Title

Long read sequencing reveals genomic diversity and associated plasmid movement of carbapenemase-producing bacteria in a UK hospital over six years.

Authors

Leah W. Roberts^{1,2}, David A. Enoch³, Fahad Khokhar⁴, Grace A. Blackwell⁵, Hayley Wilson², Ben Warne^{2,6}, Theodore Gouliouris^{2,3,6}, Zamin Iqbal¹, M. Estée Török^{2,3,6}

Affiliations

1. European Molecular Biology Laboratory (EMBL) – European Bioinformatics

Institute (EBI), Wellcome Genome Campus, Hinxton, United Kingdom

2. Department of Medicine, University of Cambridge, United Kingdom

3. Clinical Microbiology & Public Health Laboratory, UK Health Security Agency, Cambridge, United Kingdom

4. Centre for Therapeutic Immunology and Infectious Diseases, University of Cambridge, Cambridge, United Kingdom

5. Sanger Institute, Wellcome Genome Campus, Hinxton, United Kingdom

6. Department of Infectious Diseases, Cambridge University Hospitals NHS Foundation Trust, Cambridge, United Kingdom

Correspondence

Zamin Iqbal, email zi@ebi.ac.uk

M. Estée Török, email et317@cam.ac.uk

Abstract

Background: Healthcare-associated infections (HCAIs) affect the most vulnerable persons in society and are increasingly difficult to treat in the face of mounting antimicrobial resistance (AMR). We used whole-genome sequencing (WGS) to retrospectively analyse carbapenemase-producing Gram negative bacteria from a single hospital in the United Kingdom (UK) over 6 years.

Methods: From 2014 to 2020 we collected all carbapenemase-producing Gram negative bacterial isolates processed in the diagnostic microbiology laboratory at Cambridge University Hospitals NHS Foundation Trust (CUH). Of 165 isolates, 85 underwent whole-genome sequencing with both short (Illumina) and long (Oxford Nanopore) reads, from which hybrid assemblies were constructed. In addition to characterising carbapenemase alleles and sequence types, we also assessed clonal transmission and plasmid relationships using single nucleotide polymorphisms (SNPs) and MinHash sketching, respectively.

Results: The vast majority of isolates were classified as either hospital-onset (HAI) or HCAI. Most carbapenemase-producing organisms were carriage isolates, with 71% isolated from screening (rectal) swabs. Using WGS, we identified 15 species, the most common being *Escherichia coli* and *Klebsiella pneumoniae*. Carbapenemase genes were found on plasmids in 86% of isolates, the most common types among the 20 identified being *bla*_{NDM}- and *bla*_{OXA}-type alleles. Only one significant outbreak occurred during the study period and involved a sequence type (ST)78 *K. pneumoniae* carrying *bla*_{NDM-1} on an IncFIB/IncHI1B plasmid. Interrogation of all *K. pneumoniae* assemblies in the National Centre for Biotechnology Information (NCBI), all deposited reads in the European Nucleotide Archive (ENA) to 2018 and a literature search revealed only one

other ST78 from India. Overall, we found 31 plasmid backgrounds harbouring carbapenemase genes, with 21 unique to single isolates.

Conclusions: We found considerable diversity of species in our dataset with minimal transmission and largely from carriage isolates. Our results suggest that these isolates may represent independent introductions into CUH. However, a national framework to collate more contextual data, particularly for plasmids and resistant bacteria in the community, is needed to better understand how carbapenemase genes are transmitted in the UK.

Keywords: carbapenemase, ST78, long-read sequencing, plasmid, transmission, whole genome sequencing

Background

Antimicrobial resistance (AMR) is a well-recognised global threat, exacerbated by the overuse and misuse of antibiotics and poor infection control practice (1). While resistance in bacteria can manifest in different ways, gene-based resistance is of particular concern, as horizontal gene transfer via mobilizable elements can mediate rapid spread of resistance in a population (2, 3).

Carbapenems are a class of broad-spectrum antimicrobial agents that are used to treat a range of infections. Over the last three decades, resistance to carbapenems has spread worldwide via an array of carbapenemase genes (4). The consequence of this proliferation has been problematic in healthcare settings, with outbreaks of carbapenemase-producing bacteria reported globally (5-7).

Prior to 2000, the incidence of carbapenem-non-susceptible Enterobacterales in the United Kingdom (UK) was low (8). Between 2003 and 2015, the main carbapenemase genes in circulation in the UK were *bla*_{KPC} followed by *bla*_{OXA-48} and *bla*_{NDM}. Following 2015, the number of reported carbapenemase-producers grew substantially, with the proportion shifting slightly in favour of *bla*_{OXA-48}, with reduced numbers of *bla*_{KPC} (9). *Klebsiella* spp. and *Escherichia coli* remain the two species most often associated with carbapenemase genes (10).

Standard screening for carbapenemase-producers usually relies on the use of selective chromogenic agar or media with carbapenem-containing discs followed by PCR testing to detect genes of interest (11). Beyond this, whole-genome sequencing (WGS) has been used for many years as a means of deriving further resolution of both

the carbapenemase allele and the host bacterial strain. To date, WGS has been used successfully in the UK to elucidate local outbreaks or to investigate targeted resistance mechanisms (12-16). However, little is known about the exact diversity of carbapenemase-producers across UK healthcare settings. Furthermore, carbapenemase genes are often found in complex genomic regions with a large number of repeat sequences. As such, short-read sequencing, while cost effective, is often unable to resolve the complete genomic context of the carbapenemase gene.

In this study, we collected all carbapenemase-producing Gram negative bacteria isolated in the diagnostic microbiology laboratory at CUH between 2014 and 2020. The majority of these were sequenced using both Illumina and Oxford Nanopore to elucidate the exact genetic context of the carbapenemase genes and possible transmission of strains and/or plasmids.

Methods

Study design, setting and participants:

This was a prospective observational cohort study conducted at Cambridge University Hospitals NHS Foundation Trust (CUH), a 1,000 bedded secondary and tertiary referral hospital in the United Kingdom. Clinical diagnostic samples and screening (rectal swab or stool) samples from hospital inpatients as well as samples referred from local General Practices (GPs), were collected from November 1st 2014 to December 31st 2020. In most cases, a single isolate for each unique species per patient was included in this study. Carbapenemase-producing Enterobacterales (CPE) screening policies within CUH changed during the course of the study. In 2014 CPE screening was based on risk factors as defined in the Public Health England guidance

(17). Full screening details can be found in Additional File 1 (Supplementary Methods). Bacterial species were identified using MALDI-TOF MS (Bruker Diagnostics, Bremen, Germany). Isolates with reduced susceptibility to carbapenems were detected either through routine antimicrobial susceptibility testing of clinical diagnostic samples or the use of selective media (details in Additional File 1: Supplementary Methods). Onset definitions are described in Additional File 1 (Supplementary Methods).

Bacterial culture, DNA extraction and sequencing:

Carbapenemase-producing isolates were retrieved from the diagnostic microbiology laboratory archive and cultured on selective media (CHROMID® CARBA SMART, Biomérieux, Marcy L'Etoile, France). Antimicrobial susceptibility profiles were determined using the VITEK-2 system (Biomérieux, Marcy L'Etoile, France). Genomic DNA was extracted using the QICube and the QIAamp 96 DNA QiACube HT kit (QIAGEN, Hilden, Germany). 85/165 isolates were available for sequencing with both short (Illumina HiSeq) and long (Oxford Nanopore Technologies (ONT) MinION) read technologies (details in Additional File 1: Supplementary Methods).

Sequencing quality control and assembly:

Illumina raw reads were checked for quality and contamination using fastQC (v0.11.8) (18), MultiQC (v1.9) (19) and Kraken2 (v2.0.9-beta) (20). Illumina raw reads were filtered using fastp (v0.20.1) ["--average_qual 30 --length_required 80"] (21). All Nanopore fast5 reads were basecalled and demultiplexed with Guppy (v3.6.0), using the "high-accuracy" basecalling model and the relevant barcoding kit for demultiplexing. Basecalled reads were then filtered for quality using Nanofilt (v2.7.1) (22) at a threshold of Q7 and a minimum read length of 1000 base pairs (bp). Isolates

with >110x coverage were additionally subsampled down to 100x coverage using Filtlong (v0.2.0) (with --keep_percent 90 and --target_bases corresponding to 100x coverage for that isolate) (23). All isolate Nanopore reads (original and subsampled) were *de novo* assembled using Flye (v2.8) (with --plasmids flag) (24), Unicycler (v0.4.8) (hybrid settings) (25) and Canu (v2.1.1) (default) (26). The best assembly per isolate was selected, with a preference for Flye > Unicycler > Canu when the assemblies were comparable. Full details are provided in Additional File 1 (Supplementary Methods).

Multilocus sequence type (MLST), resistance genes and plasmid profiles:

MLST was determined using the hybrid assemblies and the tool mlst (v2.19.0) (27) with the following profiles: (*A. baumannii* Pasteur, *Citrobacter freundii*, *Enterobacter*, *E. coli* Achtman, *Klebsiella pneumoniae*, *Pseudomonas aeruginosa*) (28). Resistance genes and plasmid profiles were detected using the complete assemblies and Abricate (v1.0.1) (29) against the CARD (date: 2021-Mar-27) (30) and Plasmidfinder (date: 2020-Oct-24) (31) databases using minimum coverage of 90% and minimum nucleotide identity of 90%. Tandem duplication of carbapenemase genes was confirmed by comparing the duplicated region to the raw (fastq) nanopore reads using BLASTn (v2.12.0) (32).

Species identification:

In addition to Kraken2 (v2.0.9-beta), we also evaluated species identification for *Klebsiella* isolates using Kleborate (v0.4.0-beta) (33, 34) against the complete assemblies at default settings. *Enterobacter* isolates were also manually checked

174 against reference *Enterobacter* species (NCBI) by comparing chromosomes using
175 fastANI (v1.3) (35) at default settings.

176

177 *Clonal cluster analysis:*

178 Clonal clusters were confirmed using three methods: Snippy (v4.6.0) (36) against a
179 single reference for each species, assembly-free Split Kmer Analysis (SKA v1.0.1)
180 (37) and MASH (v.2.2.2) (38) (full details in Additional File 1: Supplementary Methods).
181 Based on existing literature (39-41), the following thresholds were used: <20 SNPs
182 (Snippy), MASH threshold 0.001, <25 SNPs and nucleotide identity of 95 (SKA). The
183 final clusters were derived where SKA and Snippy results agreed.

184

185 *Plasmid clustering:*

186 Complete plasmid sequences were clustered using Mash (v2.2.2) using a sketch size
187 of 10000 and a clustering threshold of 0.001 and 0.005 mash distance.

188

189 *K. pneumoniae phylogeny:*

190 350 publicly available closely related isolates were downloaded from the European
191 Nucleotide Archive (ENA). These were identified by sketching a reference genome
192 (cpe058) using sourmash (v3.5.0) (42) at a Kmer length of 31 bp and an interval of
193 5000 bp (“-n 5000 -k 31”) and querying against the sourmash index of 661,405
194 bacterial genomes (43) (full details in Additional File 1: Supplementary Methods). All
195 isolates were trimmed using fastp (v0.20.1) using the following settings: “--
196 length_required 40 --cut_front --cut_tail -M 25 “. After trimming, 12 isolates were
197 removed for being <20x cov or >300x cov. The 338 remaining were combined with
198 trimmed reads for the 11 ST78 outbreak isolates from this study and simulated reads

for an Indian ST78 isolate (NCBI: NZ_JAJBIZ010000100.1) (44). All reads were mapped to the chromosome of cpe058 and SNPs called using Snippy (v4.6.0). The core full alignment generated from Snippy-core was then run through Gubbins (v3.0.0) (45) to remove recombination and generate a phylogeny using FastTree (v2.1.10) (46) under the GTRGAMMA model. The final tree was drawn in iTol (47).

Results

A total of 165 carbapenemase-producing Gram negative bacterial isolates from 133 patients were collected in the study period. The majority of patients were male (n=78, 59%) (**Table 1**) and the mean age for both men and women was approximately 54 years. The median length of inpatient stay was 18 days (range 0 to 156 days) and the majority of patients were alive on discharge (n=94, 71%). Nine isolates (from nine patients) were collected from general practices (GP) in the Cambridgeshire area. The remainder were collected from CUH inpatients.

The majority of patients were classified as having either hospital-onset (n=54, 41%) or healthcare-associated (n=69, 52%) infections (**Table 1**). Of these, roughly a third of patients had exposure to meropenem in the previous three months (19/54 hospital-onset, 22/69 healthcare-associated). Only seven (5%) were community-associated (two from screening swabs, five from urine samples). In the three months prior to admission, 39 patients reported travel overseas (29%) and three to other parts of the UK (2%). Overall, most patients had been hospitalised within the last year, either in CUH (n=51, 38%), in other UK hospitals (n=37, 28%), or overseas (n=23, 17%). Ten patients had both UK and overseas hospitalisations (7 in CUH and 3 in other UK hospitals).

High frequency of carbapenemase gene carriage:

A total of 15 species were identified during the study, the majority of which were *Klebsiella pneumoniae* (n=54, 33%) or *Escherichia coli* (n=55, 33%) (**Figure 1A**). Five patients had recurrent isolates that were included in the study (resulting in the following duplicates: eight *Klebsiella pneumoniae*, three *E. coli*, two *Enterobacter sp.*, one *Citrobacter freundii* and one *Klebsiella oxytoca*).

The main carbapenemase genes detected were *bla*_{NDM-1} (n=65, 39%) and *bla*_{OXA-48} (n=60, 36%) (**Figure 1B**). Most isolates (n=117, 71%) were recovered from screening swabs. The remainder were isolated from urine (n=17), blood (n=9), biliary fluid (n=8), respiratory samples (n=6), wound swabs (n=5), body fluids (n=3), tissue (n=2) and central venous catheter tip (n=1). There was no association of species or carbapenemase gene with sample site (Additional File 1: **Supplementary figure 1**). There was also little association between infection classification and species or carbapenemase gene (Additional File 1: **Supplementary figure 2**). *bla*_{NDM-1}- and *bla*_{OXA-48}-harboring *E. coli* and *K. pneumoniae* predominated in all infection categories.

Most isolates were found to have intermediate susceptibility or non-susceptibility to meropenem (Additional File 1: **Supplementary Figure 3**). Notably, *bla*_{OXA-48}-carrying isolates had the most variable levels of resistance, with a large portion, in particular *E. coli*, remaining susceptible to meropenem.

249 **Table 1: Baseline characteristics of study participants**

Patients (n= 133)	Male	Female
	(n=78)	(n=55)
Age (years)		
Average	53.7	54.5
Median	57	55
Range	0-95	2-92
Length of stay (days)¹		
Min	0	
Median	18	
Average	26.6	
Max	156	
Classification²		
Community-associated	7	
Hospital-onset	54	
Healthcare-associated	69	
CUH	23	
CUH & overseas hospital	3	
Overseas hospital	17	
Other UK hospital	24	
Other UK & overseas hospital	2	
Unknown:	3	
History of travel (previous 3 months)	Travel only	Travel and hospitalisation
Asia	3	18
Africa	1	5
Continental Europe	2	9
Middle East	0	1
UK	3	0
Not recorded (n=7)		
Hospitalisation (previous 12 months)	Single site only	Multiple hospitalisation sites
CUH	51	7 (overseas)
Other UK hospital	37	3 (overseas)
Overseas hospital	23	0
Not recorded (n=1)		
Outcome (at hospital discharge)		
Alive	94	
Deceased	13	
Not recorded	9	
Not applicable ³	17	

250 ¹ Not applicable (GP sample): 9 patients; Not applicable (no admission): 8 patients; Unknown: 9
251 patients

252 ² Definitions: Community-associated: isolated <48 hours after admission – no previous
253 hospitalisation or medical treatment in past 12 months; Hospital onset: isolated ≥48 hours after
254 admission; Healthcare associated: isolated <48 hours after admission and had previous
255 hospitalisation or medical treatment in last 12 months at [i] CUH, [ii] other UK hospital, [iii]
256 overseas (last three months), or a combination of any of the three.

257 ³ Patient was not admitted, so no discharge outcome

259

260 **Figure 1: All carbapenemase-positive isolates collected over the study period and**
 261 **carbapenemase genes (n=165).** [A] species were determined using MALDI-TOF MS. [B]
 262 carbapenemase genes were determined using PCR. Sequential isolates of the same species (from the
 263 same patient) are outlined in black.

264

265 *No significant change in the frequency of carbapenemase-producers during the study*
 266 *period.*

267 Despite an observed increase per year in carbapenemase-producers (**Figure 1**),
 268 evaluating the number of isolates obtained per month in the context of adapted
 269 sampling procedures over different periods revealed little fluctuation in isolate
 270 numbers over time (Additional File 1: **Supplementary figure 4, Supplementary**
 271 **Figure 5**). In June 2016 a prospective screening study commenced in the adult
 272 intensive care unit (ICU) at CUH and identified an outbreak of *bla*_{NDM-1} positive *K.*
 273 *pneumoniae* (discussed below). This resulted in intensified screening (on admission
 274 and weekly thereafter) of all patients admitted to adult ICUs in CUH. A very slight

decrease in the number of CPE isolates was observed from late 2019 throughout 2020, most likely as a result of the SARS-CoV-2 pandemic (Additional File 1: **Supplementary Figure 5**).

Travel had little impact on the species isolated or carbapenemase genes detected

When comparing isolates from patients with and without overseas travel, we found a preference for *bla*_{NDM} (particularly *bla*_{NDM-1} *E. coli*) in patients who travelled to Asia, and similarly *bla*_{OXA-48} in patients from central Europe (Additional File 1: **Supplementary figure 6**). Beyond this, there were very few differences in species and carbapenemase genes isolated based on travel. The most common species/carbapenemase gene combinations (*bla*_{NDM-1}- and *bla*_{OXA-48}-harboring *E. coli* and *K. pneumoniae*) were common in patients irrespective of travel history. There was a slightly larger proportion of *bla*_{NDM-1}-harboring *K. pneumoniae* in patients with no travel history; these were related to a hospital outbreak (discussed below).

Whole-genome sequencing provides enhanced discrimination

Of the 165 isolates, 85 (from 68 patients) underwent whole-genome sequencing (WGS) with both short-read Illumina and long-read nanopore platforms, from which we were able to obtain complete bacterial genomes in most cases (**Additional File 2**).

Using this genomic data, we re-evaluated the species identifications using Kraken2 as well as Kleborate (for the *Klebsiella* sp.) and fastANI (for *Enterobacter* sp.). All isolates were correct at the genus level, with 12 differences at the species level. These include two reclassifications of *K. pneumoniae* to *K. quasipneumoniae* subsp.

similipneumoniae (cpe097) and *K. oxytoca* (cpe088), one *K. oxytoca* to *K. grimontii* (cpe102), one *E. kobei* to *E. hormaechei* (cpe050), five *E. cloacae* to *E. hormaechei* (cpe002, icp157, cpe042, cpe090, cpe093), one *E. cloacae* to *E. mori* (cpe018), and two *E. cloacae* to *E. roggenkampii* (cpe104, cpe106).

Enhanced resolution of the carbapenemase genes was also possible from the sequencing data, allowing better discrimination of the *bla*_{OXA} and *bla*_{NDM} alleles (Figure 3). *bla*_{NDM-1} remained the most prevalent carbapenemase (n=19), however this was due to the *bla*_{NDM-1} *K. pneumoniae* outbreak during the study (n=11) (detailed below). *bla*_{NDM-5} (n=9) was the next most prevalent NDM carbapenemase. The *bla*_{OXA-48} group was differentiated into three main types, with *bla*_{OXA-181} (n=11) the most prevalent, followed by *bla*_{OXA-48} (n=8) and *bla*_{OXA-232} (n=8). All other carbapenemase types were found less frequently, with 13/20 (65%) found in less than or equal to two isolates. Four isolates that were originally positive for *bla*_{NDM} (cpe078 and cpe020) and *bla*_{OXA-48} (cpe100 and cpe088) based on PCR had no detectable *bla*_{NDM} or *bla*_{OXA-48} after sequencing. This may have been due to loss during sub-culturing or an error occurring during sample retrieval. These four isolates were not analysed further.

Variety of carbapenemase alleles and genetic background

Assessing carbapenemase alleles with their genetic background revealed a wide variety of combinations, which were in many cases unique to a single isolate (**Figure 2A**). Further comparison to patient onset information revealed some combinations that appeared more commonly in isolates from patients who had travelled overseas versus those from the UK. These included the IncX3 *bla*_{OXA-181} plasmids, as well as some *bla*_{OXA-23}/*bla*_{OXA-66} isolates from *A. baumannii* and roughly half of the col *bla*_{OXA-232}

isolates. The remainder were all hospital-onset or healthcare-associated in either CUH or other UK hospitals, with the large notable addition of the IncFIB/IncHI1B *bla*_{NDM-1} plasmid that was part of a known outbreak during the study (detailed below). Only one community-associated isolate was sequenced (a chromosomally encoded *bla*_{OXA-48} from an *E. coli*).

Most carbapenemase genes were located on plasmids

Of 81 isolates, 65 harboured their carbapenemase gene on a plasmid (**Figure 2B**). The remaining 16 had their carbapenemase gene on the chromosome (n=11) or on both the chromosome and a plasmid (n=5). All carbapenemase genes had 100% nucleotide identity to their assigned allele, with the exception of seven isolates carrying a *bla*_{IMP} gene differing by 7 SNPs from *bla*_{IMP-1} and 1 SNP from *bla*_{IMP-70} (labelled as *bla*_{IMP-70b} in **Figure 2**). The single non-synonymous SNP difference from *bla*_{IMP-70} corresponded to I228L.

There was a clear association of certain incompatibility types to carbapenemase genes (**Figure 2C**). These include *bla*_{OXA-48} on IncL and IncM, *bla*_{OXA-232} on col, *bla*_{NDM-1} on IncC and IncFIB/IncHI1B, and *bla*_{OXA-181} on IncX3. Several isolates were also identified as harbouring multiple copies of the same carbapenemase gene (details in Additional File 1: supplementary table 1).

Figure 2: Carbapenemase alleles by [A] onset, [B] genomic location and [C] plasmid type: **A:** x-axis displays the carbapenemase and plasmid replicon type, coloured by onset status, **B:** x-axis describes whether the carbapenemase gene was found on the chromosome (Chr), a plasmid (pla) or both (coloured by species), **C:** x-axis describes location of carbapenemase gene (Chr = chromosome, otherwise specified by plasmid type. Instances of both chromosome and plasmid type indicate where a carbapenemase was found on both).

Figure 3: Clonal transmissions identified by WGS: [A] transmission cluster timeline. Species and cluster number provided on the y-axis (pseu = *P. putida*, kleb = *K. pneumoniae*, ecoli = *E. coli*). X-axis indicates year of isolation. Isolate collections (circles) coloured by sample site. Circles outlined in black are duplicate isolates from patients with an existing earlier isolate. BAL = Bronchoalveolar lavage, CVC = central venous catheter **[B]** Minimal spanning tree of ST78 *K. pneumoniae* outbreak isolates. Isolates in the same circle were identical at core genome level. Lines represent core SNP distance (1 SNP unless otherwise noted). Circle outlines denote year of isolation. Circle fill colour represents patient identity. Isolate names in italics with asterisk(*) are duplicates isolated from patients already with this *K. pneumoniae* strain

Few clonal transmissions were detected during the study period

Between 2016 to 2020, only six clonal transmission clusters between different patients were detected from the WGS data (**Figure 3A**). The majority of these involved only two patients (n=5) and include *P. putida*, *E. coli* (ST410, ST648) and *K. pneumoniae* (ST11, ST2497). Most of these outbreaks spanned less than 10 days, with the

exception of an ST2497 *K. pneumoniae* that was isolated 205 days apart. Apart from the *P. putida* isolates, all others were sampled from different wards. Of these events, only the large ST78 *K. pneumoniae* cluster and the *P. putida* cluster were detected by the hospital infection control team at the time. The remainder had no obvious epidemiological connections.

Outbreak of bla_{NDM-1} positive K. pneumoniae during study period

In 2016, a large outbreak of multidrug-resistant *K. pneumoniae* was initially detected during a prospective surveillance study conducted in the adult ICU at CUH. Infection control investigations identified spread to several other wards. WGS identified the outbreak strain as an ST78 *K. pneumoniae*, carrying bla_{NDM-1} on a large ~300 kb IncFIB/IncHI1B plasmid. Phenotypic testing found that these isolates (n=11) were resistant to a number of antibiotics, including amoxicillin, amoxicillin-clavulanic acid, trimethoprim, ciprofloxacin, cefotaxime, nitrofurantoin, piperacillin-tazobactam, ertapenem, meropenem, 18osfomycin, gentamicin, amikacin and aztreonam. They had intermediate resistance to tigecycline and were susceptible to colistin. They carried a number of antibiotic resistance genes, including resistance to aminoglycosides (AAC(3)-IId, AAC(6')-Ib-cr, armA), cephalosporins (bla_{CTX-M-15}), quinolones (qnrB17), trimethoprim (dfrA12) macrolides (mphE, msrE), phenicols (oqxAB), sulphonamides (sul1), and tetracycline (tet(D)). All isolates carried an OmpK36GD porin mutation associated with resistance to carbapenems (48). Seven isolates had evidence of truncation of mgrB (cpe022, cpe003, cpe004, cpe033, cpe034, cpe035) or pmrB (cpe001) associated with resistance to colistin (despite phenotypic susceptibility). Four of these also carried aerobactin iuc1 (cpe004, cpe022)

or *iuc5* (cpe001, cpe003), but no other genetic markers for hypervirulence were observed (49).

Based on the WGS data obtained here, this outbreak involved at least seven patients from July 2016, with one patient testing positive again with the same organism in 2018 despite being sent home (**Figure 3B**). As not all isolates from the study period were sequenced, it is also possible that other non-sequenced isolates from our dataset were also part of the outbreak (Additional File 1: **Supplementary figure 7**).

To provide further context to the outbreak, we interrogated a large public repository of 661,405 bacterial genomes collated up to 2018, which identified 352 broadly related isolates. Only 28 of these were identified as ST78, while the vast majority were found to be ST14, a single-locus variant of ST78 (Additional File 1: **Supplementary Table 2**).

Surprisingly, all of the available ST78 strains (n=28) from the public repository were also isolated at CUH and obtained in 2016 or 2017 as part of further intensive screening of the outbreak (**Figure 4, blue box**). Only one isolate appeared to be a duplicate that was sequenced in both this study and a previous study (cpe001;ER2672819). As we did not expect to find only CUH isolates, we also downloaded all available *K. pneumoniae* assemblies from NCBI (11,811 sequences) and performed *in silico* multi-locus sequence typing (MLST). None of these additional assemblies were found to be ST78. We then performed a literature search, which uncovered a single reported ST78 *K. pneumoniae* isolate from India in 2018 (50) which we then added to our analysis. This Indian isolate was positioned outside of the main

ST78 outbreak clade in the phylogenetic tree, suggesting that it is indirectly related through a distant relative. Phylogenetic comparison to other contextual strains revealed that the closest relatives were ST14, mainly from Thailand, the United States, and Canada (**Figure 4**). The index patient identified in CUH had no history of travel or hospitalisation abroad in the 3 months prior to admission.

We were also able to search the same public repository for isolates with a high likelihood of carrying the *bla*_{NDM-1} IncFIB/IncHI1B plasmid based on >80% matching kmer identities. We found 173 isolates all belonging to *K. pneumoniae* from a variety of STs, the majority being ST15, ST14 and ST78 (Additional File 1: **Supplementary figure 8**). The isolates with the highest identity to our plasmid were all from ST14 and ST78. We checked for the presence of *bla*_{NDM-1} in these isolates and found that 141 (82%) carried the gene, which was found 100% of the time in isolates with a kmer identity $\geq$ 95%.

We then compared our plasmid results to our contextual isolates (**Figure 4**) to identify those that carried the IncFIB/IncHI1B plasmid, with or without *bla*_{NDM-1}. We found that only those strains closely related to the ST78 outbreak carried this plasmid, including the Indian ST78 isolate (**Figure 4, orange box, Additional file 1: Supplementary Figure 9**). Of these 65, only five did not carry a *bla*_{NDM} gene. All other carried *bla*_{NDM-1}. As such, acquisition of this plasmid likely occurred once in an ancestral ST14 isolate and has been stably maintained with infrequent loss (**Figure 4**).

Figure 4: Phylogeny of ST78 outbreak strain and closest relatives from repository of 661K bacteria: blue box: all ST78 isolates, isolated from CUH (with one Indian ST78 as an outgroup). Orange box: All isolates from this collection that harbour the IncFIB/IncHI1B plasmid. Phylogeny is midpoint rooted. Scale indicates number of substitutions.

Combination of clonal and horizontal plasmid transmission, mixed with high background variability:

In total, 70 isolates (86%) were found to carry a carbapenemase gene on a plasmid. By comparing plasmid sequences from these isolates and using a mash distance threshold of 0.001, we identified 44 isolates (from 31 patients) that had carbapenemase-plasmids that clustered closely. These 44 isolates spanned 10 plasmid clusters (**Table 2**). Of the remaining 26 isolates, five formed clusters when applying a less strict threshold (0.005). The other 21 remained unique, resulting in a total of 31 plasmid backgrounds over the study period.

Two clusters (clust_217 and clust_33) appeared to be the result of intra-patient plasmid transfer between different colonising species, as they were exclusive to a single patient (**Table 2**). Cluster 4 (clust_4) was entirely the result of clonal transmission of the ST78 outbreak *K. pneumoniae*, rather than independent plasmid transfer. Cluster 148 (clust_148) appears to share the same *K. pneumoniae* strain but different *E. coli*, suggesting the *K. pneumoniae* transmitted between the patients first, after which the plasmid was transferred into different *E. coli* strains in either patient. The largest cluster (clust_72) contained four clonally related isolates (two *K. pneumoniae* and two *E. coli*) alongside another four unrelated strains that all harboured a *bla*_{OXA-232}-col plasmid.

Looking across all plasmids in the dataset and clustering at the same mash threshold (0.001), we found a total of 51 clusters ranging from 2 to 11 isolates in size (median = 2). The largest clusters were other plasmids associated with the ST78 *K. pneumoniae* outbreak strain (i.e., the result of clonal transmission). These were followed by the

*bla*_{OXA-181} and *bla*_{OXA-232} plasmid clusters, as described above. There were no other non-carbapenemase-carrying plasmids that appeared highly prevalent across isolates.

There was limited evidence of smaller mobile genetic elements (MGEs) involved in horizontal transmission when examining the context of the carbapenemase genes in this dataset (data not shown). As such, we found that plasmids were the main vehicle of carbapenemase transmission in this study, both vertically (clonally) and horizontally.

Table 2: Carbapenemase plasmid clusters

Cluster	Number of isolates	Species	Patients	Carb	Plasmid type
Clust_133	2	<i>A. baumannii</i>	2	OXA-72	Untyped
Clust_148	4	<i>E. coli</i> , <i>K. pneumoniae</i>	2	NDM-1	IncC
Clust_17	4	<i>K. aerogenes</i> , <i>E. coli</i> , <i>K. pneumoniae</i>	3	OXA-48	IncL
Clust_217	4	<i>K. grimontii</i> , <i>E. roggenkampii</i> , <i>R. ornithinolytica</i>	1	OXA-181	IncX3
Clust_33	2	<i>E. coli</i> , <i>K. pneumoniae</i>	1	NDM-6	IncFIB
Clust_4	9	<i>K. pneumoniae</i>	6	NDM-1	IncHI1B/IncFIB
Clust_54	2	<i>E. coli</i>	2	NDM-5	IncFIB/IncFIA
Clust_72	8	<i>K. pneumoniae</i> , <i>E. coli</i>	8	OXA-232	Col
Clust_9	3	<i>E. hormaechei</i>	3	IMP-70b	IncC
Clust_98	6	<i>E. coli</i> , <i>E. hormaechei</i> , <i>K. pneumoniae</i>	6	OXA-181	IncX3

* yellow highlight shows clusters with isolates more likely to be the result of horizontal transmission.

Discussion

This study conducted a detailed analysis of carbapenemase-producing Gram negative bacteria in a single UK hospital over 6 years. Overall, we found low rates of carbapenemase-producing organisms, and this did not appear to increase

substantially during the study period. This is consistent with a previous point prevalence study, conducted in the same hospital, which found no carbapenemase-producing Enterobacterales amongst 540 screening samples collected over a 3-day period in 2017 (51).

Additionally, we found very little evidence for the circulation of dominant lineages, and instead identified a variety of species carrying a range of carbapenemase alleles in different genetic backgrounds. Previous healthcare exposure (mainly at CUH) was a common feature among patients in this study and could be concluded as the main risk for infection with carbapenemase-producing bacteria. Indeed, a previous study looking at carbapenem-resistant *K. pneumoniae* across Europe found high levels of within-hospital spread (39). However, if these were hospital-associated lineages, we might expect to see more frequent or consistent transmission over the 6-year timeframe, but instead we found limited transmission beyond a single patient.

Another possibility is independent introduction of these strains via patient carriage as a result of prior acquisition in other healthcare or community settings, which would explain the limited clonal transmission and high diversity we have observed. Carriage of carbapenemase genes is not uncommon, and can persist for prolonged periods of time, particularly following antibiotic exposure in hospitals (52, 53). To verify this, we would need substantially more contextual data from the community and other healthcare settings. These areas represent gaps in our understanding of carbapenemase-producers nationally.

The few clusters that were detected during this study rarely exceeded two patients and most had no epidemiological support other than close temporal links within the same institution. This could indicate that transmission occurred in the hospital via an intermediate, such as contaminated surfaces or hospital staff (54). Despite these seemingly benign transmissions, these cryptic events do reveal how easily bacteria can begin to spread undetected and in the absence of other epidemiological links. Early detection of spread is crucial to guide infection control interventions, and it can also have an enormous impact on healthcare costs (55). Routine use of WGS provides a means to detect transmission early, but the costs and logistics have hampered routine adoption in healthcare settings (56). Even in this study, we were unable to sequence all carbapenemase-producers, limiting our ability to investigate the relatedness of unsequenced isolates.

An easily searchable index of public data is crucial for outbreak analysis and public health interests. Blackwell *et al.* (43) provide a searchable index of all bacterial isolates from the ENA up to 2018, which enabled us to rapidly contextualise our ST78 *K. pneumoniae* outbreak strain. Coupled with all *K. pneumoniae* assemblies submitted to NCBI, we observed almost no publicly available ST78 isolates outside of those obtained at CUH, with only a single distant relative isolated in India (50). While we do not have ENA data submitted after 2018, we would expect public data for ST78 to have existed prior to this if it had been reported elsewhere since the ST78 outbreak strain first appeared in CUH in 2016. Here we determined that this strain is extensively resistant to antibiotics but mostly lacking biomarkers for hypervirulence. Vigilance and further surveillance is required to monitor this sequence type and its possible expansion to other parts of the UK.

539

540 Finally, we used long-read sequencing to analyse plasmids in this study. Short-read
541 sequencing is usually insufficient to confidently identify and compare plasmid
542 sequences from whole genomes. Here, we were able to determine the number of
543 shared and unique carbapenemase-carrying backgrounds between isolates, the
544 majority of which were on plasmids. We found several plasmids of importance that
545 have been identified elsewhere, including pOXA-48 (57), col_OXA-232 (58) and
546 IncX3_OXA-181 (59). By determining which isolates were related clonally, we could
547 infer which plasmids were more likely to have transmitted horizontally. However, we
548 were unable to confidently predict whether horizontal plasmid transfer had occurred
549 during or prior to hospitalisation. Plasmids do not appear to follow the same mutation
550 rate expected from clonally spreading strains, as evidenced by comparisons of pOXA-
551 48 (60). As such, distantly transferred plasmids can still have very few SNP
552 differences. We also do not have access to susceptible strain information, that could
553 be used to identify the patient's isolate prior to plasmid acquisition.

554

555 We acknowledge several limitations to our study. First, the hospital screening policy
556 and laboratory diagnostic methods for detection of carbapenemase-producing
557 organisms changed over the course of the study period. This may have resulted in an
558 underestimate of the true prevalence of these organisms, and/or an overestimation of
559 hospital onset, since not all patients were screened on admission. Second, not all
560 samples were stored in the diagnostic laboratory and/or available for sequencing.
561 Third, we did not sample healthcare workers and did not routinely sample the
562 environment as part of this study. Fourth, there was limited sampling (and associated

metadata) outside the hospital setting. All of these factors may have affected our ability to detect and analyse all potential transmission events.

Nevertheless, we found considerable diversity of species in our dataset with minimal transmission and largely from carriage isolates. Our results suggest that these isolates may represent independent introductions into CUH. Detailed resolution using WGS and contextualisation with public data allowed us to better construct clonal and plasmid relationships, particularly for an emerging extensively resistant ST78 *K. pneumoniae* lineage. However, a national framework to collate contextual data, particularly for plasmids and resistant bacteria broadly in the community, is needed to better understand how carbapenemase genes are transmitted in the UK.

List of abbreviations

HCAI: healthcare-associated infection

HAI: hospital-onset infection

WGS: whole genome sequencing

CUH: Cambridge University Hospitals NHS Foundation Trust

SNP: Single Nucleotide Polymorphism

NCBI: National Centre for Biotechnology Information

ENA: European Nucleotide Archive

ST: sequencing type

UK: United Kingdom

PCR: Polymerase Chain Reaction

GP: General Practitioner

CPE: Carbapenemase-Producing Enterobacterales

588 MALDI-TOF MS: Matrix-assisted laser desorption ionization-time of flight mass
589 spectrometry

590 ONT: Oxford Nanopore Technologies

591 ICU: Intensive Care Unit

592 MGE: Mobile Genetic Element

593

594 **Additional Files**

595 Additional File 1: Supplementary Methods and Results (docx): additional details on
596 study methods and results.

597 Additional File 2: Supplementary Dataset 1 (xlsx): individual sample accession
598 numbers and metadata.

599

600 **Declarations**

601 **Ethics approval and consent to participate**

602 This study was conducted as part of surveillance for healthcare-associated infections
603 under the auspices of Section 251 of the NHS Act 2006. It therefore did not require
604 individual patient consent or ethical approval.

605

606 **Consent for publication**

607 Not applicable.

608

609 **Availability of data and materials**

610 The datasets (Illumina fastq reads, Oxford Nanopore basecalled (fastq) reads, and
611 hybrid assemblies) supporting the conclusions of this article are available in the

612 European Nucleotide Archive (ENA) under the bioproject PRJEB30134. Individual
613 sample accessions are available in Supplementary Dataset 1.

614

615 **Competing interests**

616 The authors declare that they have no competing interests.

617

618 **Funding**

619 LWR is an EBPOD Fellow funded by the European Bioinformatics Institute and the
620 University of Cambridge. BW is supported by the NIHR Cambridge Biomedical
621 Research Centre. ZI is funded by EMBL core funding. MET is supported by a Clinician
622 Scientist Fellowship (funded by the Academy of Medical Sciences and the Health
623 Foundation) and the NIHR Cambridge Biomedical Research Centre.

624

625 **Author contributions**

626 Conceptualization: LWR, DAE, ZI, MET

627 Methodology: LWR, DAE, ZI, MET

628 Formal analysis: LWR, GAB, ZI

629 Investigation: DAE, FAK, HW, BW, TG, MET

630 Resources: DAE, FAK, HW, ZI, MET

631 Data Curation: LWR, FAK, GAB, HW, BW

632 Writing - Original Draft: LWR

633 Writing - Review & Editing: all authors

634 Visualization: LWR

635 Supervision: ZI, MET

636 Project administration: ZI, MET

Funding acquisition: ZI, MET

Acknowledgements

The authors would like to thank the infection control team and the laboratory staff of the diagnostic microbiology laboratory at CUH. We also acknowledge Michael Hall and Martin Hunt for nanopore assembly advice.

References

1. Browne AJ, Chipeta MG, Haines-Woodhouse G, Kumaran EPA, Hamadani BHK, Zaraan S, et al. Global antibiotic consumption and usage in humans, 2000-18: a spatial modelling study. *Lancet Planet Health*. 2021.
2. Sheppard AE, Stoesser N, Wilson DJ, Sebra R, Kasarskis A, Anson LW, et al. Nested Russian Doll-Like Genetic Mobility Drives Rapid Dissemination of the Carbapenem Resistance Gene blaKPC. *Antimicrob Agents Chemother*. 2016;60(6):3767-78.
3. David S, Cohen V, Reuter S, Sheppard AE, Giani T, Parkhill J, et al. Integrated chromosomal and plasmid sequence analyses reveal diverse modes of carbapenemase gene spread among *Klebsiella pneumoniae*. *Proc Natl Acad Sci U S A*. 2020;117(40):25043-54.
4. Queenan AM, Bush K. Carbapenemases: the versatile beta-lactamases. *Clin Microbiol Rev*. 2007;20(3):440-58, table of contents.
5. Mathers AJ, Cox HL, Kitchel B, Bonatti H, Brassinga AK, Carroll J, et al. Molecular dissection of an outbreak of carbapenem-resistant enterobacteriaceae reveals Intergenous KPC carbapenemase transmission through a promiscuous plasmid. *mBio*. 2011;2(6):e00204-11.

661 6. Naas T, Levy M, Hirschauer C, Marchandin H, Nordmann P. Outbreak of carbapenem-
662 resistant *Acinetobacter baumannii* producing the carbapenemase OXA-23 in a tertiary care
663 hospital of Papeete, French Polynesia. *J Clin Microbiol.* 2005;43(9):4826-9.

664 7. Souli M, Galani I, Antoniadou A, Papadomichelakis E, Poulakou G, Panagea T, et al.
665 An outbreak of infection due to beta-Lactamase *Klebsiella pneumoniae* Carbapenemase 2-
666 producing *K. pneumoniae* in a Greek University Hospital: molecular characterization,
667 epidemiology, and outcomes. *Clin Infect Dis.* 2010;50(3):364-73.

668 8. PHE PHE. Carbapenemase-producing Enterobacteriaceae: laboratory confirmed
669 cases, 2003 to 2015 [updated 10-10-2016. Available from:
670 [https://www.gov.uk/government/publications/carbapenemase-producing-](https://www.gov.uk/government/publications/carbapenemase-producing-enterobacteriaceae-laboratory-confirmed-cases/carbapenemase-producing-enterobacteriaceae-laboratory-confirmed-cases-2003-to-2013)
671 [enterobacteriaceae-laboratory-confirmed-cases/carbapenemase-producing-](https://www.gov.uk/government/publications/carbapenemase-producing-enterobacteriaceae-laboratory-confirmed-cases/carbapenemase-producing-enterobacteriaceae-laboratory-confirmed-cases-2003-to-2013)
672 [enterobacteriaceae-laboratory-confirmed-cases-2003-to-2013](https://www.gov.uk/government/publications/carbapenemase-producing-enterobacteriaceae-laboratory-confirmed-cases/carbapenemase-producing-enterobacteriaceae-laboratory-confirmed-cases-2003-to-2013).

673 9. PHE PHE. English surveillance programme for antimicrobial utilisation and resistance
674 (ESPAUR) report 2019-2020.

675 10. Patel B, Hopkins, K., Meunier, D., Staves, P., Hopkins, S., & Woodford, N. A Ten-Year
676 Review of Carbapenemase Producing Enterobacterales (CPE) in London, United Kingdom.
677 *Infection Control & Hospital Epidemiology.* 2020;41(S6-S7).

678 11. Hrabak J, Chudackova E, Papagiannitsis CC. Detection of carbapenemases in
679 Enterobacteriaceae: a challenge for diagnostic microbiological laboratories. *Clin Microbiol*
680 *Infect.* 2014;20(9):839-53.

681 12. Harris SR, Feil EJ, Holden MT, Quail MA, Nickerson EK, Chantratita N, et al. Evolution
682 of MRSA during hospital transmission and intercontinental spread. *Science.*
683 2010;327(5964):469-74.

684 13. Jenkins C, Dallman TJ, Launders N, Willis C, Byrne L, Jorgensen F, et al. Public Health
685 Investigation of Two Outbreaks of Shiga Toxin-Producing *Escherichia coli* O157 Associated
686 with Consumption of Watercress. *Appl Environ Microbiol.* 2015;81(12):3946-52.

687 14. Raven KE, Reuter S, Reynolds R, Brodrick HJ, Russell JE, Torok ME, et al. A decade of
688 genomic history for healthcare-associated *Enterococcus faecium* in the United Kingdom and
689 Ireland. *Genome Res.* 2016;26(10):1388-96.

690 15. Reuter S, Harrison TG, Koser CU, Ellington MJ, Smith GP, Parkhill J, et al. A pilot study
691 of rapid whole-genome sequencing for the investigation of a *Legionella* outbreak. *BMJ*
692 *Open.* 2013;3(1).

693 16. Rew V, Mook P, Trienekens S, Baker KS, Dallman TJ, Jenkins C, et al. Whole-genome
694 sequencing revealed concurrent outbreaks of shigellosis in the English Orthodox Jewish
695 Community caused by multiple importations of *Shigella sonnei* from Israel. *Microb Genom.*
696 2018;4(3).

697 17. PHE PHE. PHE carbapenemase-producing Enterobacteriaceae toolkit published 2013
698 [Available from: [https://www.gov.uk/government/news/phe-carbapenemase-producing-](https://www.gov.uk/government/news/phe-carbapenemase-producing-enterobacteriaceae-toolkit-published)
699 [enterobacteriaceae-toolkit-published](https://www.gov.uk/government/news/phe-carbapenemase-producing-enterobacteriaceae-toolkit-published)].

700 18. Andrews S. FastQC: A quality control tool for high throughput sequence data:
701 Available online at <https://www.bioinformatics.babraham.ac.uk/projects/fastqc/> [accessed
702 2020]. 2010.

703 19. Ewels P, Magnusson M, Lundin S, Kaller M. MultiQC: summarize analysis results for
704 multiple tools and samples in a single report. *Bioinformatics.* 2016;32(19):3047-8.

705 20. Wood DE, Lu J, Langmead B. Improved metagenomic analysis with Kraken 2.
706 *Genome Biol.* 2019;20(1):257.

707 21. Chen S, Zhou Y, Chen Y, Gu J. fastp: an ultra-fast all-in-one FASTQ preprocessor.
708 Bioinformatics. 2018;34(17):i884-i90.

709 22. De Coster W, D'Hert S, Schultz DT, Cruts M, Van Broeckhoven C. NanoPack:
710 visualizing and processing long-read sequencing data. Bioinformatics. 2018;34(15):2666-9.

711 23. Wick RR. Filtlong. <https://github.com/rrwick/Filtlong>. 2017.

712 24. Kolmogorov M, Yuan J, Lin Y, Pevzner PA. Assembly of long, error-prone reads using
713 repeat graphs. Nat Biotechnol. 2019;37(5):540-6.

714 25. Wick RR, Judd LM, Gorrie CL, Holt KE. Unicycler: Resolving bacterial genome
715 assemblies from short and long sequencing reads. PLoS Comput Biol. 2017;13(6):e1005595.

716 26. Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, Phillippy AM. Canu: scalable
717 and accurate long-read assembly via adaptive k-mer weighting and repeat separation.
718 Genome Res. 2017;27(5):722-36.

719 27. Seemann T. mlst. <https://github.com/tseemann/mlst>. 2013.

720 28. Jolley KA, Maiden MC. BIGSdb: Scalable analysis of bacterial genome variation at the
721 population level. BMC Bioinformatics. 2010;11:595.

722 29. Seemann T. Abricate, Github <https://github.com/tseemann/abricate>.

723 30. Jia B, Raphenya AR, Alcock B, Waglechner N, Guo P, Tsang KK, et al. CARD 2017:
724 expansion and model-centric curation of the comprehensive antibiotic resistance database.
725 Nucleic Acids Res. 2017;45(D1):D566-D73.

726 31. Carattoli A, Zankari E, Garcia-Fernandez A, Voldby Larsen M, Lund O, Villa L, et al. In
727 silico detection and typing of plasmids using PlasmidFinder and plasmid multilocus
728 sequence typing. Antimicrob Agents Chemother. 2014;58(7):3895-903.

729 32. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search
730 tool. J Mol Biol. 1990;215(3):403-10.

731 33. Wyres KL, Wick RR, Gorrie C, Jenney A, Follador R, Thomson NR, et al. Identification
732 of *Klebsiella* capsule synthesis loci from whole genome data. *Microb Genom.*
733 2016;2(12):e000102.

734 34. Lam MMC, Wick RR, Watts SC, Cerdeira LT, Wyres KL, Holt KE. A genomic
735 surveillance framework and genotyping tool for *Klebsiella pneumoniae* and its related
736 species complex. *Nat Commun.* 2021;12(1):4188.

737 35. Jain C, Rodriguez RL, Phillippy AM, Konstantinidis KT, Aluru S. High throughput ANI
738 analysis of 90K prokaryotic genomes reveals clear species boundaries. *Nat Commun.*
739 2018;9(1):5114.

740 36. Seemann T. Snippy: fast bacterial variant calling from NGS reads. 2015.

741 37. Harris SR. SKA: Split Kmer Analysis Toolkit for Bacterial Genomic Epidemiology.
742 *BioRxiv.* 2018.

743 38. Ondov BD, Treangen TJ, Melsted P, Mallonee AB, Bergman NH, Koren S, et al. Mash:
744 fast genome and metagenome distance estimation using MinHash. *Genome Biol.*
745 2016;17(1):132.

746 39. David S, Reuter S, Harris SR, Glasner C, Feltwell T, Argimon S, et al. Epidemic of
747 carbapenem-resistant *Klebsiella pneumoniae* in Europe is driven by nosocomial spread. *Nat*
748 *Microbiol.* 2019;4(11):1919-29.

749 40. Sherry NL, Lane CR, Kwong JC, Schultz M, Sait M, Stevens K, et al. Genomics for
750 molecular epidemiology and detecting transmission of carbapenemase-producing
751 *Enterobacterales* in Victoria, Australia, 2012-2016. *J Clin Microbiol.* 2019.

752 41. Schurch AC, Arredondo-Alonso S, Willems RJL, Goering RV. Whole genome
753 sequencing options for bacterial strain typing and epidemiologic analysis based on single

754 nucleotide polymorphism versus gene-by-gene-based approaches. Clin Microbiol Infect.
755 2018;24(4):350-4.

756 42. Pierce NT IL, Reiter T et al. . Large-scale sequence comparisons with sourmash.
757 F1000Res. 2019;8(1006).

758 43. Blackwell GA, Hunt M, Malone KM, Lima L, Horesh G, Alako BTF, et al. Exploring
759 bacterial diversity via a curated and searchable snapshot of archived DNA sequences. PLoS
760 Biol. 2021;19(11):e3001421.

761 44. Huang W, Li L, Myers JR, Marth GT. ART: a next-generation sequencing read
762 simulator. Bioinformatics. 2012;28(4):593-4.

763 45. Croucher NJ, Page AJ, Connor TR, Delaney AJ, Keane JA, Bentley SD, et al. Rapid
764 phylogenetic analysis of large samples of recombinant bacterial whole genome sequences
765 using Gubbins. Nucleic Acids Res. 2015;43(3):e15.

766 46. Price MN, Dehal PS, Arkin AP. FastTree: computing large minimum evolution trees
767 with profiles instead of a distance matrix. Mol Biol Evol. 2009;26(7):1641-50.

768 47. Letunic I, Bork P. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic
769 tree display and annotation. Nucleic Acids Res. 2021;49(W1):W293-W6.

770 48. Fajardo-Lubian A, Ben Zakour NL, Agyekum A, Qi Q, Iredell JR. Host adaptation and
771 convergent evolution increases antibiotic resistance without loss of virulence in a major
772 human pathogen. PLoS Pathog. 2019;15(3):e1007218.

773 49. Russo TA, Olson R, Fang CT, Stoesser N, Miller M, MacDonald U, et al. Identification
774 of Biomarkers for Differentiation of Hypervirulent *Klebsiella pneumoniae* from Classical *K.*
775 *pneumoniae*. J Clin Microbiol. 2018;56(9).

776 50. Paul M, Narendrakumar L, A RV, Joseph I, Thomas S. Genome sequence of a
777 multidrug-resistant *Klebsiella pneumoniae* ST78 with high colistin resistance isolated from a
778 patient in India. *J Glob Antimicrob Resist*. 2019;17:187-8.

779 51. Wilson HJ, Khokhar F, Enoch DA, Brown NM, Ahluwalia J, Dougan G, et al. Point-
780 prevalence survey of carbapenemase-producing Enterobacteriaceae and vancomycin-
781 resistant enterococci in adult inpatients in a university teaching hospital in the UK. *J Hosp*
782 *Infect*. 2018;100(1):35-9.

783 52. Lubbert C, Lippmann N, Busch T, Kaisers UX, Ducomble T, Eckmanns T, et al. Long-
784 term carriage of *Klebsiella pneumoniae* carbapenemase-2-producing *K pneumoniae* after a
785 large single-center outbreak in Germany. *Am J Infect Control*. 2014;42(4):376-80.

786 53. Evain S, Bourigault C, Juvin ME, Corvec S, Lepelletier D. Carbapenemase-producing
787 Enterobacteriaceae digestive carriage at hospital readmission and the role of antibiotic
788 exposure. *J Hosp Infect*. 2019;102(1):25-30.

789 54. Dancer SJ. How do we assess hospital cleaning? A proposal for microbiological
790 standards for surface hygiene in hospitals. *J Hosp Infect*. 2004;56(1):10-5.

791 55. Otter JA, Burgess P, Davies F, Mookerjee S, Singleton J, Gilchrist M, et al. Counting
792 the cost of an outbreak of carbapenemase-producing Enterobacteriaceae: an economic
793 evaluation from a hospital perspective. *Clin Microbiol Infect*. 2017;23(3):188-96.

794 56. Kwong JC, McCallum N, Sintchenko V, Howden BP. Whole genome sequencing in
795 clinical and public health microbiology. *Pathology*. 2015;47(3):199-210.

796 57. Poirel L, Bonnin RA, Nordmann P. Genetic features of the widespread plasmid coding
797 for the carbapenemase OXA-48. *Antimicrob Agents Chemother*. 2012;56(1):559-62.

798 58. Doi Y, Hazen TH, Boitano M, Tsai YC, Clark TA, Korlach J, et al. Whole-genome
799 assembly of *Klebsiella pneumoniae* coproducing NDM-1 and OXA-232 carbapenemases

800 using single-molecule, real-time sequencing. *Antimicrob Agents Chemother.*
801 2014;58(10):5947-53.

802 59. Mouftah SF, Pal T, Darwish D, Ghazawi A, Villa L, Carattoli A, et al. Epidemic IncX3
803 plasmids spreading carbapenemase genes in the United Arab Emirates and worldwide.
804 *Infect Drug Resist.* 2019;12:1729-42.

805 60. Blackwell GA, Doughty EL, Moran RA. Evolution and dissemination of L and M
806 plasmid lineages carrying antibiotic resistance genes in diverse Gram-negative bacteria.
807 *Plasmid.* 2021;113:102528.

808