Vahey R, Cicero A, Inglesby T, Beyrer C, et al. COVID-19 and the US Criminal Justice System: Evidence for Public Health Measures to Reduce Risk. Baltimore, MD: Johns Hopkins Center for Health Security; 2020.
9. Waltenburg MA, Victoroff T, Rose CE, Butterfield M, Jervis RH, Fedak KM, et al. Update: COVID-19 Among Workers in Meat and Poultry Processing Facilities - United States, April-May 2020. MMWR Morb Mortal Wkly Rep. 2020 Jul 10;69(27):887-892. doi: 10.15585/mmwr.mm6927e2. PMID: 32644986.
10. Steinberg J, Kennedy ED, Basler C, Grant MP, Jacobs JR, Ortbahn D, et al. COVID-19 Outbreak Among Employees at a Meat Processing Facility - South Dakota, March-April 2020. MMWR Morb Mortal Wkly Rep. 2020 Aug 7;69(31):1015-1019. doi: 10.15585/mmwr.mm6931a2. PMID: 32759914; PMCID: PMC7454899.
11. Waltenburg MA, Rose CE, Victoroff T, Butterfield M, Dillaha JA, Heinzerling A, et al. Coronavirus Disease among Workers in Food Processing, Food Manufacturing, and Agriculture Workplaces. Emerg Infect Dis. 2020 Oct 19;27(1). doi: 10.3201/eid2701.203821. Epub ahead of print. PMID: 33075274.

12. Thielen BK, Bye E, Wang X, Maroushek S, Friedlander H, Bistodeau S, et al. Summer Outbreak of Severe RSV-B Disease, Minnesota, 2017 Associated with Emergence of a Genetically Distinct Viral Lineage. *J Infect Dis.* 2020 Jun 29;222(2):288-297. doi: 10.1093/infdis/jiaa075. PMID: 32083677; PMCID: PMC7323494.
13. Longmire AG, Sims S, Rytsareva I, Campo DS, Skums P, Dimitrova Z, et al. GHOST: global hepatitis outbreak and surveillance technology. *BMC Genomics.* 2017 Dec 6;18(Suppl 10):916. doi: 10.1186/s12864-017-4268-3. PMID: 29244005; PMCID: PMC5731493.
14. Quick J, Loman NJ, Duraffour S, Simpson JT, Severi E, Cowley L, et al. Real-time, portable genome sequencing for Ebola surveillance. *Nature.* 2016 Feb 11;530(7589):228-232. doi: 10.1038/nature16996. Epub 2016 Feb 3. PMID: 26840485; PMCID: PMC4817224.
15. Rounds JM, Taylor AJ, Eikmeier D, Nichols MM, Lappi V, Wirth SE, et al. Prospective *Salmonella* Enteritidis surveillance and outbreak detection using whole genome sequencing, Minnesota 2015-2017. *Epidemiol Infect.* 2020 Jun 16;148:e254. doi: 10.1017/S0950268820001272. PMID: 32539900.
16. To KK, Hung IF, Ip JD, Chu AW, Chan WM, Tam AR, et al. COVID-19 re-infection by a phylogenetically distinct SARS-coronavirus-2 strain confirmed by whole genome sequencing. *Clin Infect Dis.* 2020 Aug 25:ciaa1275. doi: 10.1093/cid/ciaa1275. Epub ahead of print. PMID: 32840608; PMCID: PMC7499500.
17. Larson D, Brodniak SL, Voegtly LJ, Cer RZ, Glang LA, Malagon FJ, et al. A Case of Early Re-infection with SARS-CoV-2. *Clin Infect Dis.* 2020 Sep 19:ciaa1436. doi: 10.1093/cid/ciaa1436. Epub ahead of print. PMID: 32949240; PMCID: PMC7543357.

18. Van Elslande J, Vermeersch P, Vandervoort K, Wawina-Bokalanga T, Vanmechelen B, Wollants E, et al. Symptomatic SARS-CoV-2 reinfection by a phylogenetically distinct strain. *Clin Infect Dis*. 2020 Sep 5:ciaa1330. doi: 10.1093/cid/ciaa1330. Epub ahead of print. PMID: 32887979; PMCID: PMC7499557.
19. Gupta V, Bhojar RC, Jain A, Srivastava S, Upadhyay R, Imran M, et al. Asymptomatic reinfection in two healthcare workers from India with genetically distinct SARS-CoV-2. *Clin Infect Dis*. 2020 Sep 23:ciaa1451. doi: 10.1093/cid/ciaa1451. Epub ahead of print. PMID: 32964927; PMCID: PMC7543380.
20. Lucey M, Macori G, Mullane N, Sutton-Fitzpatrick U, Gonzalez G, Coughlan S, et al. Whole-genome sequencing to track SARS-CoV-2 transmission in nosocomial outbreaks. *Clin Infect Dis*. 2020 Sep 19:ciaa1433. doi: 10.1093/cid/ciaa1433. Epub ahead of print. PMID: 32954414; PMCID: PMC7543366.
21. Moreno GK, Braun KM, Riemersma KK, Martin MA, Halfmann PJ, Crooks CM, et al. Distinct patterns of SARS-CoV-2 transmission in two nearby communities in Wisconsin, USA. *medRxiv [Preprint]*. 2020 Jul 10:2020.07.09.20149104. doi: 10.1101/2020.07.09.20149104. PMID: 32676620; PMCID: PMC7359545.
22. Oude Munnink BB, Nieuwenhuijse DF, Stein M, O'Toole Á, Haverkate M, Mollers M, et al. Rapid SARS-CoV-2 whole-genome sequencing and analysis for informed public health decision-making in the Netherlands. *Nat Med*. 2020 Sep;26(9):1405-1410. doi: 10.1038/s41591-020-0997-y. Epub 2020 Jul 16. Erratum in: *Nat Med*. 2020 Oct 20;: PMID: 32678356.
23. Seemann T, Lane CR, Sherry NL, Duchene S, Gonçalves da Silva A, Caly L, et al. Tracking the COVID-19 pandemic in Australia using genomics. *Nat Commun*. 2020 Sep

1;11(1):4376. doi: 10.1038/s41467-020-18314-x. PMID: 32873808; PMCID:

PMC7462846.

24. Rockett RJ, Arnott A, Lam C, Sadsad R, Timms V, Gray KA, et al. Revealing COVID-19 transmission in Australia by SARS-CoV-2 genome sequencing and agent-based modeling. *Nat Med*. 2020 Sep;26(9):1398-1404. doi: 10.1038/s41591-020-1000-7. Epub 2020 Jul 9. PMID: 32647358.

25. Centers for Disease Control and Prevention, Division of Viral Diseases; Atlanta, GA: 2020. CDC 2019-Novel Coronavirus (2019-nCoV) Real-Time RT-PCR Diagnostic Panel. CDC-006-00019, Revision 3. <https://www.fda.gov/media/134922/download>

26. Quick J. Artic amplicon sequencing protocol for MinION for nCoV-2019. [dx.doi.org/10.17504/protocols.io.bbmuik6w](https://doi.org/10.17504/protocols.io.bbmuik6w).

27. Sevinsky J. SARS-CoV-2 Sequencing on Illumina MiSeq Using ARTIC Protocol: Part 2 - Illumina DNA Flex Protocol. [dx.doi.org/10.17504/protocols.io.bffyjppw](https://doi.org/10.17504/protocols.io.bffyjppw)

28. https://staph-b.github.io/staphb_toolkit/workflow_docs/monroe/

29. <https://www.geneious.com>

30. Hadfield J, Megill C, Bell SM, Huddleston J, Potter B, Callender C, et al. Nextstrain: real-time tracking of pathogen evolution. *Bioinformatics*. 2018 Dec 1;34(23):4121-4123. doi: 10.1093/bioinformatics/bty407. PMID: 29790939; PMCID: PMC6247931.

31. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol*. 2013 Apr;30(4):772-80. doi: 10.1093/molbev/mst010. Epub 2013 Jan 16. PMID: 23329690; PMCID: PMC3603318.

32. Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol*.

2015 Jan;32(1):268-74. doi: 10.1093/molbev/msu300. Epub 2014 Nov 3. PMID:
25371430; PMCID: PMC4271533.

33. Letunic I, Bork P. Interactive Tree Of Life (iTOL) v4: recent updates and new
developments. Nucleic Acids Res. 2019 Jul 2;47(W1):W256-W259. doi:
10.1093/nar/gkz239. PMID: 30931475; PMCID: PMC6602468.

34. <https://github.com/cov-lineages/pangolin>

35. Elbe S, Buckland-Merrett G. Data, disease and diplomacy: GISAID's innovative
contribution to global health. Glob Chall. 2017 Jan 10;1(1):33-46. doi:
10.1002/gch2.1018. PMID: 31565258; PMCID: PMC6607375.

36. Gonzalez-Reiche AS, Hernandez MM, Sullivan MJ, Ciferri B, Alshammary H, Obla A, et
al. Introductions and early spread of SARS-CoV-2 in the New York City area. Science.
2020 Jul 17;369(6501):297-301. doi: 10.1126/science.abc1917. Epub 2020 May 29.
PMID: 32471856; PMCID: PMC7259823.

37. Grabowski DC, Mor V. Nursing Home Care in Crisis in the Wake of COVID-19. JAMA.
2020 Jul 7;324(1):23-24. doi: 10.1001/jama.2020.8524. PMID: 32442303.

Address for correspondence: Nicholas B. Lehnertz, Minnesota Department of Health, 625
Robert Street North, St. Paul, MN 55164; email: nick.lehnertz@state.mn.us

Table 1.

483 Features of outbreaks and convenience samples of specimens collected and characterized by
484 whole genome sequencing at long-term care facilities (LTCFs), correctional facilities and meat
485 processing plants in Minnesota: March 6 – June 30, 2020⁶

Outbreak Facility	Outbreak Dates	Total Confirmed Outbreak Cases at Facility	Total Samples Successfully Sequenced from Facility N (%)	Role at Facility	Total Confirmed Outbreak Cases at Facility by Role	Total Samples Successfully Sequenced by Role N (%)
LTCF A	April 15 – June 11	89	27 (30.3%)	Staff	38	10 (26.3%)
				Residents	51	17 (33.3%)
LTCF B	April 29 – June 11	190	82 (43.2%)	Staff	76	5 (6.6%)
				Residents	114	77 (67.5%)
LTCF C	April 24 – June 30	139	32 (23.0%)	Staff	56	23 (41.0%)
				Residents	83	9 (10.8%)
LTCF D	April 17 – May 15	74	39 (52.7%)	Staff	21	3 (14.2%)
				Residents	53	36 (67.9%)
Correctional Facility A	March 25 – July 20	128	49 (38.3%)	Staff	82	15 (18.3%)
				Inmates	46	34 (73.9%)
Correctional Facility B	May 13 – August 5	416	61 (14.7%)	Staff	210	1 (0.5%)
				Inmates	206	60 (29.1%)
Processing Plant A	March 15 – Ongoing	432	16 (3.7%)	Employees	432	16 (3.7%)

Processing Plant B	April 11 – Ongoing	724	5 (0.7%)	Employees	724	5 (0.7%)
Community Samples A	May 15 – June 1	17	8 (47.1%)	Known Contact	9	2 (22.2%)
				No Known Contact	8	6 (75.0%)
TOTALS		2,222	319 (14.4%)			

Figure 1. Phylogenetic tree of SARS-CoV-2 viruses associated with selected outbreaks in Minnesota, March 6 – June 30, 2020. IQ-TREE was used with the GTR substitution model for tree generation. Branch lengths were scaled to represent number of single nucleotide mutations as shown in the scale key.

Figure 2A-D. Phylogenetic tree of SARS-CoV-2 genome sequences associated with LTCF A from April 15 to June 11, 2020 (Figure 2A), LTCF B from April 29 to June 11, 2020 (Figure 2B), LTCF C from April 24 to June 30, 2020 (Figure 2C), and LTCF D from April 17 to May 15, 2020 (Figure 2D). Filled circles represent sequences taken from residents, open circles represent sequences from healthcare workers. IQ-TREE was used with the general time reversible substitution model for tree generation. Branch lengths were scaled to represent number of single nucleotide mutations as shown in the scale key.

Figure 2A

Figure 2C

Figure 2D

Figure 3A-B. Phylogenetic tree of SARS-CoV-2 genome sequences associated with Correctional Facility A from March 25 to June 30, 2020 (Figure 3A), and Correctional Facility B from May 13 to June 30, 2020 (Figure 3B). Filled circles represent sequences taken from inmates, open circles represent sequences from facility staff. IQ-TREE was used with the general time reversible substitution model for tree generation. Branch lengths were scaled to represent number of single nucleotide mutations as shown in the scale key.

Figure 3A

Figure 4. Phylogenetic tree of SARS-CoV-2 genome sequences associated with both LTCF C and Correctional Facility B. Filled circles represent sequences taken from inmates or residents, open circles represent sequences from facility staff or healthcare workers. Sequences from LTCF C are shown on a white background, and sequences from Correctional Facility B on a grey background. Sequences from two household contacts are noted with stars. IQ-TREE was used with the general time reversible substitution model for tree generation. Branch lengths were scaled to represent number of single nucleotide mutations as shown in the scale key.

Figure 4

Tree scale: 1 I

Figure 5A-B. Phylogenetic tree of SARS-CoV 2 genome sequences associated with Processing Plant A and the surrounding community from March 15 to June 30, 2020 (Figure 5A), and Processing Plant B from April 11 to June 30, 2020 (Figure 5B). Open circles represent sequences from staff at Processing Plant A and squares represent sequences from the surrounding community. IQ-TREE was used with the general time reversible substitution model for tree generation. Branch lengths were scaled to represent number of single nucleotide mutations as shown in the scale key.

Figure 5A

Figure 5B

Tree scale: 1