

1 **Title:**

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

4 **Authors:**

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

9 **Affiliations:**

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

15 **Abstract:**

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

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

28 **Text**

29 ***Introduction:***

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

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

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

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

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

63 ***Methods:***

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

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

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

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

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

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

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

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

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

116 ***Results:***

117 WGS and subsequent phylogenetic analysis was successfully conducted on 319 samples
118 associated with the eight outbreaks, constituting 14.4% of the 2,222 total positive cases identified
119 from outbreaks in Minnesota through June 2020. Specimens were obtained from staff and
120 residents from four different LTCFs (180 specimens sequenced out of 505 identified cases,
121 35.6%), staff and inmates from 2 different correctional facilities (110 specimens sequenced out
122 of 544 identified cases, 20.2%) and among employees at 2 different meat processing plants,
123 along with identified community cases (29 samples sequenced out of 1,173 identified cases,
124 2.5%) (Table 1). Among most of the specimens that were sequenced, three LTCFs and both
125 correctional facilities had spread associated with a single genetic sequence unique to each
126 outbreak facility. A fourth LTCF had outbreak cases associated with two distinct sequences. In
127 contrast, cases associated with outbreaks in the two meat processing plants represented multiple
128 SARS-CoV-2 sequences. (Figure 1)

129 *Single Cluster in Long-Term Care Facilities:*

130 During the COVID-19 outbreak at LTCF A⁷ between April 15 and June 11 (Figure 2A),
131 51 of 77 residents, and 38 of 108 health care personnel (HCP) tested were confirmed to be
132 infected, following identification of positive HCP. Specimens from 17 residents (33.3% of cases)
133 and 10 staff (26.3% of cases) were available for WGS. SARS-CoV-2 viral sequences from these
134 27 persons were genetically closely related (Pangolin lineage B.1.2). Viral genomes from two

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

138 In LTCF B from April 29 - June 11 (Figure 2B), 114 of 182 tested residents and 76 of
139 233 tested HCP were confirmed to be SARS-CoV-2 positive, following identification of a
140 SARS-CoV-2 positive resident on April 29. All 82 sequenced specimens from this facility,
141 including 77 residents (67.5% of cases) and 5 HCP (6.6% of cases), were closely related
142 (Pangolin lineage B.1.116).

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

151 *Single Cluster in Correctional Facilities:*

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

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

159 In early June, 2020, an outbreak was identified in Correctional Facility B (Figure 3B).
160 The investigation revealed that an employee had developed symptoms consistent with COVID-
161 19 on May 13, tested positive for SARS-CoV-2 on May 14, and was subsequently excluded from
162 work and isolated at home. Approximately two weeks later, three additional cases, one staff
163 member and two inmates from the same unit as the index patient, tested positive for SARS-CoV-
164 2. A point-prevalence survey (PPS) on June 1 revealed 63 COVID-19 positive inmates among
165 the 87 tested. Subsequent facility-wide testing of both staff and inmates identified cases in other
166 units, 83 new cases in inmates and one new staff case identified among the approximately 2,200
167 persons tested. Phylogenetic analysis of this outbreak among the 1 staff member (0.5% of cases)
168 and 60 inmates (29.1% of cases) at Correctional Facility B shows that all cases were closely
169 related (Pangolin lineage B.1.2) and genetically identical to, or plausibly descended from, the
170 sequence of SARS-CoV-2 from the index case.

171 *Linking LTCF C with Correctional Facility B:*

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

179 *Two Distinct Clusters in a Long-Term Care Facility:*

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

189 *Multiple Clusters in Meat Processing Plants:*

190 In early April 2020, an outbreak was detected at Processing Plant A (Figure 5A), a large
191 primary and secondary meat processor. The outbreak at Processing Plant A continued for several
192 weeks until mid-May when the number of cases among workers began to increase rapidly. From
193 March 15 to July 1, there were 446 confirmed cases who reported working at Processing Plant A,
194 including 4 (1%) cases who tested positive in March (management and office staff), 5 (1%) in
195 April, 211 (47%) in May, and 226 (51%) in June. Analysis demonstrated that of the 16 samples
196 (3.7% of cases) sequenced between March 15 and June 3, there were at least 6 unrelated clusters
197 or singleton cases. While the majority of genomes sequenced from Processing Plant B belong to
198 Pangolin lineages B.1, B1.2, B.1.26, one early case is genetically quite different (Pangolin
199 lineage A.1). An interview confirmed that this early case had traveled out of the state during their
200 exposure period (14 days before symptom onset).

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

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

213 ***Discussion:***

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

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

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

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

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

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

253 WGS has shown to improve understanding of an outbreak upon retrospective analysis (3,
254 20-22), to provide justification for specific public health measures (22-23), and to add insight to
255 transmission patterns in high-risk settings. Our work provides further support for the use of WGS
256 in these situations, while identifying several additional public health implications. WGS has
257 demonstrated that outbreaks in LTCFs and correctional facilities can be the result of a single
258 introduction. Continued vigilance with facility-wide staff screening, and subsequent exclusion of
259 symptomatic HCPs or staff and those with known or suspected contacts is imperative. WGS has
260 demonstrated extensive intra-facility spread, with closely related sequences comprising all or the
261 vast majority of cases contributing to the outbreak. Measures, such as IPC, consistent and correct
262 use of PPE, cohorting of known positive residents and exclusion of positive HCPs must be
263 maintained. WGS has also illuminated the transmission patterns in meat processing plants,
264 including the multiple introductions identified through the multiple closely related sequences
265 identified, and the related community strains. WGS has illustrated the need for community-level
266 mitigation to prevent introductions in high-density worksites, including accessible community-
267 wide testing, housing and transportation strategies, in addition to facility-level measures to
268 prevent unintended introduction into the workplace.

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

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

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

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

291 to close habitation and/or work environment, and delayed case detection and access to care (8,
292 11, 37). WGS has demonstrated that many outbreaks in Minnesota were caused by single
293 introductions of SARS-CoV-2, highlighting the importance of consistent and correct usage of
294 PPE, rigorous and systematic IPC, environmental control measures, and systematic testing of
295 residents and staff to identify asymptomatic cases. As this pandemic continues, community
296 mitigation strategies and strong enforcement of policies to reduce the risk of introducing SARS-
297 CoV-2 virus into congregate settings are more important than ever. Similarly, IPC and
298 aggressive containment practices are vital to mitigate the spread of SARS-CoV-2 upon
299 introduction into a facility. WGS can be a useful tool to supplement epidemiological information
300 and examine the role of facility and community factors contributing to SARS-COV-2 outbreaks
301 in high risk settings.

302 **Acknowledgments:**

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

308 **Disclaimers:**

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

311 **Author Bio:**

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

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

319 **Footnotes:**

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

321 ² An epidemiologic link among inmates is defined as residing on the same unit or ward within a
322 14-day time period of each other.

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

329 ⁴ Facility-acquired COVID-19 infection in an inmate/resident is defined as a confirmed diagnosis
330 14 days or more after entry to the facility, without an exposure during the previous 14 days to
331 another setting where an outbreak was known or suspected to be occurring.

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

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

341 ⁷ A full accounting of the outbreaks at LTCF A and LTCF B were previously published. *See*
342 Taylor J, Carter RJ, Lehnertz N, et al. Serial Testing for SARS-CoV-2 and Virus Whole Genome
343 Sequencing Inform Infection Risk at Two Skilled Nursing Facilities with COVID-19 Outbreaks
344 — Minnesota, April–June 2020. *MMWR Morb Mortal Wkly Rep* 2020;69:1288–1295. DOI:
345 <http://dx.doi.org/10.15585/mmwr.mm6937a3>

346 **References**

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

- 349 N Engl J Med. 2020 May 28;382(22):2081-2090. doi: 10.1056/NEJMoa2008457. Epub
350 2020 Apr 24. PMID: 32329971; PMCID: PMC7200056.
- 351 2. Kimball A, Hatfield KM, Arons M, James A, Taylor J, Spicer K, et al. Asymptomatic and
352 Presymptomatic SARS-CoV-2 Infections in Residents of a Long-Term Care Skilled
353 Nursing Facility - King County, Washington, March 2020. MMWR Morb Mortal Wkly
354 Rep. 2020 Apr 3;69(13):377-381. doi: 10.15585/mmwr.mm6913e1. PMID: 32240128;
355 PMCID: PMC7119514.
- 356 3. Taylor J, Carter RJ, Lehnertz N, Kazazian L, Sullivan M, Wang X, et al. Serial Testing
357 for SARS-CoV-2 and Virus Whole Genome Sequencing Inform Infection Risk at Two
358 Skilled Nursing Facilities with COVID-19 Outbreaks - Minnesota, April-June 2020.
359 MMWR Morb Mortal Wkly Rep. 2020 Sep 18;69(37):1288-1295. doi:
360 10.15585/mmwr.mm6937a3. PMID: 32966272; PMCID: PMC7498172.
- 361 4. Dora AV, Winnett A, Jatt LP, Davar K, Watanabe M, Sohn L, et al. Universal and Serial
362 Laboratory Testing for SARS-CoV-2 at a Long-Term Care Skilled Nursing Facility for
363 Veterans - Los Angeles, California, 2020. MMWR Morb Mortal Wkly Rep. 2020 May
364 29;69(21):651-655. doi: 10.15585/mmwr.mm6921e1. PMID: 32463809; PMCID:
365 PMC7269604.
- 366 5. Ladhani SN, Chow JY, Janarthanan R, Fok J, Crawley-Boevey E, Vusirikala A, et al.
367 Increased risk of SARS-CoV-2 infection in staff working across different care homes:
368 enhanced CoVID-19 outbreak investigations in London care Homes. J Infect. 2020
369 Oct;81(4):621-624. doi: 10.1016/j.jinf.2020.07.027. Epub 2020 Jul 29. PMID: 32735893;
370 PMCID: PMC7387283.

- 371 6. Wallace M, Hagan L, Curran KG, Williams SP, Handanagic S, Bjork A, et al. COVID-19
372 in Correctional and Detention Facilities - United States, February-April 2020. MMWR
373 Morb Mortal Wkly Rep. 2020 May 15;69(19):587-590. doi: 10.15585/mmwr.mm6919e1.
374 PMID: 32407300.
- 375 7. Montoya-Barthelemy AG, Lee CD, Cundiff DR, Smith EB. COVID-19 and the
376 Correctional Environment: The American Prison as a Focal Point for Public Health. Am J
377 Prev Med. 2020 Jun;58(6):888-891. doi: 10.1016/j.amepre.2020.04.001. Epub 2020 Apr
378 17. PMID: 32387174; PMCID: PMC7164863.
- 379 8. Watson C, Warmbrod L, Vahey R, Cicero A, Inglesby T, Beyrer C, et al. COVID-19 and
380 the US Criminal Justice System: Evidence for Public Health Measures to Reduce Risk.
381 Baltimore, MD: Johns Hopkins Center for Health Security; 2020.
- 382 9. Waltenburg MA, Victoroff T, Rose CE, Butterfield M, Jervis RH, Fedak KM, et al.
383 Update: COVID-19 Among Workers in Meat and Poultry Processing Facilities - United
384 States, April-May 2020. MMWR Morb Mortal Wkly Rep. 2020 Jul 10;69(27):887-892.
385 doi: 10.15585/mmwr.mm6927e2. PMID: 32644986.
- 386 10. Steinberg J, Kennedy ED, Basler C, Grant MP, Jacobs JR, Ortbahn D, et al. COVID-19
387 Outbreak Among Employees at a Meat Processing Facility - South Dakota, March-April
388 2020. MMWR Morb Mortal Wkly Rep. 2020 Aug 7;69(31):1015-1019. doi:
389 10.15585/mmwr.mm6931a2. PMID: 32759914; PMCID: PMC7454899.
- 390 11. Waltenburg MA, Rose CE, Victoroff T, Butterfield M, Dillaha JA, Heinzerling A, et al.
391 Coronavirus Disease among Workers in Food Processing, Food Manufacturing, and
392 Agriculture Workplaces. Emerg Infect Dis. 2020 Oct 19;27(1). doi:
393 10.3201/eid2701.203821. Epub ahead of print. PMID: 33075274.

- 394 12. Thielen BK, Bye E, Wang X, Maroushek S, Friedlander H, Bistodeau S, et al. Summer
395 Outbreak of Severe RSV-B Disease, Minnesota, 2017 Associated with Emergence of a
396 Genetically Distinct Viral Lineage. *J Infect Dis.* 2020 Jun 29;222(2):288-297. doi:
397 10.1093/infdis/jiaa075. PMID: 32083677; PMCID: PMC7323494.
- 398 13. Longmire AG, Sims S, Rytsareva I, Campo DS, Skums P, Dimitrova Z, et al. GHOST:
399 global hepatitis outbreak and surveillance technology. *BMC Genomics.* 2017 Dec
400 6;18(Suppl 10):916. doi: 10.1186/s12864-017-4268-3. PMID: 29244005; PMCID:
401 PMC5731493.
- 402 14. Quick J, Loman NJ, Duraffour S, Simpson JT, Severi E, Cowley L, et al. Real-time,
403 portable genome sequencing for Ebola surveillance. *Nature.* 2016 Feb 11;530(7589):228-
404 232. doi: 10.1038/nature16996. Epub 2016 Feb 3. PMID: 26840485; PMCID:
405 PMC4817224.
- 406 15. Rounds JM, Taylor AJ, Eikmeier D, Nichols MM, Lappi V, Wirth SE, et al.
407 Prospective *Salmonella* Enteritidis surveillance and outbreak detection using whole
408 genome sequencing, Minnesota 2015-2017. *Epidemiol Infect.* 2020 Jun 16;148:e254. doi:
409 10.1017/S0950268820001272. PMID: 32539900.
- 410 16. To KK, Hung IF, Ip JD, Chu AW, Chan WM, Tam AR, et al. COVID-19 re-infection by
411 a phylogenetically distinct SARS-coronavirus-2 strain confirmed by whole genome
412 sequencing. *Clin Infect Dis.* 2020 Aug 25:ciaa1275. doi: 10.1093/cid/ciaa1275. Epub
413 ahead of print. PMID: 32840608; PMCID: PMC7499500.
- 414 17. Larson D, Brodniak SL, Voegtly LJ, Cer RZ, Glang LA, Malagon FJ, et al. A Case of
415 Early Re-infection with SARS-CoV-2. *Clin Infect Dis.* 2020 Sep 19:ciaa1436. doi:
416 10.1093/cid/ciaa1436. Epub ahead of print. PMID: 32949240; PMCID: PMC7543357.

- 417 18. Van Elslande J, Vermeersch P, Vandervoort K, Wawina-Bokalanga T, Vanmechelen B,
418 Wollants E, et al. Symptomatic SARS-CoV-2 reinfection by a phylogenetically distinct
419 strain. *Clin Infect Dis*. 2020 Sep 5:ciaa1330. doi: 10.1093/cid/ciaa1330. Epub ahead of
420 print. PMID: 32887979; PMCID: PMC7499557.
- 421 19. Gupta V, Bhojar RC, Jain A, Srivastava S, Upadhayay R, Imran M, et al. Asymptomatic
422 reinfection in two healthcare workers from India with genetically distinct SARS-CoV-2.
423 *Clin Infect Dis*. 2020 Sep 23:ciaa1451. doi: 10.1093/cid/ciaa1451. Epub ahead of print.
424 PMID: 32964927; PMCID: PMC7543380.
- 425 20. Lucey M, Macori G, Mullane N, Sutton-Fitzpatrick U, Gonzalez G, Coughlan S, et al.
426 Whole-genome sequencing to track SARS-CoV-2 transmission in nosocomial outbreaks.
427 *Clin Infect Dis*. 2020 Sep 19:ciaa1433. doi: 10.1093/cid/ciaa1433. Epub ahead of print.
428 PMID: 32954414; PMCID: PMC7543366.
- 429 21. Moreno GK, Braun KM, Riemersma KK, Martin MA, Halfmann PJ, Crooks CM, et al.
430 Distinct patterns of SARS-CoV-2 transmission in two nearby communities in Wisconsin,
431 USA. *medRxiv [Preprint]*. 2020 Jul 10:2020.07.09.20149104. doi:
432 10.1101/2020.07.09.20149104. PMID: 32676620; PMCID: PMC7359545.
- 433 22. Oude Munnink BB, Nieuwenhuijse DF, Stein M, O'Toole Á, Haverkate M, Mollers M, et
434 al. Rapid SARS-CoV-2 whole-genome sequencing and analysis for informed public
435 health decision-making in the Netherlands. *Nat Med*. 2020 Sep;26(9):1405-1410. doi:
436 10.1038/s41591-020-0997-y. Epub 2020 Jul 16. Erratum in: *Nat Med*. 2020 Oct 20;:
437 PMID: 32678356.
- 438 23. Seemann T, Lane CR, Sherry NL, Duchene S, Gonçalves da Silva A, Caly L, et al.
439 Tracking the COVID-19 pandemic in Australia using genomics. *Nat Commun*. 2020 Sep

- 440 1;11(1):4376. doi: 10.1038/s41467-020-18314-x. PMID: 32873808; PMCID:
441 PMC7462846.
- 442 24. Rockett RJ, Arnott A, Lam C, Sadsad R, Timms V, Gray KA, et al. Revealing COVID-19
443 transmission in Australia by SARS-CoV-2 genome sequencing and agent-based
444 modeling. *Nat Med.* 2020 Sep;26(9):1398-1404. doi: 10.1038/s41591-020-1000-7. Epub
445 2020 Jul 9. PMID: 32647358.
- 446 25. Centers for Disease Control and Prevention, Division of Viral Diseases; Atlanta, GA:
447 2020. CDC 2019-Novel Coronavirus (2019-nCoV) Real-Time RT-PCR Diagnostic Panel.
448 CDC-006-00019, Revision 3.<https://www.fda.gov/media/134922/download>
- 449 26. Quick J. Artic amplicon sequencing protocol for MinION for nCoV-2019.
450 dx.doi.org/10.17504/protocols.io.bbmuik6w.
- 451 27. Sevinsky J. SARS-CoV-2 Sequencing on Illumina MiSeq Using ARTIC Protocol: Part 2
452 - Illumina DNA Flex Protocol. dx.doi.org/10.17504/protocols.io.bffyjjpw
- 453 28. https://staph-b.github.io/staphb_toolkit/workflow_docs/monroe/
- 454 29. <https://www.geneious.com>
- 455 30. Hadfield J, Megill C, Bell SM, Huddleston J, Potter B, Callender C, et al. Nextstrain:
456 real-time tracking of pathogen evolution. *Bioinformatics.* 2018 Dec 1;34(23):4121-4123.
457 doi: 10.1093/bioinformatics/bty407. PMID: 29790939; PMCID: PMC6247931.
- 458 31. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7:
459 improvements in performance and usability. *Mol Biol Evol.* 2013 Apr;30(4):772-80. doi:
460 10.1093/molbev/mst010. Epub 2013 Jan 16. PMID: 23329690; PMCID: PMC3603318.
- 461 32. Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ. IQ-TREE: a fast and effective
462 stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol.*

- 463 2015 Jan;32(1):268-74. doi: 10.1093/molbev/msu300. Epub 2014 Nov 3. PMID:
464 25371430; PMCID: PMC4271533.
- 465 33. Letunic I, Bork P. Interactive Tree Of Life (iTOL) v4: recent updates and new
466 developments. *Nucleic Acids Res.* 2019 Jul 2;47(W1):W256-W259. doi:
467 10.1093/nar/gkz239. PMID: 30931475; PMCID: PMC6602468.
- 468 34. <https://github.com/cov-lineages/pangolin>
- 469 35. Elbe S, Buckland-Merrett G. Data, disease and diplomacy: GISAID's innovative
470 contribution to global health. *Glob Chall.* 2017 Jan 10;1(1):33-46. doi:
471 10.1002/gch2.1018. PMID: 31565258; PMCID: PMC6607375.
- 472 36. Gonzalez-Reiche AS, Hernandez MM, Sullivan MJ, Ciferri B, Alshammary H, Obla A, et
473 al. Introductions and early spread of SARS-CoV-2 in the New York City area. *Science.*
474 2020 Jul 17;369(6501):297-301. doi: 10.1126/science.abc1917. Epub 2020 May 29.
475 PMID: 32471856; PMCID: PMC7259823.
- 476 37. Grabowski DC, Mor V. Nursing Home Care in Crisis in the Wake of COVID-19. *JAMA.*
477 2020 Jul 7;324(1):23-24. doi: 10.1001/jama.2020.8524. PMID: 32442303.

478

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

481

482 **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.

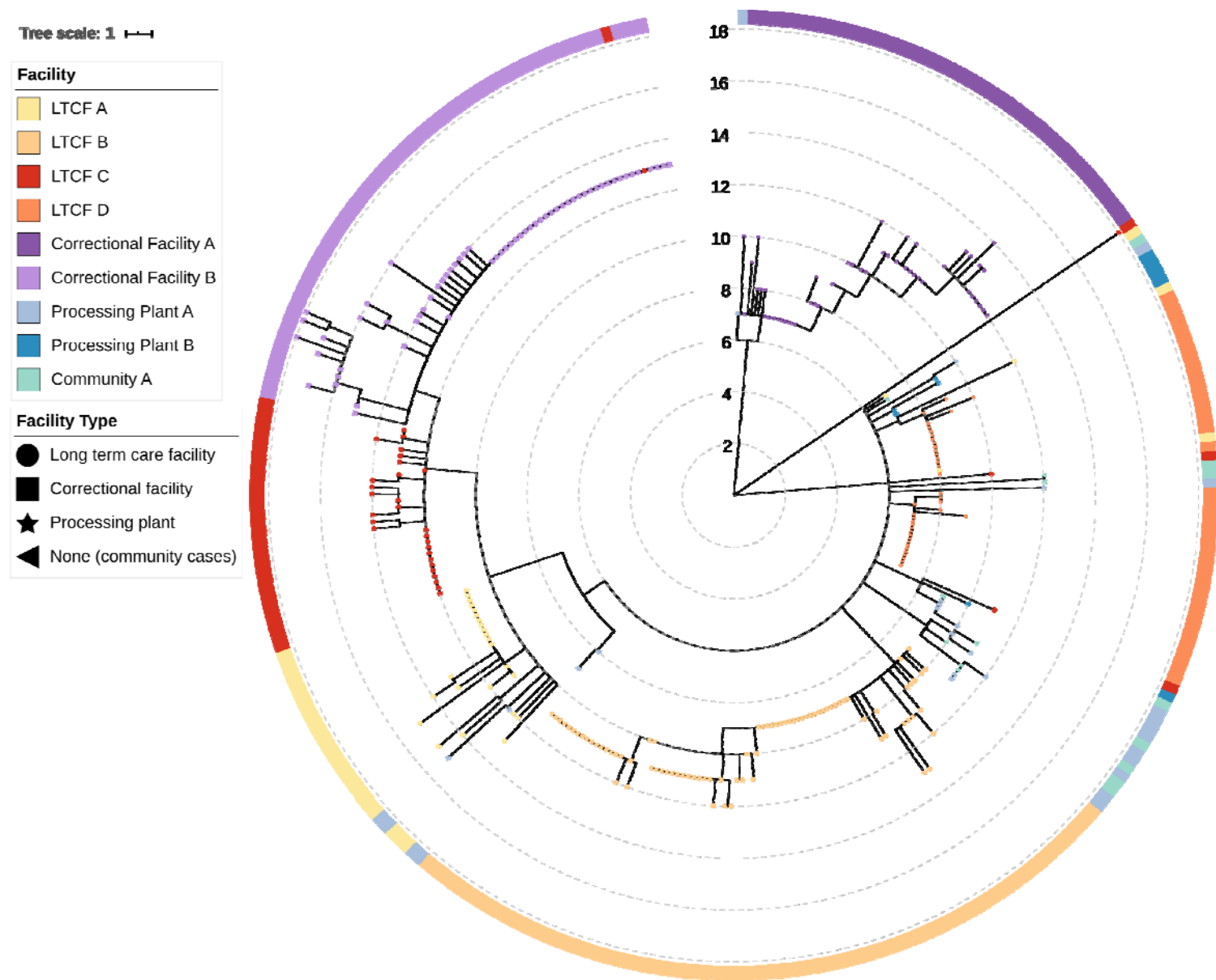


Figure 2A-D. Phylogenetic tree of SARS-CoV-2 genome sequences associated with LTCF A from April 15 to June 11, 2020 (Figure 2A), LTCF B from April 29 to June 11, 2020 (Figure 2B), LTCF C from April 24 to June 30, 2020 (Figure 2C), and LTCF D from April 17 to May 15, 2020 (Figure 2D). Filled circles represent 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

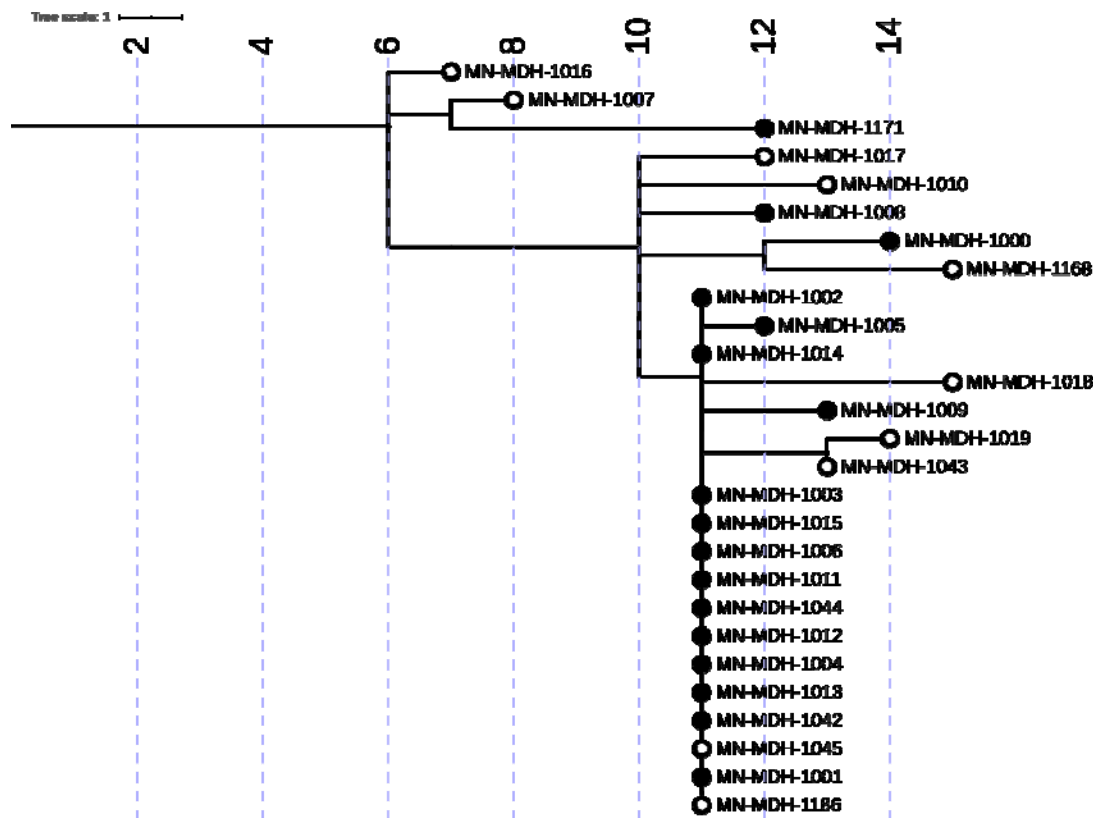


Figure 2B

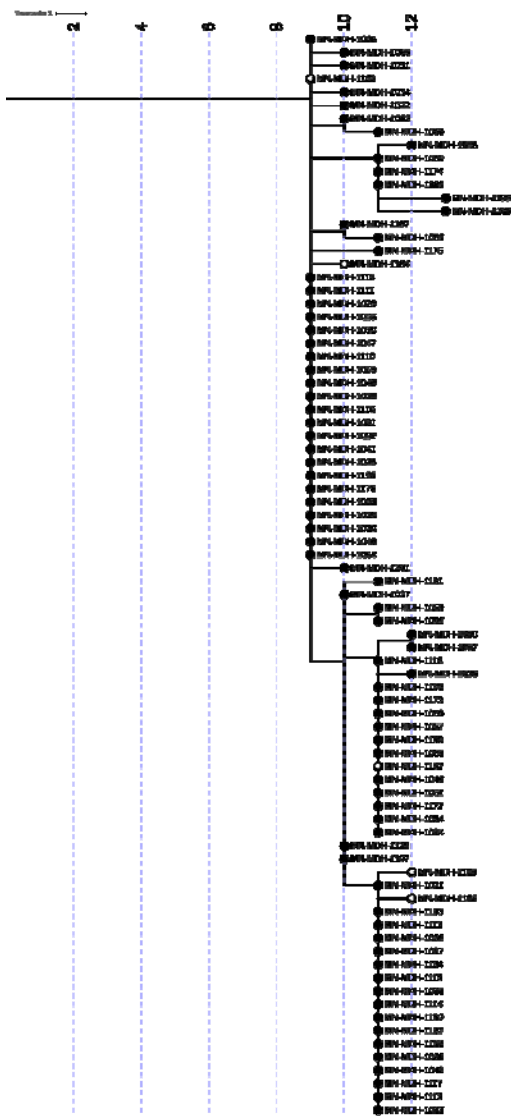


Figure 2C

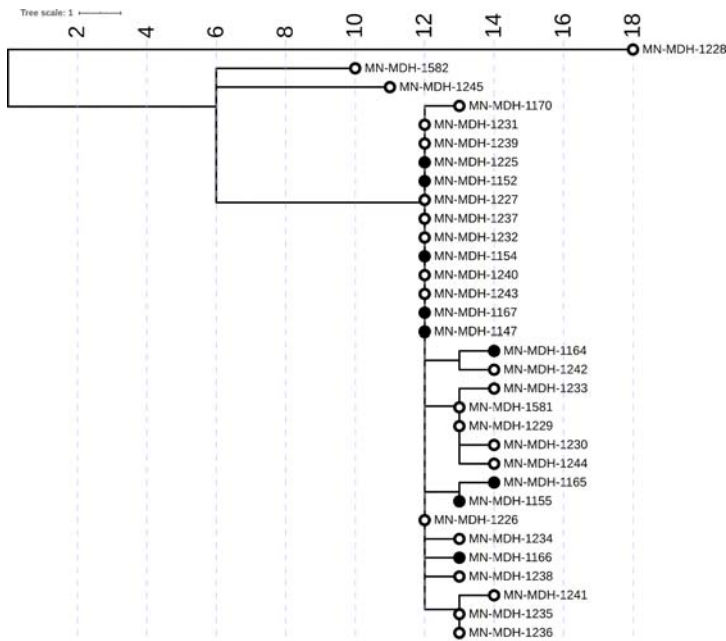


Figure 2D

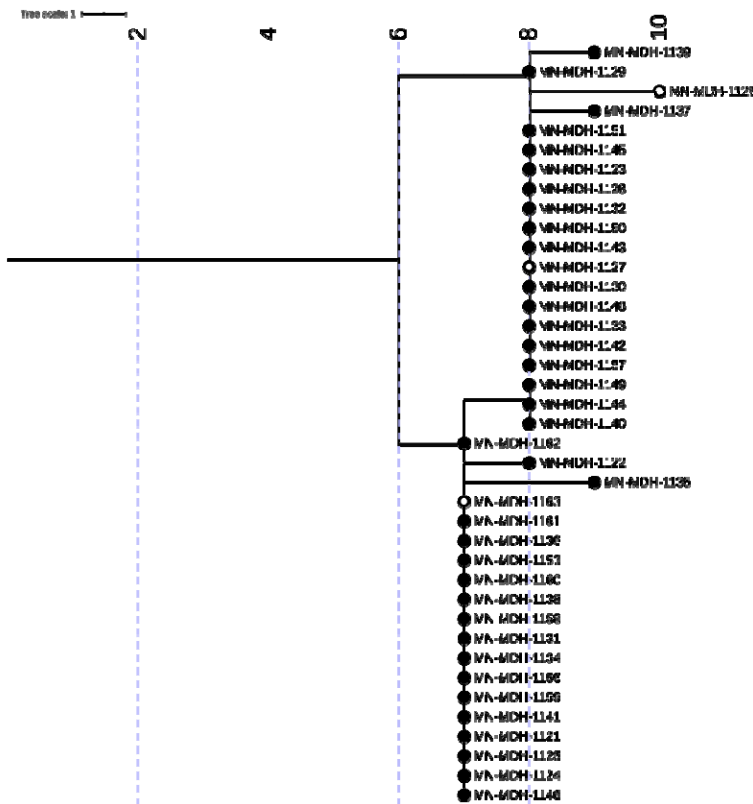


Figure 3A-B. Phylogenetic tree of SARS-CoV-2 genome sequences associated with Correctional Facility A from March 25 to June 30, 2020 (Figure 3A), and Correctional Facility B from May 13 to June 30, 2020 (Figure 3B). Filled circles represent sequences taken from inmates, open circles represent sequences 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

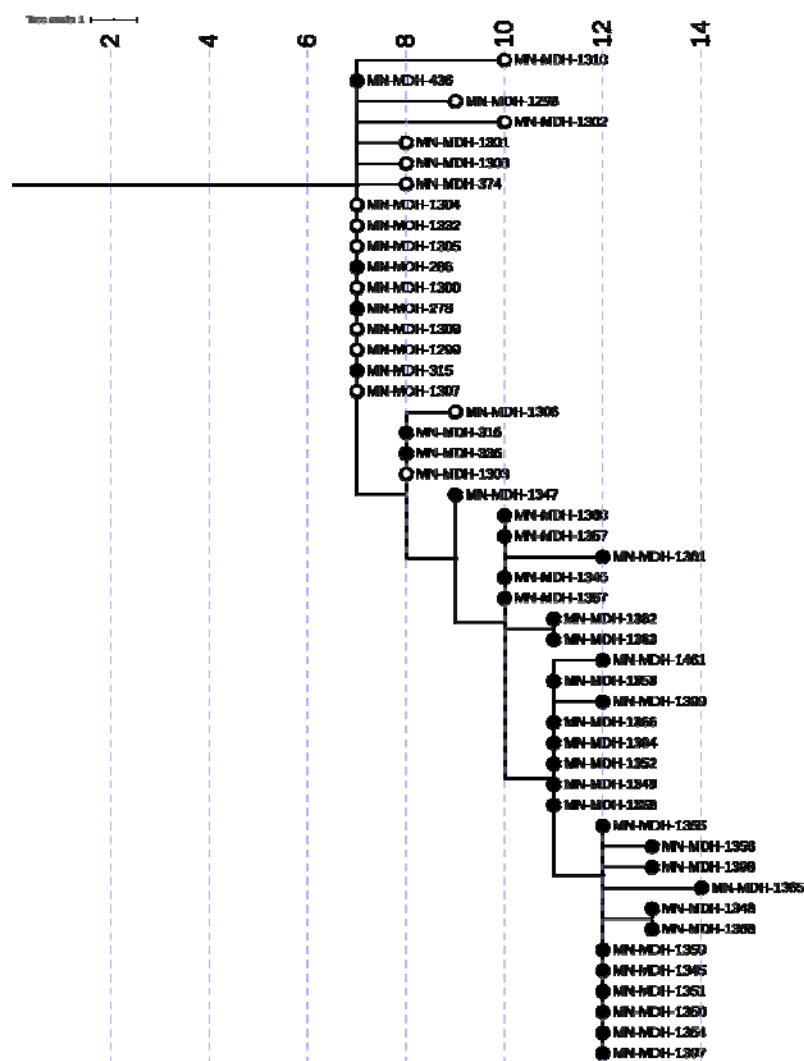


Figure 3B

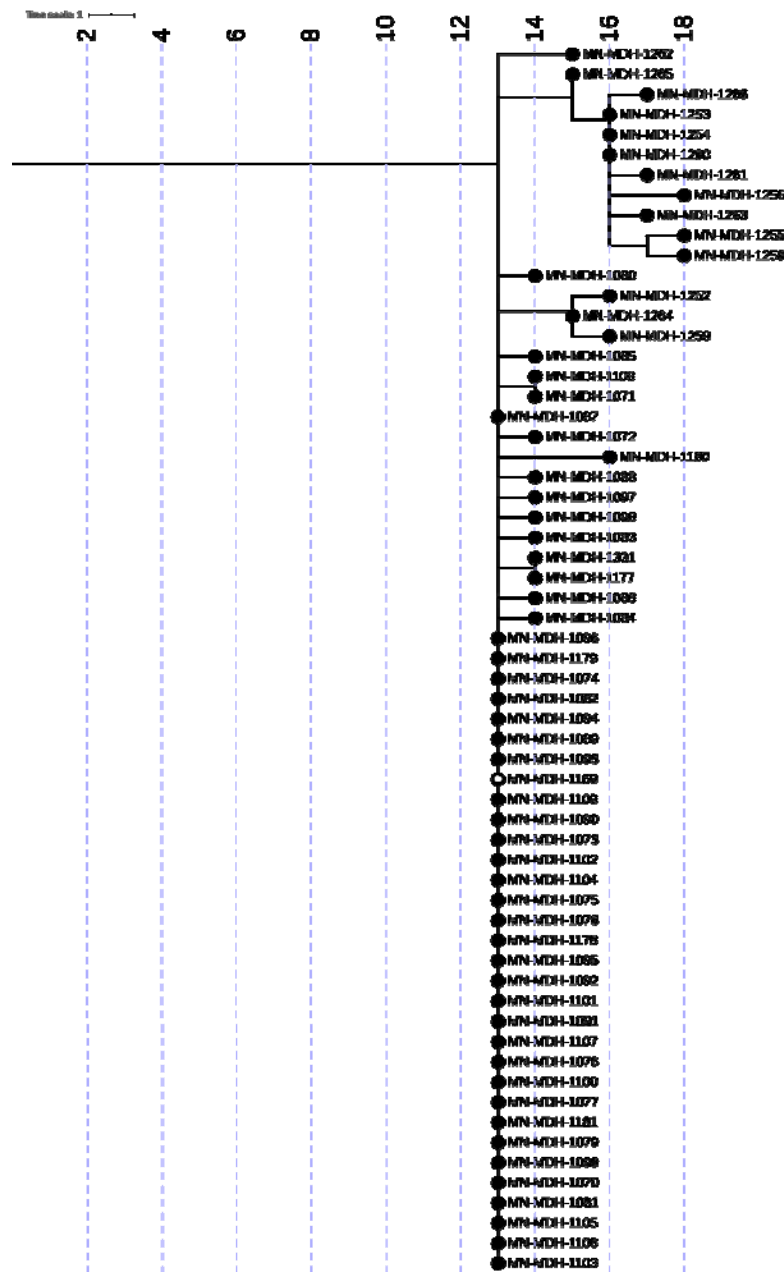


Figure 4. Phylogenetic tree of SARS-CoV-2 genome sequences associated with both LTCF C and Correctional Facility B. Filled circles represent sequences taken from inmates or residents, open circles represent sequences from facility staff or healthcare workers. Sequences from LTCF C are shown on a white 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

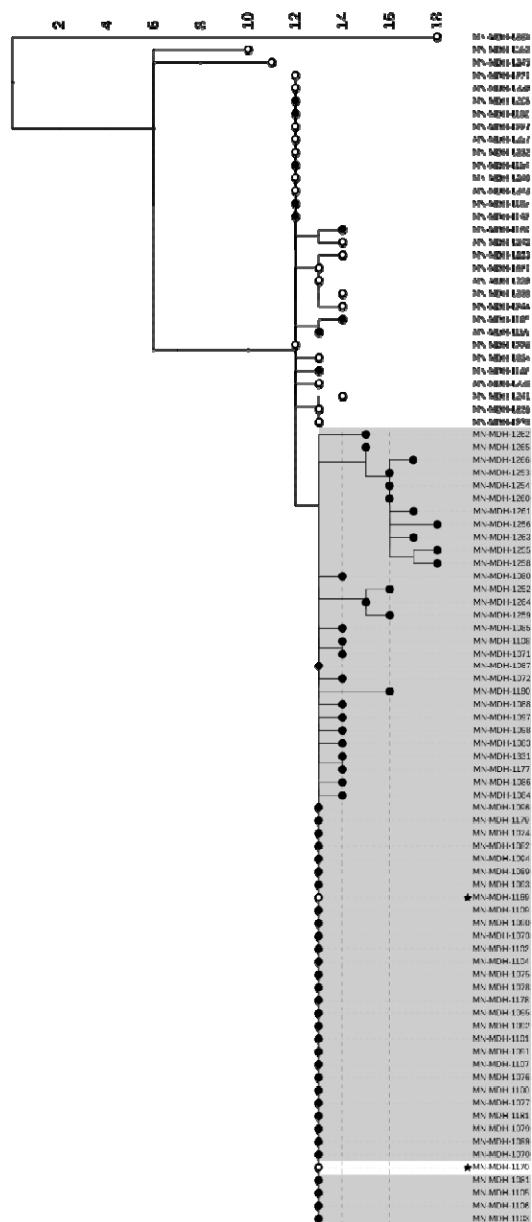


Figure 5A-B. Phylogenetic tree of SARS-CoV 2 genome sequences associated with Processing Plant A and the surrounding community from March 15 to June 30, 2020 (Figure 5A), and Processing Plant B from April 11 to June 30, 2020 (Figure 5B). Open circles represent sequences from staff at Processing Plant 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

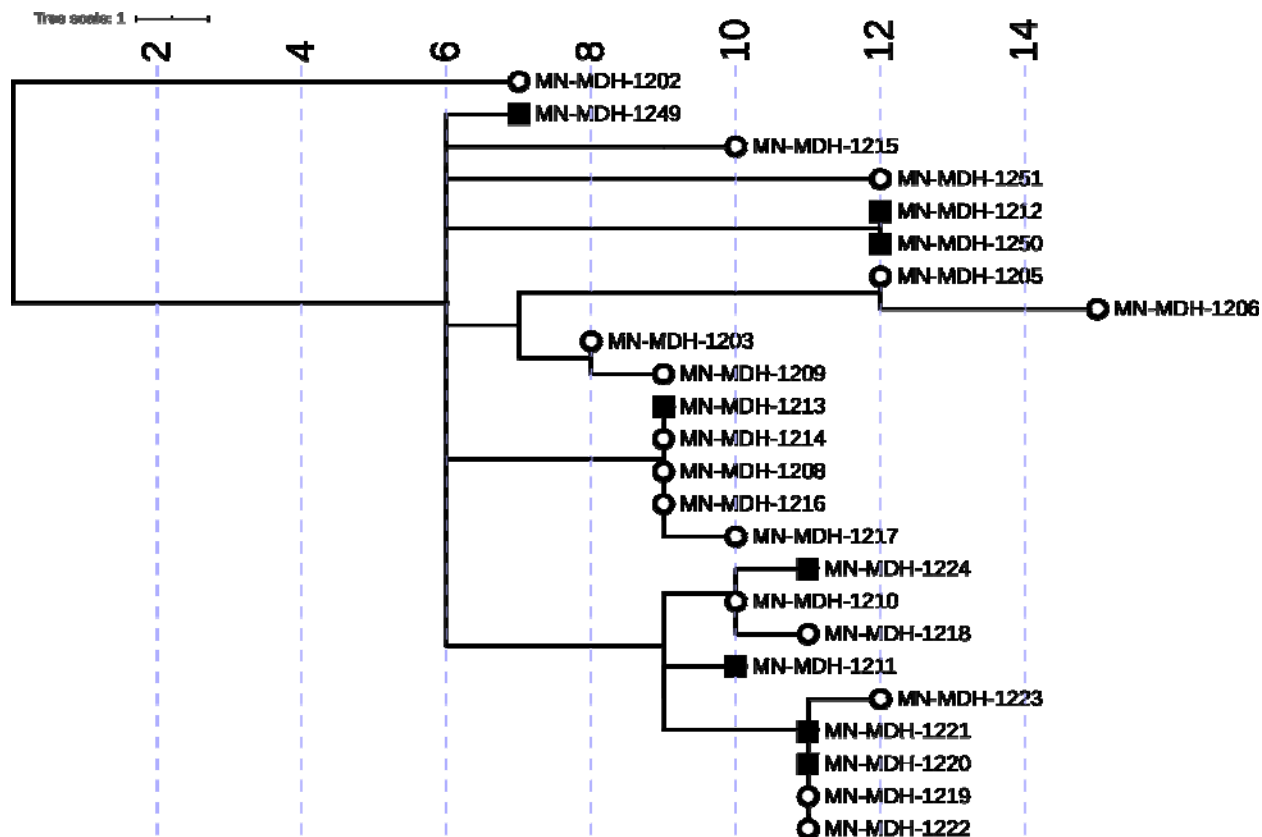


Figure 5B

Tree scale: 1 

