Distinct origins and transmission pathways of *bla*$_{KPC}$ Enterobacterales across three U.S. states

Zena Lapp[1], Rany Octaria[2,3], Sean M. O'Malley[4], Tu Ngoc Nguyen[5], Hannah Wolford[6], Ryan Crawford[1], Christina Moore[3], Paula Snippes Vagnone[4], Diane Noel[5], Nadezhda Duffy[6], Ali Pirani[7], Linda S. Thomas[3], Brittany Pattee[4], Claire Pearson[5], Sandra N. Bulens[6], Sophie Hoffman[1], Marion Kainer[3], Melissa Anacker[4], James Meek[8], Isaac See[6], Allison Chan[3], Ruth Lynfield[4], Meghan Maloney[5], Mary K. Hayden[9,10], Evan Snitkin[7], Rachel B. Slayton[6]

[1] Department of Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, Michigan, USA

[2] Department of Medicine, Division of Epidemiology, Vanderbilt University, Nashville, Tennessee, USA

[3] Tennessee Department of Health, Nashville, Tennessee, USA

[4] Minnesota Department of Health, USA

[5] Connecticut Department of Public Health, USA

[6] Centers for Disease Control and Prevention, USA

[7] Department of Microbiology and Immunology, University of Michigan, Ann Arbor, Michigan, USA

[8] Connecticut Emerging Infections Program, Yale School of Public Health, New Haven, CT

[9] Department of Medicine, Division of Infectious Diseases, Rush University Medical Center, Chicago, Illinois, USA

[10] Department of Pathology, Rush University Medical Center, Chicago, Illinois, USA

Keywords:

Transmission, carbapenem-resistant Enterobacterales (CRE), importation, horizontal gene transfer (HGT), patient transfer

Running title:

CRE origin and transmission in 3 states

Corresponding author:

Evan Snitkin

esnitkin@med.umich.edu

1520D MSRB I, 1150 W. Medical Center Dr., Ann Arbor, MI, 48109-5680

(734) 647-6472

Alternate corresponding author:

Rachel B. Slayton

via3@cdc.gov

1600 Clifton Rd NE, MS H16-3, Atlanta, GA 30329

404-636-4566

Summary:

Analysis of whole-genome sequences of *bla*$_{KPC}$ Enterobacterales from three U.S. states revealed that the underlying processes driving CRE burden in different regions can differ substantially, and that within an individual region, different STs can spread via different pathways.

51 **Abstract**

52 **Background**

53 Carbapenem-resistant Enterobacterales (CRE) are among the most concerning antibiotic
54 resistance threats due to high rates of multidrug resistance, transmissibility in healthcare settings,
55 and high mortality rates. We evaluated the potential for regional genomic surveillance to track
56 bla_{KPC}-carrying CRE transmission across healthcare facilities in three U.S. states.

57 **Methods**

58 Clinical isolates were collected from Connecticut (CT; 2017-2018), Minnesota (MN; 2012-
59 2018), and Tennessee (TN; 2016-2017) through the U.S. Centers for Disease Control and
60 Prevention's Multi-site Gram-negative Surveillance Initiative and additional surveillance.
61 bla_{KPC}-containing *Enterobacter hormaechei* and *Klebsiella pneumoniae* isolates were whole-
62 genome sequenced, and case report data on patient comorbidities, healthcare utilization, and
63 interfacility patient transfer were extracted.

64 **Findings**

65 In CT, most CRE isolates showed evidence of importation from outside the state. In MN, cases
66 were primarily from clonal expansion of epidemic lineages ST258 *K. pneumoniae* and ST171 *E.*
67 *hormaechei*. In TN, ST258, ST307 *K. pneumoniae*, and ST114 *E. hormaechei* were most
68 common; the latter two showed evidence of local acquisition of bla_{KPC} followed by onward
69 dissemination. Patients from different facilities with genetically similar isolates were enriched in
70 shared facility exposures, consistent with patient transfer driving regional spread. Placing
71 genomic data in the context of statewide patient transfer networks in MN and TN revealed
72 largely non-overlapping facility networks mediating spread of ST258 versus other STs.

73 **Conclusions**

74 These findings demonstrate that the underlying processes driving CRE burden in different
75 regions can differ substantially, and that within a region, different STs can spread via different
76 pathways. Integrating genomic and epidemiological data, including interfacility patient transfers,
77 can provide insights to target interventions.
78
79
80

Introduction

Carbapenem-resistant Enterobacterales (CRE) accounted for an estimated 13,100 infections among hospitalized patients in the United States (U.S.) in 2017 [1], and likely colonized upwards of 100,000 additional patients [2]. Infections with CRE are considered an urgent public health threat as they can be difficult to treat due to their namesake resistance to carbapenems, as well as increasing rates of resistance to other last-resort antibiotics [3,4]. Of concern are CRE that contain a carbapenemase gene found on a plasmid or other mobile genetic element due to their high propensity to spread [1]. Due to the mobile nature of the genetic elements harboring these genes, carbapenem resistance can disseminate within and between different species via horizontal gene transfer (HGT). Once acquired by a strain sufficiently adapted to the healthcare environment, carbapenem-resistant organisms can spread through clonal dissemination [5]. Others have noted that different carbapenemases, which typically reside on distinct mobile elements, have varying tendencies for HGT vs. clonal spread. While HGT and clonal dissemination have been observed to play a role in CRE dynamics at both local and regional levels [5], the relative contributions of these processes in seeding and maintaining CRE burden across different geographic regions is still unclear.

Understanding the transmission pathways for CRE in hospitalized patients is critical for effective and efficient surveillance and intervention strategies. The CRE burden in some regions appears to be driven by the evolution or importation of lineages with epidemic potential, such as bla_{KPC}-carrying ST258 *Klebsiella pneumoniae* (e.g. California) [5] and ST171 *Enterobacter hormaechei* (e.g. Minnesota) [6] in the United States. Following initial regional emergence, these strains can be amplified via transmission within healthcare facilities and can spread between facilities via

104 the transfer of infected and colonized patients [5,7,8]. In regions with sustained transmission of
105 epidemic clones, effective regional control requires identification of locations where
106 transmission is occurring and impacting the burden in the region more broadly, as well as
107 monitoring the movement of CRE carriers between healthcare facilities [9,10]. In contrast, others
108 have observed a significant role for HGT of mobile carbapenemase elements in driving CRE
109 spread within individual healthcare facilities [11–13]. If HGT is a significant contributor to
110 overall regional CRE burden, effective prevention and control requires identification of
111 reservoirs of these resistance elements and an understanding of the propensity of different
112 strain/mobile element combinations to spread between patients or act as HGT donors.
113
114 Here, we evaluate the potential for regional genomic surveillance to inform the pathways leading
115 to CRE cases across the regional healthcare networks in three U.S. states involved in the U.S.
116 Centers for Disease Control and Prevention's (CDC) Multi-site Gram-negative Surveillance
117 Initiative (MuGSI). In particular, we focus on the most commonly observed CRE in the MuGSI
118 catchment areas, *K. pneumoniae* and *E. hormaechei* harboring the *K. pneumoniae*
119 carbapenemase (KPC). This builds off of the MuGSI aim to quantify the burden of certain
120 resistant gram-negative bacteria in the U.S. by supplementing preexisting data with whole-
121 genome sequencing of CRE isolates collected through MuGSI and supplemental state-wide
122 surveillance. By integrating genomic data from densely sampled CRE cases and healthcare
123 exposure data, and comparing local strains with global isolates, we gained insight into local
124 strain origins and regional transmission pathways.
125
126

127 **Methods**

128 This retrospective study was a collaboration between the Emerging Infections Program (EIP) and
129 CDC Prevention Epicenters. The EIP is a collaboration between 10 state health departments and
130 their partnering academic institutions and the Centers for Disease Control and Prevention and
131 other federal agencies [14]. The Healthcare-Associated Infections Community Interface (HAIC)
132 is one of the core components of the EIP and conducts active population- and laboratory-based
133 surveillance for CRE [15]. Three EIP sites participated in this study (Connecticut, CT;
134 Minnesota, MN; Tennessee, TN).

135

136 <u>CRE surveillance</u>

137 Each of the participating states conducted statewide public health surveillance for CRE. In
138 addition, MN and TN conducted active surveillance for CRE through the HAIC MuGSI [16] in
139 selected counties within the state and CT conducted active surveillance statewide. In both
140 instances, carbapenem-resistant isolates were submitted by clinical laboratories to each state's
141 respective State Public Health Laboratory (SPHL; see details below). Details about each state's
142 CRE definition and MuGSI catchment area can be found in the Supplementary material. Multiple
143 isolates can be obtained from the same patient if the new isolate was obtained at least 30 days
144 after a previous case isolate was identified.

145

146 For this study, CRE are defined as any organism in the Enterobacterales order isolated from any
147 clinical specimen and resistant to doripenem, meropenem, or imipenem (minimum inhibitory
148 concentrations [MIC] of ≥4 μg/ml), ertapenem (MIC ≥2 μg/ml), or demonstrates production of a
149 carbapenemase [16]. Specifically, isolates from residents of the surveillance area identified at the

150 clinical laboratory as CRE from *Escherichia coli*, *Enterobacter cloacae* complex species (i.e., *E.*
151 *cloacae*, *E. asburiae*, *E. bugandensis*, *E. hormaechei*, *E. kobei*, *E. ludwigii*, and *E.*
152 *nimipressuralis*), and *Klebsiella* species (i.e., *K. aerogenes*, *K. oxytoca*, and *K. pneumoniae*)
153 were included. All CRE isolates identified through the described public health surveillance
154 systems were characterized at the respective SPHL, which tested for carbapenemase genes (e.g.,
155 bla_{KPC}, bla_{NDM}, bla_{OXA-48}, bla_{VIM}, bla_{IMP}; see Supplemental Methods for more details). The
156 percentage of CRE tested for carbapenemases was 81.5% in CT, 90% in MN, and 78% in TN.
157
158 CRE considered for whole-genome sequencing were collected in CT from 2017-2018, in MN
159 from 2012-2018, and in TN from 2016-2017. For CT and MN, isolates from the entire state were
160 included. In TN, only isolates from TN Emergency Medical Services (EMS) Region 2-7, which
161 includes all TN counties except those in the Memphis-Delta and Northeast Tennessee Region,
162 were included. These were excluded due to significant healthcare utilization across state lines.
163 **Table S1** provides information on the number of each type of healthcare facility in each state, as
164 well as the number of each type for which we have at least one whole-genome sequenced isolate.
165
166 <u>Isolate selection for whole-genome sequencing</u>
167 Over 90% of CRE isolates considered for whole-genome sequencing were bla_{KPC}-containing *K.*
168 *pneumoniae* or *E. hormaechei*. We elected to sequence this subset of isolates as the sample size
169 for the other carbapenemases and species would be too small to make any meaningful inferences.
170
171
172

173 Isolate and genomic data processing

174 DNA extraction and Illumina whole-genome sequencing were performed on all isolates. See
175 Supplemental Methods for state-specific details. We identified species and sequence types
176 [17,18], called single-nucleotide variants using species-specific reference genomes (*K.*
177 *pneumoniae*: KPNIH1, GenBank accession number CP008827.1, 5,394,056 base pairs; *E.*
178 *hormaechei*: MNCRE9, GenBank accession number JZDE00000000.1, 4,911,317 base pairs)
179 [19–25], built reference-based phylogenetic trees [24–28], and generated and annotated
180 assemblies [19,20,29,30]. Public CRE assemblies were downloaded from PATRIC [18,31], core
181 genome phylogenies were generated [24,32,33], and bla_{KPC} was identified [31,34]. See
182 Supplemental Methods for more details about the genomic data processing. Isolates from
183 BioProject numbers PRJNA272863 [6] and PRJNA873034 (new from this study) were used (see
184 Supplemental File S1 for isolate information).

185

186 Epidemiologic data

187 Epidemiologic data were obtained through medical record review and in TN supplemented by
188 additional data sources (see Appendix). Variables of interest collected for all isolates in this
189 study include state, treatment facility ID, patient ID, age, sex, hospitalizations in the previous
190 year, and culture source. In addition, MuGSI surveillance data included information on infection
191 type and underlying conditions, which was only obtained for selected non-MuGSI isolates. The
192 intra-facility transmission analysis investigating shared underlying conditions was limited to
193 samples for which we had data on these conditions; other analyses included all isolates (i.e.,
194 MuGSI and non-MuGSI). See Supplemental Methods for details on state-specific clinical case

195 data collection, facility identification, and matching facility names to Centers for Medicare &
196 Medicaid Services (CMS) facility IDs.
197
198 Generation of aggregate patient transfer networks
199 Aggregate patient transfer networks (i.e., data about patient flow across healthcare facilities [5])
200 for each state for the year 2017 were derived from CMS fee-for-service beneficiary claims data
201 linked to the CMS Minimum Data Set by Medicare beneficiary ID. The number of transfers
202 between two facilities includes transfers directly from one facility to another as well as transfers
203 with an intervening stay in the community of less than 365 days. See Supplemental Methods for
204 additional details.
205
206 Public CRE isolates
207 Public KPC-positive and KPC-negative *K. pneumoniae* and *E. hormaechei* isolates were
208 downloaded from the PATRIC database on 04/23/2021 [18] (see supplemental methods for more
209 details). These isolates were used to contextualize study isolates geographically and with respect
210 to KPC presence/absence. See below and supplementary methods for more details.
211
212 Integrated genomic and epidemiologic data analysis
213 To prevent overcounting of intra-facility transmission events, isolates were limited to the first
214 isolate per patient for each sequence type (ST)/facility combination. We then identified putative
215 groups of isolates related by geographic importation or HGT using maximum likelihood
216 ancestral reconstruction of core genome trees including study and public isolates. The discrete
217 trait used for ancestral reconstruction was geographic region or KPC presence/absence,

218 respectively. See supplementary methods for details. Next, we computed pairwise single
219 nucleotide variant (SNV) distances and patristic distances for study isolates from each sequence
220 alignment and classified isolates pairs as intra- or inter-facility pairs [35–37]. For most pairwise
221 SNV distance analyses, rather than choosing a specific pairwise SNV distance threshold, we
222 highlight various thresholds (≤ 5, 10, 15) as a sensitivity analysis. A threshold of 10 was chosen
223 where needed based on a decrease in the fraction of intra-facility pairs at various SNV distances
224 (**Figure S4**). For each intra-facility patient pair, we determined whether they had at least one
225 shared comorbidity and used a Fisher's exact test [38] to compare sharing of comorbidities
226 between patients with isolates ≤10 SNVs and >10 SNVs. This was done based on the assumption
227 that patients with shared comorbidities would be provided care in more proximate inpatient
228 locations. For each inter-facility patient pair, we determined whether they had a shared
229 healthcare exposure in the past year (data collection method described in the Supplemental
230 Methods). Then we compared the extent of exposure for pairs of patients from different facilities
231 with isolates ≤10 SNVs and >10 SNVs using a Fisher's Exact test [38]. We also compared paths
232 of maximum patient flow computed from aggregate patient transfers and compared it to the
233 minimum pairwise SNV distance of isolate pairs from each facility pair using a Wilcox test
234 comparing facilities where the minimum pairwise SNV distance was ≤10 with those where it was
235 >10. An empirical p-value was also computed to compare the distribution of transmission
236 between ST171 and ST258 in MN. See Supplemental Methods for more details on these
237 analyses.
238
239 <u>Data analysis and visualization</u>

240 R v4.1.1 was used for all data analysis and visualization using the following packages: tidyverse
241 v1.3.1 [39], ggtree v2.5.2 [40,41], igraph v1.2.6 [42], regentrans v0.1 [37], tidygraph v1.2.0 [43],
242 ggraph v2.0.5 [44], exact2x2 v1.6.5 [38], cowplot v1.1.1 [45], readxl v1.3.1 [46], ape v5.5 [36].
243 The code used to perform all analyses can be found here: https://github.com/Snitkin-Lab-
244 Umich/eip-cre-transmission-ms.
245
246 Human subjects research
247 The study was reviewed and approved by the Tennessee Department of Health institutional
248 review board (IRB). See 45 CFR. part 46; 21 CFR part 56. The Minnesota and Connecticut
249 Department of Health IRB determined that the study was exempt from IRB review in accordance
250 with 45 CFR 46.101(b)(4)." The study was approved by the CDC IRB with a waiver of HIPAA
251 authorization under the Privacy Rule as per 45 CFR 46.512(i). The University of Michigan
252 Medical School IRB approved this protocol.
253
254 **Results**
255 We performed whole-genome sequencing on *bla*$_{KPC}$-positive *K. pneumoniae* and *E. hormaechei*
256 clinical isolates from CT, MN, and TN and included the first isolate of each unique
257 ST/facility/patient combination for downstream analyses (**Table 1**). For the purposes of
258 readability, we refer to this set of isolates as CRE throughout the manuscript. CRE were present
259 across numerous healthcare facilities in each state (CT: 22, MN: 60, TN: 40; **Table S1**; **Figure**
260 **S1**). The dominant STs across these three states were *K. pneumoniae* ST258 in all three states, *E.*
261 *hormaechei* ST171 in MN, and *E. hormaechei* ST114 and *K. pneumoniae* ST307 in TN.
262

263 **Different extents of clonal dissemination after importation or HGT across states and STs**

264 We were first interested in investigating the origins and distribution of *bla*$_{KPC}$-positive CRE STs
265 across states. To do this we used previously sequenced *Enterobacter* spp. and *Klebsiella* spp.
266 genomes deposited in the PATRIC database (**Figure S2**) to provide geographic and genetic
267 context. We then classified phylogenetic clusters of isolates within each region as originating
268 from either importation of a preexisting strain harboring *bla*$_{KPC}$-into the state or horizontal gene
269 transfer of *bla*$_{KPC}$ into a previously *bla*$_{KPC}$-negative strain within the state (**Figure 1**; **Figures S3-**
270 **4**). We found that *bla*$_{KPC}$-positive *K. pneumoniae* ST258 is prevalent across all three states and,
271 except for 3/66 isolates of unknown origin, exclusively originated from importation from outside
272 the state. These importation events were followed by varying degrees of onward dissemination.
273 This is consistent with *K. pneumoniae* ST258 being commonly disseminated across U.S.
274 healthcare facilities and circulating lineages having a common ancestor dating back to the early
275 2000's [47]. Additionally, we found extensive dissemination of *E. hormaechei* ST171 in MN
276 after an inferred single introduction into the state, consistent with previous reports of *E.*
277 *hormaechei* ST171 transmission in MN following importation from North Dakota sometime
278 prior to 2009 [6]. We also observed clonal dissemination of the emerging epidemic clones *K.*
279 *pneumoniae* ST307 [48] and *E. hormaechei* ST114 [49] in TN following HGT of plasmids
280 harboring *bla*$_{KPC}$ into those respective strains as recently as 2016. In contrast, we found little
281 evidence of dissemination after HGT in MN or CT, with many sporadic *bla*$_{KPC}$ acquisitions
282 found in one or a few patients.

283

284 **Genomic analysis of regional clinical isolate collections allowed for the detection of local**
285 **transmission in MN and TN, and lack thereof in CT**

286 Having characterized the origin of circulating strains, we next sought to understand whether the
287 regional genomic surveillance of clinical isolates was dense enough to allow insight into
288 transmission pathways. To this end we investigated the extent to which regional phylogenetic
289 clusters that represented regional emergence events contained putative recent local transmission
290 events. We defined recent local transmission events as pairs of isolates with small pairwise SNV
291 distances from different patients and performed a sensitivity analysis with pairwise SNV distance
292 thresholds of ≤5, ≤10, and ≤15 (**Figure 2**; **Figures S5-6**) as well as with patristic distance
293 (**Figure S7**). Within an ST, smaller pairwise SNV distances indicate potential local transmission,
294 while larger pairwise SNV distances are not indicative of local transmission [50]. We captured
295 very few local transmission events in CT, suggesting that importation from other geographic
296 regions may be driving case load there. In contrast, we captured some local transmission events
297 in MN for *K. pneumoniae* ST258 and *E. hormaechei* ST171 and some local transmission of *K.*
298 *pneumoniae* ST258, *K. pneumoniae* ST307 and *E. hormaechei* ST114 in TN, consistent with our
299 finding that there was importation or HGT followed by dissemination of these STs. Closely
300 related inter-facility isolate pairs were most likely to occur between acute care hospitals (**Table**
301 **S2**).
302
303 To examine whether WGS of clinical isolate collections can shed light on which facilities had
304 transmission occur, we also investigated the relatedness of isolates within each facility (**Figure**
305 **S8**). We observed very little putative intra-facility transmission in CT. Conversely, we found that
306 facilities in MN tended to have closely related ST171 and ST258 isolates for all three pairwise
307 SNV distance thresholds, consistent with detection of intra-facility transmission. In TN, we saw
308 a similar trend, but for ST114 and ST307. Taken together, these findings suggest that clinical

309 isolate collections indeed capture recent transmission events, although the amount of
310 transmission appears different across regions, consistent with the phylogenetic analysis.
311
312 **Patients with closely related intra-facility pairs do not share underlying conditions**
313 Having detected putative intra-facility transmission in regional surveillance collections, we next
314 investigated whether these patients shared any underlying conditions, which may be indicative of
315 an outbreak related to a certain location or procedure in the hospital. However, we found that
316 patients residing in the same facility with closely related isolates did not consistently share
317 underlying conditions relative to patients with distantly related isolates (**Table S3**). This may be
318 due to extensive intermingling of patients with different chronic conditions in acute care
319 hospitals in the United States.
320
321 **Aggregate- and individual-level patient transfer data support patient transfer mediating**
322 **regional spread**
323 We next explored evidence for patient transfer mediating inter-facility spread of different CRE
324 STs using two patient sharing metrics: an aggregate metric based on CMS-derived patient
325 transfer frequencies between facilities and an individual metric based on patient-level shared
326 facility exposures (see Methods for details). A positive correlation between patient sharing and
327 isolate relatedness would suggest that the detection of closely related isolates in different
328 facilities is due to inter-facility spread via the transfer of colonized or infected patients. The
329 relationship between the aggregate and individual-level patient sharing metrics (aggregate,
330 **Figure S9**; and individual, **Table 2**, **Figure S10**), with isolate pair relatedness were generally
331 consistent for a given ST and state (**Figure S11**). For ST258 in all states ($p < 0.05$), facility pairs

332 with closely related isolate pairs tend to be ones with more patient sharing, consistent with our
333 knowledge that ST258 often traverses the healthcare network via patient transfer [5,7].
334 Moreover, we noted that in all three states, as compared to other STs, genetic linkages for ST258
335 were most strongly associated with shared facility exposures among linked patients. In MN we
336 also observe a significant association between genetic relatedness and shared exposures for
337 ST307, ST171 and ST114 (all $p < 0.05$), suggesting that patient transfer may be driving local
338 transmission in these cases as well. In TN we do not observe a significant association between
339 genetic linkage among non-ST258 isolates and shared exposures, due to a combination of small
340 numbers (e.g. ST171), and high frequencies of shared exposures among patients with isolates
341 with greater than 10 SNVs (e.g. ST114, ST307 and other STs).
342
343 **Comparison of clonal spread across MN facilities reveals different patterns of transmission**
344 **across STs**
345 As we observed evidence of substantial inter-facility transmission of both *E. hormaechei* ST171
346 and *K. pneumoniae* ST258 in MN, we next investigated the geographic spread of these STs
347 across the healthcare network. We found very different patterns of spread between these two
348 STs, with putative *K. pneumoniae* ST258 inter-facility transmission occurring sporadically
349 across the healthcare network (empirical $p = 0.3$), while putative *E. hormaechei* ST171
350 transmission was concentrated around a single facility we call Facility W (empirical $p < 0.001$;
351 **Figure 3**; **S12**). Further investigation of the connectedness of Facility W in the patient transfer
352 network revealed that it is highly connected to other facilities (**Figure S13**). Additionally, 10/24
353 (41.7%) Facility W patients with *E. hormaechei* ST171 had a shared prior healthcare exposure at
354 a specific other facility (Facility X). These findings indicate that, while both *E. hormaechei*

355 ST171 and *K. pneumoniae* ST258 may traverse the healthcare network via patient transfer, the
356 pattern by which these two STs disseminated across facilities, and the potential role of specific
357 facilities, is very different.
358
359 **Distinct facility sub-networks in TN harbor different lineages**
360 The examination of shared facility exposures in TN revealed an unexpected amount of shared
361 facility exposures among patients with *E. hormaechei* ST114 and *K. pneumoniae* ST307,
362 regardless of the precise genetic relatedness of their isolates (**Table 2**). To gain insight into this
363 pattern, we investigated what facilities patients with these STs had been to in the previous year
364 and found a cluster of seven facilities where patients with these STs, as well as STs categorized
365 as Other, had been (**Figure 4A**). Visualizing these seven facilities in the context of the aggregate
366 patient transfer network revealed that they are closely connected in the patient transfer network
367 (**Figure 4B**) and have more patient sharing among themselves than with other facilities (**Figure**
368 **4C**). Additionally, this facility cluster is distinct from a cluster containing predominately *K.*
369 *pneumoniae* ST258 (**Figure 4B**).
370
371 **Discussion**
372 As public health laboratories increase their capacity for genomic surveillance, it is critical to
373 understand if and how genomic data can be used to guide regional intervention efforts. Here, we
374 evaluated whether genomic analysis of clinical bla_{KPC}-positive CRE isolates from three U.S.
375 states could inform our understanding of the origins and transmission pathways of circulating
376 strains. We found that, in addition to detecting differences in strain composition across regions,
377 we could discern the relative contributions of importation, clonal dissemination and HGT to each

378 state's *bla*$_{KPC}$-*positive K. pneumoniae* and *E. hormaechei* case burden, as well as highlight
379 individual facilities and connected regional sub-networks where putative transmission occurred.
380 Notably, these insights did not rely on detailed clinical meta-data from affected patients,
381 indicating that other regions interested in identifying hotspots and pathways for regional spread
382 may be able to do so with only knowledge of recent healthcare exposures. These types of
383 insights can enable both targeting of regional infection prevention efforts, as well as prioritizing
384 surveillance to monitor transmission hotspots and track the reservoirs of circulating mobile
385 elements.
386
387 Despite focusing on the same two Enterobacterales species harboring the same carbapenemase
388 gene, we observed stark differences in the underlying drivers of each state's *bla*$_{KPC}$-positive CRE
389 burden. In CT, neither established epidemic lineages nor emergent CRE strains showed evidence
390 of onward transmission. Frequent importation of *bla*$_{KPC}$-positive CRE into CT is unsurprising,
391 given the proximity to other major CRE hotspots in the Northeast U.S. [51]. However, this still
392 leaves unanswered why imported *bla*$_{KPC}$-positive CRE have apparently not gained a foothold.
393 One reason may be that structural differences in regional healthcare networks influence local
394 propensity for *bla*$_{KPC}$-positive CRE amplification. For example, recent work has shown that post-
395 acute care settings such as long-term acute care hospitals and ventilator skilled nursing facilities
396 can act as initiators and amplifiers of regional CRE epidemics [52,53]. Thus, infection
397 prevention practices in these high-risk facilities, and their patient transfer connections with other
398 regional facilities may have large impacts on overall regional CRE spread. Regardless, our
399 observations suggest that focusing resources on testing admission cultures from out of state may
400 allow CT to continue having a low burden of these CRE strains.

401

402 Conversely, in MN, *bla*$_{KPC}$-positive CRE burden was driven by importation events of the major
403 epidemic lineages ST258 *K. pneumoniae* and ST171 *E. hormaechei*, followed by subsequent
404 dissemination within the region. However, there may be distinct reservoirs and/or transmission
405 pathways for ST258 and ST171 in MN as virtually all ST171 can be traced back to a single
406 importation event, while ST258 has entered the region numerous times. Furthermore,
407 examination of healthcare exposures for patients with genetically linked ST258 isolates was
408 consistent with patient transfer between facilities driving regional spread. In contrast, despite a
409 larger number of genetically linked ST171 isolate pairs, we observed a weaker enrichment in
410 shared patient exposures among genetically linked patients from different facilities. However,
411 despite a lack of individual-level shared healthcare exposures, we observed that genetically
412 linked isolates from different facilities were largely connected to isolates from a single hub
413 facility, Facility W. This facility is a hospital with many referrals that provides specialty care.
414 The lack of direct patient overlap or exposure to Facility W could indicate a large unobserved
415 pool of transmission intermediates at feeder healthcare facilities (e.g. Facility X), and/or
416 potentially an environmental reservoir within or proximate to Facility W. This suggests that MN
417 should focus their infection control efforts more on clonal dissemination rather than HGT. In
418 particular, Facility W could prioritize patients admitted from connected facilities, and state
419 epidemiologists could investigate facilities, such as Facility X, that are connected to Facility W
420 to identify other potential reservoirs and locations where intervention may have a broader
421 positive impact on regional CRE burden.

422

423 In contrast to MN where the *bla*$_{KPC}$-positive CRE burden was due to spread of two major
424 epidemic lineages, TN showed a much greater diversity in circulating strains. Most notably, in
425 addition to a significant burden of ST258, we observed large numbers of isolates from two more
426 recently emerged CRE lineages - ST307 *K. pneumoniae* and ST114 *E. hormaechei*. Although
427 their recent emergence limits genetic context provided by public databases, it is of note that the
428 most recent ancestor of clones disseminating in the region were *bla*$_{KPC}$-negative, consistent with
429 intra-region HGT initiating regional spread. In addition to putative regional HGT events in
430 ST307 and ST114, we also observed evidence of HGT events in less common CRE strains that
431 also seeded onward transmission. This pattern is in stark contrast with MN, where there were
432 very few cases of onward transmission detected after a non-epidemic strain acquired *bla*$_{KPC}$. It is
433 unclear whether these differences in the fate of novel plasmid acquisitions are due to the lack of
434 promiscuity of circulating plasmids, differences in the propensity of regional non-CRE strains to
435 maintain a *bla*$_{KPC}$ plasmid, or some combination of the two. Prior work has suggested co-
436 adaptation of epidemic strains and their plasmids [54], which could point to the dominant
437 presence of ST171 and ST258 hindering onward dissemination of their *bla*$_{KPC}$ plasmids, whereas
438 the relatively recent associations of dominant CRE strains and their plasmids in TN allows for
439 increased chance of productive HGT. This hypothesis is supported by the co-association of lower
440 frequency STs with ST307 and ST114 in a highly connected regional subnetwork in TN, as
441 compared to regional subnetworks where ST258 is dominant. Therefore, in TN we believe that
442 regional CRE prevention should focus on containing ST307 and ST114 as wel as continued
443 surveillance for new HGT events leading to broader clonal dissemination.
444

445 In addition to differences in drivers of CRE burden in different regions, we also observed some
446 similarities that are of note. In particular, despite significant differences across states in the
447 overall burden of ST258 and the number of observed genomic transmission linkages, we
448 consistently observed that patients with genetically linked ST258 isolates from different facilities
449 were strongly enriched for having histories of common healthcare exposures. This observation is
450 consistent with several previous studies showing evidence that ST258 spreads across regions via
451 the transfer of colonized and infected patients between healthcare facilities [5,55]. In contrast,
452 there were weaker, and sometimes no, association between shared facility exposures and
453 transmission for other STs, even in several cases where large numbers of genetically linked
454 patients were captured. One possible explanation for this is that non-ST258 sequence types have
455 reservoirs outside of hospitals that influence patterns of regional spread. An alternative
456 hypothesis is that the iceberg effect is much larger for non-ST258 STs, such that we are missing
457 more links in the transmission chain, and thereby failing to detect relevant exposures [2].
458
459 Our study has several limitations. First, we only included clinical isolates in our analysis and not
460 isolates that attempt to capture the burden of colonization, and we have an incomplete sampling
461 of isolates from TN. Both of these factors may lead to an underestimation of clonal
462 dissemination, that may be different for different STs depending on their virulence and
463 associated colonization-to-infection ratio (i.e. ST-specific iceberg effect). However, it is
464 noteworthy that we were still able to uncover patterns of clonal dissemination using these data,
465 suggesting that regional clinical isolate collection can provide actionable insights for regional
466 public health labs. Moreover, given that performing any kind of active surveillance at a regional
467 level is likely beyond the capacity of most local health departments, understanding how to

468 extract insights from clinical isolate collections, and potentially guide more targeted surveillance
469 efforts, is critical. While transmission may not have occurred at the facilities where clinical
470 isolates were collected, this nevertheless provides a location at which to initiate an investigation.
471 A second limitation is that the isolate collection periods were different for each state. This may
472 result in different levels of contextual information provided by public databases, through which
473 importation and HGT were inferred. However, our observations are consistent with prior studies
474 in that KPC is known to be stably associated with ST258, ST171 emerged in the upper Midwest
475 and ST307 and ST114 only recently became associated with KPC. Another limitation is that the
476 public isolates we used from PATRIC to inform the origin of locally circulating strains are not
477 equally representative of all geographic regions, which may underestimate the extent of
478 importation, as multiple importations from an unsampled reservoir may be merged. While the
479 analyses performed here would be enhanced by additional context, we note that inferences made
480 into importation using phylogenetic clustering are largely consistent with SNV-distance based
481 analyses, supporting the robustness of our conclusions. Furthermore, we only investigated HGT
482 of KPC rather than the entire bla_{KPC} plasmid, which could lead to an underestimate of the
483 number of HGT events. However, identification of plasmids with short-read data is fraught, and
484 we were still able to identify several putative HGT events using our simpler definition of HGT.
485 Furthermore, the aggregate patient transfer data used to understand the connectivity between
486 regional healthcare facilities was derived from Medicare and Medicaid patient transfers, which
487 may lead to biases in patient transfer connections. However, we note that, despite this, we
488 observed a strong relationship for many CRE strains, whereby closely related isolates tended to
489 come from pairs of facilities with high patient flow. We also observed differences between the
490 individual-level and aggregate-level patient transfer analyses, which may be due to smaller

491 sample sizes for these STs, leading to insufficient power to identify associations in the aggregate
492 patient transfer analysis.
493
494 In conclusion, our results suggest that, not only can the origins and transmission patterns of
495 *bla*$_{KPC}$-positive CRE be investigated using genomic surveillance of sufficiently comprehensive
496 regional clinical isolate collections, but also that these transmission dynamics can vary across
497 strains and regions. Importantly, the differences underlying the *bla*$_{KPC}$-positive CRE burden in
498 these three states would not have been discernible without genomic data, and in fact would be
499 strengthened further by the existence of larger genomic surveillance initiatives that could provide
500 more granular context for the origin of strains and their mobile elements to guide prevention
501 efforts. Similarly, our understanding of how *bla*$_{KPC}$-positive CRE proliferates once in a region
502 would be extremely limited without knowledge of patient healthcare exposures and transfer
503 networks, which enable a more nuanced understanding of where transmission is occurring, where
504 reservoirs might exist and where additional surveillance or intervention is required. Taken
505 together, these observations support the value of robust regional genomic surveillance for
506 antibiotic resistance threats and the need for analytic platforms capable of integrating genomic
507 and patient movement data to guide local and state infection prevention efforts.
508
509 **Acknowledgements**
510 Pamela Talley and Matt Estes from the Tennessee Department of Health. The MN-PHL
511 Microbiology and Sequencing & Bioinformatics laboratories for their work in identifying and
512 sequencing all of the MN isolates in the project. Gwendolyn Hughes and Christina Sancken from
513 CDC.

Funding

This work was supported by the CDC through the Emerging Infections Program cooperative agreement (grant number CK17-1701). It was also supported in part by CDC Prevention Epicenter Program Cooperative Agreement No. U54CK000481-S1. ZL received support from the National Science Foundation Graduate Research Fellowship Program under Grant No. DGE 1256260. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

Conflicts of interest

M.K.H. was a member of a clinical adjudication panel for an investigational SARS-CoV-2 vaccine developed by Sanofi for a year and a half, ending in July 2022.

Disclaimer: The findings and conclusions in this report are those of the authors and do not necessarily represent the views of the Centers for Disease Control and Prevention. The use of trade names is for identification only and does not imply endorsement by the Centers for Disease Control and Prevention.

References

1. Centers for Disease Control and Prevention (U.S.). Antibiotic resistance threats in the United States, 2019. Centers for Disease Control and Prevention (U.S.), 2019. Available at: https://stacks.cdc.gov/view/cdc/82532. Accessed 11 June 2020.

537 2. Lee BY, Bartsch SM, Wong KF, et al. Tracking the spread of carbapenem-resistant
538 Enterobacteriaceae (CRE) through clinical cultures alone underestimates the spread of CRE
539 even more than anticipated. Infection Control & Hospital Epidemiology. **2019**; 40: 731–
540 734.
541 3. World Health Organization, editor. Antimicrobial resistance: global report on surveillance.
542 Geneva, Switzerland: World Health Organization, 2014.
543 4. Potter RF, D'Souza AW, Dantas G. The rapid spread of carbapenem-resistant
544 Enterobacteriaceae. Drug Resistance Updates **2016**; 29:30–46.
545 5. Han JH, Lapp Z, Bushman F, et al. Whole-genome sequencing to identify drivers of
546 carbapenem-resistant Klebsiella pneumoniae transmission within and between regional
547 long-term acute-care hospitals. Antimicrobial agents and chemotherapy **2019**; 63:e01622-
548 19.
549 6. Hargreaves ML, Shaw KM, Dobbins G, et al. Clonal Dissemination of Enterobacter cloacae
550 Harboring blaKPC-3 in the Upper Midwestern United States. Antimicrobial Agents and
551 Chemotherapy **2015**; 59:7723–7734.
552 7. Snitkin ES, Won S, Pirani A, et al. Integrated genomic and interfacility patient-transfer data
553 reveal the transmission pathways of multidrug-resistant Klebsiella pneumoniae in a regional
554 outbreak. Science Translational Medicine **2017**; 9:eaan0093.
555 8. Lapp Z, Crawford R, Miles-Jay A, et al. Regional Spread of blaNDM-1-Containing
556 Klebsiella pneumoniae ST147 in Post-Acute Care Facilities. Clinical Infectious Diseases
557 **2021**; Available at: https://doi.org/10.1093/cid/ciab457. Accessed 15 July 2021.
558 9. Slayton RB, Toth D, Lee BY, et al. Vital Signs: Estimated Effects of a Coordinated
559 Approach for Action to Reduce Antibiotic-Resistant Infections in Health Care Facilities —

560 United States. MMWR Morb Mortal Wkly Rep **2015**; 64:826–831.
561 10. Toth DJA, Khader K, Slayton RB, et al. The Potential for Interventions in a Long-term
562 Acute Care Hospital to Reduce Transmission of Carbapenem-Resistant Enterobacteriaceae
563 in Affiliated Healthcare Facilities. Clin Infect Dis **2017**; 65:581–587.
564 11. Mathers AJ, Cox HL, Kitchel B, et al. Molecular Dissection of an Outbreak of
565 Carbapenem-Resistant Enterobacteriaceae Reveals Intergenus KPC Carbapenemase
566 Transmission through a Promiscuous Plasmid. mBio **2011**; Available at:
567 https://journals.asm.org/doi/abs/10.1128/mBio.00204-11. Accessed 21 February 2022.
568 12. Mathers AJ, Stoesser N, Sheppard AE, et al. Klebsiella pneumoniae Carbapenemase
569 (KPC)-Producing K. pneumoniae at a Single Institution: Insights into Endemicity from
570 Whole-Genome Sequencing. Antimicrob Agents Chemother **2015**; 59:1656–1663.
571 13. León-Sampedro R, DelaFuente J, Díaz-Agero C, et al. Pervasive transmission of a
572 carbapenem resistance plasmid in the gut microbiota of hospitalized patients. Nat Microbiol
573 **2021**; 6:606–616.
574 14. About EIP | DPEI | CDC. 2021. Available at: https://www.cdc.gov/ncezid/dpei/eip/eip-
575 about.html. Accessed 2 April 2022.
576 15. Magill SS, Dumyati G, Ray SM, Fridkin SK. Evaluating Epidemiology and Improving
577 Surveillance of Infections Associated with Health Care, United States - Volume 21,
578 Number 9—September 2015 - Emerging Infectious Diseases journal - CDC. Available at:
579 https://wwwnc.cdc.gov/eid/article/21/9/15-0508_article. Accessed 16 February 2022.
580 16. Guh AY, Bulens SN, Mu Y, et al. Epidemiology of Carbapenem-Resistant
581 Enterobacteriaceae in 7 US Communities, 2012-2013. JAMA **2015**; 314:1479–1487.
582 17. Hunt M, Mather AE, Sánchez-Busó L, et al. ARIBA: rapid antimicrobial resistance

583 genotyping directly from sequencing reads. Microb Genom **2017**; 3. Available at:
584 https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5695208/. Accessed 26 March 2021.
585 18. Davis JJ, Wattam AR, Aziz RK, et al. The PATRIC Bioinformatics Resource Center:
586 expanding data and analysis capabilities. Nucleic Acids Res **2020**; 48:D606–D612.
587 19. Andrews S. s-andrews/FastQC. 2021. Available at: https://github.com/s-andrews/FastQC.
588 Accessed 10 March 2021.
589 20. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence
590 data. Bioinformatics **2014**; 30:2114–2120.
591 21. Li H, Durbin R. Fast and accurate short read alignment with Burrows–Wheeler transform.
592 Bioinformatics **2009**; 25:1754–1760.
593 22. broadinstitute/picard. Broad Institute, 2021. Available at:
594 https://github.com/broadinstitute/picard. Accessed 20 February 2021.
595 23. Li H, Handsaker B, Wysoker A, et al. The Sequence Alignment/Map format and SAMtools.
596 Bioinformatics **2009**; 25:2078–2079.
597 24. Croucher NJ, Page AJ, Connor TR, et al. Rapid phylogenetic analysis of large samples of
598 recombinant bacterial whole genome sequences using Gubbins. Nucleic Acids Res **2015**;
599 43:e15–e15.
600 25. Poplin R, Ruano-Rubio V, DePristo MA, et al. Scaling accurate genetic variant discovery to
601 tens of thousands of samples. bioRxiv **2018**; :201178.
602 26. Kurtz S, Phillippy A, Delcher AL, et al. Versatile and open software for comparing large
603 genomes. Genome Biology **2004**; 5:R12.
604 27. Arndt D, Grant JR, Marcu A, et al. PHASTER: a better, faster version of the PHAST phage
605 search tool. Nucleic Acids Res **2016**; 44:W16–W21.

606 28. Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. IQ-TREE: A Fast and Effective
607 Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. Mol Biol Evol
608 **2015**; 32:268–274.
609 29. Bankevich A, Nurk S, Antipov D, et al. SPAdes: A New Genome Assembly Algorithm and
610 Its Applications to Single-Cell Sequencing. Journal of Computational Biology **2012**;
611 19:455–477.
612 30. Walker BJ, Abeel T, Shea T, et al. Pilon: An Integrated Tool for Comprehensive Microbial
613 Variant Detection and Genome Assembly Improvement. PLOS ONE **2014**; 9:e112963.
614 31. Brettin T, Davis JJ, Disz T, et al. RASTtk: A modular and extensible implementation of the
615 RAST algorithm for building custom annotation pipelines and annotating batches of
616 genomes. Sci Rep **2015**; 5:8365.
617 32. Crawford RD, Snitkin ES. cognac: rapid generation of concatenated gene alignments for
618 phylogenetic inference from large, bacterial whole genome sequencing datasets. BMC
619 Bioinformatics **2021**; 22:70.
620 33. Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. IQ-TREE: a fast and effective
621 stochastic algorithm for estimating maximum-likelihood phylogenies. Mol Biol Evol **2015**;
622 32:268–274.
623 34. Fu L, Niu B, Zhu Z, Wu S, Li W. CD-HIT: accelerated for clustering the next-generation
624 sequencing data. Bioinformatics **2012**; 28:3150–3152.
625 35. R Core Team. R: A Language and Environment for Statistical Computing. **2020**; Available
626 at: https://www.R-project.org/.
627 36. Paradis E, Schliep K. ape 5.0: an environment for modern phylogenetics and evolutionary
628 analyses in R. Bioinformatics **2019**; 35:526–528.

629 37. Hoffman S, Lapp Z, Wang J, Snitkin ESY 2022. regentrans: a framework and R package
630 for using genomics to study regional pathogen transmission. Microbial Genomics 8:000747.
631 38. Fay MP. Confidence intervals that match Fisher's exact or Blaker's exact tests. Biostatistics
632 **2010**; 11:373–374.
633 39. Wickham H, Averick M, Bryan J, et al. Welcome to the Tidyverse. Journal of Open Source
634 Software **2019**; 4:1686.
635 40. Yu G, Smith DK, Zhu H, Guan Y, Lam TT-Y. ggtree: an r package for visualization and
636 annotation of phylogenetic trees with their covariates and other associated data. Methods in
637 Ecology and Evolution **2017**; 8:28–36.
638 41. Yu G, Lam TT-Y, Zhu H, Guan Y. Two Methods for Mapping and Visualizing Associated
639 Data on Phylogeny Using Ggtree. Mol Biol Evol **2018**; 35:3041–3043.
640 42. Csardi G, Nepusz T. The Igraph Software Package for Complex Network Research.
641 InterJournal **2005**; Complex Systems:1695.
642 43. Pedersen TL. tidygraph: A Tidy API for Graph Manipulation. 2020. Available at:
643 https://CRAN.R-project.org/package=tidygraph. Accessed 16 February 2022.
644 44. Pedersen TL, RStudio. ggraph: An Implementation of Grammar of Graphics for Graphs and
645 Networks. 2021. Available at: https://CRAN.R-project.org/package=ggraph. Accessed 16
646 November 2021.
647 45. Wilke CO. cowplot: Streamlined Plot Theme and Plot Annotations for 'ggplot2'. 2019.
648 Available at: https://CRAN.R-project.org/package=cowplot. Accessed 15 April 2020.
649 46. Wickham H, Bryan J, attribution) Rs (Copyright holder of all R code and all C code without
650 explicit copyright, et al. readxl: Read Excel Files. 2019. Available at: https://CRAN.R-
651 project.org/package=readxl. Accessed 16 February 2022.

652 47. Bowers JR, Kitchel B, Driebe EM, et al. Genomic Analysis of the Emergence and Rapid
653 Global Dissemination of the Clonal Group 258 Klebsiella pneumoniae Pandemic. PLoS
654 ONE **2015**; 10:e0133727.
655 48. Peirano G, Chen L, Kreiswirth BN, Pitout JDD. Emerging Antimicrobial-Resistant High-
656 Risk Klebsiella pneumoniae Clones ST307 and ST147. Antimicrobial Agents and
657 Chemotherapy **2020**; 64. Available at: http://aac.asm.org/content/64/10/e01148-20.
658 Accessed 26 February 2021.
659 49. Izdebski R, Baraniak A, Herda M, et al. MLST reveals potentially high-risk international
660 clones of Enterobacter cloacae*. Journal of Antimicrobial Chemotherapy **2015**; 70:48–56.
661 50. Worby CJ, Chang H-H, Hanage WP, Lipsitch M. The Distribution of Pairwise Genetic
662 Distances: A Tool for Investigating Disease Transmission. Genetics **2014**; 198:1395–1404.
663 51. Logan LK, Weinstein RA. The Epidemiology of Carbapenem-Resistant Enterobacteriaceae:
664 The Impact and Evolution of a Global Menace. J Infect Dis **2017**; 215:S28–S36.
665 52. Lin MY, Lyles-Banks RD, Lolans K, et al. The Importance of Long-term Acute Care
666 Hospitals in the Regional Epidemiology of Klebsiella pneumoniae Carbapenemase–
667 Producing Enterobacteriaceae. Clin Infect Dis **2013**; 57:1246–1252.
668 53. Pacilli M, Kerins JL, Clegg WJ, et al. Regional Emergence of Candida auris in Chicago and
669 Lessons Learned From Intensive Follow-up at 1 Ventilator-Capable Skilled Nursing
670 Facility. Clin Infect Dis **2020**; 71:e718–e725.
671 54. Harrison E, Brockhurst MA. Plasmid-mediated horizontal gene transfer is a coevolutionary
672 process. Trends in Microbiology **2012**; 20:262–267.
673 55. Snitkin ES, Won S, Pirani A, et al. Integrated genomic and interfacility patient-transfer data
674 reveal the transmission pathways of multidrug-resistant Klebsiella pneumoniae in a regional

675 outbreak. Science Translational Medicine **2017**; 9.

676

677 **Figures and tables**

678

679 **Table 1: Overview of whole-genome sequenced isolates from each state.**

680

State	Surveillance area	Time period	Species	Number of isolates
CT	Entire state	2017-2018	*E. hormaechei*	8
			K. pneumoniae	44
MN	Entire state[1]	2012-2018	*E. hormaechei*	95
			K. pneumoniae	103
TN	83 counties[2]	2016-2017	*E. hormaechei*	58
			K. pneumoniae	78

681 [1]The MuGSI catchment area is Hennepin and Ramsey counties.

682 [2]4 counties in the Memphis-Delta Emergency Medical Services (EMS) and 8 in Northeastern TN

683 were excluded due to extensive healthcare utilization across state lines; the MuGSI catchment

684 area is TN EMS Region 5, an 8-county region encompassing the metropolitan Nashville area, of

685 TN.

686

687 **Table 2: Individual-level healthcare exposure data provides some insight into potential sites**

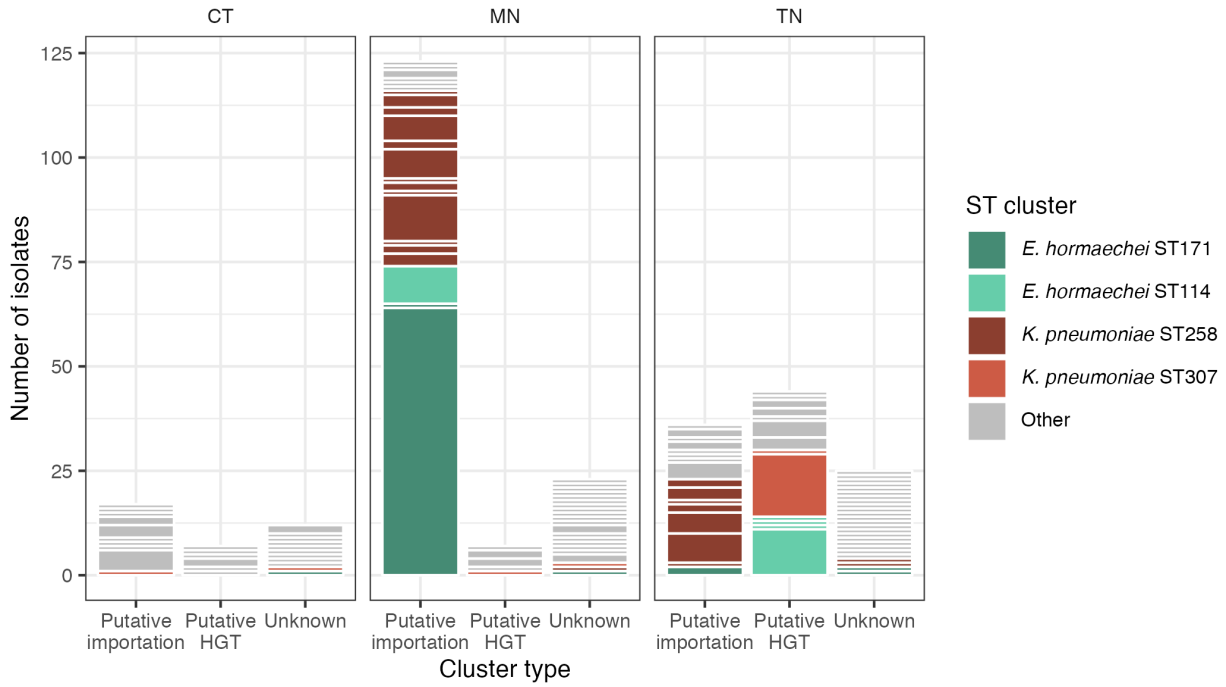
688 **of inter-facility transmission.**

State	ST	Shared facility exposures[1] (%)		Fisher's exact p
		Pairwise SNV distance ≤10	Pairwise SNV distance > 10	
CT	*K. pneumoniae* ST258	1/2 (50)	1/197 (1)	0.02
	K. pneumoniae ST307	0/1 (0)	0/2 (0)	1
	Other	0/3 (0)	0/4 (0)	1
MN	*E. hormaechei* ST114	3/9 (33)	1/43 (2)	0.02
	E. hormaechei ST171	23/430 (5)	62/1956 (3)	0.03
	K. pneumoniae ST258	10/62 (16)	23/1956 (1)	$2*10^{-8}$
	K. pneumoniae ST307	3/3 (100)	0/7 (0)	0.008
	Other	2/4 (50)	0/6 (0)	0.1

	E. hormaechei ST114	4/16 (25)	24/79 (30)	0.8
	E. hormaechei ST171	1/4 (25)	1/21 (5)	0.3
TN	*K. pneumoniae* ST258	7/21 (33)	7/369 (2)	$1*10^{-6}$
	K. pneumoniae ST307	5/17 (29)	30/104 (29)	1
	Other	7/13 (54)	4/18 (22)	0.1

[1]A shared facility exposure is when individuals with isolates from different facilities both spent time in the same facility at some point in the last year.

694



695

696 **Figure 1: Different extents of clonal dissemination after importation or horizontal gene**

697 **transfer (HGT) across states and sequence types (STs).** Clusters are separated by white lines

698 between the colored bars, where the bar height of a given cluster is the cluster size. Clusters were

699 defined as subclades that were monophyletic for a single state. Importation events were

700 considered those with an edge transition from another state to the EIP state. HGT events were

701 considered those with a KPC- to KPC+ transition regardless of the state of origin of the KPC-

702 strain. See methods for more details. We observed several importation events with limited

703 onward transmission in all states. We also see a large cluster of *E. hormaechei* ST171 in MN that

704 arose from a putative importation event, indicating importation followed by sustained

705 dissemination. Additionally, we observed two large clusters of *E. hormaechei* ST114 and *K.*

706 *pneumoniae* ST307, respectively, in TN, indicating HGT events followed by dissemination.

707

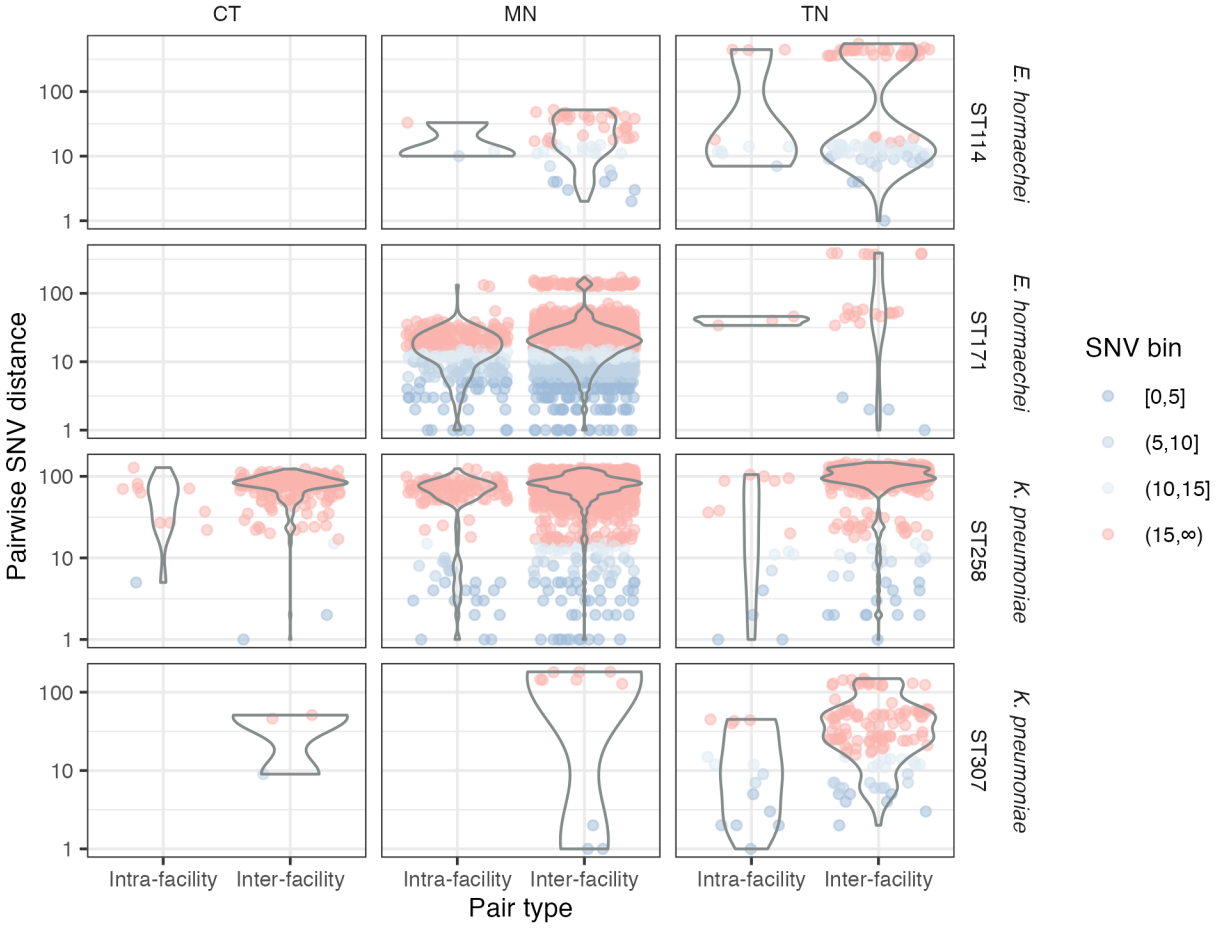


Figure 2: Small pairwise single nucleotide variant (SNV) distances between isolates suggest that local transmission was captured. Each point is an isolate pair. The y axis is log10-scaled and points are jittered horizontally for visualization purposes. Varying extents of transmission were captured for different states and STs. Putative local transmission events are those in the tail of the distribution (lower pairwise SNV distances). SNV bins were chosen to highlight closely related pairs of isolates at different SNV thresholds as a sensitivity analysis.

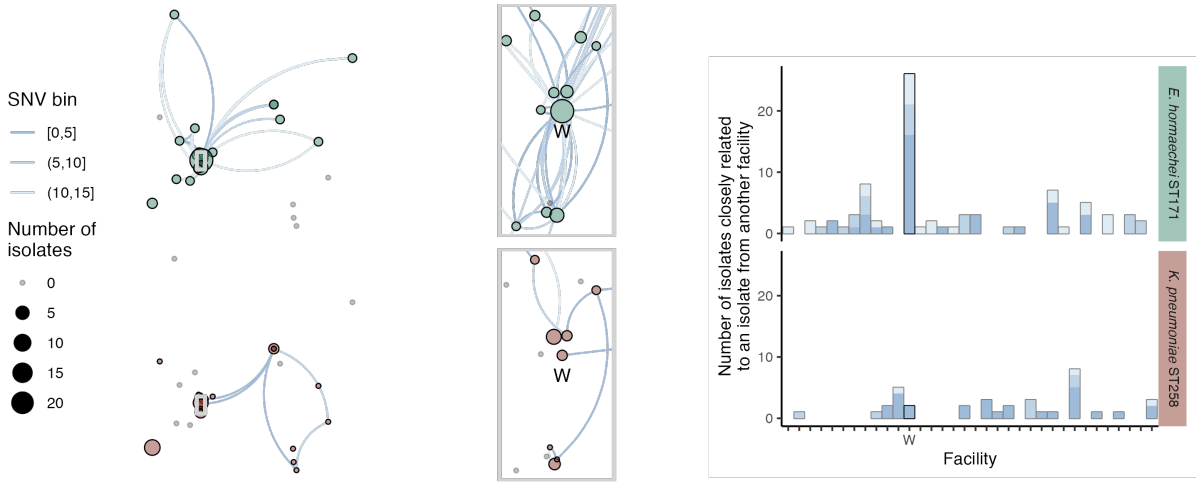


Figure 3: Different patterns of transmission across STs in MN. Facilities are oriented as they are geographically in space, but longitude and latitude are anonymized. The grey box is an insert of the left map. Each isolate was paired to their closest isolate in another facility with a subsequent collection date. *E. hormaechei* ST171 inter-facility transmission is concentrated around facility W (29/40 [72.5%] of inter-facility pairs with a pairwise SNV distance ≤10 are connected to facility W), while *K. pneumoniae* ST258 transmission is more evenly distributed between facilities (maximum 3/8 [37.5%] of inter-facility pairs with a pairwise SNV distance ≤10 are connected to any facility). SNV = single nucleotide variant.

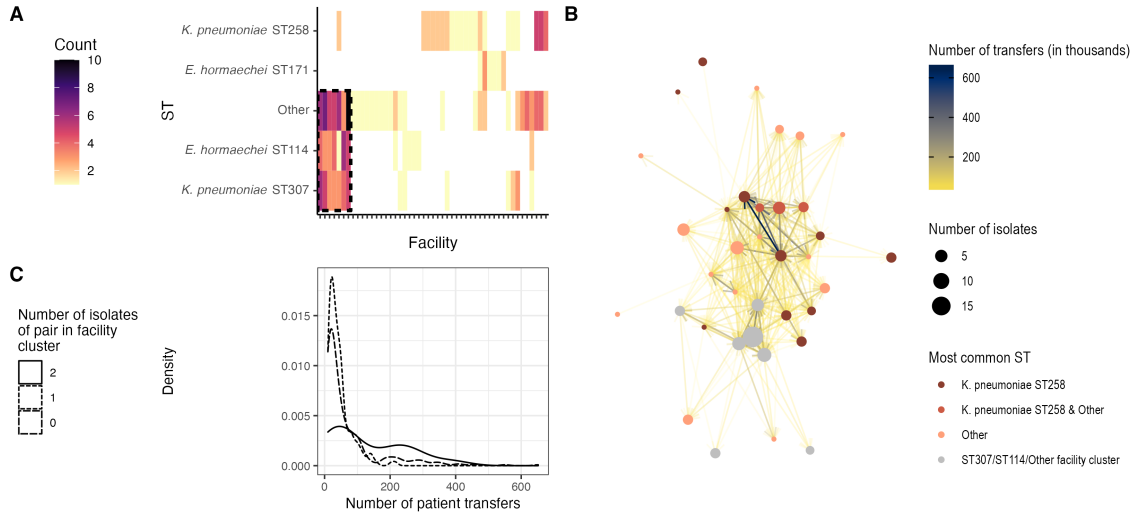


Figure 4: Distinct facility sub-networks in TN harbor different lineages. (A) Number of patient facility exposures in the prior year for each facility and ST. Each patient may be represented more than once if they have more than one facility exposure. A cluster of facilities identified using the complete linkage clustering method (dashed box) are common in patients with ST114, ST307, and other isolates. (B) Patient transfer network of facilities with at least one whole-genome sequenced isolate plotted using the Kamada-Kawai algorithm. The cluster of facilities identified in the heatmap in panel A (Kp ST307/Ec ST114/Other facility cluster) are also clustered in the network. (C) Number of patient transfers for facility pairs where 0, 1, or 2 of the facilities are in the facility cluster identified in the heatmap in panel A. Facility pairs where both facilities are in the cluster have more patient sharing than facility pairs where only 0 or 1 of the facilities is in the identified cluster.

748 **Supplemental information**

749

750 **Supplemental methods**

751 Catchment areas

752 The Multi-site Gram-negative Surveillance Initiative (MuGSI) [16] CRE surveillance catchment
753 areas for each state participating in this study are: the entire state of CT; Hennepin and Ramsey
754 counties, encompassing the metropolitan Minneapolis-St. Paul area, of MN; and TN EMS
755 Region 5, an 8-county region encompassing the metropolitan Nashville area, of TN. The TN the
756 Memphis-Delta and Northeast Regions were excluded due to healthcare utilization across state
757 lines.

758

759 CRE surveillance

760 *Connecticut*

761 Connecticut did not participate in MuGSI surveillance during the study period, so all study
762 isolates from Connecticut came from statewide passive surveillance for CRE. CRE surveillance
763 of all Enterobacterales isolates collected from all invasive clinical sites, respiratory sources, and
764 urine began on Jan 1, 2017. All CRE isolates undergo antimicrobial susceptibility testing with a
765 custom Sensititre and disk diffusion panel, phenotypic detection of carbapenemase activity using
766 the mCIM test and PCR detection of carbapenemase genes (bla_{KPC}, bla_{NDM}, bla_{OXA-48}, bla_{VIM},
767 bla_{IMP}). Isolates carrying bla_{KPC} that were identified in 2017 and 2018 from this collection were
768 included.

769

770 *Minnesota*

771 MuGSI surveillance in Minnesota is conducted in Ramsey and Hennepin counties. MuGSI
772 isolates are obtained from sterile sites or urine and include *E. coli*, *Klebsiella* spp., and
773 *Enterobacter* spp. Non-MuGSI isolates are collected from a variety of culture sources (e.g.,
774 blood, urine, respiratory) and represent multiple Enterobacterales species. Isolate identification is
775 confirmed by the MALDI Biotyper CA System (Bruker Daltonics, Inc., Billerica, MA). All CRE
776 isolates are tested for carbapenemase production using the modified carbapenem inactivation
777 method (mCIM). Isolates demonstrating phenotypic carbapenemase production undergo
778 molecular testing (i.e., polymerase chain reaction [PCR]) for detection of carbapenemase genes
779 bla_{KPC}, bla_{NDM}, $bla_{OXA\text{-}48}$, bla_{VIM}, bla_{IMP}. All bla_{KPC} -positive *K. pneumoniae* and *E. cloacae*
780 complex isolates submitted to MN PHL from 2012-2018 were analyzed for this study.
781
782 *Tennessee*
783 EMS region 2-7 were selected because it provides a good representation of patient referrals;
784 captures most isolates in patient transfer networks covering the area; and includes both the
785 Knoxville area in which CRE emerged in 2016, and the Nashville area (the MuGSI catchment
786 area: Dickson, Cheatham, Robertson, Sumner, Wilson, Rutherford, Williamson and Davidson
787 counties) where CRE increased in 2017. CRE isolates underwent organism confirmation using
788 the MALDI-TOF bioMerieux. Isolates were tested for antimicrobial resistance with the Kirby
789 Bauer method. Those isolates that were resistant to at least one carbapenem were then tested by
790 the CarbaNP Method for carbapenemase production. Those that demonstrated carbapenemase
791 production from the CarbaNP underwent a PCR assay developed by the CDC to test for the
792 bla_{KPC} and bla_{NDM}. All bla_{KPC}-positive isolates collected during 2016–2017 and submitted to the
793 TN SPHL were included.

794

795 Epidemiologic data

796 MuGSI surveillance data were collected through medical record review by EIP surveillance
797 officers in each site. Additional details on data collection for isolates identified outside of the
798 MuGSI program are listed below for each site.

799

800 *Connecticut*

801 Clinical case details were abstracted from medical records by EIP surveillance officers with the
802 Connecticut Department of Public Health.

803

804 *Minnesota*

805 At the time of culture collection, surveillance staff completed CRF through electronic or in-
806 person medical chart review.

807

808 *Tennessee*

809 For all carbapenemase-producing Enterobacterales cases, a questionnaire is completed by the
810 healthcare facility of origin. For cases that did not have a completed questionnaire,
811 hospitalization details were gathered from the Statewide Hospital Discharge data (described
812 below).

813

814 Facility identification

815 Facility identification was performed through chart review at the reporting facility. State-specific
816 details are described below.

817

818 *Connecticut*

819 Chart review was performed at the reporting facility for all inpatient and hospital network events.
820 When additional referring facilities (including nursing homes, long term acute care hospitals,
821 etc.) were identified on medical record review, follow-up investigation was performed at the
822 additional facilities. Upon identification of outpatient isolates, patient details were collected from
823 ordering providers by phone call and/or faxed data collection instrument.

824

825 *Minnesota*

826 MN staff maintain a compendium with facility name and type according to CMS. The facility
827 name and type are recorded in the MN surveillance database for each isolate and verified using
828 this compendium. For MuGSI cases, chart review at the facility where the isolate was collected
829 was also performed by MN staff.

830

831 *Tennessee*

832 TN staff linked patient identifiers of each case in the National Electronic Disease Surveillance
833 System (NEDDS)-Base System (NBS) surveillance data to the TN Hospital Discharge Data
834 System (HDDS) to identify and confirm the healthcare facility where they had the first specimen
835 culture collected for each CP-CRE case in TN. Alternatively, if the patient was not hospitalized
836 in a TDH-licensed hospital and therefore, was not found in the HDDS, TN staff used
837 surveillance data from NBS and the afore-mentioned questionnaire to identify the healthcare
838 facility of origin. Using this approach, the healthcare facility of origin was identified for 139
839 (89%) of the 157 included isolates. TN staff also gathered information on healthcare exposures

840 within one year prior to specimen collection by linking the patient identifiers in the surveillance
841 data with the inpatient and outpatient HDDS dataset. Isolates with no healthcare exposure
842 information were not used for the patient sharing network analysis.
843
844 Linkage/matching of facility names to CMS ID
845 Linkage of facility names to CMS ID was performed at each respective SPHL.
846
847 *Connecticut*
848 Facilities were matched with CMS IDs using CMS Provider of Services Files available at:
849 https://www.cms.gov/Research-Statistics-Data-and-Systems/Downloadable-Public-Use-
850 Files/Provider-of-Services
851
852 *Minnesota*
853 In MN, CMS IDs were manually entered for each isolate using Excel.
854
855 *Tennessee*
856 In TN, the name healthcare facility of origin was linked with the registered facility name by
857 CMS ID using the SPEDIS and COMPGED commands in SAS 9.4 (SAS Institute, Cary, North
858 Carolina) to allow for inexact matches that accounted for minor spelling variations between the
859 facility names from TN's surveillance data and those from the CMS dataset. Due to the data-
860 sharing agreement between TDH and the University of Michigan, the facility names and CMS
861 IDs in the WGS dataset and patient transfer data from CMS were de-identified before sharing
862 with The University of Michigan.

863

864 Generation of aggregate patient transfer networks

865 Aggregate patient transfer networks for each state for the year 2017 were derived from Centers
866 for Medicare and Medicaid Services (CMS) fee for service beneficiary claims data linked to the
867 CMS Minimum Data Set (MDS) by Medicare beneficiary ID. The dataset comprises a network
868 of medical facilities, including but not limited to short-term acute care hospitals, skilled nursing
869 facilities and long-term acute care hospitals, but may not be representative of non-Medicare
870 beneficiaries. To create an annual transfer network, CDC staff linked each unique beneficiary's
871 healthcare utilization. This included CMS hospital claims and the MDS to ascertain a
872 beneficiary's presence in a nursing home, regardless of payer for that stay. Transfer networks
873 were comprised of the year of interest and looked back to the prior year to obtain all discharging
874 facilities where patients had an opportunity to acquire CRE.

875

876 The number of transfers between two facilities includes transfers directly from one facility to
877 another as well as transfers with an intervening stay in the community of less than 365 days. We
878 chose to use a 365-day time interval rather than a smaller time interval because this gives us
879 more resolution in comparing the extent of patient flow between facilities. Any facility pairs with
880 10 or fewer patient transfers were censored, as per data use agreements between CDC and CMS.

881

882 DNA extraction

883 *Connecticut*

884 Total DNA extractions performed by CT DPHL for WGS were either by the manual method or
885 automated on the QIAcube. Manual extractions from pure single colony Gram-negative bacterial

886 isolates by QIAgen DNeasy Blood and Tissue kit and follow-up quality control of all purified
887 extracts were conducted following PulseNet Protocols for Whole Genome Sequencing on the
888 Illumina MiSeq system. Automated extractions of bacterial cell pellets were performed using the
889 QIAcube protocol for Isolation of DNA from Gram-negative bacteria and QIAGEN DNeasy
890 Blood and Tissue kit. To obtain the cell pellets, colonies from overnight plates were resuspended
891 in 1x Phosphate buffered saline (PBS, Sigma) to an OD600 of 0.35-0.40. Two milliliter cell
892 suspension were centrifuged at 5,000 *x g* for 5 minutes at 2-8 °C and used for extractions.
893
894 *Minnesota*
895 DNA extraction of the isolates was performed in MN using either the QIAcube (Qiagen,
896 Germantown, MD) extractor and QIAmp DNA Mini extraction kit (Qiagen, Germantown, MD)
897 or the Roche MagnaPure LC and the TNA Isolation extraction kit.
898
899 *Tennessee*
900 DNA extraction of the isolates was performed in TN on a Thermo Fisher, KingFisher Flex auto-
901 extraction instrument, using an Applied Biosystems MagMAX DNA Multi-Sample Ultra
902 extraction kit.
903
904 Whole-genome sequencing (WGS)
905 The extracted DNA was prepared for WGS using Illumina Nextera XT library Prep kits with
906 Illumina, Nextera XT Index kits (MN, TN, CT) or with Illumina DNA Prep kit and Illumina CD
907 Indexes (CT) (Illumina, San Diego, CA). The Library preparation process was performed on a
908 Beckman Coulter Biomek Fxp instrument (Beckman Coulter, Brea, CA) (TN), on the Eppendorf

epMotion (CT) (Eppendorf, Enfield, CT), or manually (MN, CT). The prepared libraries were sequenced on an Illumina MiSeq sequencing instrument (Illumina, San Diego, CA) using 2x250 (MN, TN, CT) or 2x150 (CT) paired end sequencing with V2 chemistry. The fastq files generated were transferred to the University of Michigan for processing and analysis.

Genomic data processing

All genomic data processing and analysis was performed by members of Evan Snitkin's lab at the University of Michigan.

Species/ST identification

MLST was assigned with Ariba v2.14.4 [17]. For PATRIC isolates mlst were extracted from the isolate metadata [18].

Sequence alignments

Single-nucleotide variants were called in study sequences using SNPKIT (https://github.com/Snitkin-Lab-Umich/snpkit). Quality of reads was assessed with FastQC v0.11.9 [19]. Adapter sequences and low-quality bases were removed with Trimmomatic v0.36 [20]. Variants were identified by mapping filtered reads to the KPNIH1 reference genome (GenBank accession number CP008827.1; 5,394,056 base pairs) for *K. pneumoniae* sequences and the MNCRE9 reference genome (GenBank accession number JZDE00000000.1; 4,911,317 base pairs) for *E. hormaechei* sequences using bwa v0.7.17 [21], removing polymerase chain reaction duplicates with Picard 2.21.7 [22], removing clipped alignments using Samclip 0.4.0, and calling variants with SAMtools v1.11and bcftools [23]. Variants were filtered from raw

932 results using GATK's VariantFiltration v3.8 (QUAL, >100; MQ, >50; ≥10 reads supporting
933 variant; and FQ, <0.025) [25]. Consensus files generated during variant calling were
934 recombination filtered using Gubbins v3.0.0 [24]. The alleles at each position that passed
935 filtering were concatenated to generate a non-core variant alignment relative to the respective
936 reference genome. Alleles that did not pass filtering were considered unknown (denoted as N in
937 the alignment).
938
939 *Reference-based phylogenies*
940 A custom Python script was used to filter out (mask) single nucleotide variants in the whole-
941 genome alignment that were: (i) <5 base pairs (bp) in proximity to indels that were identified by
942 GATK HaplotypeCaller [25], (ii) in a recombinant region identified by Gubbins v3.0.0 [24], (iii)
943 in a phage region identified by the Phaster web tool [27] or (iv) they resided in tandem repeats of
944 length greater than 20bp as determined using the exact-tandem program in MUMmer v3.23 [26].
945 This whole-genome masked variant alignment was used to reconstruct a maximum likelihood
946 phylogeny with IQ-TREE v1.6.12 using the general time reversible model GTR+G and ultrafast
947 bootstrap with 1000 replicates (-bb 1000) [33].
948
949 *Genome assemblies*
950 Quality of reads was assessed with FastQC v0.11.9 [19]. Adapter sequences and low-quality
951 bases were removed with Trimmomatic v0.36 [20]. The clean reads were assembled with Spades
952 v3.14.1 (careful mode) [29]followed by Pilon v1.23 assembly correction [30]. Contigs smaller
953 than 500bp were discarded.
954

955 *Public isolates*

956 All assemblies from the species identified within the EIP dataset were downloaded from the
957 PATRIC database on 04/23/2021 [18]. These data were merged with assemblies from
958 PRJNA603790, PRJNA690239, PRJNA401340, and PRJNA415194 for a total of 74,367
959 assemblies. Genome annotations for the dataset were generated with RAST v1.035 [31].

960

961 *Core genome phylogenies*

962 A concatenated gene alignment of core genes was generated for the entire dataset with cognac
963 v1.0 [32]. Relevant isolates for understanding the population structure and geographic
964 distribution of the isolates collected as part of the EIP study were identified as any isolate within
965 25 substitutions or the nearest phylogenetic neighbor by alignment distance. This yielded a
966 subset of the 5,346 isolates, including study isolates and public isolates; and a second,
967 concatenated, core gene alignment was generated with cognac for these most relevant isolates,
968 returning the nucleotide alignment for more precise distance comparisons. To generate sequence
969 type specific alignments, these data were divided into relevant clonal groups. Any clonal group
970 with less than 10 members was merged into a group of its respective species. Outgroups were
971 identified as the isolate with the lowest substitution distance to an isolate assigned to that clonal
972 group or species, for clonal group or species alignments, respectively. A core, concatenated gene
973 alignment for each set of isolates corresponding to species/clonal group was generated using
974 cognac with the corresponding outgroup specified, returning the nucleotide alignment.
975 Recombinant positions were removed with Gubbins v3.0.0 [24]. For alignments with more than
976 1000 isolates included, the alignments were randomly downsampled into subsets of 250
977 sequences, including the outgroup in each alignment, and gubbins was run on the downsampled

978 alignment. This was performed in triplicate, and any position identified as recombinant in the
979 three runs was masked from the alignment. The recombination filtered alignments were then
980 input to IQ-TREE v1.6.12 [33] to generate maximum likelihood phylogenies.
981
982 *bla_{KPC} identification*
983 The *bla_{KPC}* gene was identified using the CD-HIT v4.7 [34] data generated with cognac. The
984 gene cluster containing coding sequences annotated by RAST v1.035 [31] as *bla_{KPC}* was
985 identified. *bla_{KPC}* status was determined for each genome as having representation in the *bla_{KPC}*
986 cluster.
987
988 Integrated genomic and epidemiologic data analysis
989 *Identification of putative geographic importation and HGT events*
990 For each core genome tree, maximum likelihood ancestral reconstruction of each study site
991 location (MN, TN, CT) and presence of *bla_{KPC}* presence was performed individually in R v4.0.2
992 [35] using the ace() function in ape v5.5 [36]. Nodes with an ancestral reconstruction confidence
993 of < 0.875 were discarded. All transition events (for each geographic location and *bla_{KPC}*) were
994 then mapped onto the edges of the phylogeny. Each study isolate tip was traversed towards the
995 root of the tree until the first transition event was identified. This transition event was considered
996 a putative geographic importation if the transition was a geographic transition and a putative
997 HGT event if the transition was for *bla_{KPC}*. Study isolates with the same edge transition event
998 were considered part of a cluster. We consider the number of transition events for a given
999 geographic location/HGT event to be the number of identified clusters. Some study isolates were

1000 unable to be classified using this method and are considered of unknown origin. These isolates
1001 were clustered together if they appeared as a pure cluster on the phylogeny.
1002
1003 *Intra- and inter-facility transmission analysis*
1004 Pairwise single nucleotide variant (SNV) distances for isolates from each sequence alignment
1005 were calculated in R v4.0.2 [35] using the dist.dna() function in ape v5.5 [36] with
1006 pairwise.deletion = TRUE. These pairwise SNV distances were classified as intra- and inter-
1007 facility pairs using the get_pair_types() function in regentrans v0.1 [37]. This function was also
1008 used to identify closely related pairs within a given facility.
1009
1010 *Individual level patient healthcare exposure analysis*
1011 For each inter-facility patient pair, we determined whether they had a shared healthcare exposure
1012 (data collection method described above). Then we compared the extent of exposure for pairs of
1013 patients from different facilities with isolates ≤10 SNVs compared to >10 SNVs using a Fisher's
1014 Exact test [38].
1015
1016 *Aggregate patient transfer analysis*
1017 For each state, we subset the aggregate patient transfer network (described above) to only
1018 include the facilities in our dataset or ones that are connected to one of the facilities in our
1019 dataset by at least one patient transfer event over the course of a year. We then identified paths of
1020 maximum patient flow between this subset of facilities using the get_patient_flow() function in
1021 regentrans v0.1 [37]. Patient flow was then compared to the minimum pairwise SNV distance of

1022 isolate pairs from each facility pair using a Wilcox test comparing facilities where the minimum
1023 pairwise SNV distance was ≤10 with those where it was >10.
1024
1025 *Comparison of ST171 and ST258 transmission patterns in MN*
1026 Each isolate was paired to their closest isolate in another facility with a subsequent collection
1027 date. To assess whether *E. cloacae* ST171 and *K. pneumoniae* ST258 in MN had more
1028 transmission surrounding a central facility than expected by chance, we calculated empirical p-
1029 values as follows. We define *n* as 2x the total number of closely related inter-facility pairs (i.e.
1030 edges ≤10 SNVs different; see **Figure 4**), because each pair is connected to 2 facilities. We then
1031 identified facilities that were connected by at least one closely related edge. Next, we randomly
1032 sampled these facilities with replacement *n* times, counted the number of times each facility
1033 appeared, found the proportion of total edges connected to that facility, and took the maximum.
1034 We repeated this process 1000x and compare dthe maximum proportion from the sampled data
1035 to the actual maximum proportion.
1036
1037
1038

1039 **Supplemental table and figures**

1040

1041 **Table S1: Number of each facility type by state and how many facilities were represented**

1042 **in whole-genome sequencing (in parentheses).**

1043

Facility type	CT	MN	TN
ACH	30 (19)	50 (23)	92 (31)
LTACH	3 (0)	2 (2)	10 (4)
Other	7 (0)	96 (6)	48 (1)
SNF	224 (0)	366 (9)	309 (0)
Unknown	0 (3)	0 (20)	0 (4)

1044

1045 **Table S2: Facility type pairs of inter-facility isolate pairs with a pairwise SNV distance ≤10.**

1046

State	ST	ACH-ACH	ACH-LTACH	ACH-SNF	Other
CT	*K. pneumoniae* ST258	2	0	0	0
CT	*K. pneumoniae* ST307	1	0	0	0
CT	Other	1	0	0	2
MN	*E. hormaechei* ST114	2	5	1	1
MN	*E. hormaechei* ST171	109	92	56	101

MN	*K. pneumoniae* ST258	9	1	6	42
MN	*K. pneumoniae* ST307	1	0	0	2
MN	Other	1	0	0	1
TN	*E. hormaechei* ST114	9	1	0	2
TN	*E. hormaechei* ST171	1	1	0	2
TN	*K. pneumoniae* ST258	13	6	0	0
TN	*K. pneumoniae* ST307	17	0	0	0
TN	Other	6	1	0	3

1047 Abbreviations: SNV=single nucleotide variant; ST=sequence type; ACH=acute care hospital;
1048 LTACH=long-term acute care hospital; SNF=skilled nursing facility
1049
1050 **Table S3: Patients with closely related isolates from the same facility with the same ST do**
1051 **not share comorbidities more than patients with distantly related isolates.**

State	ST	Shared comorbidity (%)		Fisher's exact p
		Pairwise SNV distance ≤10	Pairwise SNV distance > 10	
CT	*K. pneumoniae* ST258	1/1 (100)	5/10 (50)	1
MN	*E. hormaechei* ST114	1/1 (100)	0/2 (0)	0.3

	E. hormaechei ST171	6/86 (0.07)	17/229 (0.07)	1
	K. pneumoniae ST258	0/25 (0)	3/102 (0.03)	1
TN	*E. hormaechei* ST114	1/1 (100)	6/9 (0.67)	1
	E. hormaechei ST171	0/0	1/3 (50)	1
	K. pneumoniae ST258	2/5 (40)	5/11 (45)	1
	K. pneumoniae ST307	5/8 (0.625)	7/7 (100)	0.2

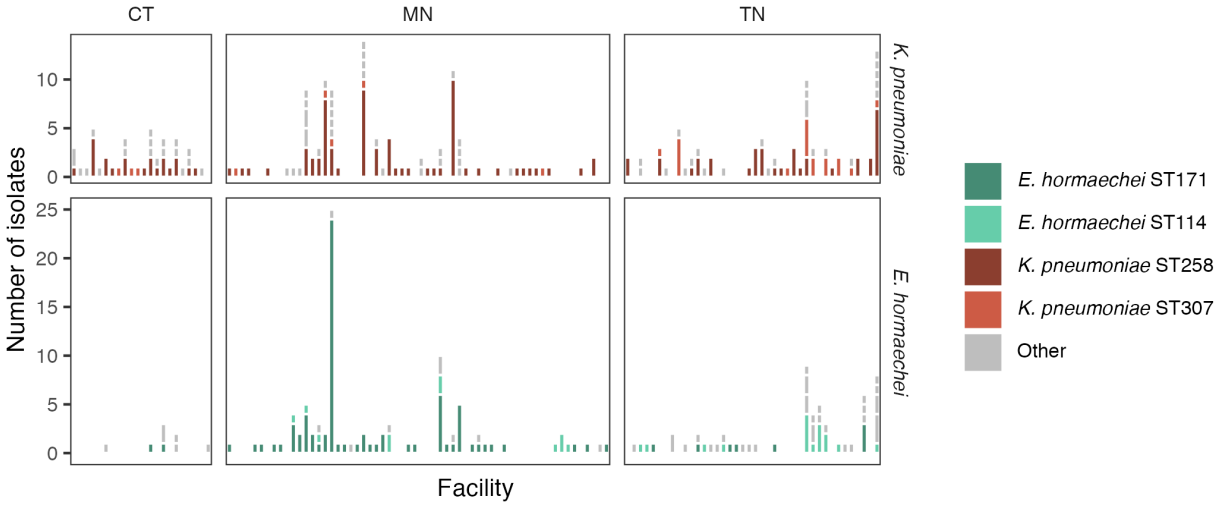
1052 Abbreviations: SNV=single nucleotide variant; ST=sequence type

1053

1054

1055

1056



1057

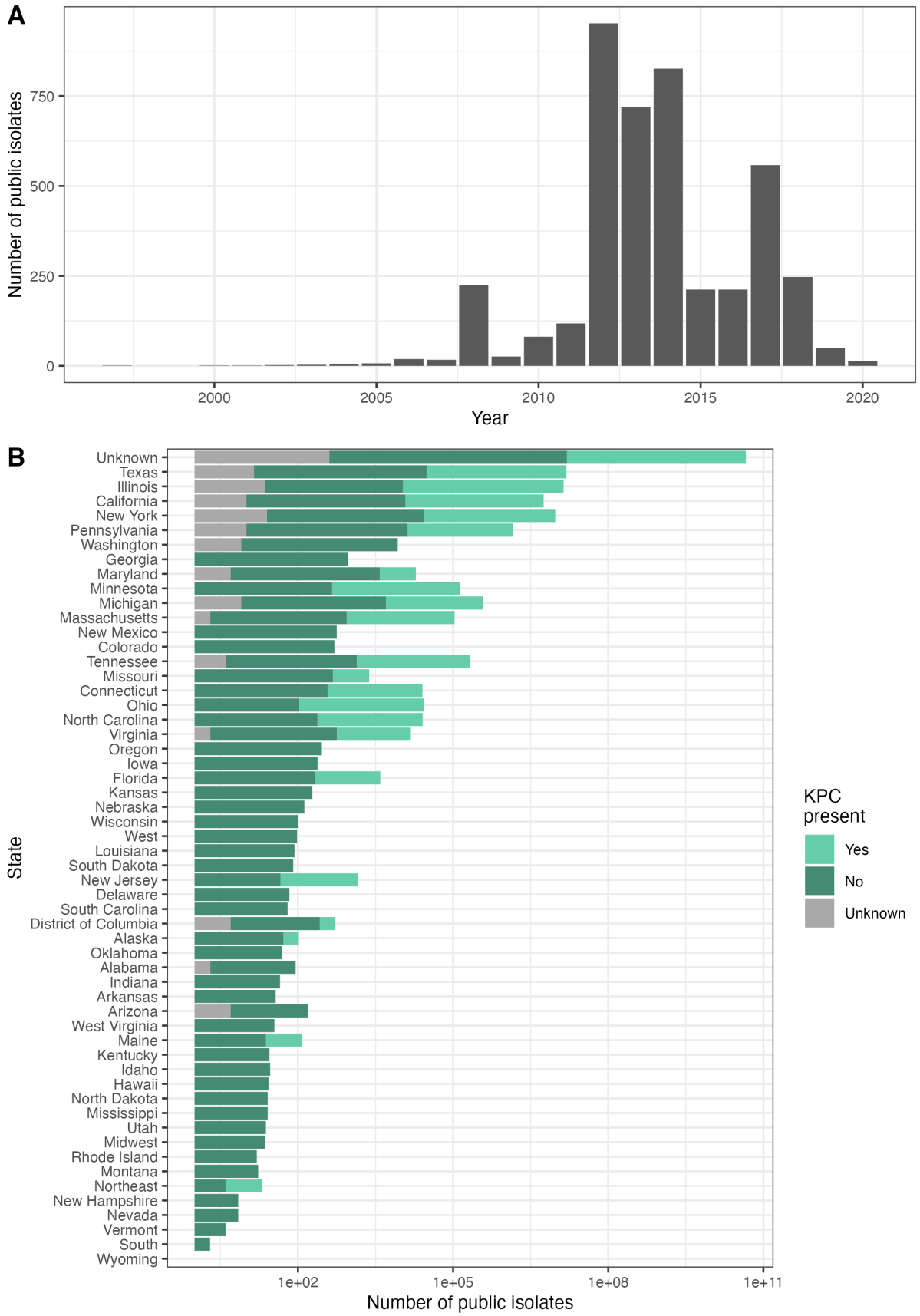
1058 **Figure S1: Distribution of species and sequence types across healthcare facilities with at**

1059 **least one identified CRE with *bla*KPC.** X axis shows facilities. Grey is any other ST of the

1060 species corresponding to that faceted plot.

1061

1062


A
Number of public isolates
750
500
250
0
2000
2005
2010
2015
2020
Year
B
State
Unknown
Texas
Illinois
California
New York
Pennsylvania
Washington
Georgia
Maryland
Minnesota
Michigan
Massachusetts
New Mexico
Colorado
Tennessee
Missouri
Connecticut
Ohio
North Carolina
Virginia
Oregon
Iowa
Florida
Kansas
Nebraska
Wisconsin
West
Louisiana
South Dakota
New Jersey
Delaware
South Carolina
District of Columbia
Alaska
Oklahoma
Alabama
Indiana
Arkansas
Arizona
West Virginia
Maine
Kentucky
Idaho
Hawaii
North Dakota
Mississippi
Utah
Midwest
Rhode Island
Montana
Northeast
New Hampshire
Nevada
Vermont
South
Wyoming
1e+02
1e+05
1e+08
1e+11
Number of public isolates
KPC present
Yes
No
Unknown

1063

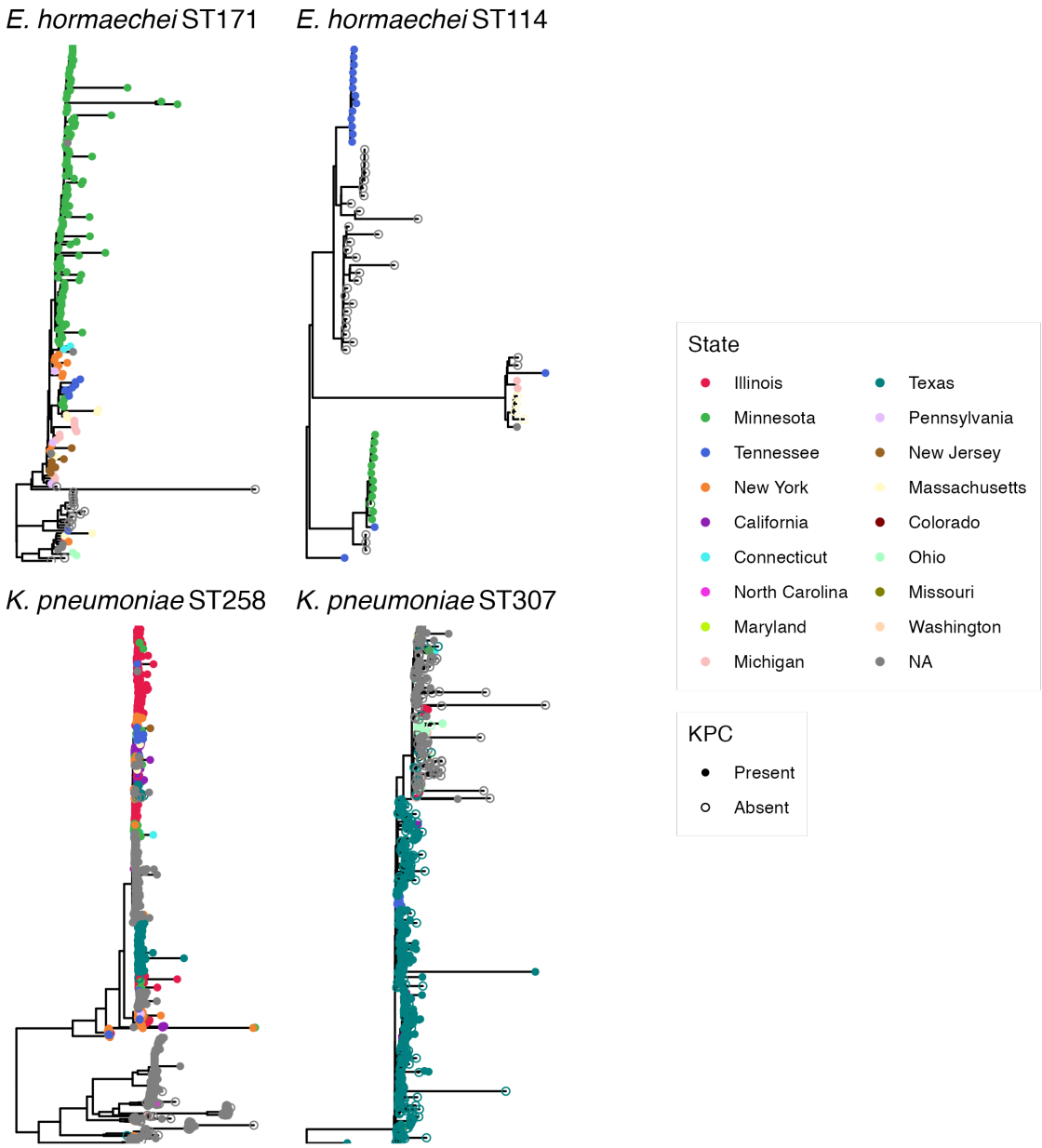
1064 **Figure S2: Information about public isolates downloaded from the PATRIC database that**

1065 **were included for context in the investigation of importation and horizontal gene transfer.**

1066 (A) Distribution of year of isolation for isolates where this information was provided. (B)

1067 Distribution of location and KPC presence/absence.

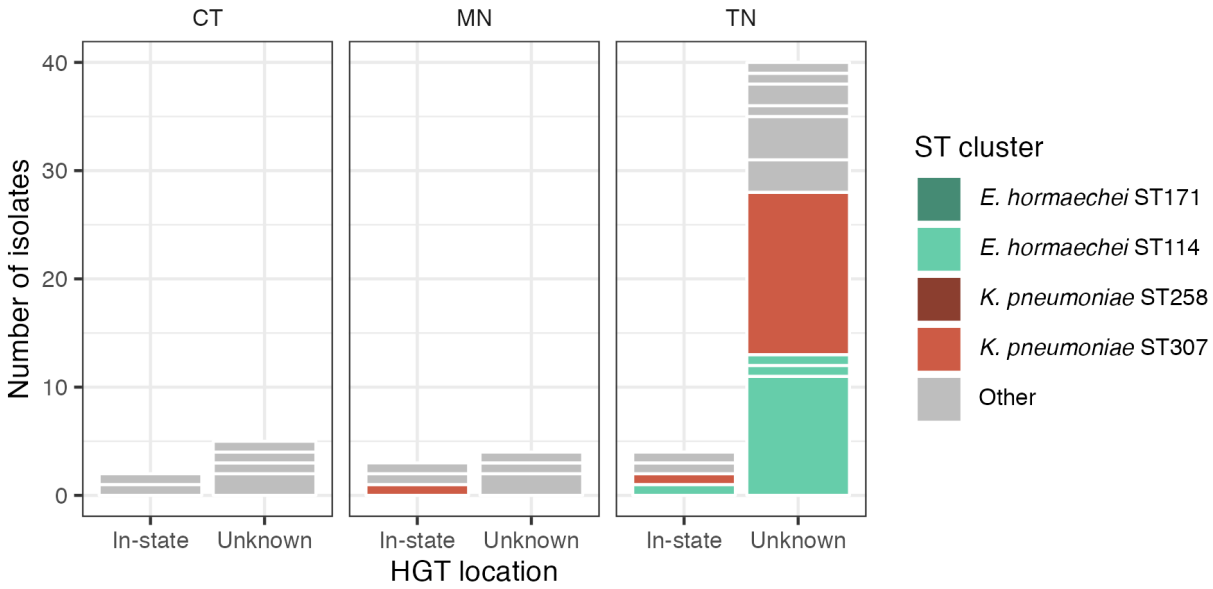
1068



1069

1070 **Figure S3: Phylogenies used to identify importation and HGT events.**

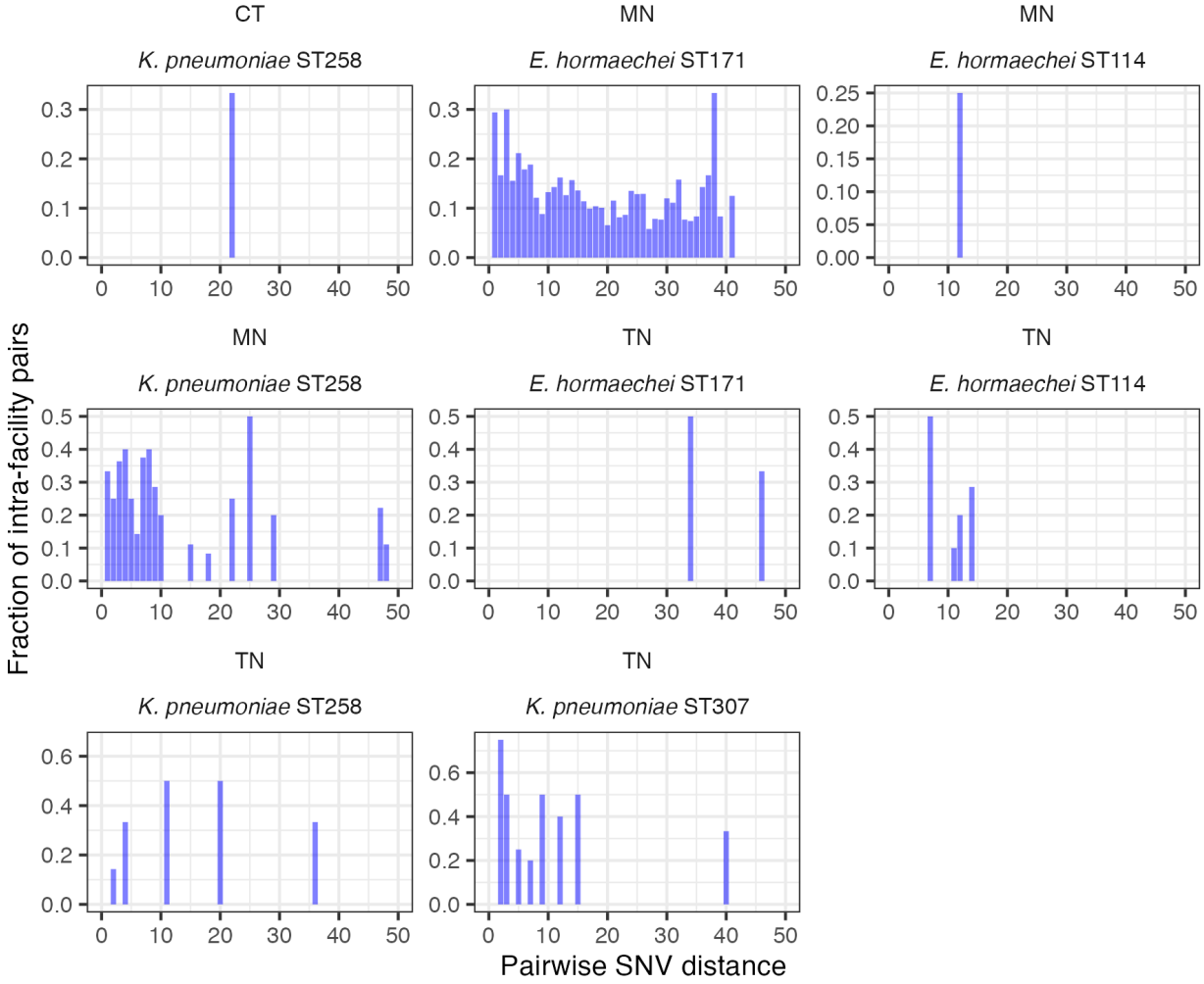
1071



1072

1073 **Figure S4: HGT locations.** In-state means that the KPC- ancestor was in the same state,

1074 unknown means the KPC- ancestor may have been in-state or out-of-state.



1075

1076 **Figure S5: Fraction of intra-facility isolate pairs for pairwise SNV distances to help identify**

1077 **pairwise SNV distance cutoffs.** We expect there to be an enrichment in closely related intra-

1078 facility pairs under the assumption that intra-facility transmission is more common than inter-

1079 facility transmission. This plot helps us identify potential pairwise SNV distance thresholds by

1080 looking for a SNV distance where there is a decrease in the fraction of intra-facility isolate pairs

1081 (around 5 and 10 appear to be reasonable thresholds for a sensitivity analysis). SNV = single

1082 nucleotide variant.

1083

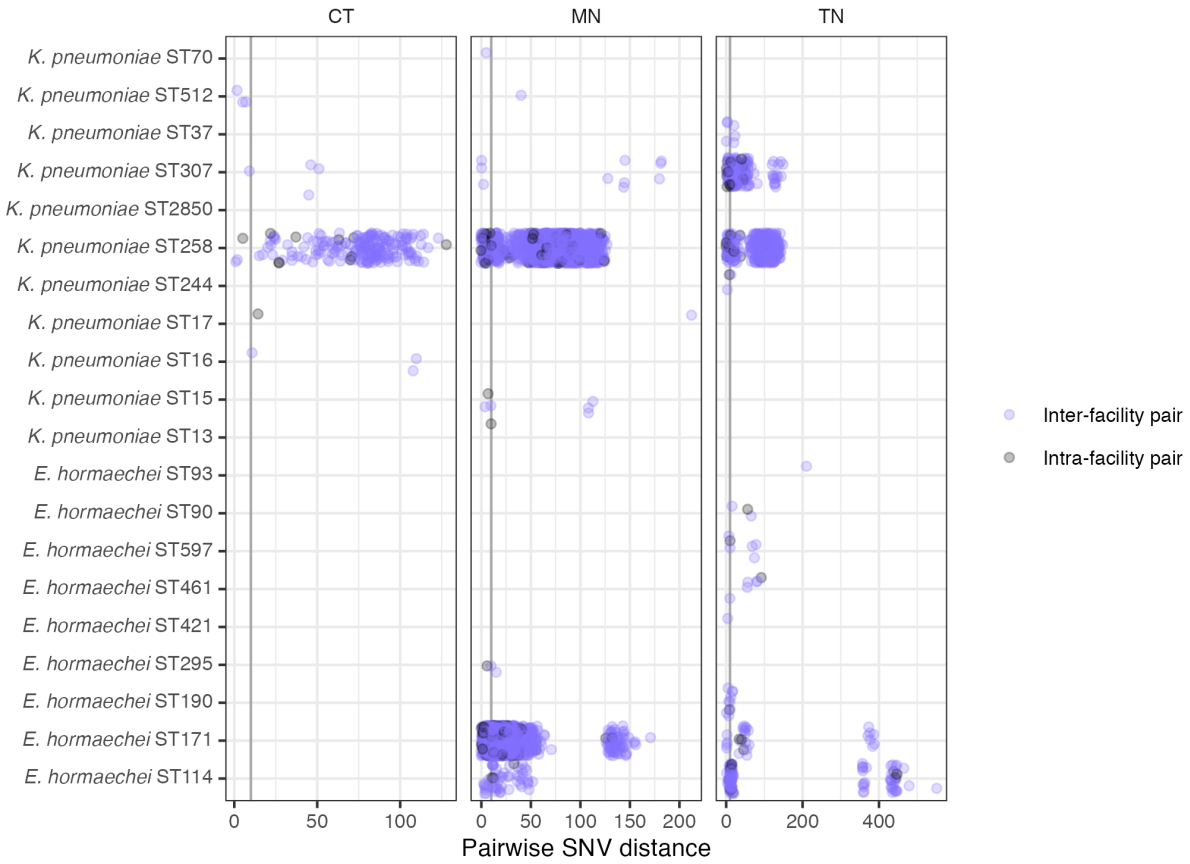


Figure S6: Pairwise single nucleotide variant (SNV) distances of all STs. SNV = single nucleotide variant. Vertical line in each plot represents 10 SNVs, used as a cutoff for determining what isolates are considered closely related (note that we also did sensitivity analyses with ≤ 5, 10, and 15 SNVs).

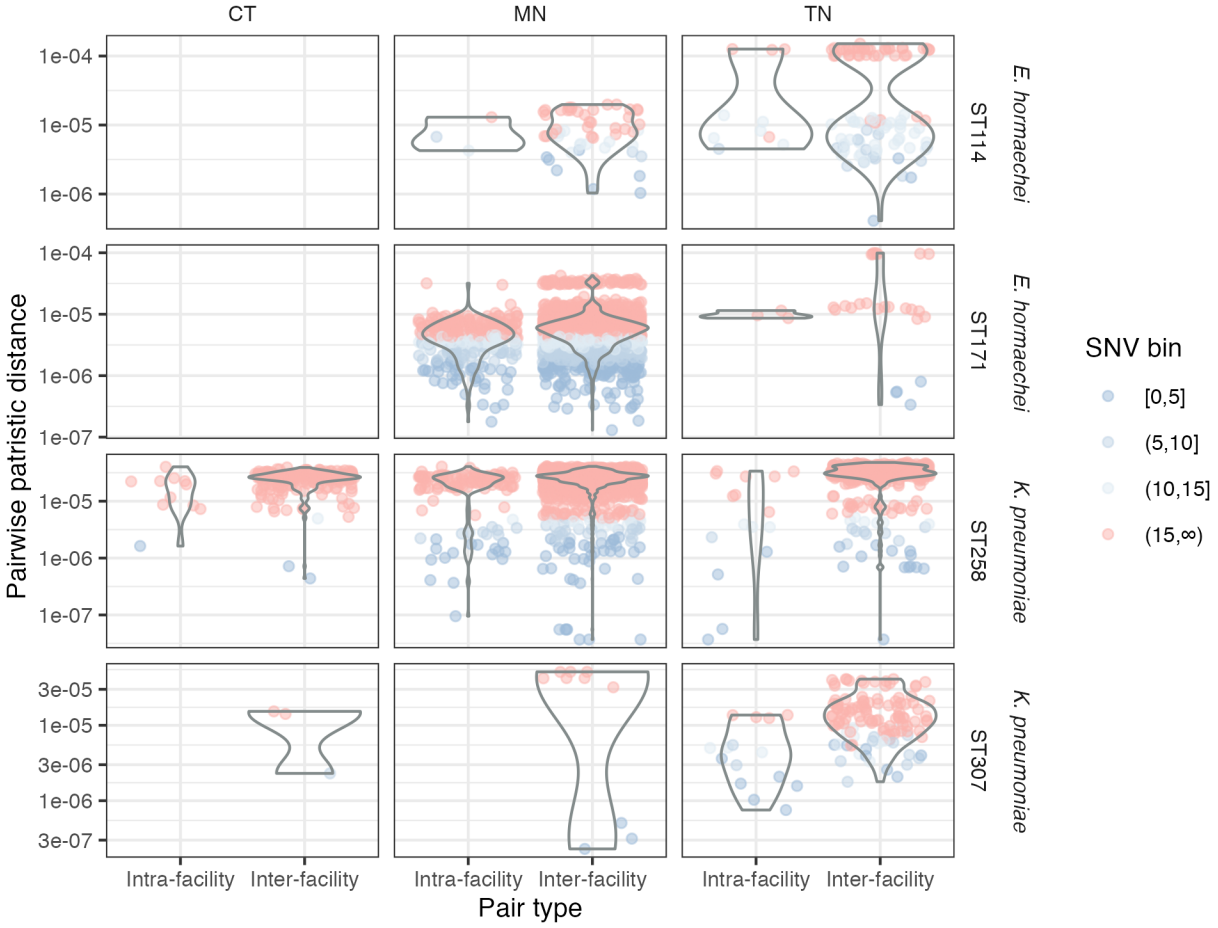
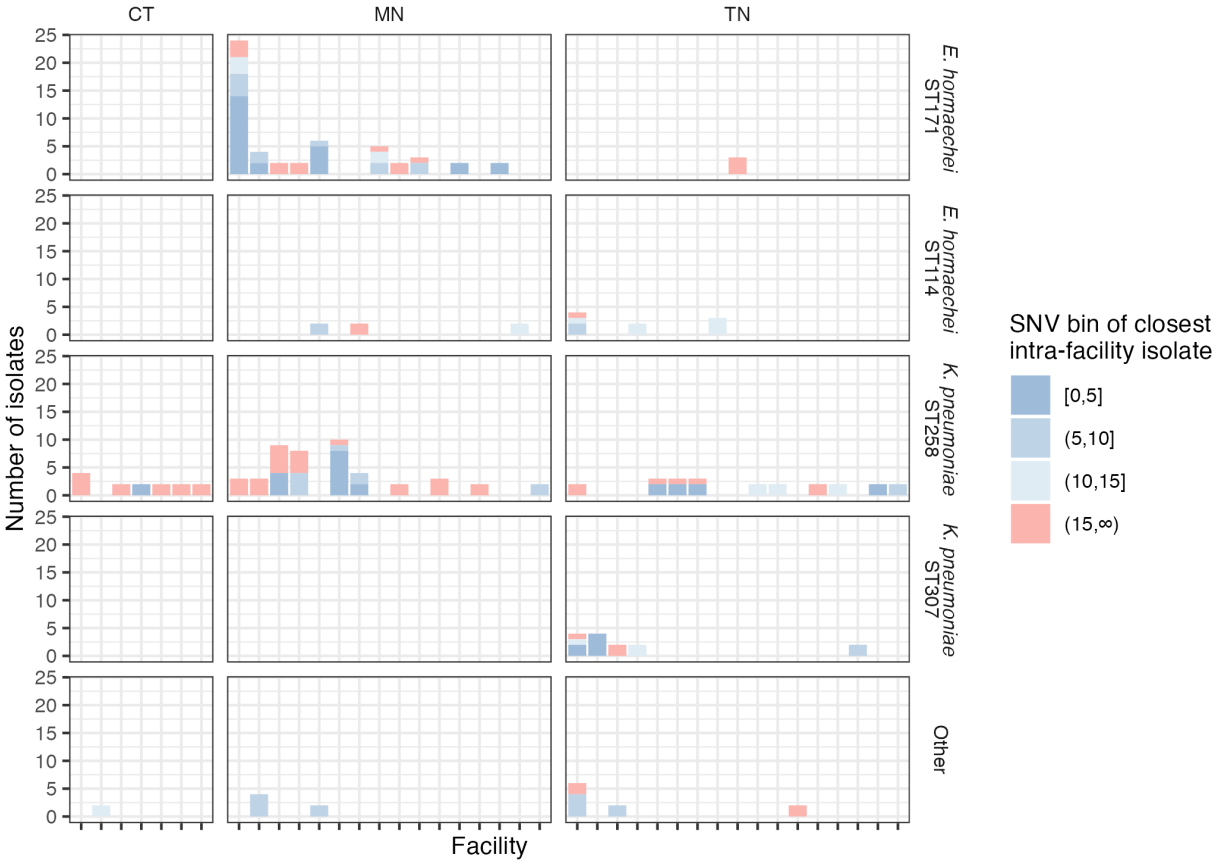


Figure S7: Pairwise patristic distances concordant with pairwise SNV distance thresholds.



1093

1094 **Figure S8: The relatedness of isolates within facilities varies.** For a given facility, sometimes

1095 the isolates are relatively closely related, while sometimes they are distantly related. SNV =

1096 single nucleotide variant.

1097

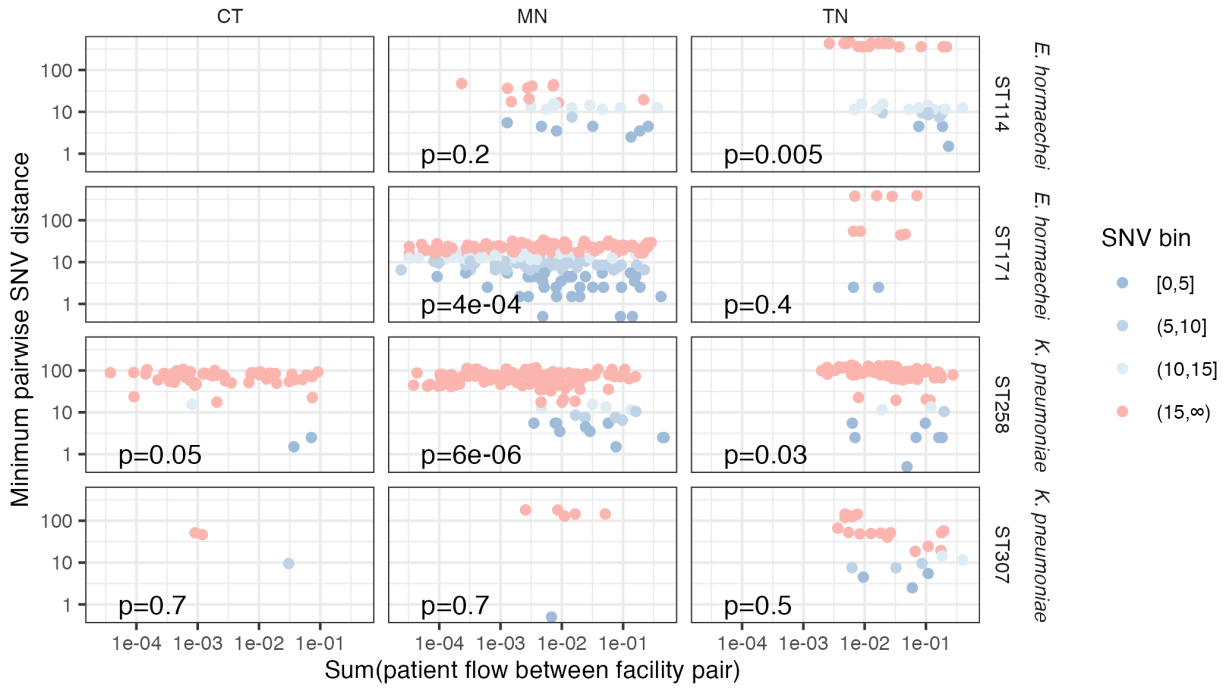


Figure S9: Aggregate-level patient transfer data provides insight into pathways of transmission. Each point is a facility pair. Patient flow was calculated using the get_patient_flow function in regentrans and is a metric of the extent of patient transfer between facilities that considers patient transfer to intermediate facilities along the path between two facilities in the patient transfer network. Both the y and x axis are shown on a log10 scale and points are jittered for visualization purposes. Wilcox p-values were calculated by comparing patient flow between facility pairs with at least one isolate pair ≤10 SNVs apart and facility pairs with no isolate pairs ≤10 SNVs apart. For most sequence types, facility pairs with closely related isolates tend to have higher patient flow between them.

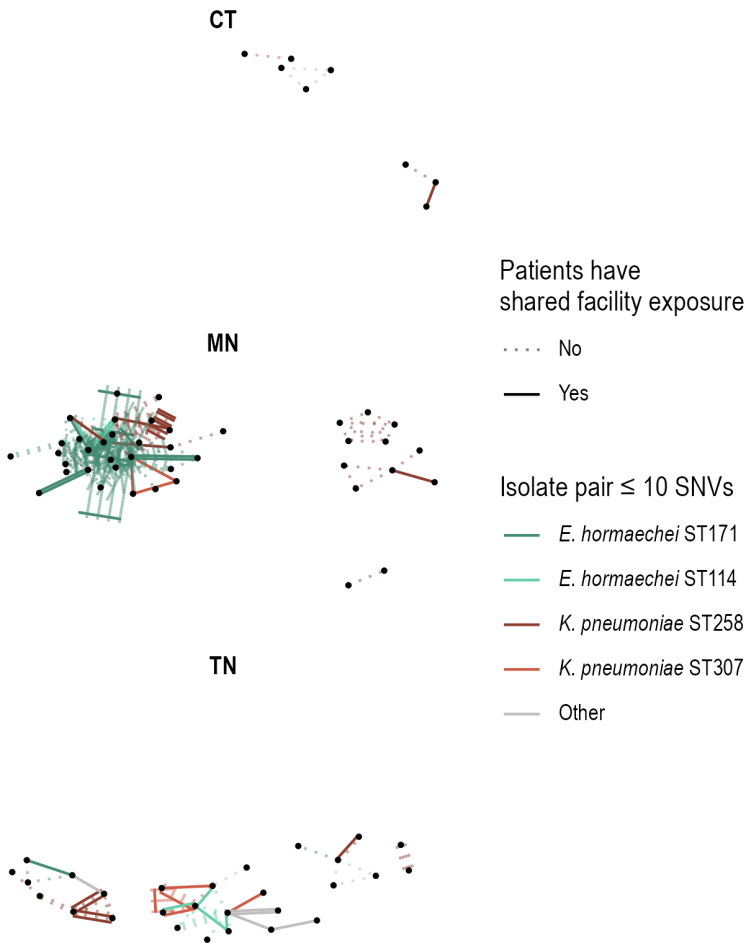


Figure S10: Patients from different facilities with closely related isolates are sometimes linked by a shared facility exposure. Nodes are facilities, line width indicates number of isolate pairs ≤10 SNVs shared by the two facilities. SNV = single nucleotide variant.

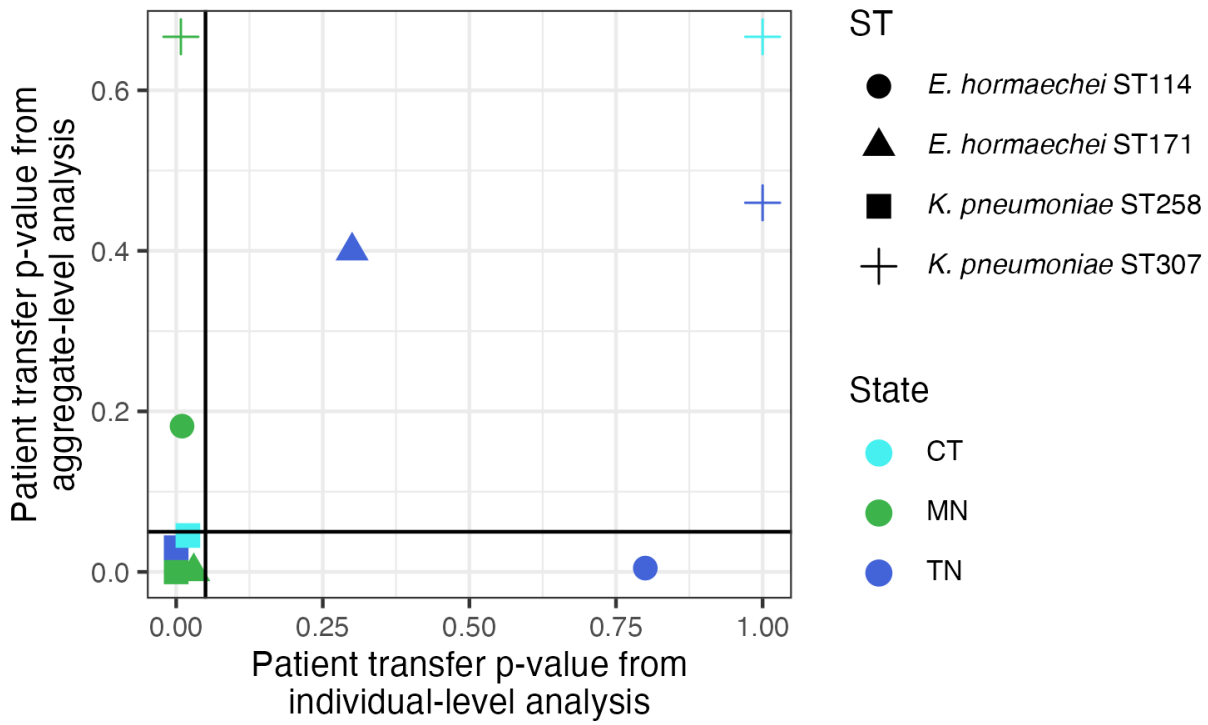
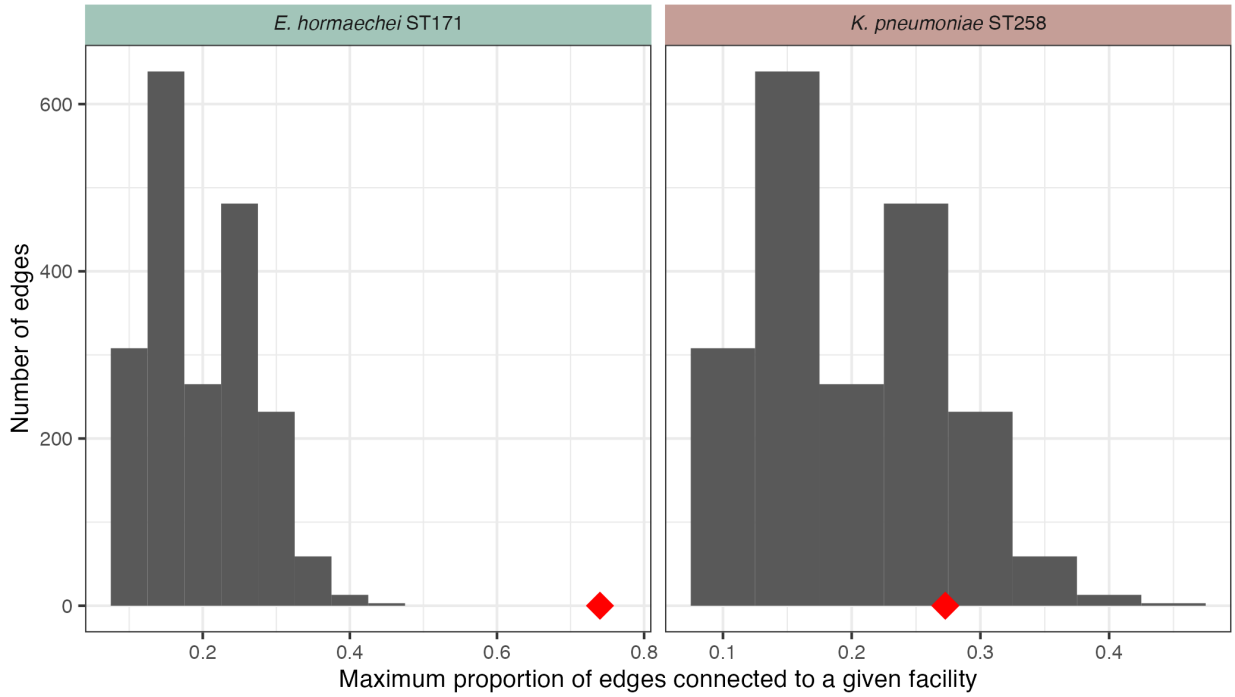


Figure S11: Differences between individual- and aggregate-level patient transfer analysis results.



1121

1122 **Figure S12: Comparison of actual and randomized distributions of facilities with closely**

1123 **related inter-facility isolate pairs.** See supplemental material for details.

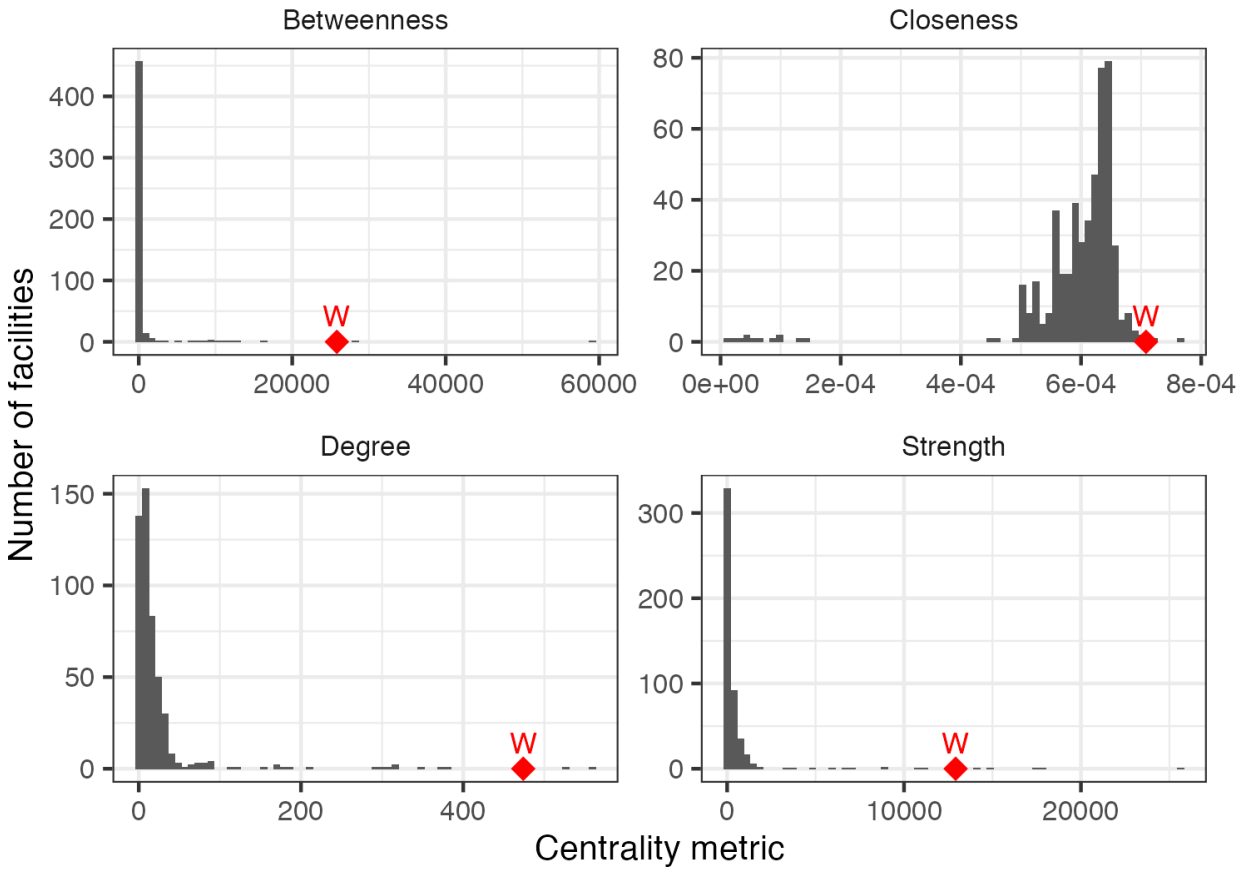


Figure S13: Facility W is more connected in the MN healthcare patient transfer network than most other healthcare facilities. Facilities are weighted by the number of patient transfers, generating a directed graph. Facility W is shown as the red diamond. For closeness, degree, and strength the in- and out- values were summed (from and to a facility).