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, 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

43 **Abstract**

44 **Background**

45 Carbapenem-resistant Enterobacterales (CRE) are among the most concerning antibiotic
46 resistance threats due to high rates of multidrug resistance, transmissibility in healthcare settings,
47 and high mortality rates. We evaluated the potential for regional genomic surveillance to track
48 *bla*$_{KPC}$-carrying CRE (KPC-CRE) transmission across healthcare facilities in three U.S. states.

49 **Methods**

50 Clinical isolates were collected from Connecticut (CT; 2017-2018), Minnesota (MN; 2012-
51 2018), and Tennessee (TN; 2016-2017) through the U.S. Centers for Disease Control and
52 Prevention's Multi-site Gram-negative Surveillance Initiative and additional surveillance. KPC-
53 CRE isolates were whole-genome sequenced, and case report data on patient comorbidities,
54 healthcare utilization, and interfacility patient transfer were extracted.

55 **Findings**

56 In CT, most KPC-CRE isolates showed evidence of importation from outside the state, with
57 limited local transmission. In MN, cases were mainly from sporadic importation and
58 transmission of *bla*$_{KPC}$-carrying *Klebsiella pneumoniae* (KPC-Kp) ST258, and clonal expansion
59 of an imported epidemic lineage of *bla*$_{KPC}$-carrying *Enterobacter hormaechei* (KPC-Ec) ST171
60 primarily at a single focal facility and its satellite facilities. In TN, KPC-Kp ST258, and more
61 recently emerged KPC-Kp ST307 and KPC-Eh ST114 were most common, with largely non-
62 overlapping facility networks mediating the spread of ST258 versus ST307 and ST114.

63 **Conclusions**

64 The underlying processes driving KPC-CRE burden can differ substantially across regions, and
65 different STs can spread via distinct pathways within a region. Integrating genomic and

66 epidemiological data from regional surveillance, and information on interfacility patient

67 transfers, can provide insights to target interventions.

68

69 **Introduction**

70 Carbapenem-resistant Enterobacterales (CRE) accounted for an estimated 13,100 infections
71 among hospitalized patients in the United States (U.S.) in 2017 [1], and likely colonized upwards
72 of 100,000 additional patients [2]. Infections with CRE are an urgent public health threat as they
73 can be difficult to treat due to resistance to carbapenems and other last-resort antibiotics [3,4]. Of
74 concern are CRE containing a carbapenemase gene found on a mobile genetic element, which
75 can disseminate within and between different species via horizontal gene transfer (HGT)
76 followed by subsequent clonal dissemination [5]. To implement interventions that effectively
77 reduce CRE infections and transmission requires not only monitoring regional CRE infections,
78 but also understanding where they were acquired. In particular, because of the capacity for CRE
79 to colonize patients for months and even years, the facility where a patient developed an
80 infection may not be where they acquired it. As active surveillance for CRE colonization at a
81 regional level is logistically infeasible, strategies are needed to leverage clinical isolate
82 collections to discern the origin of patient's CRE infections.

83

84 Understanding whether and how a hospitalized patient's CRE strain is related to previous cases
85 is critical for effective and efficient surveillance and intervention strategies. The CRE burden in
86 some regions appears to be driven by the evolution or importation of lineages with epidemic
87 potential, such as bla_{KPC}-carrying *Klebsiella pneumoniae* (KPC-Kp) ST258 (e.g. in California)
88 [5] and *Enterobacter hormaechei* (KPC-Eh) ST171 (e.g. in Minnesota) [6] in the U.S. Following
89 initial regional emergence, these strains may be transmitted within healthcare facilities and
90 spread between facilities via transfer of colonized patients [5,7,8]. In regions with sustained
91 transmission of epidemic clones, effective regional control requires identification of locations

92 where transmission is occurring and impacting the burden in the region, and monitoring the
93 movement of CRE carriers between healthcare facilities [9,10]. In contrast, others have observed
94 a role for HGT of mobile carbapenemase elements in driving CRE spread within individual
95 healthcare facilities [11–13]. If HGT is a significant contributor to overall regional CRE burden,
96 effective prevention and control requires identification of reservoirs of these resistance elements
97 and an understanding of the propensity of different strain/mobile element combinations to spread
98 between patients or act as HGT donors.
99
100 Here, we evaluate the potential for passive regional genomic surveillance to inform the pathways
101 leading to CRE cases across regional healthcare networks in three U.S. states involved in the
102 U.S. Centers for Disease Control and Prevention's (CDC) Multi-site Gram-negative Surveillance
103 Initiative (MuGSI). In particular, we focus on the most commonly observed CRE in the MuGSI
104 catchment areas, KPC-Kp and KPC-Ec. This builds off of the MuGSI aim to quantify the burden
105 of certain resistant gram-negative bacteria in the U.S. by supplementing preexisting data with
106 whole-genome sequencing (WGS) of CRE isolates collected through MuGSI and supplemental
107 state-wide surveillance. By integrating genomic data from densely sampled CRE cases and
108 healthcare exposure data, and comparing local strains with global isolates, we gained insight into
109 local strain origins and regional transmission pathways.
110
111 **Methods**
112 This retrospective study was a collaboration between the Emerging Infections Program (EIP) and
113 CDC Prevention Epicenters. The EIP is a collaboration between 10 state health departments and
114 their partnering academic institutions and the Centers for Disease Control and Prevention and

115 other federal agencies [14]. The Healthcare-Associated Infections Community Interface (HAIC)
116 is one of the core components of the EIP and conducts active population- and laboratory-based
117 surveillance for CRE [15]. Three EIP sites participated in this study: Connecticut (CT),
118 Minnesota (MN), and Tennessee (TN).
119
120 CRE surveillance
121 While only selected counties in MN and TN participated in the HAIC MuGSI [16], each of the
122 participating states conducted active statewide public health surveillance for CRE. All
123 carbapenem-resistant isolates were submitted by clinical laboratories to each state's respective
124 State Public Health Laboratory (SPHL; see Supplemental Methods). Multiple isolates from the
125 same patient were obtained if the new isolate was from at least 30 days after a previous case
126 isolate was identified.
127
128 We define CRE as any organism in the Enterobacterales order isolated from any clinical
129 specimen with resistance to doripenem, meropenem, or imipenem (minimum inhibitory
130 concentrations [MIC] of ≥4 μg/ml); with resistance to ertapenem (MIC ≥2 μg/ml); or with
131 demonstrated production of a carbapenemase [16]. All CRE isolates identified were tested for the
132 carbapenemase genes bla_{KPC}, bla_{NDM}, bla_{OXA-48}, bla_{VIM}, and bla_{IMP} (see Supplemental Methods).
133
134 CRE considered for WGS were collected in CT from 2017-2018, in MN from 2012-2018, and in
135 TN from 2016-2017. Unless otherwise noted, only isolates from 2016-2018 were analyzed to
136 enable more meaningful comparison of analyses across states. For CT and MN, isolates from the
137 entire state were included. In TN, only isolates from TN Emergency Medical Services (EMS)

138 Regions 2-7 were included (76.6% of the population). We excluded counties in the Memphis-
139 Delta and Northeast Tennessee Region due to significant healthcare utilization across state lines.
140 **Table S1** provides information on the number of each type of healthcare facility in each state and
141 the number of each type for which we have at least one sequenced isolate.
142
143 Epidemiologic data
144 Epidemiologic data were obtained through medical record review and in TN supplemented by
145 additional data sources (see Supplemental Methods). Isolate metadata includes state, treatment
146 facility ID, patient ID, age, sex, culture source, and hospitalizations in the previous year. MuGSI
147 surveillance data also included information on infection type and underlying conditions, which
148 was only obtained for selected non-MuGSI isolates. The only time we used this additional data
149 was to investigate shared underlying conditions; this analysis was limited to samples for which
150 we had data on these conditions. All other analyses included all isolates.
151
152 Generation of aggregate patient transfer networks
153 Aggregate patient transfer networks (i.e., patient flow across healthcare facilities [5]) for each
154 state for 2017 were derived from CMS fee-for-service beneficiary claims data linked to the CMS
155 Minimum Data Set by Medicare beneficiary ID. The number of transfers between two facilities
156 includes transfers directly from one facility to another and transfers with an intervening stay in
157 the community of less than 365 days. See Supplemental Methods.
158
159 Isolate and genomic data processing

160 Over 90% of clinical CRE isolates considered for WGS were KPC-Kp or KPC-Eh; we therefore
161 performed Illumina WGS on only this subset of isolates (see Supplemental Methods). Study
162 isolates included isolates from BioProject numbers PRJNA272863 [6] and PRJNA873034 (new
163 from this study) (Supplemental File S1). For this set of isolates, we identified species and
164 sequence types (STs) [17,18], called single-nucleotide variants using species-specific reference
165 genomes (*K. pneumoniae*: KPNIH1, GenBank accession number CP008827.1, 5,394,056 base
166 pairs; *E. hormaechei*: MNCRE9, GenBank accession number JZDE00000000.1, 4,911,317 base
167 pairs) [19–25], built reference-based phylogenetic trees [24–28], and generated and annotated
168 assemblies [19,20,29,30] (see Supplemental Methods). These reference-based data were used for
169 all analyses unless otherwise noted. Furthermore, to prevent overcounting of intra-facility
170 transmission, isolates were limited to the first isolate per patient for each ST/facility/patient
171 combination.
172
173 Identification of putative geographic importation events
174 5,346 public KPC-positive and KPC-negative *K. pneumoniae* and *E. hormaechei* isolates were
175 downloaded from the PATRIC database on 2021-04-23 [18,31] and used to contextualize study
176 isolates. In particular, we aimed to group study isolates into clusters that could be traced back to
177 common local emergence events based on phylogenetic clustering, and then used geographic
178 context to determine whether a cluster likely originated via importation of a bla_{KPC} carrying
179 strain from outside the state. For all public and study isolates, ST-specific core genome
180 phylogenies were generated [24,32,33] and bla_{KPC} was identified [31,34] (see Supplemental
181 Methods). For each core genome tree, maximum likelihood ancestral reconstruction of each
182 study site location (MN, TN, CT) and presence of bla_{KPC} was performed individually in R v4.0.2

183 [35] using the ace() function in ape v5.5 [36]. Nodes with an ancestral reconstruction confidence
184 of < 0.875 were discarded. All transition events (for each geographic location and *bla*$_{KPC}$) were
185 then mapped onto the edges of the phylogeny. Each KPC-positive study isolate tip was traversed
186 towards the root of the tree until the first transition event was identified. If the first transition
187 event was a geographic transition event with no *bla*$_{KPC}$ transition event on the same edge,
188 isolates were considered to originate from geographic importation of a KPC-positive strain. If the
189 first transition event was for *bla*$_{KPC}$, this supports local acquisition of *bla*$_{KPC}$ via HGT; therefore,
190 these isolates were presumed to not originate from importation of a KPC-positive strain. If a
191 geographic and *bla*$_{KPC}$ transition occurred on the same edge, or if transition edges were of low
192 confidence, we considered there to be no evidence of importation of a KPC-positive strain. Study
193 isolates traced back to the same edge transition event, regardless of the type or confidence of the
194 transition, were considered part of a local cluster traced back to a common local emergence
195 event. We consider the number of transition events for a given geographic location to be the
196 number of identified clusters.
197
198 Computing pairwise distances
199 Pairwise single nucleotide variant (SNV) distances and patristic distances for study isolates from
200 each species-specific reference-based sequence alignment were calculated in R v4.0.2 [35] using
201 the dist.dna() function in ape v5.5 [36] with pairwise.deletion = TRUE. These pairwise SNV
202 distances were classified as intra- and inter-facility pairs using the get_pair_types() function in
203 regentrans v0.1 [37]. Rather than choosing a specific pairwise SNV distance threshold, we
204 highlight various thresholds (≤ 5, 10, 15) as a sensitivity analysis. A threshold of 10 was chosen

205 where needed based on a decrease in the fraction of intra-facility pairs at various SNV distances
206 (**Figure S3A**).
207
208 Integrated genomic and epidemiologic analyses
209 *Relatedness of intra- vs. inter-facility isolates*
210 For each isolate with at least one intra-facility isolate pair and at least one inter-facility isolate
211 pair, we identified the minimum pairwise SNV distance of all intra-facility pairs and the pairwise
212 minimum pairwise SNV distance of all inter-facility isolate pairs. We then compared the
213 pairwise SNV distances of intra- vs. inter-facility isolate pairs using a paired Wilcox test.
214
215 *Intra-facility isolate pair shared comorbidity analysis*
216 For each intra-facility isolate pair, we determined whether they had at least one shared
217 comorbidity and used a Wilcox test to compare pairwise SNV distances of patients who shared
218 comorbidities and those who did not. This was done based on the assumption that patients with
219 shared comorbidities would be provided care in more proximate inpatient locations.
220
221 *Inter-facility isolate pair shared prior healthcare exposure analysis*
222 For each inter-facility patient pair, we determined whether they had a shared healthcare exposure
223 in the past year (data collection method described in the Supplemental Methods). Then we used a
224 Wilcox test compared the pairwise SNV distances of patients who shared a healthcare exposure
225 and those who did not.
226
227 Data analysis and visualization

228 R v4.1.1 was used for all data analysis and visualization using these packages: tidyverse v1.3.1
229 [39], ggtree v2.5.2 [40,41], igraph v1.2.6 [42], regentrans v0.1 [37], tidygraph v1.2.0 [43],
230 ggraph v2.0.5 [44], exact2x2 v1.6.5 [38], cowplot v1.1.1 [45], readxl v1.3.1 [46], ape v5.5 [36].
231 The code used to perform analyses is here: https://github.com/Snitkin-Lab-Umich/eip-cre-
232 transmission-ms.
233
234 Human subjects research
235 The study was reviewed and approved by the Tennessee Department of Health institutional
236 review board (IRB). See 45 CFR. part 46; 21 CFR part 56. The Minnesota and Connecticut
237 Department of Health IRB determined that the study was exempt from IRB review in accordance
238 with 45 CFR 46.101(b)(4). The study was approved by the CDC IRB with a waiver of HIPAA
239 authorization under the Privacy Rule as per 45 CFR 46.512(i). The University of Michigan
240 Medical School IRB approved this protocol.
241
242 **Results**
243 Over 90% of collected clinical CRE isolates were KPC-Kp or KPC-Eh; we therefore focus our
244 analysis on this subset of isolates to understand the pathways leading to the dominant causes of
245 infections. We included the first isolate of each unique ST/facility/patient combination. In each
246 state, over 50 KPC-Kp and KPC-Eh isolates were present across over 20 distinct healthcare
247 facilities (**Table S1**; **Figure S1**). The dominant STs were KPC-Kp ST258 in all three states,
248 KPC-Eh ST171 in MN, and KPC-Eh ST114 and KPC-Kp ST307 in TN.
249
250 **Clonal dissemination after importation or *bla*$_{KPC}$ acquisition differs across states and STs**

251 To gain insight into the origin and magnitude of spread of different KPC-Kp and KPC-Eh strains
252 in each state we performed phylogenetic analyses including 5,346 publicly available
253 *Enterobacter* spp. and *Klebsiella* spp. genomes deposited in the PATRIC database (**Figures**
254 **S2A-B**). We sought to use the genetic and geographic context provided by PATRIC isolates to
255 identify phylogenetic clusters including only isolates from a given state (**Figure 1**; **Figure S2C**).
256 Despite what we assume to be relatively sparse sampling of circulating *Klebsiella* and
257 *Enterobacter* strains in PATRIC, phylogenetic context revealed that KPC-Kp and KPC-Eh cases
258 in all three states were attributable to large numbers of independent regional emergence events.
259 In CT and MN, large phylogenetic clusters consistent with local dissemination were almost
260 exclusively restricted to those that could be traced back to the outside importation of the
261 epidemic lineages KPC-Kp ST258 and KPC-Eh ST171. In TN, we also detected evidence for
262 local dissemination of imported KPC-Kp ST258, and to a lesser extent KPC-Eh ST171.
263 However, in contrast to CT and MN, we also found large phylogenetic clusters that could not be
264 traced to an external source. Most notable were large clusters of the emerging epidemic clones
265 KPC-Kp ST307 [48] and KPC-Eh ST114 [49]. Support for these phylogenetic clusters being due
266 to local acquisition of bla_{KPC} comes from the nearest neighbors of these strains in public
267 collections being KPC-negative (**Figure S2C**).
268
269 **Genomic analysis of regional clinical isolate collections allowed for the detection of local**
270 **transmission in MN and TN, and lack thereof in CT**
271 Having characterized the origin of circulating strains, we next sought to understand whether
272 regional genomic surveillance of clinical isolates was dense enough to discern facilities and
273 regional subnetworks where transmission was occurring. To this end, we first investigated the

274 extent to which the surveillance isolates captured putative recent local transmission events,
275 defined as small pairwise SNV distances (**Figures S3B**) or small patristic distances (**Figure**
276 **S3C**). Within an ST, smaller pairwise SNV distances indicate potential local transmission, while
277 larger distances are not indicative of local transmission [50]. We captured very few local
278 transmission events in CT, suggesting that importation from other geographic regions may be
279 driving the case load there. In contrast, we captured likely local transmission events of KPC-Eh
280 ST171 in MN and of KPC-Kp ST258, KPC-Kp ST307 and KPC-Eh ST114 in TN, consistent
281 with our observation of large phylogenetic clusters for these STs (**Figure 1**).
282
283 Patients with closely related intra-facility isolate pairs did not share any underlying conditions
284 that may be indicative of an outbreak related to a certain location or procedure in the hospital (all
285 $p > 0.05$). In addition, we observed that isolates close in pairwise SNV distance to another isolate
286 from the same facility were also often closely related to an isolate from a different facility
287 (**Figure 2A).** In fact, for ST258 and ST114 in TN, isolates tended to have closer genetic
288 neighbors at different facilities versus the same facility ($p < 0.05$). These findings suggest that
289 observation of closely related isolates from within a facility does not directly equate to an intra-
290 facility transmission event having occurred, but may also reflect recent transmission at a
291 connected facility. Supporting the notion that we observed signatures of inter-facility
292 transmission is that patients from different facilities with a common prior healthcare exposure
293 tended to have more closely related isolates than those with no common prior healthcare
294 exposure (**Figure 2B**).
295

296 **Longitudinal genomic surveillance in MN revealed limited spread of ST258 and localized**
297 **transmission of ST171 in a regional sub-network**
298 With evidence for isolate pairs with small SNV distances being informative of recent intra- and
299 inter-facility transmission, we next set out to look at genomic linkages across MN and TN to
300 gain insight into where transmission of different STs is occurring. MN has been collecting CRE
301 isolates since 2012. We observed a decrease in the amount of KPC-Eh ST171 and KPC-Kp
302 ST258 (**Figure S4A**) and corresponding signatures of transmission over time (**Figure S4B**).
303 Incorporating information from this extended dataset to investigate transmission, we observed
304 stark contrasts in the relatedness of ST171 compared to ST258 isolates across time. ST258
305 isolates collected over one year apart were rarely closely related to each other; however, ST171
306 isolates collected even over 5 years apart were often quite closely related (**Figure 3A**). We
307 investigated whether this difference was driven by certain facilities, and found that one facility,
308 F38, was highly represented in closely related ST171 intra- (**Figure 3B**) and inter-facility
309 (**Figure 3C**) isolate pairs. Of all ST171 isolate pairs within 10 SNVs, 66/86 (77%) intra-facility
310 pairs were from F38 and 240/430 (56%) inter-facility pairs contained at least one isolate from
311 F38. About half of these inter-facility pairs were connected to one of three other facilities
312 (**Figure 3D**). F38 is highly connected to other facilities in the patient transfer network (**Figure**
313 **S3E**) and is well-connected by patient transfer to the other three highly represented facilities
314 (**Figure S3F**; $\geq 90^{th}$ percentile of all facility pairs). These findings indicate that the pattern by
315 which ST171 and ST258 disseminated, and the potential role of specific facilities, is very
316 different.
317
318 **Distinct facility sub-networks in TN harbor different lineages**

319 We next explored whether we could discern where transmission was occurring in TN by
320 investigating what facilities patients with each ST had been to in the previous year. We observed
321 a cluster of seven facilities where patients with ST114, ST307, and as STs categorized as Other,
322 had been in the previous year (**Figure 4A**). Visualizing these seven facilities in the context of the
323 aggregate patient transfer network revealed that they are closely connected (**Figure 4B**) and have
324 more patient sharing among themselves than with other facilities (**Figure 4C**; Wilcox $p < 0.001$).
325 Additionally, this facility cluster is distinct from a cluster containing predominately ST258
326 (**Figure 4B**).
327
328 **Discussion**
329 As public health laboratories increase their capacity for genomic surveillance, it is critical to
330 understand if and how genomic data from passive surveillance can be used to guide regional
331 intervention efforts. Statewide active surveillance is infeasible, and while that would be required
332 to infer who transmitted to whom, passive surveillance may be enough to provide practical
333 guidance for interventions including if and where transmission is happening and if there are any
334 emerging threats. Here, we evaluated whether genomic analysis of clinical bla_{KPC}-positive CRE
335 isolates from three U.S. states could inform our understanding of the origins and transmission
336 pathways of circulating strains. We could discern the relative contributions of importation and
337 clonal dissemination to each state's KPC-Kp and KPC-Eh case burden and highlight individual
338 facilities and connected regional sub-networks where putative transmission occurred. These
339 insights did not rely on detailed clinical metadata, indicating that other regions interested in
340 identifying hotspots and pathways for regional spread may be able to do so with only knowledge
341 of recent healthcare exposures. Our findings suggest that states should consider prioritizing

342 genomic analysis of clinical isolates to monitor transmission hotspots as insights gleaned from
343 these analyses may enable targeting of regional infection prevention efforts to certain facilities.
344
345 Despite focusing on the same two Enterobacterales species harboring the same carbapenemase
346 gene, we observed stark differences in the underlying drivers of each state's KPC-Kp and KPC-
347 Eh burden. In CT, strains were frequently imported but did not show evidence of onward
348 transmission. Importation is unsurprising, given the state's proximity to other major CRE
349 hotspots in the Northeast U.S. [51]. However, this still leaves unanswered why imported bla_{KPC}-
350 positive CRE have apparently not gained a foothold. One reason may be that the structure of the
351 CT regional healthcare network decreases local transmission. Recent work showed that post-
352 acute care settings can act as initiators and amplifiers of regional CRE epidemics [52,53]. Thus,
353 infection prevention practices in these facilities, and their patient transfer connections with other
354 facilities, may have large impacts on overall regional CRE spread. Regardless, our observations
355 suggest that focusing resources on testing admission cultures from out-of-state may allow CT to
356 continue having a low burden of these CRE strains.
357
358 In MN, bla_{KPC}-positive CRE burden was driven by a major importation event of the epidemic
359 lineage KPC-Eh ST171 followed by dissemination within the region. Incorporation of historical
360 data revealed that isolates far apart in time were often closely related and were largely from or
361 connected to isolates from a single hub facility with many referrals that provides specialty care,
362 F38. These observations could indicate an environmental reservoir within or proximate to F38,
363 or undetected endemic spread at one or more referral facilities. Having this data in real-time
364 generates actionable hypotheses to potentially stem regional spread. F38 could prioritize testing

365 patients admitted from facilities connected by extensive patient transfer. Similarly, state
366 epidemiologists could investigate facilities with large numbers of genomic linkages to isolates
367 from F38, such as the three identified here, to detect other potential reservoirs and locations
368 where intervention may have a broad positive impact on regional KPC-Eh burden.
369
370 TN showed a diversity in circulating strains. In addition to a significant burden of ST258, we
371 observed CRE isolates from two more recently emerged lineages – KPC-Kp ST307 and KPC-Eh
372 ST114 – and less-common STs that appear to have emerged and spread more recently. This
373 pattern is in stark contrast with CT and MN, where there were very few cases of onward
374 transmission detected after a non-epidemic strain acquired bla_{KPC}. Furthermore, these recently
375 emerged lineages traversed a distinct sub-network of healthcare facilities relative to KPC-Kp
376 ST258. These observations suggest that the propensity for potential HGT of bla_{KPC} into
377 emerging and novel receptive strains, followed by clonal dissemination, appears to be higher in a
378 subset of TN healthcare facilities. While we were unable to discern specific HGT events in this
379 study, nor why these facilities may be hotspots for successful HGT, our observations do suggest
380 that the subnetwork observed here may be an important location to monitor for emerging threats.
381
382 Our study has several limitations. First, our use of only clinical isolates may lead to an
383 underestimation of clonal dissemination, that may be different for different STs depending on
384 their virulence and associated colonization-to-infection ratio (i.e. ST-specific iceberg effect).
385 However, we were still able to uncover patterns of clonal dissemination using these data,
386 suggesting that regional clinical isolate collections can provide actionable insights for public
387 health labs. While transmission may not have occurred at the facilities where clinical isolates

388 were collected, this nevertheless provides a location at which to initiate an investigation and
389 could ultimately guide more targeted surveillance efforts. Second, the isolate collection periods
390 were different for each state. This may result in different levels of contextual information
391 provided by public databases, through which importation was inferred. However, our
392 observations are consistent with prior studies in that KPC is known to be stably associated with
393 ST258, ST171 emerged in the upper Midwest, and ST307 and ST114 only recently became
394 associated with KPC. Another limitation is that the public isolates used from PATRIC to inform
395 the origin of locally circulating strains are not equally representative of all geographic regions,
396 which may underestimate the extent of importation, as multiple importations from an unsampled
397 reservoir may be merged. While the analyses performed here would be enhanced by additional
398 context, we note that inferences made into importation using phylogenetic clustering are largely
399 consistent with the SNV distance-based analyses, supporting the robustness of our conclusions.
400 Furthermore, the aggregate patient transfer data used to understand the connectivity between
401 regional healthcare facilities was derived from Medicare patient transfers, which may lead to
402 biases in patient transfer connections. However, despite this, we were able to discern local sub-
403 networks of healthcare facilities with distinct transmission patterns.
404
405 In conclusion, our results suggest that, not only can the origins and transmission patterns of
406 bla_{KPC}-positive CRE be investigated using genomic surveillance of sufficiently comprehensive
407 regional clinical isolate collections, but also that these transmission dynamics can vary across
408 strains and regions. The differences underlying the KPC-Kp and KPC-Eh burden in these three
409 states would not have been discernible without genomic data, and in fact would be strengthened
410 further by the existence of larger genomic surveillance initiatives that could provide more

411 granular context for the origin of strains and their mobile elements to guide prevention efforts.
412 Similarly, our understanding of how KPC-Kp and KPC-Eh proliferate in a region would be
413 limited without knowledge of patient healthcare exposures and transfer networks, which enable a
414 more nuanced understanding of where transmission is occurring, where reservoirs might exist,
415 and where additional surveillance or intervention is required. Taken together, these observations
416 support the value of robust regional genomic surveillance for antibiotic resistance threats and the
417 need for analytic platforms capable of integrating genomic and patient movement data to guide
418 local and state infection prevention efforts.
419

420 **Acknowledgements**
421 Pamela Talley and Matt Estes from the Tennessee Department of Health. The MN-PHL
422 Microbiology and Sequencing & Bioinformatics laboratories for their work in identifying and
423 sequencing all of the MN isolates in the project. Gwendolyn Hughes and Christina Sancken from
424 CDC.
425

426 **Funding**
427 This work was supported by the CDC through the Emerging Infections Program cooperative
428 agreement (grant number CK17-1701). It was also supported in part by CDC Prevention
429 Epicenter Program Cooperative Agreement No. U54CK000481-S1. ZL received support from
430 the National Science Foundation Graduate Research Fellowship Program under Grant No. DGE
431 1256260. Any opinions, findings, and conclusions or recommendations expressed in this material
432 are those of the authors and do not necessarily reflect the views of the National Science
433 Foundation.

434

435 **Conflicts of interest**

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

438

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

443

444 **References**

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

448 2. Lee BY, Bartsch SM, Wong KF, et al. Tracking the spread of carbapenem-resistant
449 Enterobacteriaceae (CRE) through clinical cultures alone underestimates the spread of CRE
450 even more than anticipated. Infection Control & Hospital Epidemiology. **2019**; 40: 731–
451 734.

452 3. World Health Organization, editor. Antimicrobial resistance: global report on surveillance.
453 Geneva, Switzerland: World Health Organization, 2014.

454 4. Potter RF, D'Souza AW, Dantas G. The rapid spread of carbapenem-resistant
455 Enterobacteriaceae. Drug Resistance Updates **2016**; 29:30–46.

456 5. Han JH, Lapp Z, Bushman F, et al. Whole-genome sequencing to identify drivers of

457 carbapenem-resistant Klebsiella pneumoniae transmission within and between regional
458 long-term acute-care hospitals. Antimicrobial agents and chemotherapy **2019**; 63:e01622-
459 19.
460 6. Hargreaves ML, Shaw KM, Dobbins G, et al. Clonal Dissemination of Enterobacter cloacae
461 Harboring blaKPC-3 in the Upper Midwestern United States. Antimicrobial Agents and
462 Chemotherapy **2015**; 59:7723–7734.
463 7. Snitkin ES, Won S, Pirani A, et al. Integrated genomic and interfacility patient-transfer data
464 reveal the transmission pathways of multidrug-resistant Klebsiella pneumoniae in a regional
465 outbreak. Science Translational Medicine **2017**; 9:eaan0093.
466 8. Lapp Z, Crawford R, Miles-Jay A, et al. Regional Spread of blaNDM-1-Containing
467 Klebsiella pneumoniae ST147 in Post-Acute Care Facilities. Clinical Infectious Diseases
468 **2021**; Available at: https://doi.org/10.1093/cid/ciab457. Accessed 15 July 2021.
469 9. Slayton RB, Toth D, Lee BY, et al. Vital Signs: Estimated Effects of a Coordinated
470 Approach for Action to Reduce Antibiotic-Resistant Infections in Health Care Facilities —
471 United States. MMWR Morb Mortal Wkly Rep **2015**; 64:826–831.
472 10. Toth DJA, Khader K, Slayton RB, et al. The Potential for Interventions in a Long-term
473 Acute Care Hospital to Reduce Transmission of Carbapenem-Resistant Enterobacteriaceae
474 in Affiliated Healthcare Facilities. Clin Infect Dis **2017**; 65:581–587.
475 11. Mathers AJ, Cox HL, Kitchel B, et al. Molecular Dissection of an Outbreak of
476 Carbapenem-Resistant Enterobacteriaceae Reveals Intergenus KPC Carbapenemase
477 Transmission through a Promiscuous Plasmid. mBio **2011**; Available at:
478 https://journals.asm.org/doi/abs/10.1128/mBio.00204-11. Accessed 21 February 2022.
479 12. Mathers AJ, Stoesser N, Sheppard AE, et al. Klebsiella pneumoniae Carbapenemase

480 (KPC)-Producing K. pneumoniae at a Single Institution: Insights into Endemicity from
481 Whole-Genome Sequencing. Antimicrob Agents Chemother **2015**; 59:1656–1663.
482 13. León-Sampedro R, DelaFuente J, Díaz-Agero C, et al. Pervasive transmission of a
483 carbapenem resistance plasmid in the gut microbiota of hospitalized patients. Nat Microbiol
484 **2021**; 6:606–616.
485 14. About EIP | DPEI | CDC. 2021. Available at: https://www.cdc.gov/ncezid/dpei/eip/eip-
486 about.html. Accessed 2 April 2022.
487 15. Magill SS, Dumyati G, Ray SM, Fridkin SK. Evaluating Epidemiology and Improving
488 Surveillance of Infections Associated with Health Care, United States - Volume 21,
489 Number 9—September 2015 - Emerging Infectious Diseases journal - CDC. Available at:
490 https://wwwnc.cdc.gov/eid/article/21/9/15-0508_article. Accessed 16 February 2022.
491 16. MuGSI | HAIC Activities | HAI | CDC. 2022. Available at:
492 https://www.cdc.gov/hai/eip/mugsi.html. Accessed 2 January 2023. **2015**; 314:1479–1487.
493 17. Hunt M, Mather AE, Sánchez-Busó L, et al. ARIBA: rapid antimicrobial resistance
494 genotyping directly from sequencing reads. Microb Genom **2017**; 3. Available at:
495 https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5695208/. Accessed 26 March 2021.
496 18. Davis JJ, Wattam AR, Aziz RK, et al. The PATRIC Bioinformatics Resource Center:
497 expanding data and analysis capabilities. Nucleic Acids Res **2020**; 48:D606–D612.
498 19. Andrews S. s-andrews/FastQC. 2021. Available at: https://github.com/s-andrews/FastQC.
499 Accessed 10 March 2021.
500 20. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence
501 data. Bioinformatics **2014**; 30:2114–2120.
502 21. Li H, Durbin R. Fast and accurate short read alignment with Burrows–Wheeler transform.

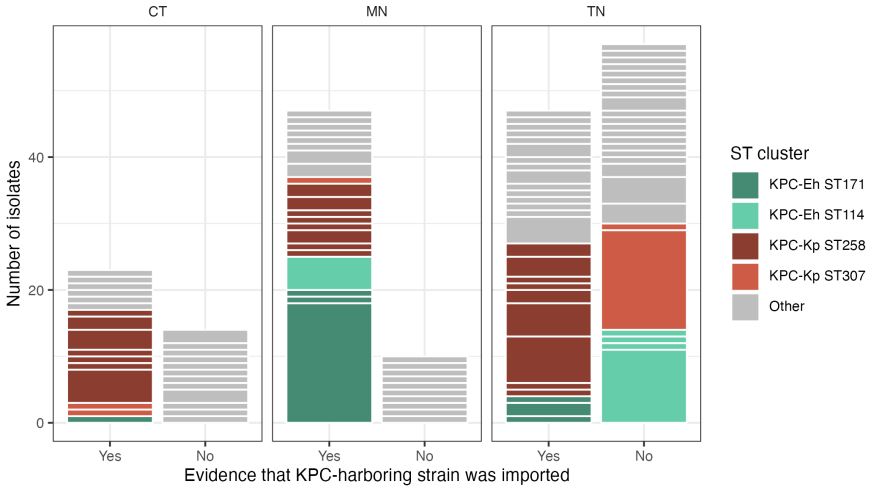
503 Bioinformatics **2009**; 25:1754–1760.
504 22. broadinstitute/picard. Broad Institute, 2021. Available at:
505 https://github.com/broadinstitute/picard. Accessed 20 February 2021.
506 23. Li H, Handsaker B, Wysoker A, et al. The Sequence Alignment/Map format and SAMtools.
507 Bioinformatics **2009**; 25:2078–2079.
508 24. Croucher NJ, Page AJ, Connor TR, et al. Rapid phylogenetic analysis of large samples of
509 recombinant bacterial whole genome sequences using Gubbins. Nucleic Acids Res **2015**;
510 43:e15–e15.
511 25. Poplin R, Ruano-Rubio V, DePristo MA, et al. Scaling accurate genetic variant discovery to
512 tens of thousands of samples. bioRxiv **2018**; :201178.
513 26. Kurtz S, Phillippy A, Delcher AL, et al. Versatile and open software for comparing large
514 genomes. Genome Biology **2004**; 5:R12.
515 27. Arndt D, Grant JR, Marcu A, et al. PHASTER: a better, faster version of the PHAST phage
516 search tool. Nucleic Acids Res **2016**; 44:W16–W21.
517 28. Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. IQ-TREE: A Fast and Effective
518 Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. Mol Biol Evol
519 **2015**; 32:268–274.
520 29. Bankevich A, Nurk S, Antipov D, et al. SPAdes: A New Genome Assembly Algorithm and
521 Its Applications to Single-Cell Sequencing. Journal of Computational Biology **2012**;
522 19:455–477.
523 30. Walker BJ, Abeel T, Shea T, et al. Pilon: An Integrated Tool for Comprehensive Microbial
524 Variant Detection and Genome Assembly Improvement. PLOS ONE **2014**; 9:e112963.
525 31. Brettin T, Davis JJ, Disz T, et al. RASTtk: A modular and extensible implementation of the

526 RAST algorithm for building custom annotation pipelines and annotating batches of
527 genomes. Sci Rep **2015**; 5:8365.
528 32. Crawford RD, Snitkin ES. cognac: rapid generation of concatenated gene alignments for
529 phylogenetic inference from large, bacterial whole genome sequencing datasets. BMC
530 Bioinformatics **2021**; 22:70.
531 33. Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. IQ-TREE: a fast and effective
532 stochastic algorithm for estimating maximum-likelihood phylogenies. Mol Biol Evol **2015**;
533 32:268–274.
534 34. Fu L, Niu B, Zhu Z, Wu S, Li W. CD-HIT: accelerated for clustering the next-generation
535 sequencing data. Bioinformatics **2012**; 28:3150–3152.
536 35. R Core Team. R: A Language and Environment for Statistical Computing. **2020**; Available
537 at: https://www.R-project.org/.
538 36. Paradis E, Schliep K. ape 5.0: an environment for modern phylogenetics and evolutionary
539 analyses in R. Bioinformatics **2019**; 35:526–528.
540 37. Hoffman S, Lapp Z, Wang J, Snitkin ESY 2022. regentrans: a framework and R package
541 for using genomics to study regional pathogen transmission. Microbial Genomics 8:000747.
542 38. Fay MP. Confidence intervals that match Fisher's exact or Blaker's exact tests. Biostatistics
543 **2010**; 11:373–374.
544 39. Wickham H, Averick M, Bryan J, et al. Welcome to the Tidyverse. Journal of Open Source
545 Software **2019**; 4:1686.
546 40. Yu G, Smith DK, Zhu H, Guan Y, Lam TT-Y. ggtree: an r package for visualization and
547 annotation of phylogenetic trees with their covariates and other associated data. Methods in
548 Ecology and Evolution **2017**; 8:28–36.

549 41. Yu G, Lam TT-Y, Zhu H, Guan Y. Two Methods for Mapping and Visualizing Associated
550 Data on Phylogeny Using Ggtree. Mol Biol Evol **2018**; 35:3041–3043.
551 42. Csardi G, Nepusz T. The Igraph Software Package for Complex Network Research.
552 InterJournal **2005**; Complex Systems:1695.
553 43. Pedersen TL. tidygraph: A Tidy API for Graph Manipulation. 2020. Available at:
554 https://CRAN.R-project.org/package=tidygraph. Accessed 16 February 2022.
555 44. Pedersen TL, RStudio. ggraph: An Implementation of Grammar of Graphics for Graphs and
556 Networks. 2021. Available at: https://CRAN.R-project.org/package=ggraph. Accessed 16
557 November 2021.
558 45. Wilke CO. cowplot: Streamlined Plot Theme and Plot Annotations for 'ggplot2'. 2019.
559 Available at: https://CRAN.R-project.org/package=cowplot. Accessed 15 April 2020.
560 46. Wickham H, Bryan J, attribution) Rs (Copyright holder of all R code and all C code without
561 explicit copyright, et al. readxl: Read Excel Files. 2019. Available at: https://CRAN.R-
562 project.org/package=readxl. Accessed 16 February 2022.
563 47. Bowers JR, Kitchel B, Driebe EM, et al. Genomic Analysis of the Emergence and Rapid
564 Global Dissemination of the Clonal Group 258 Klebsiella pneumoniae Pandemic. PLoS
565 ONE **2015**; 10:e0133727.
566 48. Peirano G, Chen L, Kreiswirth BN, Pitout JDD. Emerging Antimicrobial-Resistant High-
567 Risk Klebsiella pneumoniae Clones ST307 and ST147. Antimicrobial Agents and
568 Chemotherapy **2020**; 64. Available at: http://aac.asm.org/content/64/10/e01148-20.
569 Accessed 26 February 2021.
570 49. Izdebski R, Baraniak A, Herda M, et al. MLST reveals potentially high-risk international
571 clones of Enterobacter cloacae*. Journal of Antimicrobial Chemotherapy **2015**; 70:48–56.

572 50. Worby CJ, Chang H-H, Hanage WP, Lipsitch M. The Distribution of Pairwise Genetic
573 Distances: A Tool for Investigating Disease Transmission. Genetics **2014**; 198:1395–1404.
574 51. Logan LK, Weinstein RA. The Epidemiology of Carbapenem-Resistant Enterobacteriaceae:
575 The Impact and Evolution of a Global Menace. J Infect Dis **2017**; 215:S28–S36.
576 52. Lin MY, Lyles-Banks RD, Lolans K, et al. The Importance of Long-term Acute Care
577 Hospitals in the Regional Epidemiology of Klebsiella pneumoniae Carbapenemase–
578 Producing Enterobacteriaceae. Clin Infect Dis **2013**; 57:1246–1252.
579 53. Pacilli M, Kerins JL, Clegg WJ, et al. Regional Emergence of Candida auris in Chicago and
580 Lessons Learned From Intensive Follow-up at 1 Ventilator-Capable Skilled Nursing
581 Facility. Clin Infect Dis **2020**; 71:e718–e725.
582 54. Harrison E, Brockhurst MA. Plasmid-mediated horizontal gene transfer is a coevolutionary
583 process. Trends in Microbiology **2012**; 20:262–267.
584 55. Snitkin ES, Won S, Pirani A, et al. Integrated genomic and interfacility patient-transfer data
585 reveal the transmission pathways of multidrug-resistant Klebsiella pneumoniae in a regional
586 outbreak. Science Translational Medicine **2017**; 9.
587

588 **Figures**



589

590 **Figure 1: Evidence of importation and clonal dissemination differs across states and**
591 **sequence types (STs).** Clusters are separated by white lines between the colored bars, where the
592 bar height of a given cluster is the cluster size. Clusters were defined as subclades that were
593 monophyletic for a single state. Importation events were considered those with a KPC+ edge
594 transition from another state to the EIP state (see Methods). We observed events with limited
595 onward transmission in all states. We also see a large cluster of *E. hormaechei* ST171 in MN that
596 arose from a putative importation event, indicating importation followed by sustained
597 dissemination. Additionally, we observed large clusters of *E. hormaechei* ST114 and *K.*
598 *pneumoniae* ST307 in TN with no evidence of importation from another state. CT=Connecticut;
599 MN=Minnesota; TN=Tennessee; ST=sequence type; Eh=*E. hormaechei*; Kp=*K. pneumoniae*.
600

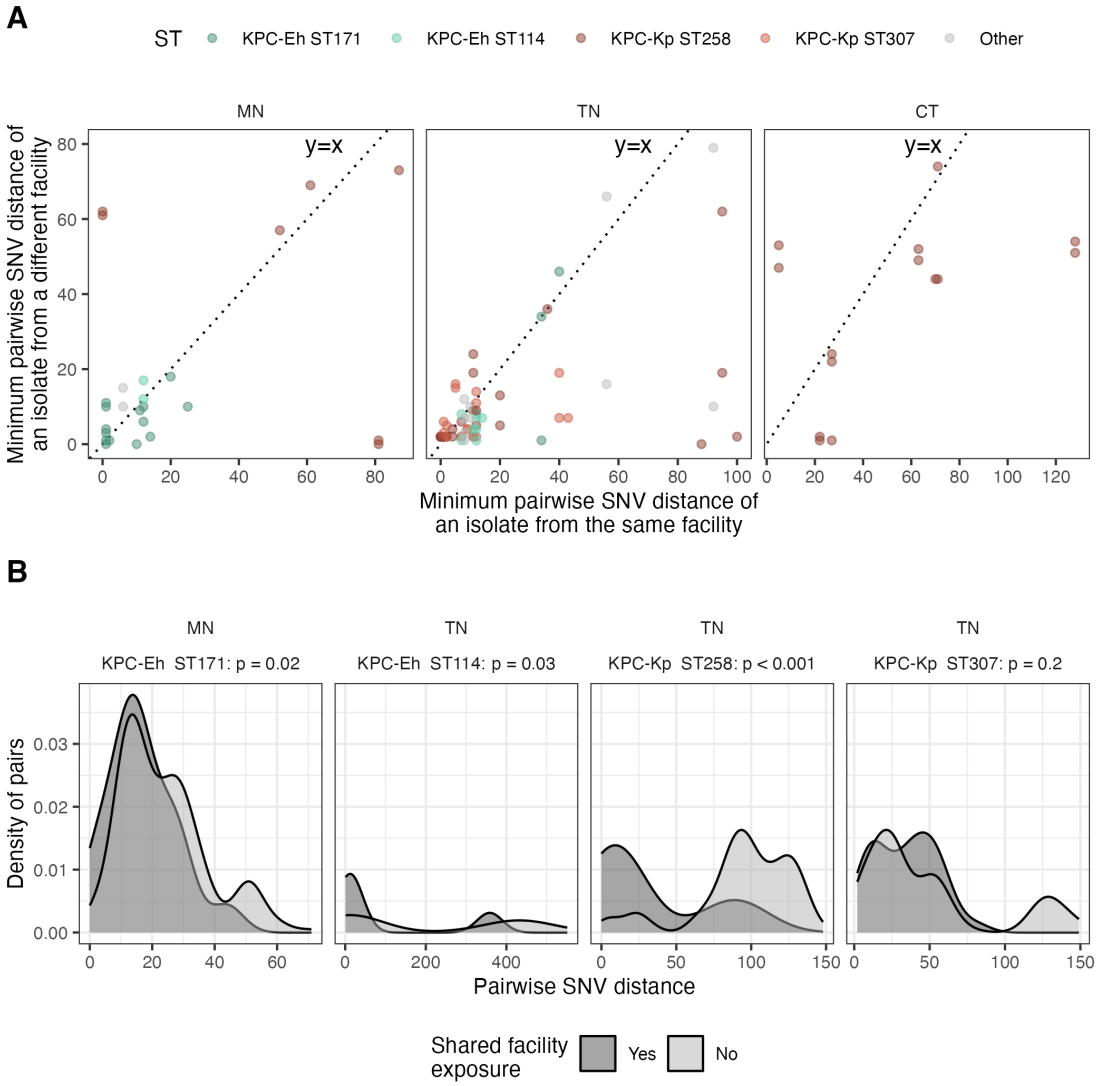


Figure 2: Small pairwise single nucleotide variant (SNV) distances between isolates suggest that local transmission was captured. (A) For each isolate with at least one intra-facility isolate pair and at least one inter-facility isolate pair, pairwise SNV distance of the most closely related intra-facility compared to the most closely related inter-facility isolate. (B) Pairwise SNV distance of inter-facility isolate pairs from patients with a shared facility exposure compared to those not linked by a shared facility exposure. 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. Only STs with >10 isolate pairs with a pairwise SNV distance of ≤15 are shown. One-sided

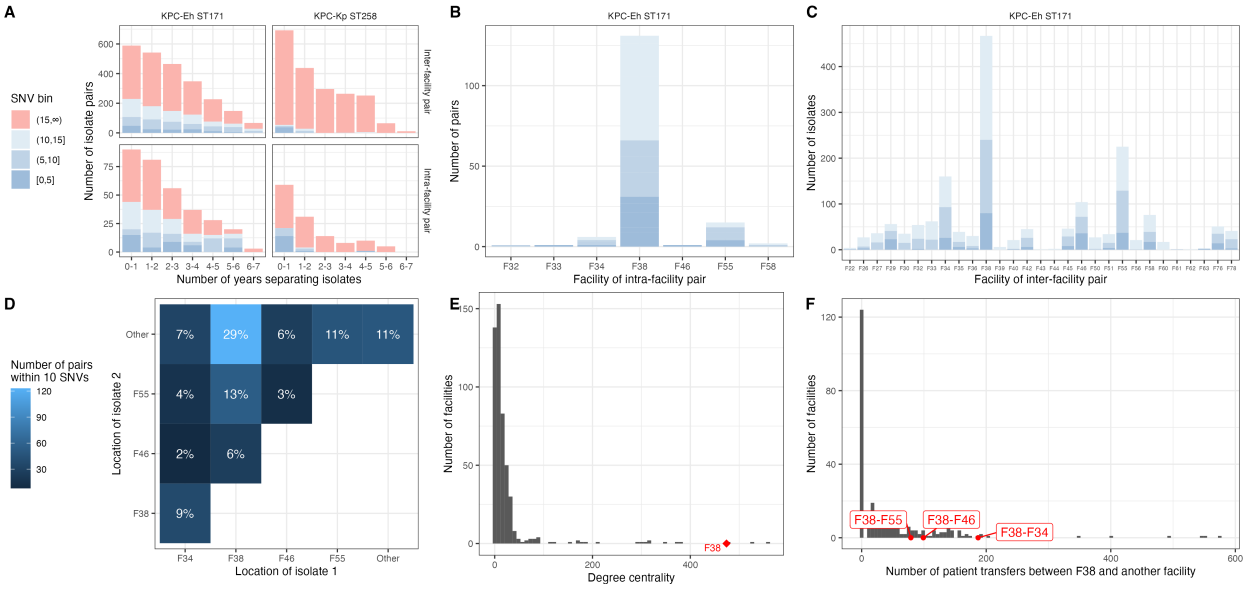
610 Wilcox p-values compare pairwise SNV distances of isolates with a shared facility exposure (N

611 ranges from 14 to 35) to isolates without a shared facility exposure (N ranges from 67 to 376).

612 CT=Connecticut; MN=Minnesota; TN=Tennessee; SNV=single nucleotide variant;

613 ST=sequence type; Eh=*E. hormaechei*; Kp=*K. pneumoniae*.

614



615

616 **Figure 3: Transmission of KPC-Eh ST171 in Minnesota (MN).** (A) Number of years

617 separating ST171 and ST258 isolate pairs. (B) Intra-facility ST171 isolate pairs by facility. Each

618 pair is represented once. (C) Inter-facility isolate pairs by facility. Each pair is represented twice

619 (once per facility). (D) Facilities connected by ST171 isolates within 10 SNVs of each other. (E)

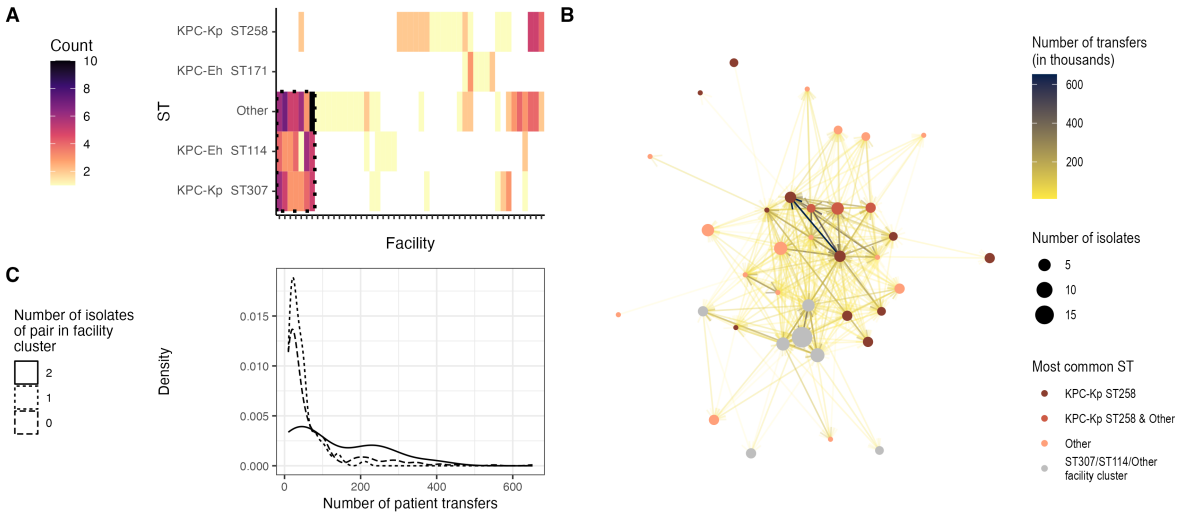
620 Degree centrality of all facilities, with F38 highlighted. The in- and out- degree values were

621 summed (from and to a facility). (F) Number of patient transfers between F38 and each other

622 facility. The values to and from each facility were summed. SNV=single nucleotide variant;

623 ST=sequence type; Eh=*E. hormaechei*; Kp=*K. pneumoniae*.

624



625

626 **Figure 4: Distinct facility sub-networks in Tennessee (TN) harbor different lineages.** (A)
627 Number of patient facility exposures in the prior year for each facility and ST. Each patient may
628 be represented more than once if they have more than one facility exposure. A cluster of
629 facilities identified using the complete linkage clustering method (dashed box) are common in
630 patients with ST114, ST307, and other isolates. (B) Patient transfer network of facilities with at
631 least one whole-genome sequenced isolate plotted using the Kamada-Kawai algorithm. The
632 cluster of facilities identified in the heatmap in panel A (Kp ST307/Ec ST114/Other facility
633 cluster) are also clustered in the network. (C) Number of patient transfers for facility pairs where
634 0, 1, or 2 of the facilities are in the facility cluster identified in the heatmap in panel A. Facility
635 pairs where both facilities are in the cluster have more patient sharing than facility pairs where
636 only 0 or 1 of the facilities is in the identified cluster. ST=sequence type; Eh=*E. hormaechei*;
637 Kp=*K. pneumoniae*.

638

639 **Supplemental information**

640

641 **Supplemental methods**

642 Catchment areas

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

649

650 CRE surveillance

651 CRE are defined as any organism in the Enterobacterales order isolated from any clinical
652 specimen and resistant to doripenem, meropenem, or imipenem (minimum inhibitory
653 concentrations [MIC] of ≥4 μg/ml), ertapenem (MIC ≥2 μg/ml), or demonstrates production of a
654 carbapenemase [16]. Specifically, isolates from residents of the surveillance area identified at the
655 clinical laboratory as CRE from *Escherichia coli*, *Enterobacter cloacae* complex species (i.e., *E.*
656 *cloacae*, *E. asburiae*, *E. bugandensis*, *E. hormaechei*, *E. kobei*, *E. ludwigii*, and *E.*
657 *nimipressuralis*), and *Klebsiella* species (i.e., *K. aerogenes*, *K. oxytoca*, and *K. pneumoniae*)
658 were included. All CRE isolates identified through the described public health surveillance
659 systems were characterized at the respective SPHL, which tested for carbapenemase genes (e.g.,
660 bla_{KPC}, bla_{NDM}, bla_{OXA-48}, bla_{VIM}, bla_{IMP}; see below for more details). While all CRE are

661 supposed to be tested for carbapenemases, this does not always happen in practice; the
662 percentage of CRE tested for carbapenemases was 81.5% in CT, 90% in MN, and 78% in TN.
663
664 *Connecticut*
665 CT did not participate in MuGSI surveillance during the study period, so all study isolates from
666 CT came from statewide passive surveillance for CRE. CRE surveillance of all Enterobacterales
667 isolates collected from all invasive clinical sites, respiratory sources, and urine began on Jan 1,
668 2017. All CRE isolates undergo antimicrobial susceptibility testing with a custom Sensititre and
669 disk diffusion panel, phenotypic detection of carbapenemase activity using the mCIM test and
670 PCR detection of carbapenemase genes (bla_{KPC}, bla_{NDM}, bla_{OXA-48}, bla_{VIM}, bla_{IMP}). Isolates
671 carrying bla_{KPC} that were identified in 2017 and 2018 from this collection were included.
672
673 *Minnesota*
674 MuGSI surveillance in Minnesota is conducted in Ramsey and Hennepin counties. MuGSI
675 isolates are obtained from sterile sites or urine and include *E. coli*, *Klebsiella* spp., and
676 *Enterobacter* spp. Non-MuGSI isolates are collected from a variety of culture sources (e.g.,
677 blood, urine, respiratory) and represent multiple Enterobacterales species. Isolate identification is
678 confirmed by the MALDI Biotyper CA System (Bruker Daltonics, Inc., Billerica, MA). All CRE
679 isolates are tested for carbapenemase production using the modified carbapenem inactivation
680 method (mCIM). Isolates demonstrating phenotypic carbapenemase production undergo
681 molecular testing (i.e., polymerase chain reaction [PCR]) for detection of carbapenemase genes
682 bla_{KPC}, bla_{NDM}, bla_{OXA-48}, bla_{VIM}, bla_{IMP}. All bla_{KPC} -positive *K. pneumoniae* and *E. cloacae*
683 complex isolates submitted to MN PHL from 2012-2018 were analyzed for this study.

684

685 *Tennessee*

686 EMS region 2-7 were selected because it provides a good representation of patient referrals;
687 captures most isolates in patient transfer networks covering the area; and includes both the
688 Knoxville area in which CRE emerged in 2016, and the Nashville area (the MuGSI catchment
689 area: Dickson, Cheatham, Robertson, Sumner, Wilson, Rutherford, Williamson and Davidson
690 counties) where CRE increased in 2017. CRE isolates underwent organism confirmation using
691 the MALDI-TOF bioMerieux. Isolates were tested for antimicrobial resistance with the Kirby
692 Bauer method. Those isolates that were resistant to at least one carbapenem were then tested by
693 the CarbaNP Method for carbapenemase production. Those that demonstrated carbapenemase
694 production from the CarbaNP underwent a PCR assay developed by the CDC to test for the
695 bla_{KPC} and bla_{NDM}. All bla_{KPC}-positive isolates collected during 2016–2017 and submitted to the
696 TN SPHL were included.

697

698 <u>Epidemiologic data</u>

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

702

703 *Connecticut*

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

706

707 *Minnesota*

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

710

711 *Tennessee*

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

716

717 Facility identification

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

720

721 *Connecticut*

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

727

728 *Minnesota*

729 MN staff maintain a compendium with facility name and type according to CMS. The facility
730 name and type are recorded in the MN surveillance database for each isolate and verified using
731 this compendium. For MuGSI cases, chart review at the facility where the isolate was collected
732 was also performed by MN staff.
733
734 *Tennessee*
735 TN staff linked patient identifiers of each case in the National Electronic Disease Surveillance
736 System (NEDDS)-Base System (NBS) surveillance data to the TN Hospital Discharge Data
737 System (HDDS) to identify and confirm the healthcare facility where they had the first specimen
738 culture collected for each CP-CRE case in TN. Alternatively, if the patient was not hospitalized
739 in a TDH-licensed hospital and therefore, was not found in the HDDS, TN staff used
740 surveillance data from NBS and the afore-mentioned questionnaire to identify the healthcare
741 facility of origin. Using this approach, the healthcare facility of origin was identified for 139
742 (89%) of the 157 included isolates. TN staff also gathered information on healthcare exposures
743 within one year prior to specimen collection by linking the patient identifiers in the surveillance
744 data with the inpatient and outpatient HDDS dataset. Isolates with no healthcare exposure
745 information were not used for the patient sharing network analysis.
746
747 <u>Linkage/matching of facility names to CMS ID</u>
748 Linkage of facility names to CMS ID was performed at each respective SPHL.
749
750 *Connecticut*

751 Facilities were matched with CMS IDs using CMS Provider of Services Files available at:
752 https://www.cms.gov/Research-Statistics-Data-and-Systems/Downloadable-Public-Use-
753 Files/Provider-of-Services
754
755 *Minnesota*
756 In MN, CMS IDs were manually entered for each isolate using Excel.
757
758 *Tennessee*
759 In TN, the name healthcare facility of origin was linked with the registered facility name by
760 CMS ID using the SPEDIS and COMPGED commands in SAS 9.4 (SAS Institute, Cary, North
761 Carolina) to allow for inexact matches that accounted for minor spelling variations between the
762 facility names from TN's surveillance data and those from the CMS dataset. Due to the data-
763 sharing agreement between TDH and the University of Michigan, the facility names and CMS
764 IDs in the WGS dataset and patient transfer data from CMS were de-identified before sharing
765 with The University of Michigan.
766
767 Generation of aggregate patient transfer networks
768 Aggregate patient transfer networks for each state for the year 2017 were derived from Centers
769 for Medicare and Medicaid Services (CMS) fee for service beneficiary claims data linked to the
770 CMS Minimum Data Set (MDS) by Medicare beneficiary ID. The dataset comprises a network
771 of medical facilities, including but not limited to short-term acute care hospitals, skilled nursing
772 facilities and long-term acute care hospitals, but may not be representative of non-Medicare
773 beneficiaries. To create an annual transfer network, CDC staff linked each unique beneficiary's

774 healthcare utilization. This included CMS inpatient claims for acute care stays and the MDS to
775 ascertain a beneficiary's presence in a nursing home, regardless of payer for that stay. Transfer
776 networks were comprised of the year of interest and looked back to the prior year to obtain all
777 discharging facilities where patients had an opportunity to acquire CRE.
778
779 Transfers are defined as a patient discharged from a facility and then subsequently admitted to a
780 facility. The number of transfers between two facilities includes transfers directly from one
781 facility to another as well as transfers with an intervening stay in the community of less than 365
782 days. We chose to use a 365-day time interval rather than a smaller time interval because this
783 gives us more resolution in comparing the extent of patient flow between facilities. Any facility
784 pairs with 10 or fewer patient transfers were censored, as per data use agreements between CDC
785 and CMS.
786
787 For each state, we subset the aggregate patient transfer network to only include the facilities for
788 which we have at least one WGS isolate or ones that are connected to one of these facilities by at
789 least one patient transfer event over the course of a year. We then identified paths of maximum
790 patient flow between this subset of facilities using the get_patient_flow() function in regentrans
791 v0.1 [37].
792
793 <u>DNA extraction</u>
794 *Connecticut*
795 Total DNA extractions performed by CT DPHL for WGS were either by the manual method or
796 automated on the QIAcube. Manual extractions from pure single colony Gram-negative bacterial

797 isolates by QIAgen DNeasy Blood and Tissue kit and follow-up quality control of all purified
798 extracts were conducted following PulseNet Protocols for Whole Genome Sequencing on the
799 Illumina MiSeq system. Automated extractions of bacterial cell pellets were performed using the
800 QIAcube protocol for Isolation of DNA from Gram-negative bacteria and QIAGEN DNeasy
801 Blood and Tissue kit. To obtain the cell pellets, colonies from overnight plates were resuspended
802 in 1x Phosphate buffered saline (PBS, Sigma) to an OD600 of 0.35-0.40. Two milliliter cell
803 suspension were centrifuged at 5,000 *x g* for 5 minutes at 2-8 °C and used for extractions.
804
805 *Minnesota*
806 DNA extraction of the isolates was performed in MN using either the QIAcube (Qiagen,
807 Germantown, MD) extractor and QIAmp DNA Mini extraction kit (Qiagen, Germantown, MD)
808 or the Roche MagnaPure LC and the TNA Isolation extraction kit.
809
810 *Tennessee*
811 DNA extraction of the isolates was performed in TN on a Thermo Fisher, KingFisher Flex auto-
812 extraction instrument, using an Applied Biosystems MagMAX DNA Multi-Sample Ultra
813 extraction kit.
814
815 Whole-genome sequencing (WGS)
816 The extracted DNA was prepared for WGS using Illumina Nextera XT library Prep kits with
817 Illumina, Nextera XT Index kits (MN, TN, CT) or with Illumina DNA Prep kit and Illumina CD
818 Indexes (CT) (Illumina, San Diego, CA). The Library preparation process was performed on a
819 Beckman Coulter Biomek Fxp instrument (Beckman Coulter, Brea, CA) (TN), on the Eppendorf

820 epMotion (CT) (Eppendorf, Enfield, CT), or manually (MN, CT). The prepared libraries were
821 sequenced on an Illumina MiSeq sequencing instrument (Illumina, San Diego, CA) using 2x250
822 (MN, TN, CT) or 2x150 (CT) paired end sequencing with V2 chemistry. The fastq files
823 generated were transferred to the University of Michigan for processing and analysis.
824
825 Genomic data processing
826 All genomic data processing and analysis was performed by members of Evan Snitkin's lab at
827 the University of Michigan.
828
829 *Species/ST identification*
830 MLST was assigned with Ariba v2.14.4 [17]. For PATRIC isolates mlst were extracted from the
831 isolate metadata [18].
832
833 *Sequence alignments*
834 Single-nucleotide variants were called in study sequences using SNPKIT
835 (https://github.com/Snitkin-Lab-Umich/snpkit). Quality of reads was assessed with FastQC
836 v0.11.9 [19]. Adapter sequences and low-quality bases were removed with Trimmomatic v0.36
837 [20]. Variants were identified by mapping filtered reads to the KPNIH1 reference genome
838 (GenBank accession number CP008827.1; 5,394,056 base pairs) for *K. pneumoniae* sequences
839 and the MNCRE9 reference genome (GenBank accession number JZDE00000000.1; 4,911,317
840 base pairs) for *E. hormaechei* sequences using bwa v0.7.17 [21], removing polymerase chain
841 reaction duplicates with Picard 2.21.7 [22], removing clipped alignments using Samclip 0.4.0,
842 and calling variants with SAMtools v1.11and bcftools [23]. Variants were filtered from raw

843 results using GATK's VariantFiltration v3.8 (QUAL, >100; MQ, >50; ≥10 reads supporting
844 variant; and FQ, <0.025) [25]. Consensus files generated during variant calling were
845 recombination filtered using Gubbins v3.0.0 [24]. The alleles at each position that passed
846 filtering were concatenated to generate a non-core variant alignment relative to the respective
847 reference genome. Alleles that did not pass filtering were considered unknown (denoted as N in
848 the alignment).
849
850 *Reference-based phylogenies*
851 A custom Python script was used to filter out (mask) single nucleotide variants in the whole-
852 genome alignment that were: (i) <5 base pairs (bp) in proximity to indels that were identified by
853 GATK HaplotypeCaller [25], (ii) in a recombinant region identified by Gubbins v3.0.0 [24], (iii)
854 in a phage region identified by the Phaster web tool [27] or (iv) they resided in tandem repeats of
855 length greater than 20bp as determined using the exact-tandem program in MUMmer v3.23 [26].
856 This whole-genome masked variant alignment was used to reconstruct a maximum likelihood
857 phylogeny with IQ-TREE v1.6.12 using the general time reversible model GTR+G and ultrafast
858 bootstrap with 1000 replicates (-bb 1000) [33].
859
860 *Genome assemblies*
861 Quality of reads was assessed with FastQC v0.11.9 [19]. Adapter sequences and low-quality
862 bases were removed with Trimmomatic v0.36 [20]. The clean reads were assembled with Spades
863 v3.14.1 (careful mode) [29]followed by Pilon v1.23 assembly correction [30]. Contigs smaller
864 than 500bp were discarded.
865

866 *Public isolates*

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

871

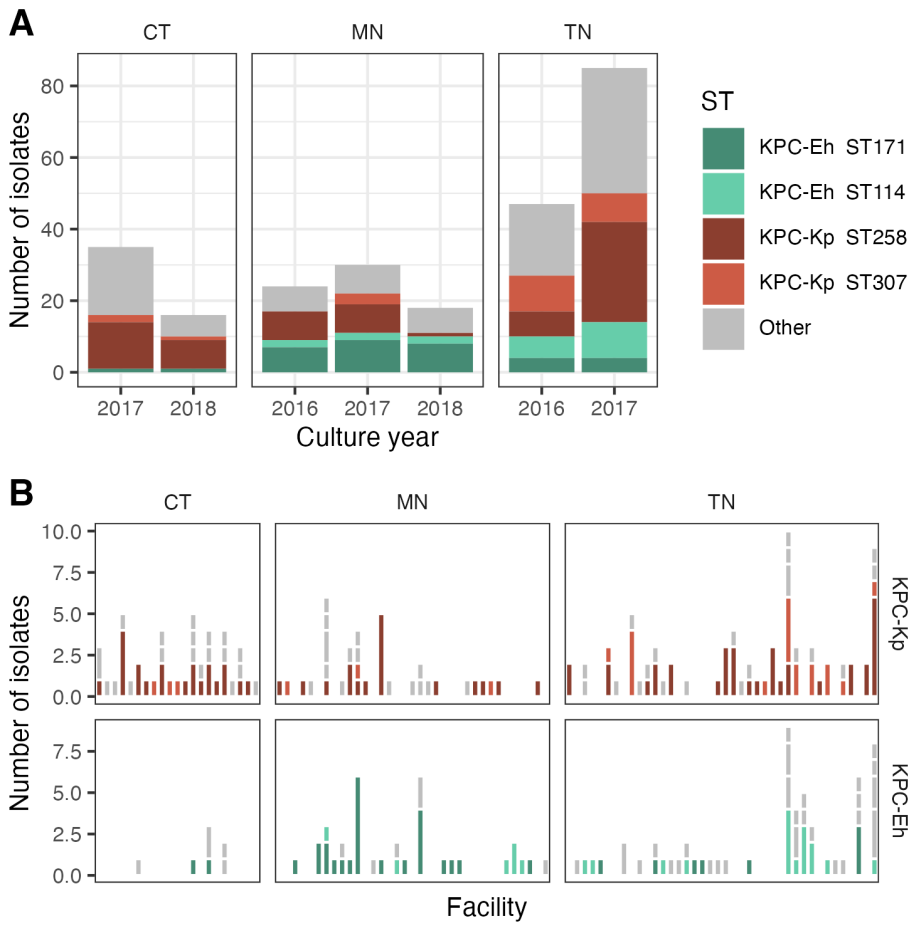
872 *Core genome phylogenies*

873 A concatenated gene alignment of core genes was generated for the entire dataset with cognac
874 v1.0 [32]. Relevant isolates for understanding the population structure and geographic
875 distribution of the isolates collected as part of the EIP study were identified as any isolate within
876 25 substitutions or the nearest phylogenetic neighbor by alignment distance. This yielded a
877 subset of the 5,346 isolates, including study isolates and public isolates; and a second,
878 concatenated, core gene alignment was generated with cognac for these most relevant isolates,
879 returning the nucleotide alignment for more precise distance comparisons. To generate sequence
880 type specific alignments, these data were divided into relevant clonal groups. Any clonal group
881 with less than 10 members was merged into a group of its respective species. Outgroups were
882 identified as the isolate with the lowest substitution distance to an isolate assigned to that clonal
883 group or species, for clonal group or species alignments, respectively. A core, concatenated gene
884 alignment for each set of isolates corresponding to species/clonal group was generated using
885 cognac with the corresponding outgroup specified, returning the nucleotide alignment.
886 Recombinant positions were removed with Gubbins v3.0.0 [24]. For alignments with more than
887 1000 isolates included, the alignments were randomly downsampled into subsets of 250
888 sequences, including the outgroup in each alignment, and gubbins was run on the downsampled

889 alignment. This was performed in triplicate, and any position identified as recombinant in the
890 three runs was masked from the alignment. The recombination filtered alignments were then
891 input to IQ-TREE v1.6.12 [33] to generate maximum likelihood phylogenies.
892
893 *bla_{KPC} identification*
894 The *bla_{KPC}* gene was identified using the CD-HIT v4.7 [34] data generated with cognac. The
895 gene cluster containing coding sequences annotated by RAST v1.035 [31] as *bla_{KPC}* was
896 identified. *bla_{KPC}* status was determined for each genome as having representation in the *bla_{KPC}*
897 cluster.
898
899 **Supplemental table and figures**
900
901 **Table S1: Number of each facility type by state and how many facilities were represented**
902 **in whole-genome sequencing (in parentheses).**
903

Facility type	CT	MN	TN
ACH	30 (19)	50 (16)	92 (31)
LTACH	3 (0)	2 (1)	10 (4)
SNF	224 (0)	366 (5)	309 (0)
Other	7 (0)	96 (1)	48 (1)
Unknown	0 (2)	0 (12)	0 (4)

904



905

906 **Figure S1: Distribution of species and sequence types of sequenced isolates from each state.**

907 (A) Across time. (B) Across healthcare facilities with at least one identified CRE with *bla*KPC.

908 X axis shows facilities. Grey is any other ST of the species corresponding to that faceted plot.

909 Notes: The MuGSI catchment area for MN is Hennepin and Ramsey counties. Four counties in

910 the Memphis-Delta Emergency Medical Services (EMS) and 8 in Northeastern TN were

911 excluded due to extensive healthcare utilization across state lines; the MuGSI catchment area is

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

913 CT=Connecticut; MN=Minnesota; TN=Tennessee; ST=sequence type; Eh=*E. hormaechei*;

914 Kp=*K. pneumoniae*.

915

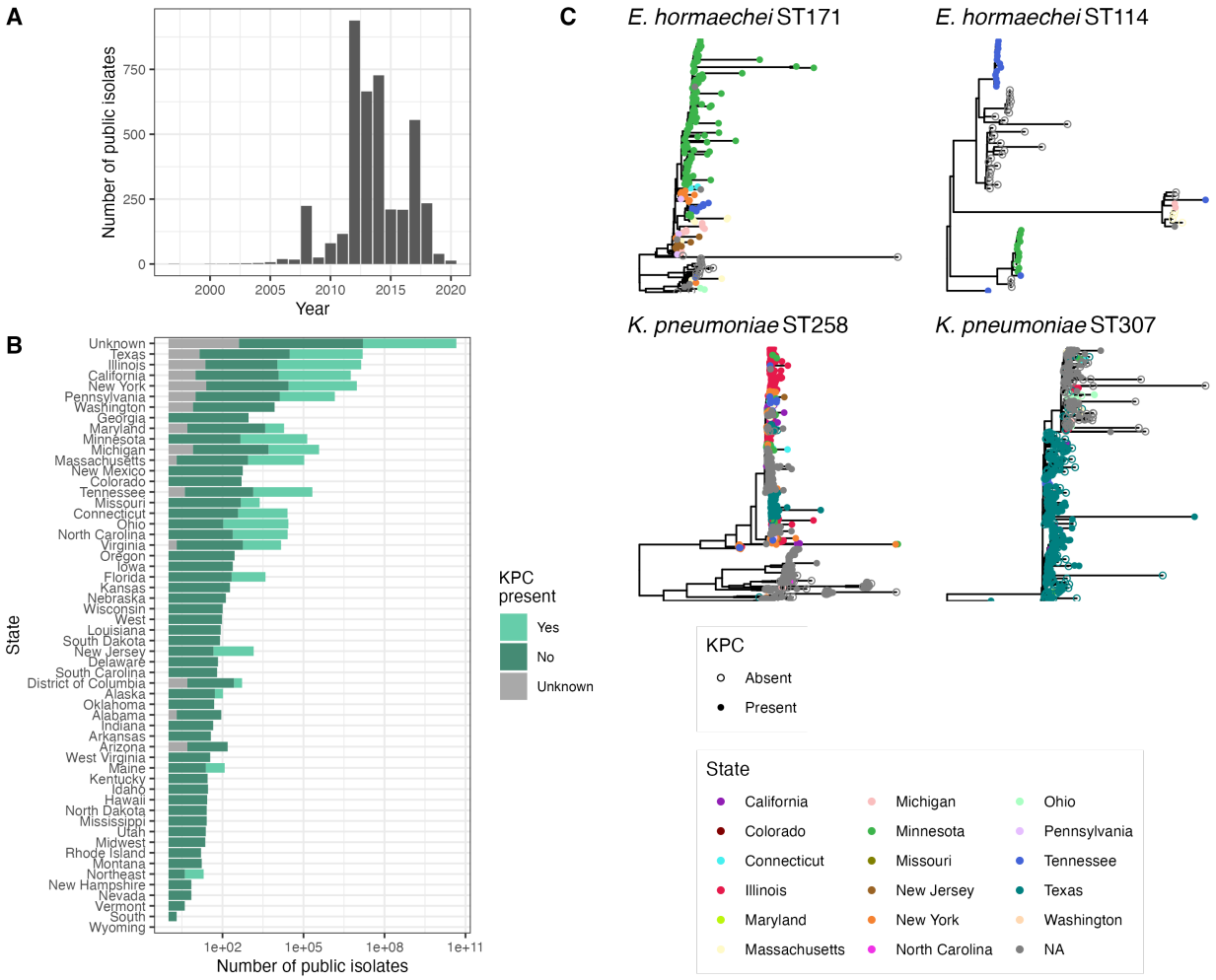
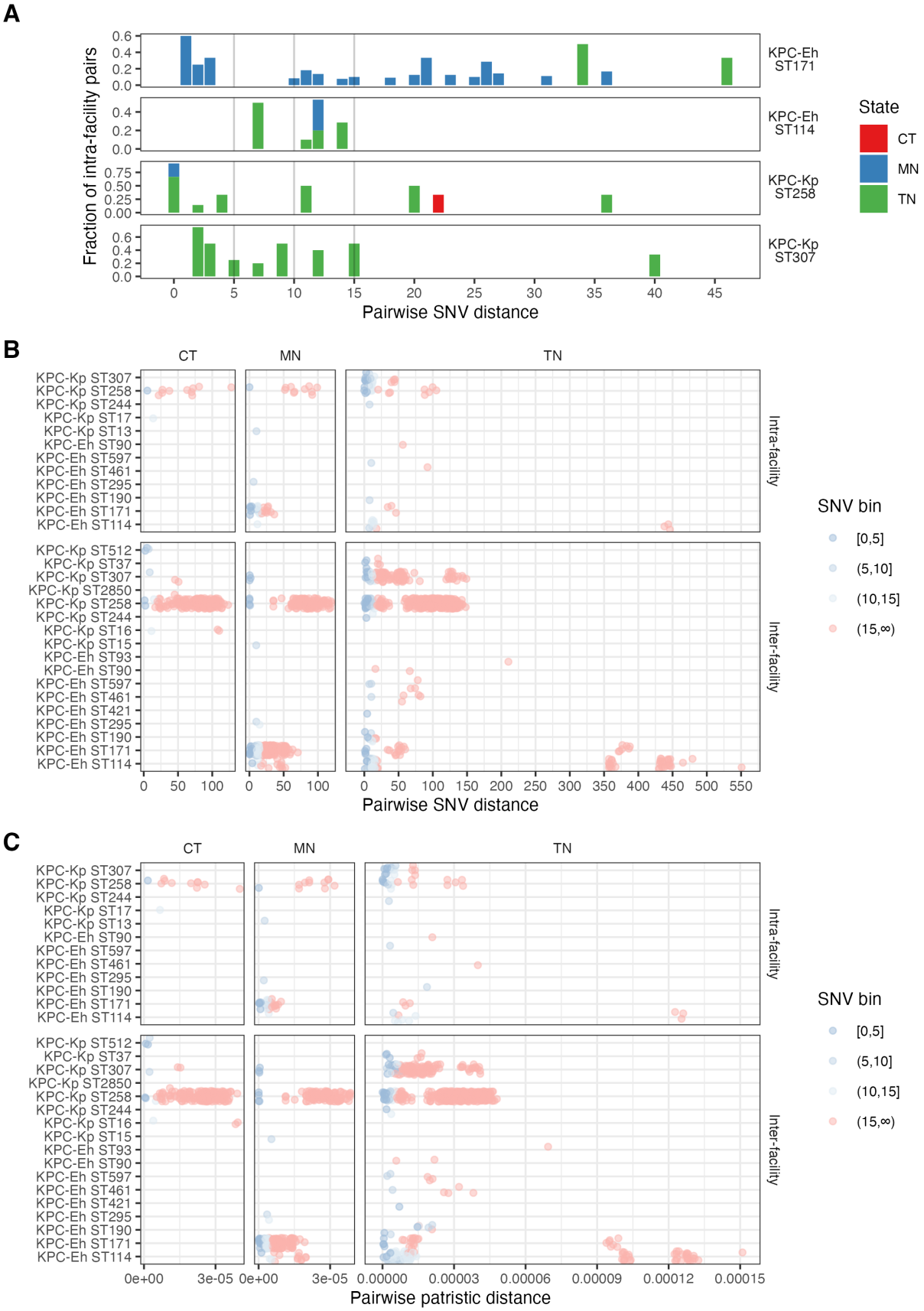


Figure S2: Information about public isolates downloaded from the PATRIC database that were included for context in the investigation of importation. (A) Distribution of year of isolation for isolates where this information was provided. (B) Distribution of location and KPC presence/absence. (C) Phylogenies used to identify importation events, including isolates from this study. ST=sequence type.


A
Fraction of intra-facility pairs
KPC-Eh ST171
KPC-Eh ST114
KPC-Kp ST258
KPC-Kp ST307
Pairwise SNV distance
State
CT
MN
TN
B
CT
MN
TN
KPC-Kp ST307
KPC-Kp ST258
KPC-Kp ST244
KPC-Kp ST17
KPC-Kp ST13
KPC-Eh ST90
KPC-Eh ST597
KPC-Eh ST461
KPC-Eh ST295
KPC-Eh ST190
KPC-Eh ST171
KPC-Eh ST114
KPC-Kp ST512
KPC-Kp ST37
KPC-Kp ST2850
KPC-Kp ST16
KPC-Kp ST15
KPC-Eh ST93
KPC-Eh ST421
Intra-facility
Inter-facility
Pairwise SNV distance
SNV bin
[0,5]
(5,10]
(10,15]
(15,∞)
C
Pairwise patristic distance

923

924 **Figure S3: Inter-facility genomic analysis.** (A) Fraction of intra-facility isolate pairs for
925 pairwise single nucleotide variant (SNV) distances to help identify pairwise SNV distance
926 cutoffs. We expect there to be an enrichment in closely related intra-facility pairs under the
927 assumption that intra-facility transmission is more common than inter-facility transmission. This
928 plot helps us identify potential pairwise SNV distance thresholds by looking for a SNV distance
929 where there is a decrease in the fraction of intra-facility isolate pairs (around 5 and 10 appear to
930 be reasonable thresholds for a sensitivity analysis). Vertical lines at 5, 10, and 15 SNVs represent
931 potential cutoffs for determining what isolates are considered closely related; 3 thresholds were
932 chosen as a sensitivity analysis. (B) Pairwise SNV distances of all STs. Within an ST, smaller
933 pairwise SNV distances indicate potential local transmission, while larger pairwise SNV
934 distances are not indicative of local transmission. Varying extents of transmission were captured
935 for different states and STs. Putative local transmission events are those in the tail of the
936 distribution (lower pairwise SNV distances). (C) Pairwise patristic distances are concordant with
937 pairwise SNV distance thresholds. CT=Connecticut; MN=Minnesota; TN=Tennessee;
938 ST=sequence type; Eh=*E. hormaechei*; Kp=*K. pneumoniae*; SNV = single nucleotide variant.
939

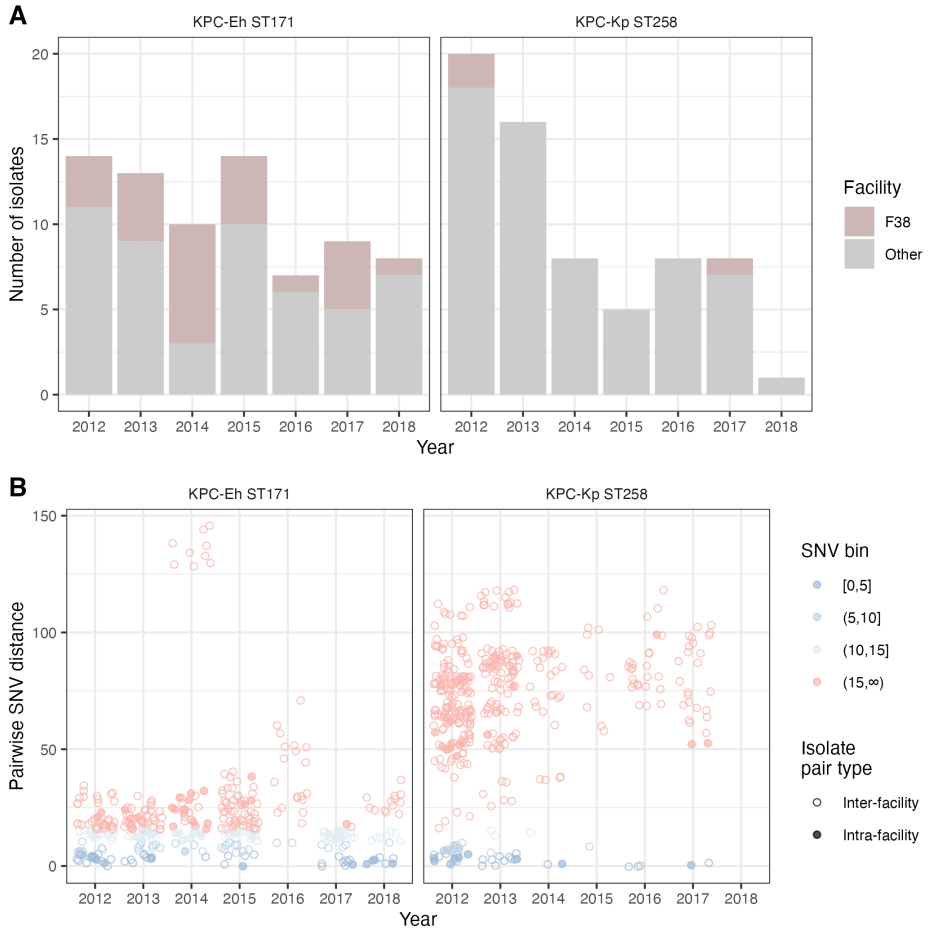


Figure S4: Minnesota (MN) isolates over time. (A) Number of ST171 and ST258 isolates over time. (B) Pairwise SNV distance of isolates over time. ST=sequence type; Eh=*E. hormaechei*; Kp=*K. pneumoniae*.