Next-generation sequencing based hospital outbreak investigation yields insight into *Klebsiella aerogenes* population structure and determinants of carbapenem resistance and virulence

Adel Malek[1], Kelly McGlynn[1], Samantha Taffner[1], Lynn Fine[3], Brenda Tesini[4], Jun Wang[1], Heba Mostafa[1], Sharon Petry[1], Archibald Perkins[1], Paul Graman[4], Dwight Hardy[1,2], and Nicole Pecora[1,#]

[1]Department of Pathology and Laboratory Medicine, University of Rochester Medical Center, New York, USA

[2]Department of Microbiology and Immunology, University of Rochester Medical Center, New York, USA

[3]Infection Prevention Program, University of Rochester Medical Center, New York

[4]Department of Medicine, Infectious Diseases Division, University of Rochester Medical Center, New York, USA

Running title: Carbapenem-resistant *Klebsiella aerogenes* outbreak

[#]Correspondence to:

Nicole Pecora, MD, PhD

University of Rochester Medical Center

601 Elmwood Avenue, Rochester, NY-14642, USA

Phone: (585) 276-4674, Email: nicole_pecora@urmc.rochester.edu

KEY WORDS

Carbapenem resistant *Klebsiella aerogenes*, whole-genome sequencing, cardiothoracic intensive care unit, outbreak, AmpC β-lactamases, porins, Omp36, integrative conjugative elements, AmpD, yersiniabactin, colibactin, ST4, MLST, genomic epidemiology.

ABSTRACT

Klebsiella aerogenes is a nosocomial pathogen associated with drug resistance and outbreaks in intensive care units. In a 5-month period in 2017, we experienced an increased incidence of cultures for carbapenem-resistant *K. aerogenes* (CR-KA*)* from an adult cardiothoracic intensive care unit (CICU) involving 15 patients. Phylogenomic analysis following whole-genome sequencing (WGS) identified the outbreak CR-KA isolates to group together as a tight clonal cluster (<7 SNPs apart), suggestive of a protracted intra-ward transmission event. No clonal relationships were identified between the CICU CR-KA strains and additional hospital CR-KA patient isolates from different wards and/or previous years. Genes encoding carbapenemases or drug-resistant plasmids were absent in the outbreak strains, and carbapenem resistance was attributed to mutations impacting AmpD activity and membrane permeability. The CICU outbreak strains harbored an integrative conjugative element (ICE*Kp*10), which has been associated with pathogenicity in hypervirulent *Klebsiella pneumoniae* lineages. Comparative genomics with global *K. aerogenes* genomes showed our outbreak strains to group closely with global ST4 strains, which along with ST93 likely represent dominant *K. aerogenes* lineages associated with human infections. WGS is a powerful tool that goes beyond high-resolution tracking of transmission events into identifying the genetic basis of drug-resistance and virulence, which are not part of conventional diagnostic workflows. With an increasing availability of sequenced genomes from across the globe, population structure analysis offers opportunities to identify emerging trends and dominant clones associated with specific syndromes and geographical locations for poorly characterized pathogens.

INTRODUCTION

Klebsiella aerogenes (formerly described as *Enterobacter aerogenes*) is a ubiquitous member of the *Enterobacteriaceae* family and a significant nosocomial pathogen associated with drug resistance and a wide variety of infections including pneumonia, bacteremia, urinary tract and surgical site infections (1, 2). In vulnerable patients, *K. aerogenes* infections can arise endogenously (gastrointestinal flora) or be acquired from surroundings in the facility where the patient is admitted (horizontal transmission through colonized healthcare workers, contaminated devices/shared equipment, other patients etc.), with the most critical risk factor for acquiring infection being prolonged broad-spectrum antibiotic administration. Additional risk factors for *K. aerogenes* infections include prolonged stay at healthcare facilities especially ICUs and neonatal wards, complex underlying medical conditions, immunosuppression and mechanical ventilation or the presence of foreign devices. Numerous hospital ward outbreaks in both pediatric and adult populations due to *K. aerogenes* have been described due to a common source or spread via patient-to-patient transmission (1, 2). A particularly high frequency of hospital ICU outbreaks was continually reported from Western Europe in a period between the 1990s and early 2000s, that were largely attributed to the spread and endemic establishment of a clonal *K. aerogenes* strain harboring the extended-spectrum B-lactamase TEM-24 ($bla_{\text{TEM-24}}$) (2).

Within the US and other regions across the globe, *K. aerogenes* has also been reported along with *Klebsiella pneumoniae, Enterobacter cloacae* and *Escherichia coli* to be among the frequently isolated carbapenem resistant *Enterobacteriaceae* (3-5). Clinical CR-KA strains harboring plasmid-borne serine carbapenemases have been

96 described in the US and worldwide, while metallo-β-lactamases and OXA- 48 have
97 been reported in Europe, Asia and Brazil (2). However, the primary mechanisms
98 underlying carbapenem non-susceptibility in *K. aerogenes* are thought to be
99 carbapenemase independent and mediated by mutations affecting regulation of
100 chromosomal AmpC β-lactamase expression and membrane permeability (2). The
101 latter has been well documented in *K. aerogenes* with reports describing mutations
102 impacting porin function/expression that can arise *in vivo* during antibiotherapy (6), be
103 reversible (2), and present complex diagnostic and therapeutic management challenges
104 (7).

105 Despite the role of *K. aerogenes* as an important opportunistic pathogen and its
106 epidemic potential, the clinical relevance of intra-species genetic diversity and
107 significance of specific sequence types (STs) remains unknown. In comparison, in
108 genomically closely related *K. pneumoniae* and to some extent in *E. cloacae*, clonal
109 complexes and sequence types (STs) associated with geographical distribution, multi-
110 drug resistance, hospital outbreaks and disease syndromes have been defined (8, 9).
111 Recently, a multi-locus sequence-typing (MLST) scheme has been developed for *K.*
112 *aerogenes* (10) that can help explore the above-mentioned issues, however, its
113 performance in evaluating and discriminating clinical/environmental isolates has not yet
114 been reported.

115 We pursued this study to investigate an outbreak of CR-KA in a cardiothoracic
116 intensive care unit (CICU) at our hospital, which persisted for 5 months despite
117 aggressive infection control measures. The primary goals of our study included whole-

genome sequencing (WGS) based investigation of the clonal relationships among the CR-KA strains isolated from patients in our hospital and defining putative loci associated with carbapenem resistance and virulence. In addition, the recently developed publicly available *K. aerogenes* MLST scheme afforded us the opportunity to delineate the population structure of CR-KA stains isolated from patients at our hospital. Our initial findings led us to broadly investigate the origin and significance of specific *K. aerogenes* sequence types identified in our hospital CR-KA strains by performing comparative genomics using publicly available global *K. aerogenes* genomes.

RESULTS

Epidemiological and genomic characterization of the CICU CR-KA cluster

In July 2017, five CICU patients had CR-KA isolated from respiratory tract specimens (room occupancy and patient demographics detailed in Fig. 1 and Table S1). The first identified case (Patient A) had a past medical history significant for intravenous drug use and recurrent methicillin-resistant *Staphylococcus aureus* infections. Patient A's CICU course is described in Fig. S1. The temporal association of subsequent positive cultures in the CICU prompted an outbreak investigation. Between August to November 2017, despite extensive infection prevention interventions, 10 additional CICU patients had positive CR-KA cultures (Table S1). Of the 15 positive cultures from unique patients, 6 were clinical specimens (5 respiratory, 1 blood) and 9 were rectal surveillance cultures. Interventions included contact precautions for all patients on the unit, cohorting of staff, and weekly surveillance cultures of rectal swabs and tracheal aspirates. Environmental surveillance cultures during this period were negative. An

141 epidemiological investigation into the possible risk factors among the CICU patients
142 developing CR-KA infections was non-contributory.

143 Antibiotic susceptibility testing (AST) of the CICU CR-KA strains revealed some
144 differences in phenotype (Table 1) among the carbapenems and cephalosporins. Sixty
145 percent of the CICU cluster strains were resistant to three carbapenems (ertapenem,
146 imipenem and meropenem), while 93% showed phenotypic resistance to ertapenem.
147 Susceptibility to cefepime also varied, with a subset of outbreak strains displaying
148 phenotypic resistance (Table 1).

149 To address the possibility of an outbreak event, particularly in light of variable
150 antibiotic susceptibilities, a WGS investigation was undertaken to identify phylogenetic
151 relationships and transmission links. A total of 26 URMC *K. aerogenes* isolates were
152 sequenced using Illumina WGS. The primary strains investigated in this study were 15
153 CR-KA isolates from patients admitted in the CICU, between June and Nov 2017 (Table
154 S1). For context and comparison, an additional set of 9 CR-KA strains
155 epidemiologically unlinked to the CICU-cluster strains (patient isolates between years
156 2015-17), were included in the study (Table S2). Two clinical isolates were also
157 included as controls, URMC 201 (intermediate susceptible to imipenem) and URMC 223
158 (susceptible to all carbapenems tested). The AST profiles of the non-CICU CR-KA study
159 strains are described in Table 1.

160 All of the 26 sequenced URMC *K. aerogenes* genomes showed high coverage
161 (>88%) relative to the *K. aerogenes* reference KCTC 2190 strain (ATCC 13048^T),
162 Dataset 1. Single nucleotide polymorphisms (SNPs) were identified across the study
163 genomes relative to the reference sequence (pair-wise SNP differences ranged from 1-

28,170, see Dataset 2). MLST assignment indicated that all of the CICU clinical and surveillance isolates belonged to ST4, and SNP-based analyses grouped them in a tight cluster separately and distantly from non-outbreak isolates (Fig. 2, Dataset 2). Within the CICU cluster, strains differed from URMC 205 (first case, patient A isolate) by less than 7 SNPs. In addition, these isolates bore an identical plasmid profile (based on replicon and plasmid typing) (Table S3). In contrast, all of the 2017 non-CICU isolates were significantly distant from the CICU outbreak isolates (>20,000 SNPs). The most closely related non-CICU CR-KA strain was URMC 201 (isolated in 2015). This strain was also ST4, with 433 SNPs in a pairwise comparison to URMC 205. (Fig. 2, Dataset 2). These relationships, coupled with the epidemiological data, indicated that the 2017 CICU CR-KA strains were of clonal origin.

Carbapenem resistance in the URMC CR-KA strains was driven by adaptive chromosomal gene alterations

A. WGS based identification of acquired antibiotic resistance genes

Despite phenotypic carbapenem resistance, the CICU CR-KA did not harbor genes for carbapenemases, extended-spectrum β-lactamases, or plasmid-borne AmpC cephalosporinases. A single non-outbreak CR-KA isolate (URMC 203) harbored a carbapenemase gene (*bla*$_{nmcA}$). No horizontally acquired genes conferring resistance to non-β-lactam antibiotics were identified among the study strains, consistent with their susceptibility profiles (Table 1).

B. Non-synonymous sequence alterations in key chromosomal loci implicated in carbapenem resistance

In the absence of genes encoding carbapenemases and ESBLs in the CR-KA outbreak strains, variations in other genetic loci associated with carbapenemase-independent resistance mechanisms were investigated (11-13). Focusing on the AmpC cephalosporinase and outer membrane porins, sequence variations in *ampD, ampG, ampR, omp35*, *omp36* and *ompR* genes in the study strains were assessed relative to the 'wild-type' allele in the carbapenem-susceptible reference type strain KCTC 2190 (14). The sequences were also compared to alleles in URMC 223 (carbapenem susceptible) and URMC 201 (intermediate susceptibility). For the *omp* genes, the upstream DNA sequences were also assessed. The *ampG* and *ompR* genes were wild-type in all study strains. Variants were identified in all other loci and are described below:

ampD. Mutations were identified in the *ampD* gene for each of the 24 CR-KA isolates in this study (Table 2), while the control strains URMC 223 and URMC 201 bore the 'wild-type' *ampD* allele. The outbreak strains harbored single missense SNPs (either 284G>T or 482G>A) in *ampD* resulting in Trp95Leu or Arg161His substitutions. Six non-outbreak CR-KA strains harbored independent non-synonymous single substitutions while 2 missense substitutions were identified in URMC 221. A single non-outbreak strain, URMC 202, harbored a nonsense mutation resulting in a truncated AmpD protein (Table 2). DNA sequence corresponding to *ampD* allele was absent in isolate URMC 203, (the only study strain to possess a carbapenemase-encoding gene $bla_{\text{NMC-A}}$) due to a large deletion in the genomic region harboring *ampD* and the neighboring *ampE* gene.

The potential impact of Trp95Leu or Arg161His substitutions on AmpD activity in the outbreak strains was investigated by homology-based structural modeling of *K. aerogenes* AmpD using a high-resolution crystal structure of *Citrobacter freundii* AmpD (15), a close homolog (83.33% amino acid sequence identity). AmpD contains a hydrophobic surface to accommodate its GlcNAc-anh-MurNAc ligand, which is made up of tripeptide and glycan moieties. The tripeptide portion of GlcNAc-anh-MurNAc is coordinated through three salt bridges between carboxyl groups on the tripeptide and residues Arg71, Arg161 and Arg107, while the peptide backbone is oriented across the hydrophobic surface. Trp95 forms a planar surface at the end of the ligand-binding channel to position the diaminopimelate moiety at the distal end of the tripeptide.

Based on the *in silico K. aerogenes* AmpD model, the positively charged guanidinium group of Arg161 forms two strong electrostatic interactions with the carboxyl group of D-glutamine on the tripeptide portion of the ligand (Fig. 3). Mutation of this residue to histidine was predicted to weaken ligand binding, likely affecting the positioning of the ligand in the active site. In the Trp95Leu mutation, the hydrophobicity in the region is preserved, but the shorter length of the leucine side chain leaves a gap at the end of the binding channel likely affecting the positioning of the entire ligand (Fig. 3). These observations lend support that the missense mutations observed in the CICU outbreak strains would alter ligand docking on AmpD likely reducing/inhibiting its activity. Four out of the seven CR-KA non-outbreak strains had substitutions within glycan and peptide interacting regions of AmpD (Table 2), suggesting that these alterations might also impact activity.

ampR. All of the outbreak strains harbored the reference *ampR* allele. Several non-outbreak CR-KA strains harbored 2-5 substitutions that likely represent variant alleles as they were also observed in the control carbapenem-susceptible strain URMC 223. A single non-outbreak CR-KA strain, URMC 210, bore a nonsense mutation in the *ampR* gene resulting in a premature stop codon (Trp117X), likely resulting in a non-functional truncated AmpR protein.

***omp35* and *omp36*.** Among the 15 outbreak CR-KA strains, three different *omp36* variants were identified relative to the wild-type allele (Table 3). An identical profile of missense SNPs in these strains and the control strain, URMC 201 (intermediate carbapenem susceptibility) were observed relative to the reference genome allele. Seven outbreak strains had additional mutations that resulted in severely truncated proteins. Several non-outbreak CR-KA strains also harbored missense SNPS of unclear significance. Two strains URMC 202 and URMC 204 harbored distinct frame-shift mutations yielding truncated Omp36 protein variants. A 42 base-pair region of high variation (nucleotides 680-724), corresponding to 15/16 amino acid substitutions in loop L5 was observed in all the clinical isolates relative to the reference sequence (Table 3, Fig. S2). The hypervariable region results in a different charge profile in the region and has been previously reported in a *K. aerogenes* study describing imipenem resistant clinical isolates (harboring ESBL TEM-24) from patients in France (16). The predicted Omp36 protein in clinical strain URMC 221 and the carbapenem susceptible control strain (URMC 223) bore 87% identity relative to the reference genome Omp36 and were considered significantly distant variants (not included in the comparative analyses).

All but one of our 24 CR-KA strains had the wild-type allele of *omp35*. A single strain (URMC 202) bore a deletion in the N-terminal encoding region. The DNA sequence upstream of the *omp35* gene was investigated to identify mutations in the promoter sites of the strains, of which one (URMC 204) had a nucleotide difference of unclear significance at the -22 position.

Virulome analysis identifies a large pathogenicity-associated integrative and conjugative element (ICE), ICE*Kp*10 in the CICU outbreak CR-KA strains

The prolonged nature of the clonal CR-KA outbreak in our CICU led us to search for putative virulence genes or pathogenicity loci, which could have promoted persistence and transmission. A cluster of chromosomally encoded genes encoding yersiniabactin (*ybt*) metallophore and colibactin (*clb*) genotoxin systems was identified in all of the outbreak strains (Fig. 4). These loci have been implicated in invasive infections of *K. pneumoniae* (17).

The *ybt* locus harbored putative genes involved in regulation as well as synthesis of the siderophore, corresponding transport associated proteins, and a receptor protein for the uptake of metal-bound siderophore. The *clb* locus included putative homologs encoding enzymes, transferases and transport proteins involved in production and secretion of the polyketide colibactin (Fig. 4, Dataset 3). Investigation of the genomic loci associated with the virulence factor gene cluster identified them to be present on a mobilizable integrative conjugative element (ICE) inserted in a tRNA-Asn site adjacent to a gene encoding the glycine cleavage system (Fig. 4). The ICE element bore a modular arrangement of gene clusters encoding mobile elements, P4 like integrase, type IV secretion system conjugation machinery and mobilization genes (Fig. 4, Dataset

277 3). The element was identified to be ICE*Kp*10, using a recently described typing and
278 interpretation scheme of *ybt* and *clb* sequences (17).

279 Among the URMC non-outbreak *K. aerogenes* strains, these putative virulence
280 genes were identified in only 4/10 isolates, which were either ST4 or ST93. The control
281 strain URMC 223 harbored the yersiniabactin locus exclusively. Detailed descriptions of
282 these elements in the URMC *K. aerogenes* strains are described in Table S4.

283 **Comparative analyses of URMC CR-KA and publicly available *K. aerogenes***
284 **genomes**

285 To gain insights into the emergence and epidemiology of the CICU outbreak
286 clones and to place our hospital CR-KA strains in the broader context of global *K.*
287 *aerogenes* strains, comparative phylogenomic analyses were performed using Harvest
288 genomics suite (18). Publicly available *K. aerogenes* genome assemblies (n=110) were
289 included in the analyses. These included 70 clinical and surveillance strains isolated
290 from human specimens, 3 'environmental' strains, and 36 strains of unknown origin
291 (Dataset 4). Based on the newly described MLST scheme, ST4 and ST93 strains were
292 found to be markedly overrepresented in the available genomes (51.8%, 57/110).

293 Excellent correlation was observed between Harvest generated tree topologies
294 (Fig. 5) as well as pairwise SNP differences (Dataset 5) as compared to those
295 generated by the CFSAN SNP pipeline, for the URMC CR-KA study strains. Based on
296 Harvest analyses, the CICU outbreak strains clustered closely with each other and to
297 other global ST4 genomes, as compared to the other URMC CR-KA strains (outbreak-
298 unrelated), which were distantly dispersed throughout the phylogenomic distribution
299 (Fig. 5). The MLST based sequence types of URMC CR-KA and the global *K.*

300 *aerogenes* genomes also correlated tightly with HARVEST generated core-genome
301 based topologies (Fig. 5).
302 Six publicly available assembled genomes grouped closely with the CICU
303 outbreak strain genomes (< 200 SNPs apart). These included ST4 strains, UCI 27, UCI
304 28, UCI 45, which have been described in a carbapenem resistance surveillance study
305 by Cerqueria et al. (19), and GN04794, GN05662 and GN02525; strains derived from
306 varied US patients' clinical specimens (blood, sputum, wound drainage) Fig. 5, Dataset
307 4). The above-mentioned strains had fewer SNP differences in relation to the CICU
308 outbreak strains as compared to URMC 201, the closest and only non-outbreak ST4
309 strain isolated in our hospital (Fig. 5, Dataset 5).
310 Carbapenemase encoding genes were identified in a total 13/111 global *K.*
311 *aerogenes* genome assemblies (11.8%) (Fig. 5). These included genes encoding KPC-
312 2 (n=7), KPC-3 (n=1), OXA-48 (n=4) and NDM-6 (n=1). The chromosomal serine
313 carbapenemase, $bla_{\text{NMC-A,}}$ was solely present in URMC 203, a CR-KA strain isolated
314 from a patient in our hospital in 2015. Among the thirty ST4 genomes, 4 strains (13%)
315 harbored genes encoding carbapenemases (4/4, KPC-2), and these were clinical
316 strains isolated from non-USA patients. The relative contribution of carbapenemase
317 mediated versus non-carbapenemase mediated mechanisms of resistance to
318 carbapenems in the global *K. aerogenes* could not be assessed due to the absence of
319 antibiotic susceptibility metadata for most strains.
320 A characteristic genomic feature of the URMC CICU outbreak strains and a
321 subset of non-outbreak associated strains was the presence of the yersiniabactin
322 siderophore and colibactin systems. Using Kleborate (17), we investigated the

distribution and organization of these systems in the global *K. aerogenes* genomes to identify associations, if any, with specific ST types and geographical regions (Fig. 5, Datasets 6). The prevalence of yersiniabactin and colibactin encoding systems in the global *K. aerogenes* genomes was found to be 53.64% (59/110) and 52.72% (58/111) respectively. A 100 % association was found between the presence of colibactin in the genomes and concurrent presence of yersiniabactin. Higher prevalence of the virulence cluster was observed in ST4 and ST93 types; 85% (12/14) and 95% (42/44) respectively, although these two ST types were also the most abundantly represented in the available set of genomes. The other ST types were less well represented in the study set (n<3), so the prevalence of these systems in them could not be accurately established. Interestingly, the strains exclusively designated 'environmental' isolates (B3, FGI35, and B) did not harbor genes encoding the above-mentioned virulence systems (Fig. 5, Dataset 6). These trends correlated with analyses of our 26 hospital strains, where 75% of the ST 93 strains (3/4) were positive for yersiniabactin and colibactin systems, while none of the strains with novel/unassigned ST types harbored genes encoding the same (0%, 5/5).

DISCUSSION

Whole genome sequencing presents a powerful resource that can be deployed for prospective and comprehensive outbreak investigations offering tremendous resolution in tracking transmission events and delineating genomic determinants associated with drug resistance and virulence (20). This WGS study was initiated in order to establish the molecular epidemiology of CR-KA strains isolated from patients in

346 our hospital. Beyond the outbreak investigation, whole genome data enabled us to
347 address additional fundamental questions associated with carbapenem resistance,
348 virulence attributes and population structure of *K. aerogenes* that are poorly understood.
349 Fifteen patients were associated with the CICU outbreak event in the period
350 between July-November 2017. Based on the differences in AST patterns in a subset of
351 study strains, it was initially questioned whether the strains were clonal. WGS
352 investigations demonstrated that the 15 CICU CR-KA isolates differed from each other
353 by less than 7 SNPs and grouped distantly from the other hospital CR-KA isolates,
354 which were used a baseline for context and comparison (Fig. 2). Based on these
355 findings, the CICU CR-KA strains were concluded to be part of a single clonal cluster
356 indicating protracted intra-ward transmission. Environmental sampling within the CICU
357 remained negative and a common source of the outbreak or contributory risk factors
358 could not be identified. After the last positive patient surveillance culture (Nov 2017), no
359 CR-KA strains were isolated in a 2-month period of continued surveillance and the
360 outbreak was deemed to have subsided. Phylogenomic analysis (albeit limited) of the
361 non-CICU CR-KA isolates did not identify specific dominant clones circulating in the
362 hospital (Fig. 2, Dataset 2).
363 Infections due to carbapenem-resistant organisms present complex diagnostic
364 and therapeutic management challenges and a better understanding of how adaptive or
365 acquired resistance emerges in the healthcare environment is needed (21). In clinical
366 CR-KA strains, carbapenem resistance has been associated with carbapenemase
367 production or adaptive mutations following antibiotic exposure (2). Among our 24 study
368 CR-KA isolates, a single non-CICU strain harbored a carbapenemase-encoding gene,

369 *bla*$_{NMC-A}$. NmcA has been reported in *E. cloacae* (22), but to our knowledge this is the
370 first report of this carbapenemase being identified in *K. aerogenes*. All but one of the 24
371 CR-KA strains were found to harbor mutations in genes involved in the synthesis or
372 regulation of the inducible AmpC cephalosporinase (*ampD, ampR*) and outer membrane
373 porins (*omp 35, omp36*). A majority of mutations were found to be within the open-
374 reading frames of *ampD* and *omp36* (catalogued in Tables 2 and 3).
375 There are limited reports regarding the role of AmpD in adaptive carbapenem
376 resistance in *K. aerogenes* strains (23, 24), although the role of AmpD in AmpC
377 expression in *E. cloacae* has been well established (25). *In silico* modeling of *K.*
378 *aerogenes* AmpD using the *C. freundii* AmpD crystal structure (15) predicted that
379 Arg161His and Trp95Leu substitutions due to SNPs in the outbreak strains would likely
380 impact enzymatic activity (Fig. 3). Other *ampD* mutations in several non-CICU CR-KA
381 isolates were found to result in substitutions that could affect substrate-enzyme
382 interactions (Table 2).
383 Mutations in *omp36* have been described in clinical CR-KA strains and functional
384 studies investigating their impact have been reported (6, 26). A diverse array of
385 mutations was identified in the *omp36* gene among the CR-KA study strains, including
386 non-synonymous mutations resulting in frame-shifts or premature stop codons resulting
387 in truncated and likely non-functional Omp36 variants (Table 3). Additional SNPs
388 relative to the carbapenem-susceptible reference genome that resulted in substitutions
389 of unknown significance in predicted β-sheets and extracellular loop regions (Table 3,
390 Fig. S2).

Even within the CICU CR-KA clonal cluster, heterogeneity was identified in *ampD* and *omp36* genes (Table 2, Table 3). These genetic loci likely represent mutational hotspots associated with adaptive carbapenem resistance in *K. aerogenes.* It is likely that the CICU CR-KA cluster represents a population of clonal origin, albeit with micro-heterogeneity in the above regions. The testing and archiving of single isolated carbapenem resistant colonies instead of multiples during individual patient specimen workup likely represents a limitation that did not allow us to capture the full complement of CR-KA strain microdiversity associated with individual patients during the outbreak event.

The detailed significance of the alleles and mutations described above in development of carbapenem resistance in *K. aerogenes* needs to be verified by additional genetic (allelic exchange and complementation) and biochemical approaches. Moreover, complex regulatory networks including transcriptional activators, sensor kinases and two-component systems have been implicated in adaptive drug resistance in *Enterobacteriaceae spp*. (13), and their roles in contributing towards carbapenem resistance in our study strains cannot be ruled out.

Despite their lack of horizontally acquired drug resistance elements, the outbreak strains in this study managed to persist for several months in the face of an active infection prevention effort prompting us to assess their virulence determinants. The outbreak strains and a subset of non-outbreak CR-KA strains harbored an ICE encoding the metallophore yersiniabactin (Ybt) and genotoxin colibactin (Clb) systems (Fig. 4, Table S4). A Ybt encoding pathogenicity island has been described in association with a prolonged nation-wide outbreak in the Netherlands involving multiple

414 hospitals and >100 patients due to a multi-drug resistant *Enterobacter hormaechei*
415 clone (27). A recent study by Lam *et al.* described the prevalence of *ybt* to be
416 particularly high (>80%) in certain hypervirulent *K. pneumoniae* clonal-groups (CG23)
417 (17). The study also reported a significant association of Ybt with an increased risk of
418 invasive infections (bacteremia, liver abscesses etc.). Ybt was first described in
419 pathogenic *Yersinia spp.*, encoded by a chromosomal gene cluster termed the High
420 Pathogenicity Island (HPI), with a critical role in iron scavenging during infection (28).
421 Subsequent studies reported acquisition of the HPI in other clinical *Enterobacteriaceae*
422 strains (17, 29) and additional functions ascribed to Ybt include evasion of host
423 lipocalin-2 (30) and sequestration/import of heavy metals (31). The polyketide colibactin
424 is frequently associated with yersiniabactin and has been shown to induce
425 chromosomal instability and DNA damage in eukaryotic cells (32). Our findings present
426 new avenues for research investigating the role of ICE elements encoding Ybt and Clb
427 in the virulence of clinical *K. aerogenes* strains. While horizontal transmission of
428 carbapenemases via mobile elements is increasingly recognized as a major public
429 health issue (33), the transmission of mobilizable virulence factors present an
430 underappreciated threat in the healthcare environment that warrants more surveillance.
431 In order to set our hospital strains (outbreak and non-outbreak) into a broader
432 context, core-genome comparisons and MLST were used to examine the population
433 structure of our strains relative to global *K. aerogenes* strains pulled from public
434 databases (Fig. 5, Datasets 5, 6). This analysis is the first evaluation of the nascent *K.*
435 *aerogenes* MLST scheme in discriminating clinical *K. aerogenes* isolates. The scheme
436 was found to be robust with distribution of STs correlating closely with topologies based

437 on *K. aerogenes* strains core-genomes. Our preliminary analyses suggest that ST4 and
438 ST93 might be dominant global clones associated with *K. aerogenes* infections.
439 Outbreak CR-KA strains clustered closest to other clinical US ST4 strains, suggesting a
440 clonal expansion of this ST (Fig. 5). It is noteworthy that the ST4 group also included
441 carbapenemase-producing *K. aerogenes* isolates from international sites. These
442 included CR-KA strains associated with drug-resistant intra-abdominal and urinary tract
443 infections in patient samples from Brazil, Canada, Colombia and China that had been
444 sequenced as a part of large surveillance study, SMART (Study for Monitoring
445 Antimicrobial Resistance Trends) (34). ST93 was the most prevalent sequence type in
446 the global *K. aerogenes* assembled genomes (43/110, 39%), with a wide geographical
447 distribution including the US (Fig. 5). Four of the eleven non-CICU CR-KA strains from
448 our hospital belonged to this group. Two closely related ST93 isolates, *K. aerogenes*
449 1509E and G7, have been described as representatives of clonal strains associated
450 with multiple multi-drug resistant *K. aerogenes* outbreaks in France (35, 36).
451 Incidentally, global ST4 and ST93 isolates also had a higher prevalence of HPI
452 encoding *ybt* and *clb* (85% and 95% respectively). The success of these potentially
453 "high-risk" clones needs to be examined more closely by undertaking large-scale
454 studies with strains from diverse global sites and patient populations. These studies will
455 help examine niche adaptation, emergence of antibiotic resistance and evolution of
456 pathogenicity leading to a better understanding of *K. aerogenes.*

457 In summary, genomic approaches for surveillance and outbreak investigations
458 are emerging as critical functions for infection prevention and diagnostic microbiology
459 laboratories. Apart from evaluating the effectiveness of infection measures and

dimension of transmission events, WGS applied across sets of isolates is a powerful tool for assessing virulence factors and identifying clinically relevant and emerging sequence types.

MATERIALS AND METHODS

Setting, study design, *K. aerogenes* strains and metadata.

The University of Rochester Medical Center (URMC) is an 830-bed tertiary-care medical center, with a 14-bed cardiothoracic intensive care unit, serving the Greater Rochester Area, New York. Following approval by the University of Rochester Institutional Review Board (RSRB00068143), a total of 26 *K. aerogenes* strains isolated from patients at URMC in the course of regular clinical care and/or surveillance efforts were selected for the study. Each isolate corresponded to a single first CR-KA strain isolated during the course of hospitalization. For context and comparison, an additional set of *K. aerogenes* strains epidemiologically unlinked to the CICU outbreak was included in the study. Ward occupancy and pertinent clinical and epidemiological information were obtained through review of patient medical records and the laboratory information system, and are described in Tables S1 and S2.

Antibiotic susceptibility testing (AST) of the study *K. aerogenes* strains

AST of the study strains was performed as part of routine diagnostics using the VITEK®2 (BioMérieux, France) system and/or Kirby Bauer disk-diffusion methods. AST interpretations were based on interpretive criteria defined by the Clinical and Laboratory Standards Institute (M100 document, 27th Edition).

483 **Sequencing library preparation and raw data acquisition**

484 The isolates were cultured on standard laboratory media from archived frozen stocks,
485 examined for purity, and re-identified by Vitek MALDI-TOF MS (BioMérieux, France).
486 Single colony genomic DNA extractions were performed using MagNA Pure Compact
487 instrument (Roche Diagnostics, Indianapolis, IN); the DNA quality was analyzed using
488 QuantiFluor dsDNA system (Promega). Nextera XT kit (Illumina, San Diego, CA) was
489 used for preparing dual-indexed WGS libraries. For QC dsDNA quality of individual
490 samples was analyzed using the 4200 TapeStation system (Agilent). Following library
491 cleanup, individual libraries were pooled at equimolar ratios and denatured; DNA
492 concentration was determined using Qubit ssDNA kit (Thermo fisher Scientific). The
493 libraries were combined with20ρM PhiX control and sequenced on the Illumina Miseq™
494 benchtop sequencer (Illumina, San Diego, CA) at the URMC Genomic Core Facilities,
495 using V3 kit and 2X300 bp paired-end protocol.
496
497 **Genomic analyses**
498 Analyses were performed using an in-house bioinformatics pipeline 'URMC Bacterial
499 Genomic Analysis Pipeline' (v2.0.6), run on a high-performance computer cluster at the
500 Center for Integrated Research Computing at University of Rochester. **A) Quality**
501 **control:** For each sample, the read quality scores across all bases were assessed
502 using FastQC v0.11.5 (37) and low quality reads were trimmed using Trimmomatic
503 v0.36 (38). Genomic coverage of the sequencing reads relative to the reference *K.*
504 *aerogenes* KCTC 2190 genome (ATCC 13048^T; Refseq accession
505 number: NC_015663.1) were determined using Bowtie 2 v2.2.9 (39). Across the study

isolates, alignments showed average genomic coverage of 90.84% (minimum 89%) with an average depth of coverage: 100X (excluding regions below 12X). Quality metrics are detailed in Dataset 1. **B) Core-genome SNP calling workflow:** A read mapping approach was used to assess SNPs in the genomes of the study CR-KA strains relative to the reference genome *K. aerogenes* KCTC 2190. Mapping, variant calling and phylogenetic analysis were performed by locally installed CFSAN SNP Pipeline v1.0.0 (40), an analysis workflow developed by the U.S Food and Drug Administration (FDA). CFSAN pipeline employs a 2-phase variant calling workflow and the 'optimized' version of the pipeline with criteria as applied by Saltykova et al (higher stringency in allele frequency thresholds and coverage) (41) was applied. In the first phase, variants were called based on *mpileup function* of SAMTools and *mpileup2snp tool* from VarScan (minimum average base quality=20, minimum read depth of coverage at site=12, minimum allele frequency=90%). Densely clustered SNPs that could arise due to recombination were excluded (>/=3 SNPs in a 50 bp window). High-confidence SNP variants meeting the criteria were composed to a list. In the 2nd phase, nucleotide sites at the listed positions were determined for all sites (minimum allele frequency threshold for SNP filtering=90%). SNPs located in mobile elements as annotated by RAST were excluded. The final nucleotide sites post filtering corresponding to the listed positions were used to build a SNP matrix. A multi FASTA file with concatenated SNP matrix entries were used for inferring phylogeny by FastTree (42), which were visualized using ITOL (43). **C) Sequence assembly, scaffolding, annotation and analyses:** De novo genome assemblies (average N_{50}: 344943 across all isolates) were generated by SPAdes genome Assembler v3.11.1 (44), and the assembly quality was assessed by

529 QUAST v4.5 (45). Ordering and orientation of contigs was performed using Medusa .
530 Draft genomes were annotated with RAST (46) and Prokka (47). MLST on the *K.*
531 *aerogenes* genomes was performed using a newly developed publicly available scheme
532 (10). Alleles and genetic markers corresponding to acquired antibiotic resistance,
533 virulence and plasmid replicons were identified using PlasmidFinder (48), ResFinder
534 (49), and Virulence Factor Database (50). Typing of genetic loci associated with
535 mobilizable yersiniabactin siderophore and genotoxin colibactin systems was performed
536 using Kleborate (17), and the loci were visualized with MacVector software (MacVector,
537 Cary, NC). Sequence variations in chromosomal genes associated with carbapenem
538 resistance in the study strains were assessed relative to alleles in control strains
539 (URMC 201, URMC 223) and 'wild-type' alleles in carbapenem susceptible (14)
540 reference genome KCTC 2190 [Genbank ID-*ampD*: 10792472 (EAE_11350), *ampG*:
541 10792757 (EAE_12735), *ampR*: 10792632 (EAE_12115), *omp35*: 10793271
542 (EAE_15245), *omp36*: 10795060 (EAE_24205) and *ompR*: 10791222 (EAE_05245)].
543 Multiple sequence alignments of gene alleles and corresponding proteins was
544 performed using Vector NTI software (Invitrogen, Carlsbad, CA) and the alignments
545 were manually inspected to identify substitutions that would likely impact function. **D)**
546 **Comparative genomics of publicly available global *K. aerogenes* assembled**
547 **genomes with URMC CR-KA assemblies**: Using a custom shell script, publicly
548 available *K. aerogenes* genomes available as of July 2018 in the NCBI genome
549 database were downloaded from the NCBI FTP site. Harvest genomics suite was used
550 to perform intraspecific core-genome alignments as described before (18); phylogenies
551 were visualized using ITOL (43). **E) Accession number(s):** WGS and metadata

552 corresponding to the study URMC *K. aerogenes* isolates were deposited at NCBI under
553 BioProject accession number PRJNA504784. Accession numbers for individual isolates
554 are listed in Table S1.
555
556 ***In silico* protein analyses**
557 For homology modeling, SWISS-MODEL (51) was used to thread *K. aerogenes*
558 AmpD (acc WP_015704411.1, aa1-187) through the structure of *C. freundii* AmpD (15)
559 (PDB 2y2c, aa1-187). The overall quaternary structure of *K. aerogenes* AmpD was
560 predicted with high precision (95% confidence, 99% coverage). Comparative analyses
561 and imaging of protein structures were performed with PyMOL (52). JalView was used
562 to create alignments (53).
563
564 **ACKNOWLEDGEMENTS**
565 We are grateful to the URMC Clinical Microbiology Laboratories and Infection
566 Prevention staff in specimen processing, data collection and epidemiological
567 investigations. We wish to acknowledge the URMC Genomics Research Center for
568 support with WGS. We also thank Steve Gill (URMC, Genomics Research Center) for
569 reviewing the manuscript draft. Internal funding from University of Rochester
570 Department of Pathology and Laboratory Medicine supported this study. We report no
571 conflicts of interest relevant to this article.
572
573
574
575

576 **REFERENCES**

577

578 1. **Sanders WE, Jr., Sanders CC.** 1997. *Enterobacter spp*.: pathogens poised to
579 flourish at the turn of the century. Clin Microbiol Rev **10:**220-241.
580 2. **Davin-Regli A, Pages JM.** 2015. *Enterobacter aerogenes* and *Enterobacter*
581 *cloacae*; versatile bacterial pathogens confronting antibiotic treatment. Front
582 Microbiol **6:**392.
583 3. **Guh AY, Bulens SN, Mu Y, Jacob JT, Reno J, Scott J, Wilson LE, Vaeth E,**
584 **Lynfield R, Shaw KM, Vagnone PM, Bamberg WM, Janelle SJ, Dumyati G,**
585 **Concannon C, Beldavs Z, Cunningham M, Cassidy PM, Phipps EC, Kenslow**
586 **N, Travis T, Lonsway D, Rasheed JK, Limbago BM, Kallen AJ.** 2015.
587 Epidemiology of Carbapenem-Resistant *Enterobacteriaceae* in 7 US
588 Communities, 2012-2013. JAMA **314:**1479-1487.
589 4. **Lee HJ, Choi JK, Cho SY, Kim SH, Park SH, Choi SM, Lee DG, Choi JH, Yoo**
590 **JH.** 2016. Carbapenem-resistant *Enterobacteriaceae:* Prevalence and Risk
591 Factors in a Single Community-Based Hospital in Korea. Infect Chemother
592 **48:**166-173.
593 5. **Robert J, Pantel A, Merens A, Lavigne JP, Nicolas-Chanoine MH, Group**
594 **ONsCRS.** 2014. Incidence rates of carbapenemase-producing
595 *Enterobacteriaceae* clinical isolates in France: a prospective nationwide study in
596 2011-12. J Antimicrob Chemother **69:**2706-2712.
597 6. **Philippe N, Maigre L, Santini S, Pinet E, Claverie JM, Davin-Regli AV, Pages**
598 **JM, Masi M.** 2015. In Vivo Evolution of Bacterial Resistance in Two Cases of

Enterobacter aerogenes Infections during Treatment with Imipenem. PLOS One **10:**e0138828.

7. **Boucher HW, Talbot GH, Bradley JS, Edwards JE, Gilbert D, Rice LB, Scheld M, Spellberg B, Bartlett J.** 2009. Bad bugs, no drugs: no ESKAPE! An update from the Infectious Diseases Society of America. Clin Infect Dis **48:**1-12.

8. **Wyres KL, Holt KE.** 2016. *Klebsiella pneumoniae* Population Genomics and Antimicrobial-Resistant Clones. Trends Microbiol **24:**944-956.

9. **Gomez-Simmonds A, Annavajhala MK, Wang Z, Macesic N, Hu Y, Giddins MJ, O'Malley A, Toussaint NC, Whittier S, Torres VJ, Uhlemann AC.** 2018. Genomic and Geographic Context for the Evolution of High-Risk Carbapenem-Resistant *Enterobacter cloacae* Complex Clones ST171 and ST78. MBio **9**.

10. https://pubmlst.org/kaerogenes/

11. **Jacoby GA.** 2009. AmpC beta-lactamases. Clin Microbiol Rev **22:**161-182.

12. **Dupont H, Choinier P, Roche D, Adiba S, Sookdeb M, Branger C, Denamur E, Mammeri H.** 2017. Structural Alteration of OmpR as a Source of Ertapenem Resistance in a CTX-M-15-Producing *Escherichia coli* O25B:H4 Sequence Type 131 Clinical Isolate. Antimicrob Agents Chemother **61**.

13. **Pages JM, James CE, Winterhalter M.** 2008. The porin and the permeating antibiotic: a selective diffusion barrier in Gram-negative bacteria. Nat Rev Microbiol **6:**893-903.

14. **Lavigne JP, Sotto A, Nicolas-Chanoine MH, Bouziges N, Bourg G, Davin-Regli A, Pages JM.** 2012. Membrane permeability, a pivotal function involved in

621 antibiotic resistance and virulence in *Enterobacter aerogenes* clinical isolates.
622 Clin Microbiol Infect **18:**539-545.
623 15. **Carrasco-Lopez C, Rojas-Altuve A, Zhang W, Hesek D, Lee M, Barbe S,**
624 **Andre I, Ferrer P, Silva-Martin N, Castro GR, Martinez-Ripoll M, Mobashery**
625 **S, Hermoso JA.** 2011. Crystal structures of bacterial peptidoglycan amidase
626 AmpD and an unprecedented activation mechanism. J Biol Chem **286:**31714-
627 31722.
628 16. **Thiolas A, Bornet C, Davin-Regli A, Pages JM, Bollet C.** 2004. Resistance to
629 imipenem, cefepime, and cefpirome associated with mutation in Omp36
630 osmoporin of *Enterobacter aerogenes*. Biochem Biophys Res Commun **317:**851-
631 856.
632 17. **Lam MMC, Wick RR, Wyres KL, Gorrie CL, Judd LM, Jenney AWJ, Brisse S,**
633 **Holt KE.** 2018. Genetic diversity, mobilisation and spread of the yersiniabactin-
634 encoding mobile element ICE*Kp* in *Klebsiella pneumoniae* populations. Microb
635 Genom.
636 18. **Treangen TJ, Ondov BD, Koren S, Phillippy AM.** 2014. The Harvest suite for
637 rapid core-genome alignment and visualization of thousands of intraspecific
638 microbial genomes. Genome Biol **15:**524.
639 19. **Cerqueira GC, Earl AM, Ernst CM, Grad YH, Dekker JP, Feldgarden M,**
640 **Chapman SB, Reis-Cunha JL, Shea TP, Young S, Zeng Q, Delaney ML, Kim**
641 **D, Peterson EM, O'Brien TF, Ferraro MJ, Hooper DC, Huang SS, Kirby JE,**
642 **Onderdonk AB, Birren BW, Hung DT, Cosimi LA, Wortman JR, Murphy CI,**
643 **Hanage WP.** 2017. Multi-institute analysis of carbapenem resistance reveals

remarkable diversity, unexplained mechanisms, and limited clonal outbreaks. Proc Natl Acad Sci U S A **114:**1135-1140.

20. **Quainoo S, Coolen JPM, van Hijum S, Huynen MA, Melchers WJG, van Schaik W, Wertheim HFL.** 2017. Whole-Genome Sequencing of Bacterial Pathogens: the Future of Nosocomial Outbreak Analysis. Clin Microbiol Rev **30:**1015-1063.

21. **Goodman KE, Simner PJ, Tamma PD, Milstone AM.** 2016. Infection control implications of heterogeneous resistance mechanisms in carbapenem-resistant *Enterobacteriaceae* (CRE). Expert Rev Anti Infect Ther **14:**95-108.

22. **Pottumarthy S, Moland ES, Juretschko S, Swanzy SR, Thomson KS, Fritsche TR.** 2003. NmcA carbapenem-hydrolyzing enzyme in *Enterobacter cloacae* in North America. Emerging infectious diseases **9:**999-1002.

23. **Tzouvelekis LS, Tzelepi E, Kaufmann ME, Mentis AF.** 1994. Consecutive mutations leading to the emergence in vivo of imipenem resistance in a clinical strain of *Enterobacter aerogenes*. J Med Microbiol **40:**403-407.

24. **Babouee Flury B, Ellington MJ, Hopkins KL, Turton JF, Doumith M, Woodford N.** 2016. The differential importance of mutations within AmpD in cephalosporin resistance of *Enterobacter aerogenes* and *Enterobacter cloacae*. Int J Antimicrob Agents **48:**555-558.

25. **Babouee Flury B, Ellington MJ, Hopkins KL, Turton JF, Doumith M, Loy R, Staves P, Hinic V, Frei R, Woodford N.** 2016. Association of Novel Nonsynonymous Single Nucleotide Polymorphisms in *ampD* with Cephalosporin Resistance and Phylogenetic Variations in *ampC*, *ampR*, *ompF*, and *ompC* in

Enterobacter cloacae Isolates That Are Highly Resistant to Carbapenems. Antimicrob Agents Chemother **60:**2383-2390.

26. **De Gheldre Y, Maes N, Rost F, De Ryck R, Clevenbergh P, Vincent JL, Struelens MJ.** 1997. Molecular epidemiology of an outbreak of multidrug-resistant *Enterobacter aerogenes* infections and in vivo emergence of imipenem resistance. J Clin Microbiol **35:**152-160.

27. **Paauw A, Caspers MP, Leverstein-van Hall MA, Schuren FH, Montijn RC, Verhoef J, Fluit AC.** 2009. Identification of resistance and virulence factors in an epidemic *Enterobacter hormaechei* outbreak strain. Microbiology **155:**1478-1488.

28. **Heesemann J, Hantke K, Vocke T, Saken E, Rakin A, Stojiljkovic I, Berner R.** 1993. Virulence of *Yersinia enterocolitica* is closely associated with siderophore production, expression of an iron-repressible outer membrane polypeptide of 65,000 Da and pesticin sensitivity. Molecular Microbiology **8:**397-408.

29. **Putze J, Hennequin C, Nougayrede JP, Zhang W, Homburg S, Karch H, Bringer MA, Fayolle C, Carniel E, Rabsch W, Oelschlaeger TA, Oswald E, Forestier C, Hacker J, Dobrindt U.** 2009. Genetic structure and distribution of the colibactin genomic island among members of the family *Enterobacteriaceae*. Infect Immun **77:**4696-4703.

30. **Bachman MA, Lenio S, Schmidt L, Oyler JE, Weiser JN.** 2012. Interaction of lipocalin 2, transferrin, and siderophores determines the replicative niche of *Klebsiella pneumoniae* during pneumonia. MBio **3**.

688 31. **Robinson AE, Lowe JE, Koh EI, Henderson JP.** 2018. Uropathogenic
689 enterobacteria use the yersiniabactin metallophore system to acquire nickel. J
690 Biol Chem.
691 32. **Fais T, Delmas J, Barnich N, Bonnet R, Dalmasso G.** 2018. Colibactin: More
692 Than a New Bacterial Toxin. Toxins (Basel) **10**.
693 33. **Gupta N, Limbago BM, Patel JB, Kallen AJ.** 2011. Carbapenem-resistant
694 *Enterobacteriaceae*: epidemiology and prevention. Clin Infect Dis **53:**60-67.
695 34. **Morrissey I, Hackel M, Badal R, Bouchillon S, Hawser S, Biedenbach D.**
696 2013. A Review of Ten Years of the Study for Monitoring Antimicrobial
697 Resistance Trends (SMART) from 2002 to 2011. Pharmaceuticals (Basel)
698 **6:**1335-1346.
699 35. **Diene SM, Merhej V, Henry M, El Filali A, Roux V, Robert C, Azza S, Gavory**
700 **F, Barbe V, La Scola B, Raoult D, Rolain JM.** 2013. The rhizome of the
701 multidrug-resistant *Enterobacter aerogenes* genome reveals how new "killer
702 bugs" are created because of a sympatric lifestyle. Mol Biol Evol **30:**369-383.
703 36. **Thiolas A, Bollet C, La Scola B, Raoult D, Pages JM.** 2005. Successive
704 emergence of *Enterobacter aerogenes* strains resistant to imipenem and colistin
705 in a patient. Antimicrob Agents Chemother **49:**1354-1358.
706 37. http://www.bioinformatics.babraham.ac.uk/projects/fastqc/ .
707 38. **Bolger AM, Lohse M, Usadel B.** 2014. Trimmomatic: a flexible trimmer for
708 Illumina sequence data. Bioinformatics **30:**2114-2120.
709 39. **Langmead B, Salzberg SL.** 2012. Fast gapped-read alignment with Bowtie 2.
710 Nat Methods **9:**357-359.

40. **Davis S, Pettengill JB, Luo Y, Payne J, Shpuntoff A, Rand H, Strain E.** 2015. CFSAN SNP Pipeline: an automated method for constructing SNP matrices from next-generation sequence data. PeerJ Computer Science **1:**e20.

41. **Saltykova A, Wuyts V, Mattheus W, Bertrand S, Roosens NHC, Marchal K, De Keersmaecker SCJ.** 2018. Comparison of SNP-based subtyping workflows for bacterial isolates using WGS data, applied to *Salmonella enterica* serotype Typhimurium and serotype 1,4,[5],12:i. PLOS One **13:**e0192504.

42. **Price MN, Dehal PS, Arkin AP.** 2010. FastTree 2-approximately maximum-likelihood trees for large alignments. PLOS One **5:**e9490.

43. **Letunic I, Bork P.** 2016. Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. Nucleic Acids Res **44:**W242-245.

44. **Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA.** 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol **19:**455-477.

45. **Gurevich A, Saveliev V, Vyahhi N, Tesler G.** 2013. QUAST: quality assessment tool for genome assemblies. Bioinformatics **29:**1072-1075.

46. **Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA, Formsma K, Gerdes S, Glass EM, Kubal M, Meyer F, Olsen GJ, Olson R, Osterman AL, Overbeek RA, McNeil LK, Paarmann D, Paczian T, Parrello B, Pusch GD, Reich C, Stevens R, Vassieva O, Vonstein V, Wilke A, Zagnitko O.** 2008. The

RAST Server: rapid annotations using subsystems technology. BMC genomics **9:**75.

47. **Seemann T.** 2014. Prokka: rapid prokaryotic genome annotation. Bioinformatics **30:**2068-2069.

48. **Carattoli A, Zankari E, Garcia-Fernandez A, Voldby Larsen M, Lund O, Villa L, Moller Aarestrup F, Hasman H.** 2014. In silico detection and typing of plasmids using PlasmidFinder and plasmid multilocus sequence typing. Antimicrob Agents Chemother **58:**3895-3903.

49. **Zankari E, Hasman H, Cosentino S, Vestergaard M, Rasmussen S, Lund O, Aarestrup FM, Larsen MV.** 2012. Identification of acquired antimicrobial resistance genes. J Antimicrob Chemother **67:**2640-2644.

50. **Chen L, Zheng D, Liu B, Yang J, Jin Q.** 2016. VFDB 2016: hierarchical and refined dataset for big data analysis--10 years on. Nucleic Acids Res **44:**D694-697.

51. **Waterhouse A, Bertoni M, Bienert S, Studer G, Tauriello G, Gumienny R, Heer FT, de Beer TAP, Rempfer C, Bordoli L, Lepore R, Schwede T.** 2018. SWISS-MODEL: homology modelling of protein structures and complexes. Nucleic Acids Res **46:**W296-W303.

52. **Seeliger D, de Groot BL.** 2010. Ligand docking and binding site analysis with PyMOL and Autodock/Vina. J Comput Aided Mol Des **24:**417-422.

53. **Waterhouse AM, Procter JB, Martin DM, Clamp M, Barton GJ.** 2009. Jalview Version 2--a multiple sequence alignment editor and analysis workbench. Bioinformatics **25:**1189-1191.

FIGURE LEGENDS

Fig. 1. Patient occupancy and overlap in CICU ward during the *K. aerogenes* outbreak. Black bars-first positive CR-KA clinical cultures, blue bars-first positive CR-KA surveillance cultures

Fig. 2. Dendogram showing pairwise SNP differences based phylogenetic relatedness of URMC *K. aerogenes* strains. Whole genome sequence of *K. aerogenes* KCTC 2190 (ATCC 13048) was used for reference mapping. A total of 28,170 discriminatory high-quality SNPs in the core genomes obtained by the CFSAN SNP pipeline were used to plot the tree (excluding mobile elements and putative recombination sites). CICU outbreak strains: IDs highlighted in red, clinical isolates in bold, patient A isolate (1st case) with yellow background. Year strain isolated: orange-2015, yellow-2016, blue-2017. SNP differences relative to patient A shown. Scale bar indicates nucleotide substitutions per site.

Fig. 3. Computational modeling of the impact of *ampD* mutations on AmpD in URMC outbreak CR-KA strains. A) *K. aerogenes* AmpD modeled on *C. freundii* AmpD structure. Key residues interacting with glycan and peptide portions of ligand are shown. B, C) Surface models of *K. aerogenes* AmpD depicting the wild-type binding surface for the diaminopimelate moiety (W95, top) versus the binding surface of AmpD containing the W95L mutation (blue surfaces indicate amino acid positions from panel A; bottom, altered surface highlighted in red).

Fig. 4. Yersiniabactin and colibactin encoding gene loci on integrative conjugative element ICE*Kp*10 in the CICU outbreak CR-KA strains. Blue arrows-integrase encoding genes, brown arrows-$Zn^{2+/}Mn^{2+}$modules grey arrows-genes encoding mobilization proteins, pink arrows: vir-T4SS system.

Fig. 5. Harvest based phylogenomic comparisons of URMC *K. aerogenes* genomes with global *K. aerogenes* genomes. Discriminatory SNPs based on core genome comparisons were used to plot the tree. URMC study *K. aerogenes* strains in purple (outbreak strains in bold, index patient strain highlighted yellow). Presence of genes encoding yersiniabactin siderophore system (green squares), colibactin synthesis cluster (blue triangles), and carbapenamases (red circles) in assembled genomes shown. Scale bar indicates nucleotide substitutions per site.

Table 1. Phenotypic antibiotic susceptibility profiles of *K. aerogenes* strains in this study

	Study Isolates	Amikacin	Gentamicin	Tobramycin	Ciprofloxacin	Moxifloxacin	Trimethoprim-Sulfa	Piperacillin-Tazobactam	Ceftriaxone	Cefepime	Ertapenem	Imipenem	Meropenem
2017 URMC CICU associated CR-KA strains	**URMC 205***	S	S	S	S	S	S	R	R	S	R	R	R
	URMC 206	S	S	S	S	S	S	R	R	S	R	R	R
	URMC 207	S	S	S	S	S	S	R	R	S	R	R	R
	URMC 208	S	S	S	S	S	S	R	R	S	S	S	R
	URMC 209	S	S	S	S	S	S	R	R	S	R	R	R
	URMC 211	S	S	S	S	S	S	R	R	D	R	R	R
	URMC 212	S	S	S	S	S	S	R	R	R	R	I	R
	URMC 213	S	S	S	S	S	S	R	R	R	R	I	R
	URMC 215	S	S	S	S	S	S	R	R	S	R	R	R
	URMC 216	S	S	S	S	S	S	R	R	S	R	S	S
	URMC 218	S	S	S	S	S	S	R	R	S	R	R	R
	URMC 219	S	S	S	S	S	S	R	R	S	R	R	R
	URMC 224	nd	S	S	S	S	S	R	R	R	R	nd	R
	URMC 225	S	S	S	S	S	S	R	R	R	R	R	R
	URMC 226	S	S	S	S	S	S	R	R	R	R	I	I
Other URMC CR-KA strains	URMC 200	S	S	S	S	nd	S	R	R	R	R	nd	R
	URMC 202	S	S	S	S	S	S	R	R	S	R	R	R
	URMC 203	S	S	S	S	S	S	R	R	S	R	R	R
	URMC 204	S	S	S	S	nd	S	R	R	S	R	nd	R
	URMC 210	S	S	S	S	nd	S	R	R	S	R	nd	S
	URMC 214	S	S	S	S	nd	S	I	R	S	R	I	S
	URMC 217	S	S	S	R	nd	S	R	R	S	R	I	S
	URMC 221	S	S	S	S	nd	S	S	S	S	R	nd	S
	URMC 222	S	S	S	S	S	S	R	R	R	R	I	S
Con	URMC 201	S	S	S	S	S	S	S	S	S	S	I	S
	URMC 223	S	S	S	S	nd	S	S	S	S	S	nd	S

R=Resistant, S= Sensitive, I =Intermediate, D= Dose Dependent, nd=not determined; *Patient A (1st case)-URMC 205; Con: control strains: URMC 223, carbapenem susceptible and URMC 201, intermediate carbapenem resistance

Table 2. Non-synonymous SNPs in *ampD* gene associated with carbapenem resistance in the URMC *K. aerogenes* isolates

Study Isolates		Differences in *ampD* gene relative to wild-type allele		Effect on encoded AmpD sequence	
		MS SNPs	NS SNPs	Single AA substitutions	PS
CICU outbreak associated CR-KA strains	URMC 205*, URMC 207, URMC 209, URMC 211, URMC 212, URMC 213, URMC 215, URMC 218, URMC 219, URMC 224, URMC 224, URMC 226	482G>A		Arg161His	
	URMC 206, URMC 208, URMC 216	284G>T		Trp95Leu	
Outbreak unrelated CR-KA strains	URMC 200	338T>G		Ile113Ser	
	URMC 202		412C>T		Gln138X
	URMC 204	117C>T		Pro39Ser	
	URMC 210	335C>T		Ser112Leu	
	URMC 214	280G>A		Ala94Thr	
	URMC 217	496G>C		Gly166Arg	
	URMC 221	478A>G, 501C>A		Ile160Val, Ala168Asp	
	URMC 222	492T>G		Glu164Asp	
Control strains	URMC 201, URMC 223				

No sequencing reads mapped to the *ampD* gene in URMC 203.

*Index patient isolate.

Abbr: SNP: single nucleotide polymorphism, MS: missense, NS: nonsense, AA: amino acids, PS: premature stop codon

Table 3. Non-synonymous genetic lesions in *omp36* genes and resulting alterations in Omp36 sequence in the URMC *K. aerogenes* isolates

Study Isolates		Differences in *omp36* gene sequence relative to wild-type allele				Effect on encoded Omp36 amino acid sequence relative to wild-type sequence				
		FS	HV region (Nt 680-724)	NS SNPs	MS SNPs	FS	HV region (AA 226-241)	PS	Truncated	Single AA substitions in protein synthesized
CICU outbreak associated CR-KA strains	URMC 205*, URMC 207	+	+		175A>G, 564T>C, 566A>G	pAsp91ThrfsX12		102X	+	Ile59Val
	URMC 206, URMC 208, URMC 212, URMC 213, URMC 216, URMC 224, URMC 225, URMC 226		+		175A>G, 564T>C, 566A>G		+			Ile59Val, Asp189Gly
	URMC 209, URMC 211, URMC 215, URMC 218, URMC 219		+	184A>T	175A>G, 564T>C, 566A>G			62X	+	Ile59Val
Outbreak unrelated CR-KA strains	URMC 200, URMC 203, URMC 210, URMC 214		+		564T>C, 566A>G, 615 T>G, 834T>C, 835A>G		+			Asp189Gly, Asp205Glu, Asn279Asp
	URMC 202	+	+		564T>C, 566A>G, 615 T>G, 834T>C, 835A>G	pTrp77LysfsX2		78X	+	Trp77Lys
	URMC 204	+	+		564T>C, 566A>G, 615 T>G, 834T>C, 835A>G	pSer304ProfsX21	+	324X	+	Asp189Gly, Asp205Glu, Asn279Asp
	URMC 217		+		564T>C, 566A>G, 569T>A		+			Asp189Gly, Phe190Tyr
	URMC 222		+		175A>G,564T>C, 566A>G, 615 T>G		+			Ile59Val, Asp189Gly, Asp205Glu
Control strain	URMC 201 (intermediate carbapenem R)		+		175A>G,564T>C, 566A>G		+			Ile59Val, Asp189Gly

*Index patient isolate.

Abbr: SNP: single nucleotide polymorphism, NS: nonsense, MS; missense, FS: frame-shift, PS: premature stop codon, HV: hypervariable region, Nt: nucleotides, AA: amino acids

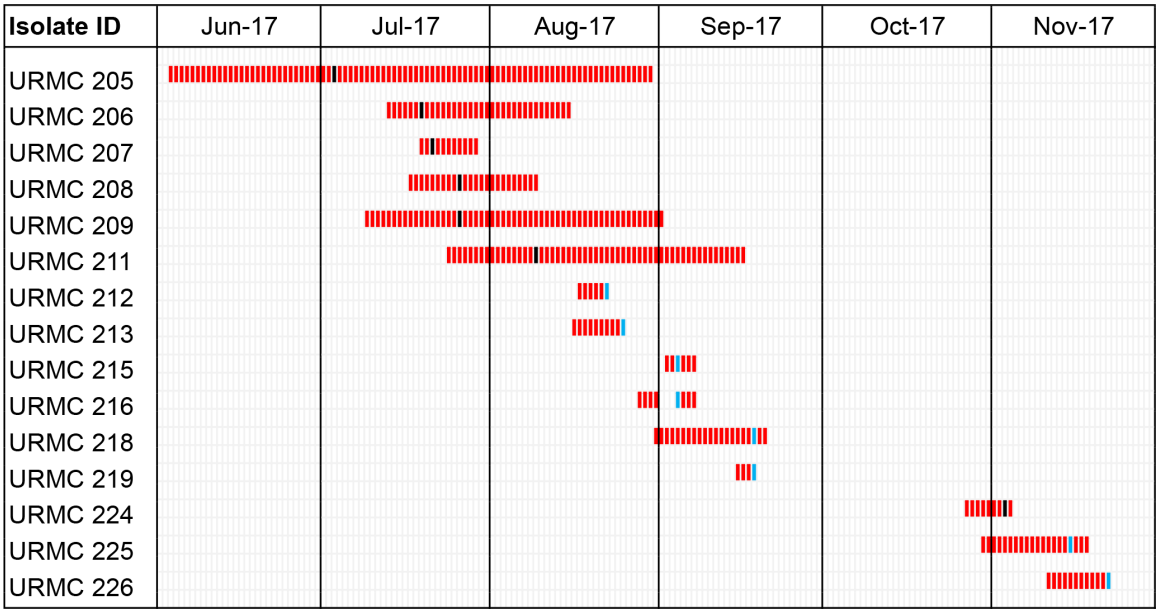

Fig. 1. Patient occupancy and overlap in CICU ward during the *K. aerogenes* outbreak. Black bars-first positive CR-KA clinical cultures, blue bars-first positive CR-KA surveillance cultures

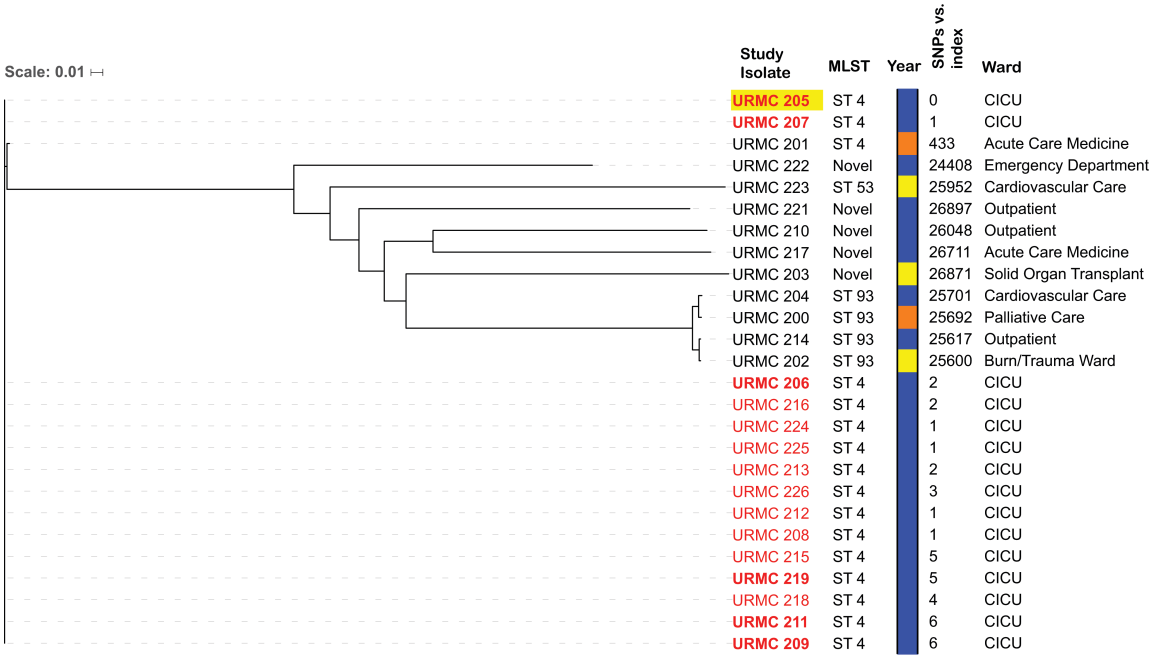

Fig. 2. Dendogram showing pairwise SNP differences based phylogenetic relatedness of URMC *K. aerogenes* strains. Whole genome sequence of *K. aerogenes* KCTC 2190 (ATCC 13048) was used for reference mapping. A total of 28,170 discriminatory high-quality SNPs in the core genomes obtained by the adapted CFSAN pipeline were used to plot the tree (excluding mobile elements and putative recombination sites). CICU outbreak strains: IDs highlighted in red, clinical isolates in bold, patient A isolate (1st case) with yellow background. Year strain isolated: orange-2015, yellow-2016, blue-2017. SNP differences relative to patient A shown. Scale bar indicates nucleotide substitutions per site.

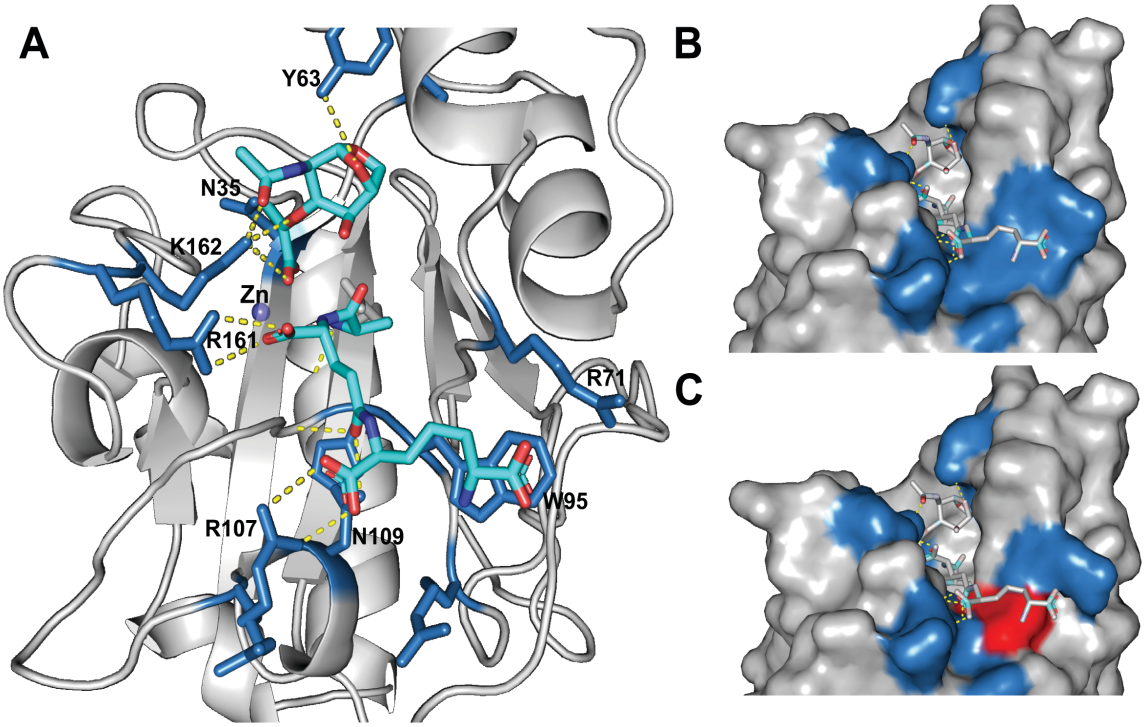

Fig. 3. Computational modeling of the impact of *ampD* mutations on AmpD in URMC outbreak CR-KA strains. A) *K. aerogenes* AmpD modeled on *C. freundii* AmpD structure. Key residues interacting with glycan and peptide portions of ligand are shown. B, C) Surface models of *K. aerogenes* AmpD depicting the wild-type binding surface for the diaminopimelate moiety (W95, top) versus the binding surface of AmpD containing the W95L mutation (blue surfaces indicate amino acid positions from panel A; bottom, altered surface highlighted in red).

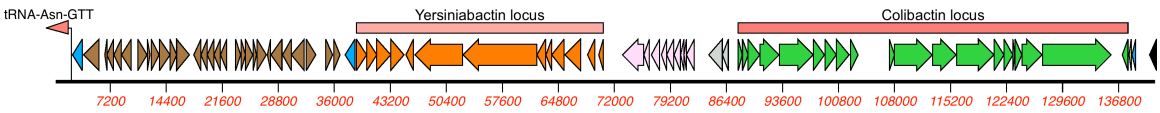

Fig. 4. Yersiniabactin and colibactin encoding gene loci on integrative conjugative element ICE*Kp*10 in the CICU outbreak CR-KA strains. Blue arrows-integrase encoding genes, brown arrows-$Zn^{2+/}Mn^{2+}$modules grey arrows-genes encoding mobilization proteins, pink arrows: vir-T4SS system.

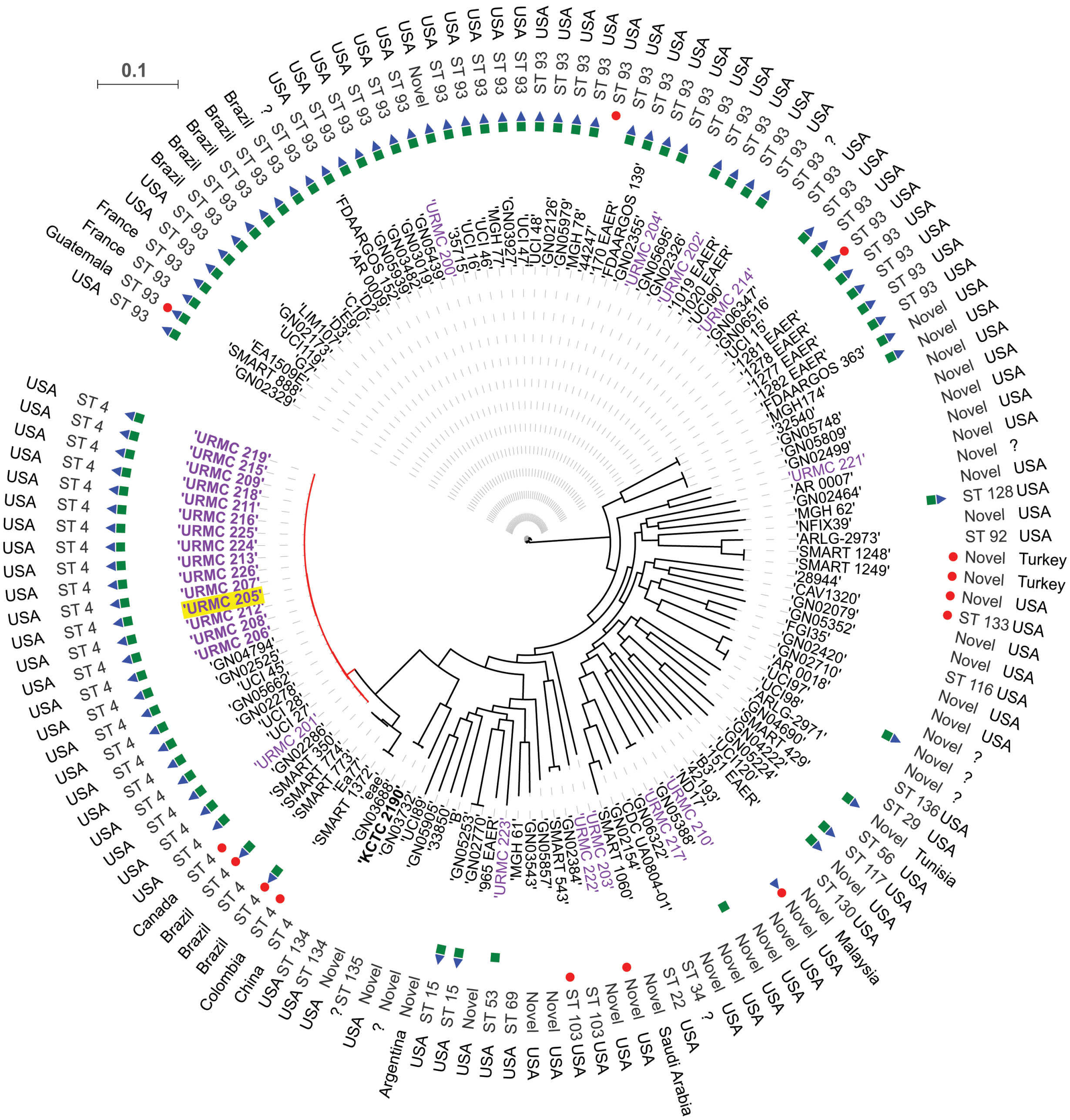

Fig. 5. Harvest based phylogenomic comparisons of URMC *K. aerogenes* genomes with global *K. aerogenes* genomes. Discriminatory SNPs based on core genome comparisons were used to plot the tree. URMC study *K. aerogenes* strains in purple (outbreak strains in bold, index patient strain highlighted yellow). Presence of genes encoding yersiniabactin siderophore system (green squares), colibactin synthesis cluster (blue triangles), and carbapenamases (red circles) in assembled genomes shown. Scale bar indicates nucleotide substitutions per site.