# Inter-species geographic signatures for tracing horizontal gene transfer and long-term persistence of carbapenem resistance

Rauf Salamzade[1,2], Abigail L. Manson[1], Bruce J. Walker[1,3], Thea Brennan-Krohn[4], Colin J. Worby[1], Peijun Ma[1], Lorrie L. He[1], Terrance P. Shea[1], James Qu[1], Sinéad B. Chapman[1], Whitney Howe[1], Sarah K. Young[1], Jenna I. Wurster[5], Mary L. Delaney[6], Sanjat Kanjilal[6,7], Andrew B. Onderdonk[6], Cassiana E. Bittencourt[8], Gabrielle M. Gussin[9], Diane Kim[9], Ellena M. Peterson[8], Mary Jane Ferraro[10^], David C. Hooper[10], Erica S. Shenoy[10], Christina A. Cuomo[1], Lisa A. Cosimi[1,6], Susan S. Huang[9], James E. Kirby[4], Virginia M. Pierce[10], Roby P. Bhattacharyya[1,10] and Ashlee M. Earl[1*]

## Abstract

**Background:** Carbapenem-resistant *Enterobacterales* (CRE) are an urgent global health threat. Inferring the dynamics of local CRE dissemination is currently limited by our inability to confidently trace the spread of resistance determinants to unrelated bacterial hosts. Whole-genome sequence comparison is useful for identifying CRE clonal transmission and outbreaks, but high-frequency horizontal gene transfer (HGT) of carbapenem resistance genes and subsequent genome rearrangement complicate tracing the local persistence and mobilization of these genes across organisms.

**Methods:** To overcome this limitation, we developed a new approach to identify recent HGT of large, near-identical plasmid segments across species boundaries, which also allowed us to overcome technical challenges with genome assembly. We applied this to complete and near-complete genome assemblies to examine the local spread of CRE in a systematic, prospective collection of all CRE, as well as time- and species-matched carbapenem-susceptible *Enterobacterales*, isolated from patients from four US hospitals over nearly 5 years.

**Results:** Our CRE collection comprised a diverse range of species, lineages, and carbapenem resistance mechanisms, many of which were encoded on a variety of promiscuous plasmid types. We found and quantified rearrangement, persistence, and repeated transfer of plasmid segments, including those harboring carbapenemases, between organisms over multiple years. Some plasmid segments were found to be strongly associated with specific locales, thus representing *geographic signatures* that make it possible to trace recent and localized HGT events. Functional analysis of these signatures revealed genes commonly found in plasmids of nosocomial pathogens, such as functions

Salamzade *et al. Genome Medicine* (2022) 14:37

Page 2 of 22

required for plasmid retention and spread, as well survival against a variety of antibiotic and antiseptics common to the hospital environment.

**Conclusions:** Collectively, the framework we developed provides a clearer, high-resolution picture of the epidemiology of antibiotic resistance importation, spread, and persistence in patients and healthcare networks.

## Background

Carbapenem-resistant *Enterobacterales* (CRE) cause difficult-to-treat infections [1–5] with high mortality rates [6–8], largely because antibiotic options for treating them are limited [9, 10]. CRE are also highly transmissible through contact [11–16], leading to nosocomial outbreaks that are costly to contain with significant patient morbidity and mortality [17–19], making CRE a leading healthcare problem [20–25]. Despite the adoption of extensive infection control measures [11, 12] that have begun to curb the incidence of CRE infection in some countries, the global incidence of CRE infections continues to rise [20, 23, 24, 26–28].

Genomic studies are providing new insights into the emergence and spread of CRE within healthcare institutions [29–31]. Comprising many species, most commonly *Klebsiella pneumoniae*, *Escherichia coli*, and *Enterobacter cloacae* complex, CRE can be found in diverse environments within hospitals [32–35], ranging from the gastrointestinal tract of asymptomatic carriers [36, 37] to contaminated hospital sinks and drains [16, 38–41]. CRE are readily acquired and spread from these reservoirs [13, 39], including hospital-adapted *high-risk lineages* (e.g., *K. pneumoniae* sequence type (ST)-258) [42] that are associated with both intra- and inter-facility clonal transmission [13, 29, 43, 44]. Known reservoirs tend to be polymicrobial and thus can act as sites for CRE diversification and are believed to have played important roles in horizontal gene transfer (HGT) of carbapenemases [39, 41].

Though bioinformatic approaches targeting the sequences surrounding carbapenemases have been used to predict the movement of carbapenemases across some CRE populations [45], tracing the movement and persistence of these genes within facilities is complicated. CRE reservoirs can be large and diverse, and comparatively few have been studied [16, 39, 41, 46, 47]. Furthermore, HGT rates are predicted to be high [48–50], and carbapenemase-containing plasmids frequently recombine at sites of repetitive sequence, leading to mosaic plasmid structures [31, 51]. Given this, carbapenemase-containing plasmid sequences are challenging to accurately assemble. Long-read sequencing technologies have provided the strongest evidence for HGT of plasmids containing carbapenemases between *Enterobacterales* within individual hospitals, as well as for transmission between patients and hospital reservoirs [39, 41]. However, it remains challenging to trace carbapenemase movement within and between different plasmid backgrounds and organisms even when using high-quality assemblies generated using long-read sequencing data.

We previously published results from a surveillance study conducted in 2012–2013 [30] that highlighted the shortcomings of existing methods for tracking CRE movement. Here, we expanded this initial study to capture all patient-derived CRE, regardless of infection site or resistance mechanism, from across the same four hospitals over an additional 3-year period from December 2013 through 2016. Our sequencing methodology, which combined short paired-end and long-insert mate-pair Illumina sequencing libraries, enabled high-quality whole-genome and plasmid assemblies for over 600 isolates. We developed a novel computational methodology to holistically screen for conserved segments within mosaic plasmids that allowed us to trace the movement and persistence of genes, including carbapenemases, within facilities. This approach revealed near-identical plasmid segments, including carbapenemase encoding segments, that crossed plasmid and species boundaries. Many of these were specific to and recurrent within a single hospital site, revealing extensive linkages between patient isolates that would have been missed otherwise. From long-read sequencing of select isolates, we also observed rapid plasmid mosaicism, including the shuffling of segments into new genomic locations, occurring on the same timescale as single nucleotide variation.

## Methods

### Isolate collection and drug susceptibility testing

Our sample collection represented a continuation of our previous prospective study, conducted in 2012–2013 [30]. Between December 2013 and November 2016, we collected additional samples from symptomatic patients at Beth Israel Deaconess Medical Center (BIDMC), Brigham and Women's Hospital (BWH), and Massachusetts General Hospital (MGH) in Boston, MA, and the University of California Irvine (UCI) Medical Center in Orange, CA. We collected all isolates from clinical samples sent to these hospitals' clinical microbiology laboratories that, using the laboratories' standard operating procedures [30], were identified as *Enterobacterales* with a meropenem minimum

Salamzade *et al. Genome Medicine*      (2022) 14:37

Page 3 of 22

inhibitory concentration (MIC) $\geq$ 2 μg/mL, the threshold that we used to define resistance throughout our analysis. We thus included samples categorized as both *intermediate* (2 μg/mL) and *resistant* ($\geq$ 4 μg/mL) by the Clinical and Laboratory Standards Institute interpretative criteria [52]. For each resistant isolate, we also collected date- and species-matched meropenem-susceptible isolates. UCI submitted two carbapenem-susceptible isolates per resistant isolate; the other three hospitals submitted a single susceptible isolate per resistant isolate. In total, we collected 347 new isolates (147 resistant and 200 susceptible isolates; one isolate per patient). Isolates whose meropenem resistance phenotype was discordant with the genotype were retested using a previously validated automated digital dispensing method [53] adapted from standard broth microdilution procedures [52]. Meropenem was tested against bacterial isolates at doubling dilution concentrations from 0.016 to 32 μg/mL. All isolate cultures are available from the corresponding author upon reasonable request.

Combining our newly collected samples from December 2013 to November 2016 with the 261 *Enterobacterales* isolates that we previously sequenced in 2012–2013 from these four hospitals [30] resulted in a total of 608 isolates (of which 275 were meropenem-resistant; Additional file 1: Table S1). The 261 previously reported isolates consisted of the retrospective *Boston Historical Collection*, consisting of 49 isolates collected between September 2007 and July 2012, including 17 isolates grouped into six sets of same-patient isolates, and 74 CRE and 138 carbapenem-susceptible *Enterobacterales* (CSE) isolates collected prospectively (including 11 isolates from 3 sets of same-patient isolates). In total, our prospective collection included 559 isolates collected between 2012 and 2016.

Institutional review board (IRB) approval was granted by the Massachusetts Institute of Technology Committee on the Use of Humans as Experimental Subjects. Samples were collected under study approvals of the IRB committees of the participating institutions: Massachusetts General Brigham (covering both MGH and BWH); Beth Israel Deaconess Medical Center, Boston; and University of California, Irvine (IRB approval numbers 2012P001062, 2012P000205, and 2012-8957, respectively). Waivers of informed consent were granted for each institution due to the fact that bacterial isolates were collected for the study after all clinical work was completed, including after some patients were discharged, which made obtaining informed consent logistically infeasible. This study conforms to the principles of the Belmont Report and follows the ethical principles relevant to this work as outlined in the Declaration of Helsinki.

## Genome sequencing, assembly, and annotation
### Illumina sequencing and assembly
We prepared whole-genome paired-end and mate-pair libraries for 347 isolates and paired-end sequenced them as previously described [30] on Illumina HiSeq 2500 or HiSeq X sequencers (Additional file 1: Tables S1-S2). Genome assembly and annotation were carried out as previously described, using the Broad Institute's prokaryotic annotation pipeline [30]. Sequencing reads and assemblies were submitted to GenBank under bioproject PRJNA271899 (https://www.ncbi.nlm.nih.gov/bioproject/PRJNA271899) [54].

### Long-read sequencing and assembly
We selected a subset of 19 isolates for further sequencing using Oxford Nanopore Technology (ONT) (Additional file 1: Table S3). For the 12 isolates containing signatures Sig5.1-CP or Sig5.6-CP, we used the Oxford Nanopore Rapid Barcoding kit (SQK-RBK004) and ran the samples on a Minion (Oxford Nanopore Technologies Ltd, Science Park, UK). For an additional 7 isolates, 600 ng of DNA from each sample was used as input into the Oxford Nanopore 1D ligation library construction protocol (SQK-LSK109) following the manufacturer's recommendation. Samples were barcoded using the Native Barcoding Expansion 1-12 kit to run in batches of between 1 and 4 samples per flow cell on a GridIon (Oxford Nanopore Technologies Ltd, Science Park, UK).

Oxford Nanopore (ONT) reads were demultiplexed using Deepbinner (v0.2.0) [55], trimmed of any remaining adapter using Porechop (v0.2.3; https://github.com/rrwick/Porechop) [56], and subsampled to approximately 50x depth of genome coverage. Illumina reads were trimmed of adapter using Trim Galore (v0.5.0; https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/) [57] and subsampled to approximately 100x depth of genome coverage. Two Unicycler (v0.4.3 or v0.4.4, with default settings) [58] hybrid assemblies were generated for each sample, one assembly combining the Illumina 100x data set with the 50x subsampled ONT data set and another assembly combining the Illumina 100x data set with the full set of ONT reads (if >50x).

ONT reads were aligned to Unicycler contigs using minimap2 (v2.15) [59]. Illumina reads were aligned to Unicycler contigs using BWA mem (v0.7.17) [60], and the resulting alignments were input to Pilon (v1.23) [61] for assembly polishing. Contigs were screened for adapter sequence and then input to GAEMR (v1.0.1; https://github.com/broadinstitute/GAEMR) [62] which produced chart and metric tables for use in the manual assembly analysis process. Hybrid assemblies were annotated as above. Reads were submitted to SRA under

Salamzade *et al. Genome Medicine*     (2022) 14:37

Page 4 of 22

bioproject PRJNA271899 [54] (https://www.ncbi.nlm.nih.gov/bioproject/PRJNA271899).

### *Annotation of resistance genes*
We annotated our assemblies for the presence of resistance genes as done previously [30], but we queried predicted genes with an updated set of antimicrobial resistance databases using BLASTn in megablast mode [63, 64]: (i) our database of carbapenem-hydrolyzing beta-lactamases [30]; (ii) ResFinder [65], downloaded January 22, 2018; and (iii) the antimicrobial resistance database of the National Database of Antibiotic Resistant Organisms (https://www.ncbi.nlm.nih.gov/pathogens/antimicrobial-resistance/, downloaded January 22, 2018) [66]. Genes with hits to any of these databases (*E*-value $< 10^{-10}$; query coverage $\geq$ 80%) were annotated with the name of the antibiotic resistance gene with the highest BLAST bit score; matches to multiple databases were resolved with the same order of precedence with which the databases are listed above (Additional file 1: Tables S4-S6). A subset of beta-lactamases were designated as ESBLs by our pipeline for identifying resistance genes (Additional file 1: Table S7), together with additional evidence from the literature [67].

In order to identify defects in porin sequences contributing to carbapenem resistance, we first identified all porin genes by sequence similarity search [64]. We queried our set of predicted genes against a database of *ompC* and *ompF* reference sequences [30], retaining the best hit with *E*-value $< 10^{-10}$ and $\geq$ 80% coverage of the reference sequence. We also searched for matches in the full sequences of the assemblies, in order to identify genes or gene fragments that were not part of the predicted gene set. Within the resulting set of predicted porin genes, we then identified mutations by comparing them to the best matching reference sequence with MUMmer (v3) [68]. Additionally, we pinpointed porins disrupted by insertion sequences by submitting porin sequences and their promoter regions (500 bp upstream) to the ISfinder [69] BLAST server, retaining hits with *E*-value $< 10^{-10}$. We regarded a porin gene as disrupted if (i) no BLAST hit for the gene was produced, (ii) the best matching predicted gene contained < 90% of the reference sequence, (iii) a frameshift mutation affected 30 codons of the gene or more, or (iv) an insertion sequence was found disrupting the porin gene or up to 300 bp upstream.

In order to search for evidence of genotypic resistance that may not have been captured in our assemblies, we applied ARIBA (v2.10.3) [70], a read-based gene search tool leveraging local targeted assembly, with a database comprising carbapenemases and common ESBL genes. We considered ARIBA calls with one non-synonymous mutation or less, and a coverage of 100% (Additional file 1: Table S8).

### *Plasmid annotations*
We used MOB-suite (database version from January 2, 2019) [71] to identify plasmid scaffolds, plasmid replicon types, and relaxase types. MOB-suite is a bioinformatic tool that predicts and categorizes plasmid scaffolds into discrete "plasmid groups" by clustering input scaffolds with reference plasmids if their estimated ANI [72] is at least 95%. To supplement MOB-suite's plasmid replicon database, we added 111 additional replicons from the PlasmidFinder database [73]. The tool either assigned a reference plasmid group to each plasmid scaffold or called it "novel" if there was no match. In each assembly, plasmid scaffolds with the same plasmid group were predicted to be part of the same plasmid and grouped (but "novel" scaffolds were not grouped). While the largest scaffold to be assigned a plasmid group was 481 kb in size, longer (500 kb to 5.6 Mb) scaffolds marked as "novel" were inspected and frequently found to be chromosomal and to contain plasmid replicon and relaxase genes, suggesting that they contained integrative plasmids or similar elements [74]. For this reason, 162 "novel" scaffolds with length > 500 kb were not assumed to be part of plasmids.

## Comparative genomics
### *Orthogroup clustering and construction of multi-species phylogeny*
In order to identify genes shared between isolates, we performed orthogroup clustering for our entire set of 608 genomes using Synerclust (v1.0.1) [75], a tool which provided the high level of scalability needed for this large set of genomes, and also leveraged the syntenic organization of genes to help in defining orthogroups. We generated a final set of orthogroups using an iterative, two-step process. First, we ran Synerclust with an approximate, *k*-mer-based input dendrogram generated by (i) k-merizing our set of genomes with a k-mer size of 15, (ii) computing a similarity matrix using the Jaccard index [72] to compare each pair of genomes in our dataset, and (iii) computing a dendrogram with the neighbor joining tree algorithm [76] contained in the ape [77] package (v5.0) of R [78] (v3.4.0). We then ran SynerClust with default parameters in order to identify ortholog clusters. We produced a codon-based multiple-sequence alignment for each single-copy core gene using Muscle [79] and produced a concatenated alignment of all genes by extracting alignment columns without gaps. We then computed a phylogenetic tree using FastTree2 (v2.1.3) [80] with default parameters, which we then used as input for a second iteration of Synerclust. The orthogroup output from the

Salamzade *et al. Genome Medicine* (2022) 14:37

Page 5 of 22

second iteration of Synerclust was used to establish the final single-copy core gene set to be used for downstream analysis, including construction of a single-copy core alignment (as in the first iteration). This final alignment spanned 676,371 nucleotide sites, of which 348,152 were variable, and was used to generate a phylogenetic tree using RAxML [81] (v7.3.3), using rapid analysis of 1000 bootstrap replicates. To generate phylogenetic trees containing only subsets of isolates, we used PareTree (v1.0.2; (http://emmahodcroft.com/PareTree.html)) [82].

### Species identification
We used average nucleotide identity (ANI) to obtain species designations for each isolate. For each pair of isolates, we used alignments of all shared genes (using orthogroup clusters) to compute ANI [83]. We compared to reference assemblies obtained from the NCBI taxonomy browser (https://www.ncbi.nlm.nih.gov/taxonomy) [66] to obtain species designations.

### Construction of species-specific multiple-sequence alignments of the core genome
In order to construct more detailed, species-specific SNV-based phylogenies, for each species, we selected the assembly with the smallest number of contigs as a reference. Then, we produced alignments for both the short paired-end and long-insert mate-pair sequencing reads of each isolate using BWA mem [60] and sorted the alignments with Picard SortSam (v2.20.6; http://broadinstitute.github.io/picard) [84]. Finally, we used Pilon [61] in order to call variants. We produced multiple-sequence alignments based on the variant calls for all isolates of each species, excluding alignment positions with insertions or deletions. The number of nucleotide sites in these alignments ranged from 2,931,929 sites for *E. coli* to 5,048,275 for *K. oxytoca* (Additional file 1: Table S9).

### Computing phylogenetic trees after removing effects of recombination
To construct more accurate species-specific phylogenies using our SNV-based alignments, we used ClonalFrameML (v1.11) [85] to identify and remove alignment regions with evidence for recombination. We ran ClonalFrameML with default parameters and 100 bootstrap replicates, using an input phylogenetic tree generated with FastTree [80] and 100 bootstrap replicates. We produced a recombination-removed multiple-sequence alignment by removing any site from the species-specific alignment in which recombination was detected in at least one isolate. The number of nucleotide sites in the resulting alignments ranged from 593,538 for *E. coli* to 4,978,737 for *K. oxytoca* (Additional file 1: Table S9). Using these alignments, we used RAxML (v7.3.3) [81]

with 1000 rapid bootstrap replicates to generate final phylogenies. We also used these alignments to calculate core SNV distances between each pair of isolates.

### Determination of lineages and sequence types
We computationally determined lineages in each species using the recombination-removed phylogenetic trees. We assigned isolates to the same lineage if they were connected by a path consisting entirely of branches with a length of $10^{-4}$ substitutions per nucleotide site or less. Sequence types were computationally determined as before [30]. In brief, sequence types were determined using our Broad pipeline for determining sequence types. In brief, this script uses BLAST to compare the assembly against a database of sequences from pubMLST [86] using a 95% threshold in order to predict the sequence type. For the isolates belonging to each sequence type, we identified the lineage most commonly assigned to the members of this sequence type; this mapping produced the correct lineage in 87% of isolates. Conversely, for the isolates belonging to each lineage, we identified the sequence type most commonly assigned to members of this lineage; this mapping was correct for 98% of isolates. Lineages corresponding to the following sequence types were considered *high-risk* [42]: *E. coli* ST38, ST69, ST131, ST155, ST393, ST405, and ST648 and *K. pneumoniae* ST14, ST37, ST147, and ST258.

### Assessment of established genetic markers to trace local movement of bla$_{KPC}$
We characterized Tn*4401* structural variants and isoforms, as well as target site duplication (TSD) flanking sequences of Tn*4401* using TETyper (v1.1) [45] for the 608 samples included in our study. We analyzed 5 bp surrounding the three most common Tn*4401* isoforms (Tn*4401*a, Tn*4401*b, and Tn*4401*d) in our assemblies.

### Identification of geographic signatures
The widespread geographic distribution of plasmids and their rapid rearrangement complicates tracing the movement of resistance genes within plasmids. Because specific genetic markers surrounding resistance determinants were also mostly geographically widespread, we developed a method to search for longer conserved sequences that may serve as markers of local HGT of clinically important genes. The *ConSequences* software suite (v1.0; https://github.com/broadinstitute/ConSequences) [87] identifies nearly identical ≥ 10 kb segments conserved between two or more plasmids (Additional file 2: Figs. S1-S2).

Salamzade *et al. Genome Medicine*     (2022) 14:37

Page 6 of 22

### Selection of plasmid sequences

To construct the database of plasmid sequences that we searched for geographic signatures, we included all scaffolds between 10 and 500 kb that were not classified as chromosomal by MOB-suite [71]. We excluded scaffolds classified as plasmidic and longer than 500 kb, since we found these to be often misclassified (*Plasmid annotations*). Circular and complete representations of plasmids [88, 89] were determined when a plasmid scaffold showed both significant overlap between its ends (*e*-value $< 10^{-5}$) and had at least five mate-pair sequencing reads bridging scaffold ends using Bowtie2 (v2.3.4.3) [90]. A Python program was developed for identifying scaffolds predicted to correspond to completed plasmid instances and is included in the *ConSequences* software suite [87].

### Identification of highly conserved 10-kb windows shared across pairs of plasmid backbones

All plasmid-predicted scaffolds were aligned in a pairwise manner using BLASTn (v2.7.1) in megablast mode [63]. In order to account for circularity of complete plasmids [88, 89], we duplicated 10 kb from the beginning of the scaffold and appended it to its end. For each scaffold, a sliding window approach, with a window size of 10 kb and a step size of 100 bp, was applied to identify highly conserved and contiguous windows shared with at least one other scaffold, where matches were required to exhibit $\geq$ 99% identity and coverage through single high scoring pairs (HSPs). A 10-kb window size was selected for the analysis because (i) the vast majority (98.7%) of HSPs with identity $\geq$ 99% were shorter; thus, 10 kb and longer sequences were outliers and hypothesized to share a recent ancestral origin; and (ii) it allowed us to capture whole operons or large transposable elements and their surrounding contexts. For example, isoforms of *Tn4401* typically span around 10 kb, are well conserved, and are often found on different plasmid backbones [30, 45].

### Delineating boundaries of shared segments between plasmids

We developed a novel algorithm to identify the boundaries of shared segments spanning multiple adjacent windows along a reference plasmid scaffold by first traversing blocks of adjacent windows in the forward direction and then repeating the process in the reverse direction. For each 10-kb window in the series, the *focal window*, we checked whether downstream windows showed conservation in the same set of scaffolds as the focal window, tracking how far the *segment* could potentially be expanded. This procedure was then repeated in the reverse direction for the same series of windows. After potential segments were identified from both forward and reverse traversals, they were merged if they exhibited

overlap in coordinates and shared conservation in a common set of scaffolds (Additional file 2: Fig. S1). As identical segments were often obtained by using different reference scaffolds, we used CD-HIT (v.4.8.1) [91] to cluster sequences with $\geq$ 99% global identity and $\geq$ 95% coverage of both sequences. Representative segments were selected from each cluster by maximizing for the number of samples segments were found in.

### Filtering shared segments to identify geographic signatures

In order to identify signatures, we filtered for segments which had broad host range and exhibited geographic association. Starting with the set of segments conserved across multiple species, we identified those which were found exclusively in isolates from a single city (Boston, MA or Orange, CA). These segments were then screened for uniqueness against NCBI's Nucleotide Collection database (nt; https://ftp.ncbi.nlm.nih.gov/blast/db/; downloaded in December 2019) [92], using thresholds of 98% identity and 95% query coverage. Hits matching samples in NCBI sourced from the same city, assessed using BioSample meta information and latitudinal coordinates with geopy (v1.20.0; https://github.com/geopy/geopy) [93], were retained as potential geographic signatures.

The presence of assembly errors, including incorrect copy counts for tandem repeats, could lead to the incorrect association of segments with geographies. Thus, we checked whether any of the potential geographic signatures contained such tandem repeats using Pilon [61]. Paired-end library sequencing reads from each isolate found to harbor a signature were aligned to the signature's reference sequence using BWA mem with default settings [60]. Then, Pilon [61] was run with the options: "*--vcf --fix all,breaks --mindepth 5.*" Instances which triggered a fix break report with the flag "TandemRepeat," indicating the segment likely contained a tandem repeat motif, were identified and removed to ensure geographic association was not driven by faulty estimation of the tandem repeat motif's copy count.

We next performed a more comprehensive assessment of uniqueness for each of the signatures using BIGSI [94], searching against all raw sequencing read sets in a snapshot of the ENA/SRA database taken on December 2016 [92, 95], contemporaneous with the most recent isolation dates for the 608 samples in this study. The SRA/ENA snapshot provided a broader database (455,632 read sets) compared to our original screening against the nt database, which included only a subset of assemblies available in the nt database and did not account for bacterial samples with sequencing data but no assembly. To perform this search, we used a sliding window (2-kb window; 1 kb step) across each signature to identify read sets

containing at least 99% of all *k*-mers for each window. Matching read sets were downloaded from EBI's ENA database [95] and further searched using a *k*-mer-based methodology (described below) to more stringently assess whether they harbored any of the geographic signature sequences. Coordinates for these signature sequences as they appear on our assemblies in GenBank are provided (Additional file 1: Table S10).

### *Searching for additional instances of geographic signatures directly in raw Illumina sequencing read sets*

It is possible that instances of geographic signatures were missed in our dataset of predicted plasmid segments since (i) not all assemblies contained finished, circular representations of plasmids and (ii) chromosomal scaffolds were not accounted for in our original search for signatures.

In order to recover missing signature instances, we searched the raw sequencing read sets against each multi-species geographic signature (Additional file 2: Fig. S2). First, we created reference guided multiple-sequence alignments for each geographic signature from all assemblies that contained that sequence. For each isolate, raw sequencing reads from both paired-end and mate-pair libraries were then downsampled to ~100x and Jellyfish (v2.3.0) was used to count 31-mers [96]. All 31-mers that were observed at least five times in each read set were next compared to each signature multiple-sequence alignment. A signature was considered present when all 31-mer windows along the multiple-sequence alignment had a corresponding match in the sample's set of 31-mers. As slight variations can exist between instances of a signature in the multiple-sequence alignment, a sample only needed to possess one of the possible 31-mers. Windows encompassing small deletions, insertions, or missing characters were ignored.

To further validate additional signature instances identified by the *k*-mer-based approach, we aligned a representative sequence for each signature to the draft assembly of the sample the sequence was extracted from using BLASTn [63]. Hits that achieved identity > 98% and signature coverage > 95% were retained for downstream analysis and were often captured from chromosomal scaffolds that were not part of the plasmid fraction that was originally analyzed or were missed due to assembly fragmentation. To prevent incorporating false positives into our analysis, we excluded instances where the assembly included only part of the respective signature that was embedded fully within a scaffold or had no significant alignment to a sample's assembly. To further refine the list of geographic signatures, we also checked whether smaller signatures nested within larger signatures were found in the same set of isolates (had identical

conservation profiles). For such cases, we excluded the smaller nested signature from consideration.

### *Functional annotation of signature gene content*

In order to characterize the diversity of functions encoded by plasmids and signatures, we clustered predicted protein sequences from all plasmid scaffolds larger than 10 kb using CD-HIT [91] with the parameters c=0.95, aS=0.9, and aL=0.9 (95% identity and 90% subject and query coverage). For each cluster, a representative protein was annotated by (i) using the Broad Institute's prokaryotic annotation pipeline [30]; (ii) transferring annotations from BLAST matches ($\geq$ 90% identity and $\geq$ 80% coverage) to NCBI RefSeq's non-redundant database of bacterial proteins (https://ftp.ncbi.nlm.nih.gov/refseq/release/bacteria/) [97]; and (iii) transferring protein-domain annotations from Pfam [98]. Phages were identified using ProphET [99]. To further refine our annotations of the subset of the genes found in signatures, we used keyword searches on these combined annotations, together with other gene-family-specific tools to identify genes within six broad functional categories of interest (Additional file 1: Tables S11 and S12).

Antibiotic resistance genes were predicted using the methods described above (the "Annotation of resistance genes" section). Chemical and heavy metal resistance operons, providing resistance to mercury, arsenic, tellurium, nickel, and copper, were identified by keyword searches within our combined annotations. Classification of operons required three or more functionally relevant genes located in close physical proximity to each other. Genes involved in efflux or response to stressors, including stressor efflux and transport (e.g., *silE*, *crcB*, *fieF*, *sugE*) and response genes (e.g., *dnaJ*, *usmG*, *frmR*) were identified by searching for keywords within our combined annotations. BLAST alignment [63] of proteins to representative transporter proteins in TCDB [100] (*e*-value < $10^{-10}$) was also used to flag additional proteins which might be involved in efflux, and such proteins were further examined through alignment to NCBI's comprehensive NR database. Conjugation machinery, notoriously difficult to identify and differentiate from other type IV secretion systems [101], was flagged using MacSyFinder tool [102] with CONJscan HMMs [103, 104]. To improve sensitivity, we also classified additional genes as likely related to conjugation machinery based on keywords found in our combined annotations. Plasmid uptake machinery included type I and II toxin-antitoxin systems [105] and anti-restriction proteins [106, 107]. Toxin and antitoxin genes were predicted using HMMer (v3) [108] with HMMs from the TAsmania database [109] (*e*-value <$10^{-5}$) and

Salamzade *et al. Genome Medicine*      (2022) 14:37

Page 8 of 22

filtered for the likelihood of representing a true toxin/antitoxin system through manual assessment of annotations. Anti-restriction proteins were identified by using keyword searches. Genes associated with mobile genetic elements included transposases, integrases or other recombinases, and homing endonucleases. Insertion elements and transposon genes were identified using ISFinder [69], as described above (the "Annotation of resistance genes" section). We found additional instances by searching for the keywords "transposases" and "IS" in general annotations together with manual inspection. Genes corresponding to integrases or alternate recombinases as well as homing endonucleases were also identified using keyword searches and manual validation.

To assess whether any genes found in signatures originated from sources outside the order of *Enterobacterales*, we aligned the nucleotide sequences of each gene to NCBI's Nucleotide Collection database (nt; July 2020) using BLASTn [63]. For each gene, the top 100 hits in nt were selected based on bitscore and then filtered to ensure they matched the query gene at 99% identity and 90% coverage. Next, the taxonomic information of each target sequence was extracted from the Entrez database using Biopython (v1.76) [110] to enable the calculation of what percentage belonged to bacteria from outside *Enterobacterales*.

### Statistical analysis

For assessments of statistical significance, a *p*-value threshold of 0.05 was used. The statistical significance of the differences in relaxase carriage in plasmids carrying carbapenemases vs. other plasmids was assessed using a two-sided Fisher's exact test. The statistical significance for the increase in proportion of ESBLs over time was calculated using a regression slope test.

## Results

### Comprehensive collection of clinical specimen CREs from microbiology labs over a nearly 5-year period reveals striking diversity and clusters
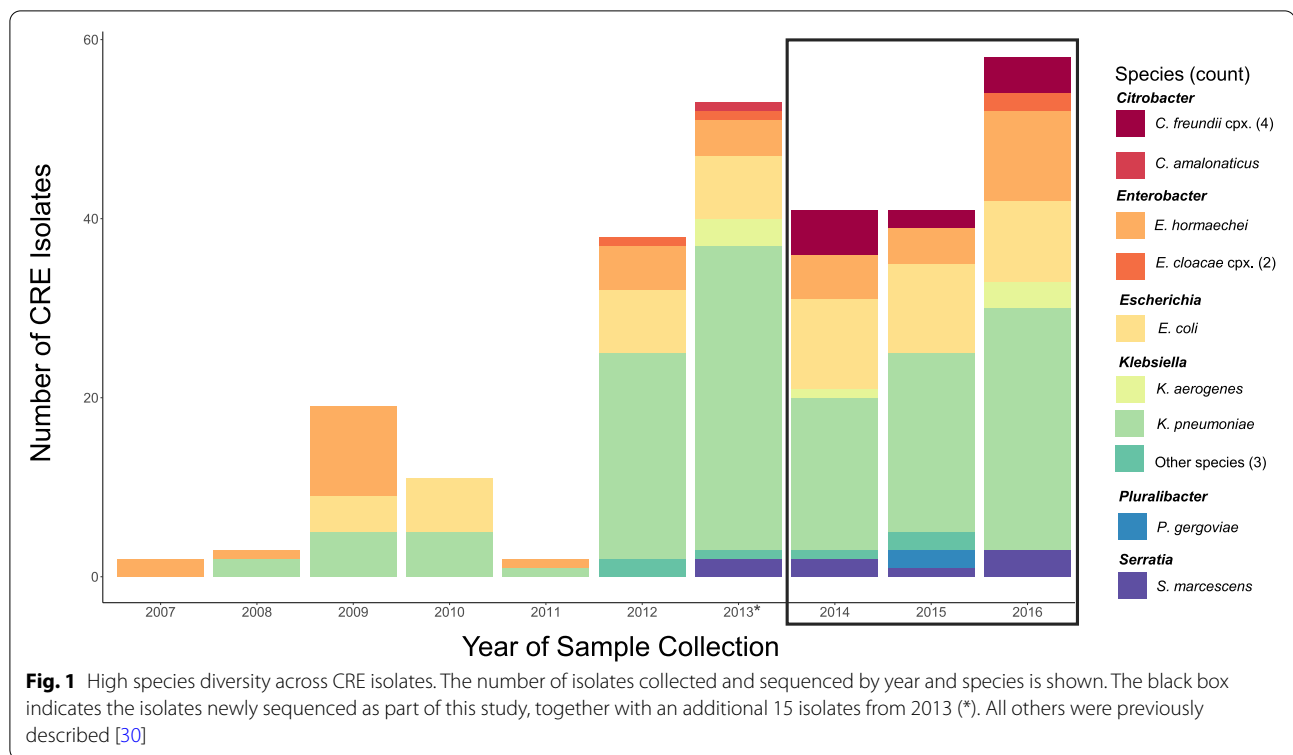
As part of our continued surveillance of CRE at four large tertiary hospitals located in Boston, MA, and Orange, CA, we collected and sequenced the genomes of all carbapenem-resistant *Enterobacterales* (CRE; defined here as meropenem MIC $\geq$ 2μg/mL; Methods) cultured from clinical specimens between December 2013 and December 2016, regardless of species or resistance mechanism. These 147 CRE were added to our earlier published dataset [30] of 74 CRE prospectively collected between August 2012 and November 2013, and 47 historical CRE isolates from the same hospitals, including 12 sets of related same-patient isolates (Table 1; Fig. 1; Additional file 1: Tables S1, S3, S13). For each CRE, we also collected and sequenced at least one species- and time-matched carbapenem-susceptible *Enterobacterales* (CSE; defined here as meropenem MIC < 2μg/mL). This collection strategy gave us access to isolates representing a wide range of carbapenem resistance mechanisms, as well as a snapshot of the sympatric susceptible population at each hospital [30] (Additional file 1: Table S1). Consistent with our previous findings, CRE were most often cultured from urine specimens (40%), followed by respiratory tract (18%), blood (9%), and bile (8%) specimens (Additional file 1: Table S1).

For each isolate, we generated highly contiguous de novo genome assemblies (average scaffold N50 of 3.8 Mb; Additional file 1: Table S2) using a combination of Illumina short paired-end and long-insert mate-pair libraries [111]. Echoing our previous work [30], and consistent with previous studies [29–31, 112], whole-genome average nucleotide identity (ANI) comparisons revealed a vast taxonomic diversity among patient CRE isolates: 16 different species were observed among this collection, though *K. pneumoniae sensu stricto* (52%), *E. coli* (18%) and *Enterobacter hormaechei* (13%; a member of *E. cloacae* complex) were most prevalent (Fig. 1; Additional file 1: Table S1). As expected by the study design, CSE

**Table 1** Number of isolates in the three different collections of our dataset, stratified by resistance status

| | Total CRE | Total CSE | Same Patient isolates | | |
| --- | --- | --- | --- | --- | --- |
| | | | Unique patients | Total same-patient CRE isolates | Total same-patient CSE isolates |
| Boston historical collection (Cerqueira et al., 2017) [30] | 47 | 2 | 7 | 17 | 1 |
| Initial prospective Collection (Cerqueira et al., 2017) [30] | 74 | 138 | 5 | 7 | 3 |
| Newly sequenced prospective collection | 147 | 200 | 0 | 0 | 0 |
| Total | 268 | 340 | 12 | 24 | 4 |

**Fig. 1** High species diversity across CRE isolates. The number of isolates collected and sequenced by year and species is shown. The black box indicates the isolates newly sequenced as part of this study, together with an additional 15 isolates from 2013 (*). All others were previously described [30]

isolates were similarly distributed by taxonomy (Additional file 1: Table S1). To further classify and explore the diversity of isolates, we generated single-copy core phylogenetic trees for each species, computationally defined lineages using phylogenetic distances as previously described [31], and mapped these lineages to existing STs, including those previously defined as high-risk [42] due to their ability to cause severe and/or recurrent drug-resistant infections and rapidly spread (Methods).

This revealed a striking intra-species diversity of organisms within our dataset (Fig. 2; Additional file 1: Table S1; Additional file 2: Fig. S3), though we also observed closely related clusters of isolates. Twenty percent of isolates were separated from another isolate by two or fewer single nucleotide variants (SNVs), based on core genome comparisons, which was the range of SNVs observed for related same-patient isolates (Additional file 1: Table S13). Another 14% were separated from another



**Fig. 2** Resistance mechanisms were phylogenetically dispersed but their carriage varied among closely related isolates. For *E. hormaechei*, *E. coli*, and *K. pneumoniae*, lineages are indicated in the inner ring with alternating shades of gray. Resistance mechanisms for each isolate are shown in the outer ring. The high-risk lineages *K. pneumoniae* ST-258 and *E. coli* ST-131 are marked. *K. pneumoniae* ST-15, which contained an example of likely nosocomial spread of porin-based CRE, is also marked with an asterisk

Salamzade *et al. Genome Medicine* (2022) 14:37

Page 10 of 22

isolate by a maximum of 10 core SNVs, a number previously used to suggest recent common ancestry [113, 114].

## Heterogeneous phylogenetic distribution of resistance determinants, and evidence for nosocomial spread of non-CP-CRE

Our analysis of the mechanisms of carbapenem resistance revealed that the majority of resistance was mediated by carbapenemases (CP-CRE), though extended-spectrum beta-lactamases coinciding with porin disruptions also contributed (Table 2; Fig. 2; Additional file 1: Table S1, S4-S7, S14-S15; Additional file 2: Figs. S3-S5; Additional file 3). Pointing to the known role of HGT in the spread of carbapenem resistance in the *Enterobacterales* [30, 41, 51], species phylogenies revealed both carriage of identical carbapenemase alleles (and their transposon Tn*4401* variants; Additional file 1: Table S14) among distantly

related isolates and heterogeneity in the carriage of resistance genes among closely related isolates (Fig. 2). Importantly, despite capturing a very small fraction of patient CSE from these hospitals, we observed CP-CRE and susceptible isolates separated by as few as 4 core SNVs, connections that would be missed by phenotyping alone (Additional file 1: Table S16), and pointing to transfer of carbapenemase-containing plasmids within locally circulating populations.

While the majority of closely related clusters (< 10 SNPs) comprised CP-CRE, we also found evidence for the nosocomial spread of non-CP-CRE. A cluster of 12 *K. pneumoniae* ST-15 isolates (separated by 0–2 SNVs) that all harbored the ESBL $bla_{CTX-M-15}$ were isolated from 12 unique patients from a single hospital across four years. Eleven of 12 members also carried inactivating mutations in both major porins, OmpK35 and OmpK36 (Additional

**Table 2** Resistome categories for carbapenem-resistant isolates in our prospective collection

| Resistance mechanism | # of CRE isolates | # of CSE isolates | % of total isolates in this category that are CRE | % of CRE isolates with this mechanism |
|---|---|---|---|---|
| **i. Carbapenemases** | 146 | 2 | 98.6% | 66.1% |
| **All$bla_{KPC}$** | 124 | 2 | 98.4% | 56.1% |
| $bla_{KPC-3}$ | 60 | 1 | 98.4% | 27.1% |
| $bla_{KPC-2}$ | 50 | 1 | 98.0% | 22.6% |
| $bla_{KPC-4}$ | 11 | 0 | 100.0% | 5.0% |
| $bla_{KPC}$(read based)[a] | 3 | 0 | 100.0% | 1.4% |
| **All$bla$NDM** | 14 | 0 | 100.0% | 6.3% |
| $bla_{NDM-1}$ | 11 | 0 | 100.0% | 5.0% |
| $bla_{NDM-5}$ | 3 | 0 | 100.0% | 1.4% |
| **All$bla$SME** | 6 | 0 | 100.0% | 2.7% |
| $bla_{SME-1}$ | 3 | 0 | 100.0% | 1.4% |
| $bla_{SME-2}$ | 3 | 0 | 100.0% | 1.4% |
| **Other carbapenemases** | 2 | 0 | 100.0% | 0.9% |
| $bla_{IMP-84}$and$bla_{KPC-4}$ | 1 | 0 | 100.0% | 0.5% |
| $bla_{IMP-4}$ | 1 | 0 | 100.0% | 0.5% |
| **ii. ESBL/AmpC + porin defect(s)[b]** | 61 | 32 | 65.6% | 27.6% |
| **Includes any ESBL** | 46 | 11 | 80.7% | 20.8% |
| **Includes$bla_{CTX-M-15}$** | 34 | 2 | 94.4% | 15.4% |
| **Includes$bla_{SHV-12}$** | 11 | 5 | 68.8% | 5.0% |
| **Includes other ESBLs** | 13 | 5 | 72.2% | 5.9% |
| **Includes any AmpC** | 31 | 23 | 57.4% | 14.0% |
| **Includes$bla_{EC}$** | 19 | 8 | 70.4% | 8.6% |
| **Includes$bla_{CMY-2}$** | 12 | 4 | 75.0% | 5.4% |
| **Includes other AmpCs** | 5 | 12 | 29.4% | 2.3% |
| **iii. ESBL/AmpC without porin defect** | 6 | 168 | 3.4% | 2.7% |
| **iv. Porin defect(s) without ESBL/AmpC** | 8 | 23 | 25.8% | 3.6% |
| **Defect in both major porins** | 5 | 5 | 50.0% | 2.3% |
| **v. No known resistant determinants** | 0 | 113 | 0.0% | 0.0% |

[a] The $bla_{KPC}$ was not present in the whole-genome assembly, but we were able to detect its presence by examining the read data (see Methods)

[b] Strains may have multiple predicted AmpC and/or ESBL genes. See Additional file 1: Tables S6-S7 for specific resistance determinants predicted for each isolate

file 1: Tables S4-S5). Though similar to an earlier report from hospitals in Greece [115], where 19 clonally related isolates carried $bla_{CTX-M-15}$, together with consistent disruptions (*ompK35*) or mutations (*ompK36*) in porins, we observed that, while each isolate carried the same, presumably vertically transmitted, frame-shifted copy of *ompK35*, all but one carried a distinct inactivated form of *ompK36* with non-identical IS element insertions [116–118].

## Geographically widespread, rapidly rearranging plasmid groups prevented accurate tracing of carbapenemase exchange between unrelated organisms

Because our assemblies were generated using data from both short paired-end and long-insert mate-pair Illumina libraries, they were highly contiguous (i.e., >20% of >2 kb plasmid scaffolds were predicted to represent complete, circular plasmids (Additional file 1: Table S17)), which allowed us to examine the distribution of plasmids across isolates, genera, and geographic locations. The vast majority of plasmids (83%) were assigned to one of 215 plasmid groups by MOB-suite (Additional file 1: Table S18; Additional file 2: Fig. S6), of which nearly 20% contained instances of plasmids encoding carbapenemases. However, most (80%) of these carbapenemase-containing plasmid groups also contained plasmids lacking a carbapenemase (Additional file 2: Fig. S7), showcasing the known genetic flexibility of plasmids [31, 112, 119]. Furthermore, plasmid groups were also remarkably geographically widespread [120], with nearly all (92%) of the most prevalent plasmid groups found in isolates from multiple hospitals, including all of those containing carbapenemases (Additional file 2: Fig. S7), with a majority (70%) found in isolates from both cities.

The widespread geographic distribution of plasmids, and the ample opportunity they have to interact with and rearrange genetic content with other plasmids (Additional file 3), complicates the tracing of clinically important resistance genes contained within plasmids. Specific genetic markers have previously been used to trace the local spread of $bla_{KPC}$ [45]. However, when applied to our dataset, these markers, including specific plasmid groups, Tn*4401* isoforms and their 5 base pair (bp) or longer flanks, were also mostly geographically widespread (Additional file 2: Fig. S8).
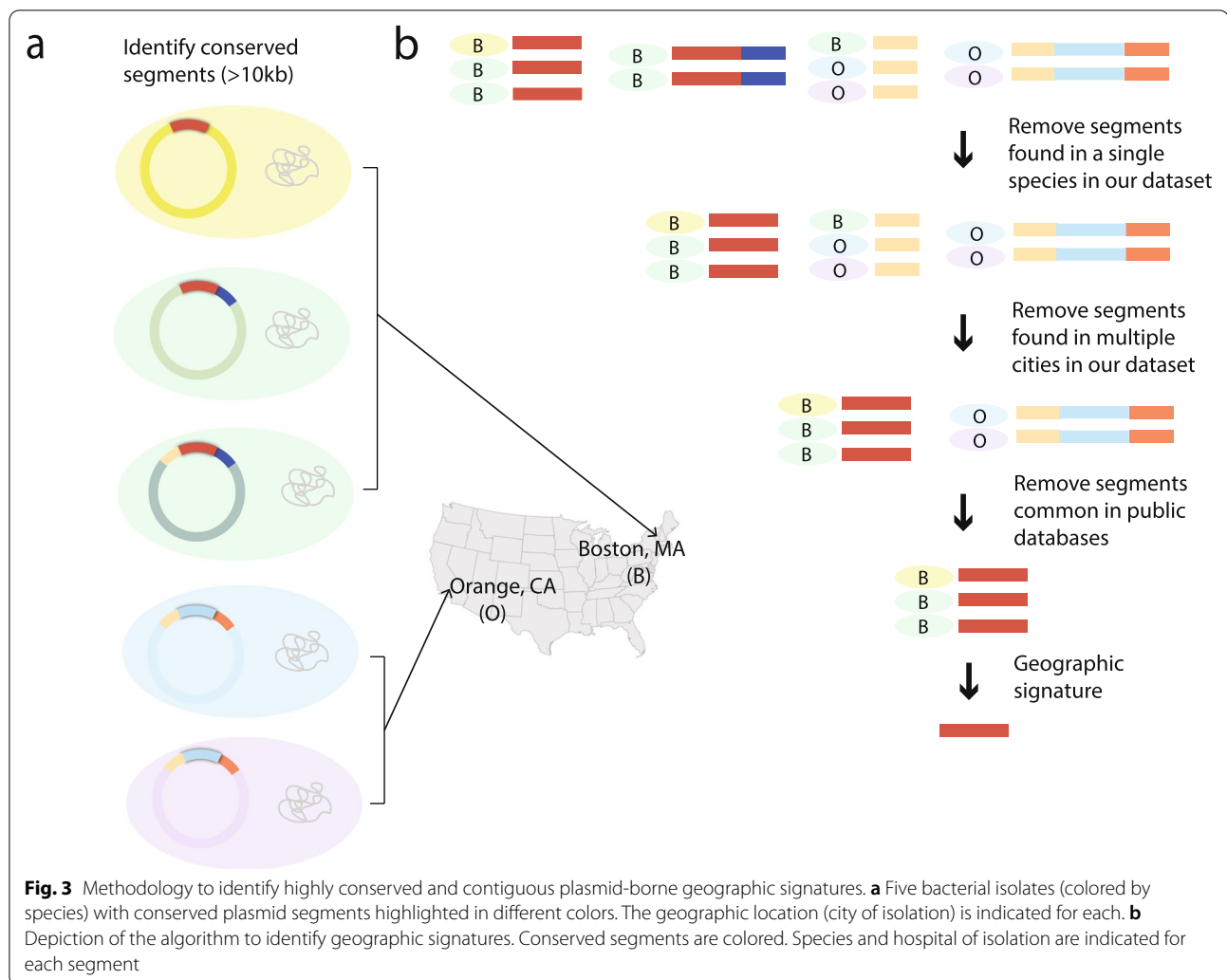
## Identification of geographic signatures allows for local tracing of HGT and inter-molecular movement of genes

Inspired by analyses investigating larger flanking regions surrounding resistance determinants [121], we developed *ConSequences*, a broader, gene-agnostic approach to identify highly conserved and contiguous segments on plasmids that may serve as markers of local HGT of clinically important genes, such as those encoding carbapenemases and other hospital-adaptive traits [87]. We first searched predicted plasmids from across our entire dataset of >600 isolates for 10 kb or larger segments that were conserved in gene order and nucleotide identity ($\geq$ 99% per 10kb block) across two or more plasmid scaffolds (Fig. 3; Additional file 2: Figs. S1-S2). Of the 4605 unique segments meeting these criteria, 95% exhibited some amount of overlap or nesting with other segments, and 58% were identified in multiple plasmid groups, highlighting the frequent recombination between plasmids and complex nesting among mobile genetic elements in *Enterobacterales* [31, 122]. The size of conserved segments ranged from 10 to 310 kb, with longer segments typically observed in fewer isolates.

We focused our analysis on segments that were likely horizontally transferred and present in more than one species, as inter-species transfers would be least suspected as nosocomially linked. To filter out segments likely to have been repeatedly imported into the hospital from elsewhere, we removed those present in isolates from both states represented in our study, as well as those appearing in publicly available genomes represented in the NCBI's Nucleotide Collection database. We elected to group the Boston-based hospitals together to account for known fluidity between some hospital staff and patient populations who could interact with and share organisms from a common reservoir. Although our assemblies were highly contiguous, we took further steps to ensure that these segments represented only high-confidence regions of our assemblies and were not missed by improper assembly or scaffold breaks (Methods). Finally, we stringently screened these for uniqueness by searching against the ENA/SRA database using BIGSI [94].

This analytical framework revealed 44 *geographic signatures*, which we define as plasmid segments found in two or more species and associated exclusively with either Boston, MA, or Orange, CA (Additional file 1: Tables S10, S19; Additional file 2: Fig. S9). These signatures represented only 2% of all segments found in multiple species, and more than half (52%) were specific to a single hospital. We found that signature prevalence was higher among CRE (23%) than CSE (6%) in our sample set, likely owing to our sampling strategy that included all CRE from hospital microbiology labs, but only a small fraction of CSE. As observed for the full set of conserved segments, many signatures exhibited substantial overlap with one another. Ten were fully nested within larger signatures, representing different conservation profiles, with each shorter signature found in more isolates than the larger signature. Signatures had a mean length of 31 kb (range 10 to 311 kb) and were observed, on average, 5 times (range 2 to 19), across 2 plasmid groups (range

**Fig. 3** Methodology to identify highly conserved and contiguous plasmid-borne geographic signatures. **a** Five bacterial isolates (colored by species) with conserved plasmid segments highlighted in different colors. The geographic location (city of isolation) is indicated for each. **b** Depiction of the algorithm to identify geographic signatures. Conserved segments are colored. Species and hospital of isolation are indicated for each segment

1 to 4; 28 total), 2 species (range 2 to 4; 15 total), and 3 sequence types (range 2 to 6; 39 total).

### Geographic signatures carry important cargo for hospital survival and dissemination

We hypothesized that the 44 geographic signatures would encode functions that enable their movement and persistence within patients (i.e., colonization) and the built environment, similar to other widely conserved plasmid sequences from nosocomial bacteria. Of the 1494 individual genes predicted within signatures, we could assign some putative function to nearly two-thirds (Additional file 1: Tables S11-S12; Additional file 2: Figs. S10-S11). All but one of the signatures were predicted to encode functions for maintenance or dissemination of DNA into new genomic contexts or hosts, including IS elements [123], integrases or other recombinases [124], conjugation machinery [104], and plasmid uptake and maintenance apparatuses [105, 109]. The prevalence of conjugation

and plasmid uptake genes expectedly pointed to the carriage of signatures on conjugative plasmids; however, one signature (Sig20) also overlapped with a predicted prophage that was situated on a circularized scaffold containing a plasmid replicon but no conjugative relaxase. Also, given their demonstrated ability to cross species boundaries, it was unsurprising that half of the signatures featured genes which were also found at high nucleotide identity ($\geq$99%) in bacteria outside the order of *Enterobacterales* (Additional file 1: Table S11).

As we also hypothesized, the majority (66%) of signatures featured genes predicted to encode for survival strategies against antimicrobials, including quaternary ammonium compounds used in standard disinfectants in healthcare settings [125] (Additional file 2: Figs. S10-S11). Eight signatures encoded enzymatic antibiotic resistance, including examples with $bla_{KPC}$, likely reflecting both our sampling strategy focused on CRE and that antibiotic resistance genes are often co-located on plasmids

Salamzade *et al. Genome Medicine*    (2022) 14:37

Page 13 of 22

[126–128]. Genes coding for metal resistance were also prevalent, occurring in more than a third of signatures, and often co-occurring with genes for antibiotic resistance, including in half of signatures containing $bla_{KPC}$, highlighting recent findings that metal resistance and antibiotic resistance are frequently co-selected [129, 130].

### Additional instances of localized carbapenemase spread were identified through tracing of geographic signatures

Six of the 44 geographic signatures encoded $bla_{KPC}$. These $bla_{KPC}$ signatures ranged in size from 18 to 147kb (the smallest of these signatures, Sig 5.1-CP, was nested inside the largest signature, Sig 5.6-CP) and were found in 21% of all $bla_{KPC}$-carrying CRE in our collection (Fig. 4; Additional file 2: Fig. S9). As expected from the filters we applied to identify geographic signatures, $bla_{KPC}$ signatures were distributed across multiple species (2 to 4) and sequence types (2 to 7), associated with multiple plasmid groups (1 to 3) and sometimes the chromosome, and observed across variable time spans of at least 4 years and up to 10 years (Fig. 4; Additional file 1: Tables S10, S19). Furthermore, some isolates carried more than one signature (e.g., Sig1-CP and Sig4-CP), sometimes on the same plasmid.

Though our inclusion criteria would have allowed for as many as 100 SNVs in a 10-kb signature, we observed much greater identity, with most $bla_{KPC}$ signatures varying by only two or fewer SNV differences between instances (Additional file 2: Fig. S12). This striking degree of similarity, extending well beyond the edges of the Tn*4401* elements, likely indicates a recent common ancestral source for these signatures, as well as their persistence and movement across species boundaries within a local environment. These instances of suspected local spread of carbapenemases between distantly related isolates would not have been picked up by traditional epidemiology or even standard whole-genome sequence analysis in most cases. Furthermore, many $bla_{KPC}$ signature instances were harbored by isolates that were very closely related (0–10 core SNVs), indicating the ability of these signatures to spread within the hospital along with their bacterial hosts (Fig. 4), in addition to their ability to move between bacterial sequence types and species.

### A multi-genus $bla_{KPC}$-containing geographic signature is highly conserved despite rapid rearrangements of its plasmid backgrounds

Although our analysis was based on highly contiguous assemblies that combined data from short- and long-insert Illumina libraries, we sought to improve the assemblies further in order to follow the details of signature evolution and spread across plasmids and taxonomic boundaries. To do this, we re-sequenced all isolates

carrying Sig5.1-CP using Oxford Nanopore Technology to generate long-read sequences for hybrid assembly, followed by manual curation. Sig5.1-CP, a $bla_{KPC-3}$-carrying signature, was present in the largest number of different genera and species in our prospective collection and was also found in historical isolates dating back to 2008 (Fig. 5; Additional file 1: Table S20). In addition, Sig5.1-CP was almost exclusively found in a single Boston hospital, including within four isolates collected over a 5-week period in 2017. These four isolates were initially sequenced because they were suspected to be part of a short-lived *C. freundii* complex (later revealed to be *Citrobacter portucalensis*) clonal case cluster that differed from one another by less than four SNVs. We found that this 2017 cluster of isolates all differed by the same 12 SNVs from a 2014 isolate carrying Sig5.1-CP from the same hospital that was already part of our collection.

The complete assembly of each individual replicon unambiguously revealed the relocation of Sig5.1-CP into the chromosome and its association with three different plasmid groups (Fig. 5). One of these plasmids encoded both IncP and IncH replicons, which are known for their broad host range [131, 132], likely contributing to this signature's ability to transfer to diverse hosts and persist. Although our draft assemblies were adequate to identify the signature using our methodology, manual inspection of Sig5.1-CP boundaries in the completed assemblies revealed that this signature could be expanded to include approximately an additional 1 kb of sequence encoding a $bla_{TEM-1A}$ beta-lactamase. Manual inspection also revealed an even longer, 26-kb conserved region in the 9 isolates occurring since 2012 (Additional file 2: Fig. S13). This extended signature included additional cargo that could be involved in adaptation to the healthcare environment, including genes predicted to encode mercury resistance, a Na+/H+ antiporter (possibly influencing cell viability at high pH [133]), and a multidrug efflux pump. Despite Sig5.1-CP's distinct genomic contexts, it was highly conserved, maintaining perfect base-level identity, with the exception of one isolate (*K. pneumoniae* MGH39 from 2012) that harbored a variant of the signature with two non-synonymous SNVs, both within the same predicted transposon.

Though possessing nearly identical chromosomes ($\leq 3$ SNVs), isolates from the *C. portucalensis* clonal case cluster had a striking variety of plasmid and chromosomal arrangements (Additional file 2: Fig. S14). While two chromosomally identical 2017 isolates also had identical plasmid profiles and Sig5.1-CP location, one of the other 2017 isolates carried the signature on a different plasmid, while the remaining 2017 isolate carried it on a cointegrate of both plasmids. Many of the rearrangements were likely mediated by IS*26* sequences (Additional file 2:
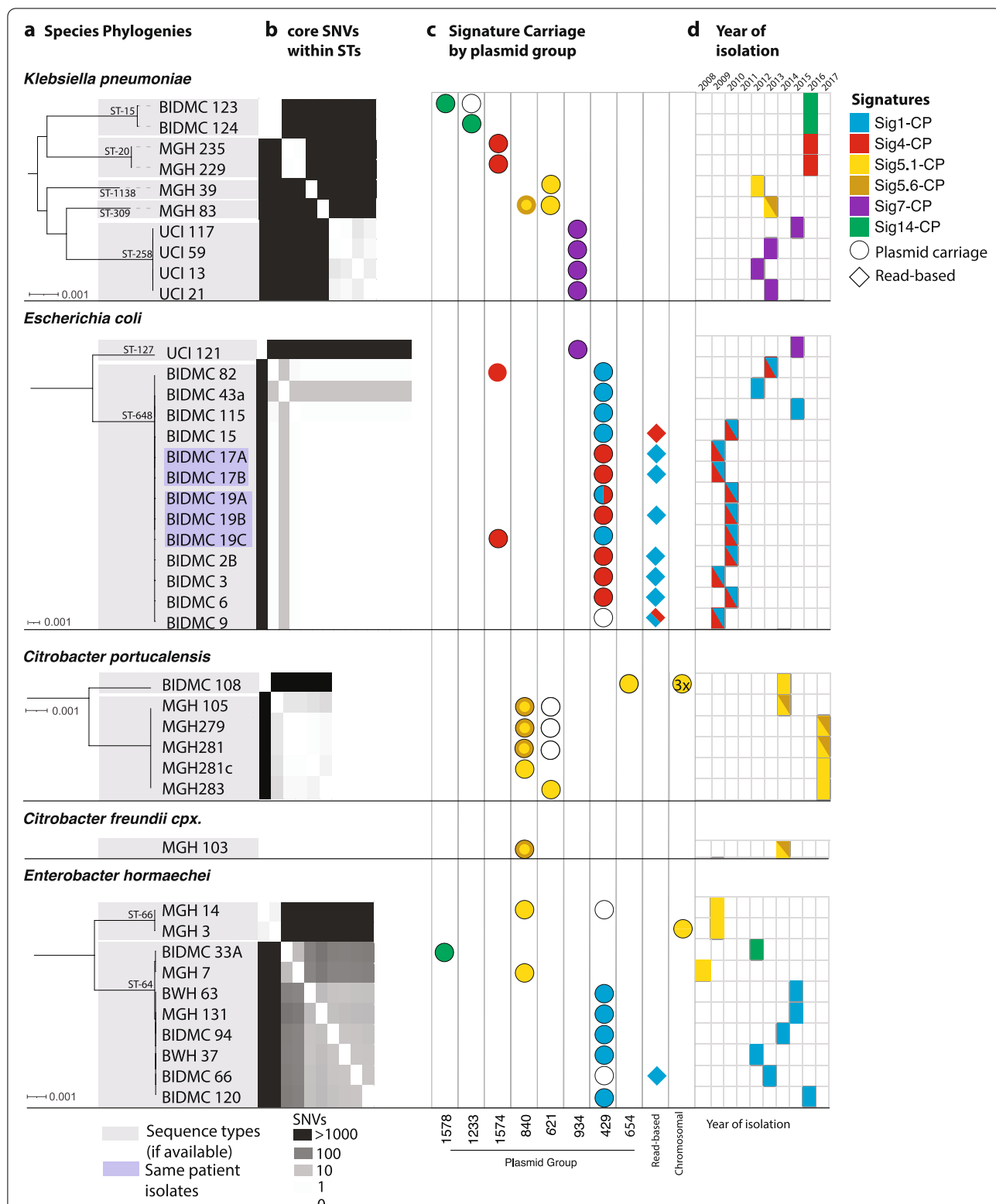
**Fig. 4** Carbapenemase-carrying signatures are found in diverse species, lineages, and plasmid backgrounds. **a** Species phylogenies for all geographic signature-containing isolates from each species, showing sequence type (STs). **b** Within each ST, core SNV distances are shown (heatmap). **c** For each isolate, columns indicate plasmid content, and colored icons indicate signatures present on these plasmids. For the nested signatures 5.1-CP and 5.6-CP, a solid yellow circle indicates the presence of 5.1-CP only, whereas the yellow circle with a darker ring indicates the presence of both 5.1-CP and 5.6-CP. **d** Year of isolation is marked for each, colored by signature content as in **c**
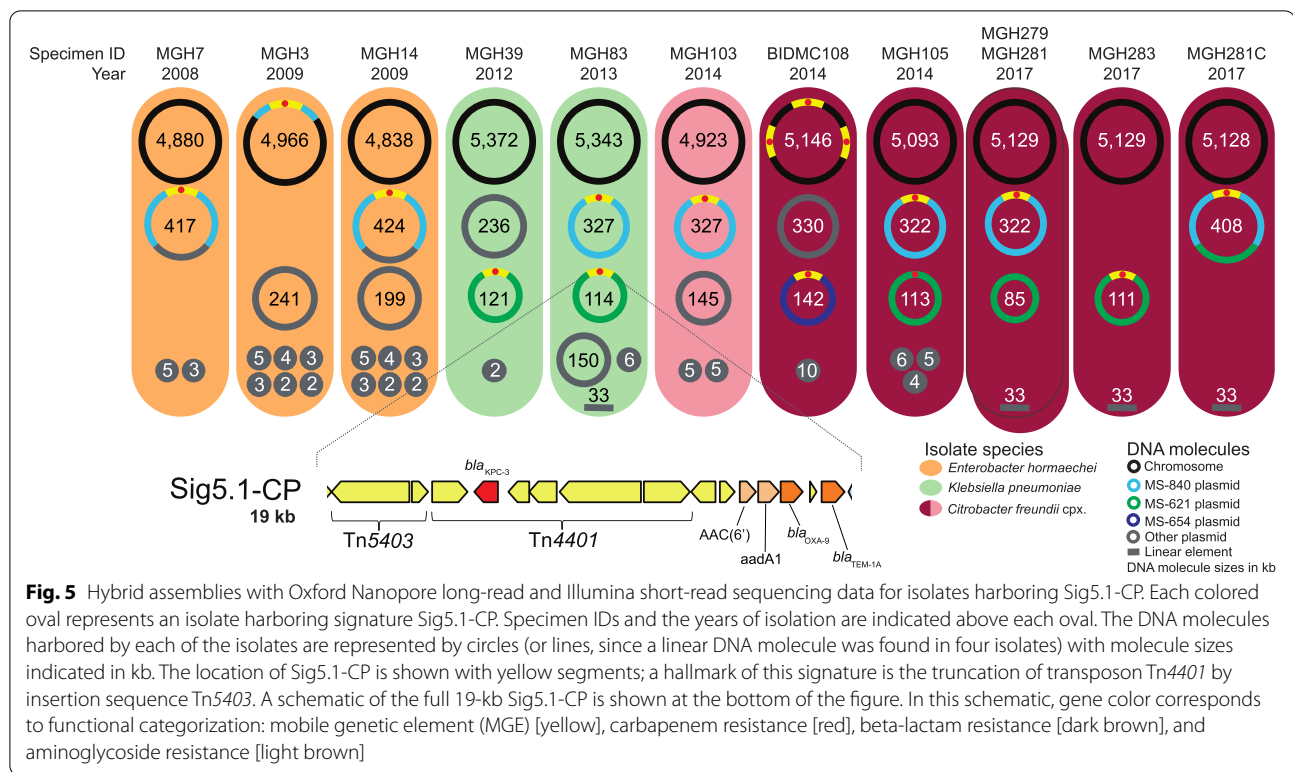
**Fig. 5** Hybrid assemblies with Oxford Nanopore long-read and Illumina short-read sequencing data for isolates harboring Sig5.1-CP. Each colored oval represents an isolate harboring signature Sig5.1-CP. Specimen IDs and the years of isolation are indicated above each oval. The DNA molecules harbored by each of the isolates are represented by circles (or lines, since a linear DNA molecule was found in four isolates) with molecule sizes indicated in kb. The location of Sig5.1-CP is shown with yellow segments; a hallmark of this signature is the truncation of transposon Tn*4401* by insertion sequence Tn*5403*. A schematic of the full 19-kb Sig5.1-CP is shown at the bottom of the figure. In this schematic, gene color corresponds to functional categorization: mobile genetic element (MGE) [yellow], carbapenem resistance [red], beta-lactam resistance [dark brown], and aminoglycoside resistance [light brown]

Fig. S14), previously shown to drive the reorganization of plasmids through replicative transposition [134]. Although we could not precisely trace the divergence of the cluster of 2017 isolates from the 2014 isolate, multiple plasmid gains, losses, and rearrangements in the 2017 isolates appear to have occurred on the same timescale as the accumulation of SNVs.

## Discussion

Using a systematic, prospective isolate collection strategy, along with a sequencing approach that enabled highly contiguous assemblies, we found diverse mechanisms for the spread of carbapenem resistance and examples of the clonal spread of both carbapenemase-carrying and porin-based CRE. We also found geographic signatures that allowed us to trace the local spread of carbapenemases across genetic backgrounds, even in cases where standard genomic epidemiology might not identify a link.

Our collection was agnostic to species and carbapenem-resistance mechanism, giving us a more complete view of patient CRE organismal and resistance mechanism diversity. As we and others observed previously [29, 30], the majority of carbapenem resistance was mediated by the presence of a carbapenemase, with the most prevalent being $bla_{KPC}$, consistent with their known prevalence in the USA [22]. We saw that another approximately quarter of resistance could

be associated with ESBL or AmpC beta-lactamases, together with at least one disrupted major porin — a resistance genotype that can be difficult to detect using targeted molecular or even whole-genome sequencing analytical approaches.

Though there have been many examples of the clonal spread of CP-CRE, few studies have pointed out the contribution of porin-based CRE in nosocomial CRE spread [115, 135]. Although strains with porin mutations are thought to be less successful in spreading to other patients, we observed a large clonal cluster of *K. pneumoniae* isolates with inactivating mutations in both major porins appearing in patients across a 4-year period. Interestingly, the mutation in *ompK35* was the same across the isolates, whereas the mutations in *ompK36* varied. This suggested that (i) the transmitted form had a single porin inactivation and, thus, may have been less impaired for long-term persistence and spread; (ii) selective pressure for inactivation of both porins was able to override any fitness defect; and/or (iii) the potential for yet to be identified compensatory mutations which could attenuate the fitness defect associated with mutations of these porins in this strain [136, 137]. The presence of porin-based CRE clonal clusters indicates that fitness phenotypes in such strains should be examined more closely and that porin-based mutants should not be overlooked as contributors to clonal case clusters.

Salamzade *et al. Genome Medicine*      (2022) 14:37

Page 16 of 22

Plasmids are key to understanding the ongoing carbapenem resistance epidemic but pose challenges for tracing resistance evolution and local spread because of their diversity, widespread nature, and tendency to rearrange [31, 51, 112, 119, 138, 139]. These processes are accelerated by the frequent co-occurrence of carbapenemase plasmids with other diverse plasmids within a single isolate [41]; in our dataset, three-quarters of all plasmid groups were observed to co-occur with a carbapenemase plasmid group. Furthermore, as observed here and by others [39, 41], carbapenemases are often carried by plasmids that can conjugate across species and genera [39, 41]. Our results showcase an extensive network of possible exchange points that carbapenemases (and the signatures that contain them) can use to transfer to new plasmid backgrounds and onward to new hosts. We also observed striking plasticity of plasmids. For instance, plasmid groups were rarely composed purely of carbapenemase-containing plasmids and were also widespread and not strongly associated with specific geographic locales. This flexibility is exemplified in the *C. portucalensis* clonal case cluster, in which plasmid rearrangements and carbapenemase transfer between different plasmid backgrounds were repeatedly documented for isolates that were likely to have diverged very recently and isolated only a few weeks apart.

While approaches for systematically tracing carbapenemases across species include analyzing transposons and their immediate flanking sequences [45], these regions in our collection were common and widespread, preventing us from confidently tracing carbapenemase movement. Additionally, the recently published alignment-based, pairwise screen of Evans et al. [122] shows promise for tracing carbapenemases; however, their clustering approach applied in this case would likely group together different Tn*4401* isoforms and does not assess whether segments are geographically associated.

In order to overcome the limitations of these approaches and to achieve a higher level of resolution for tracing the localized spread of carbapenemases as well as other hospital-adaptive traits, we developed a novel, broadly applicable, and gene-agnostic framework to identify highly conserved plasmid segments found in multiple species or lineages with strong geographic associations. Though strict filters for identifying signatures were used in the work presented here, our approach can be tuned with regard to these filters, including the ability to incorporate different levels of geographic specificity (i.e., hospital, city, etc.), phylogenetic specificity (i.e., lineage, species, etc.), and different signature lengths, up to the size of an entire plasmid. Loosening these requirements — along with the use of even more contiguous assemblies — could yield a more comprehensive profile of HGT

signals, including an understanding of the possible differences in HGT occurring within versus between different species. Although we only searched for inter-species signatures in this study, we also identified intra-species instances in 8 of the 44 signatures.

While long-read technologies are typically needed to achieve the high level of contiguity necessary to examine plasmids in detail, our strategy, involving both long and short insert Illumina libraries, together with our novel approach for identifying signatures, was able to successfully identify many highly conserved signatures involved in the predicted local movement of carbapenemases across species boundaries. Although having high-quality assemblies was key to identifying signatures, our method also includes searches against raw sequence reads to identify additional instances of motifs, which would make it possible to include read sets from lower quality assemblies in analysis. Improved genomic assemblies, including more long-read sequencing, would likely assist in identifying and tracing signatures more completely. Furthermore, given our method's dependence upon public databases, growth in the number of deposited sequences will provide additional resolution to discern true geographic signatures from segments that are more broadly geographically represented.

The links between unrelated isolates uncovered by geographic signatures add to the growing recognition that there are likely to be reservoirs of *Enterobacterales* within healthcare networks that are involved in reshuffling plasmids and their signatures across organisms, and aiding in their long-term local persistence. Remarkably, over 20% of CP-CRE isolates carried at least one $bla_{KPC}$ geographic signature, suggesting that a substantial fraction of patient CRE may have originated from HGT within the hospital (or hospital-proximal environments, including ambulatory and nursing facilities). Although we expected that selective pressure for maintaining carbapenem resistance would be mainly present within patients treated with carbapenems, our $bla_{KPC}$-containing signatures were long-lived, each being observed in patient isolates spanning periods of 4 to 10 years. This is likely because these signatures also carried other genes key to adaptation in hospital reservoirs. Genes co-localized with $bla_{KPC}$ within our signatures that could provide functions related to persistence in the hospital environment included (i) additional antibiotic resistance genes, which could lead to their joint long-term conservation through co-selection; (ii) genes conferring resistance to hospital disinfectants and metals, increasingly used in hospital touch surfaces [129, 130]; and (iii) conjugation and plasmid uptake machinery, both of which can amplify $bla_{KPC}$ and assist its persistence and spread across isolates and species [140, 141]. Genes within signatures also included several encoding

for proteins of unknown function; future work, outside the scope of this study, to characterize the functions of these genes will be key to understanding the persistence of these signatures in the hospital environment. In addition, the association of these signatures with plasmids from different incompatibility groups, particularly those adapted to different conditions such as temperature [142–144], appear adaptive for reservoir switching, e.g., from the human body to the hospital environment and back. Our views of the taxonomic distributions of signature genes suggest that some of these hospital-adaptive traits may have been recently acquired from distantly related organisms in shared reservoirs.

Our study had several limitations. Sampling of suspected hospital reservoirs, including environmental samples and those from asymptomatic carriers, would likely allow us to create a more complete view of signatures and their exchange across organisms and demonstrate the role of these reservoirs in seeding infections. Other studies, including our previous work [30], point to high levels of asymptomatic colonization, including a study that suggests that clinical testing detects only one out of nine carriers [145]. We also lacked epidemiological data about patients, which further limited our ability to quantify the extent of clonal spread. Despite the lack of epidemiological data and non-patient samples, our analysis suggested that there is persistence of resistance genes within hospital networks for years.

One additional limitation of our approach is our inability to distinguish convergent evolution or re-introductions from the community from the local spread. However, due to the high degree of similarity over long stretches (10–310kb), convergent evolution for all of these signatures is unlikely. In the case of two signatures (Sig1-CP and Sig4-CP), instances that were otherwise nearly identical differed in which $bla_{KPC}$ allele or Tn*4401* isoform they carried. Possible explanations other than de novo formation of a similar signature include mutation within a signature, or recombination taking place within the signature. The propensity for some signatures, like Sig5.1-CP, to persist for so long within a single hospital, yet to not be found more generally, including among patients treated at other nearby hospitals, strongly hints at a local reservoir, rather than re-introduction.

## Conclusions

Using a long-term systematic collection of isolates — including both CRE and CSE — together with high-quality sequencing and a novel analysis methodology, we achieved high-resolution views of mechanisms accounting for carbapenem resistance and a greater understanding of their spread across four US hospitals.

In addition to examples of the clonal spread of CP-CRE and porin-based CRE, we observed the sharing of plasmid segments containing hospital-adaptive traits — including carbapenemases — circulating among local diverse bacterial populations over long timeframes. Within these segments and the plasmids harboring them, we observed inter-molecular rearrangements over short timeframes, underscoring the complexity entailed in tracing the movement of plasmids and their component parts. However, our long-term surveillance strategy, high-quality assemblies, and novel methodology for identifying geographic signatures revealed previously unsuspected links between CP-CRE (as well as other organisms across these hospitals) that helped to clarify the epidemiology of antibiotic resistance spread and persistence in these healthcare networks.

## Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s13073-022-01040-y.

---

**Additional file 1: Table S1.** Overview of isolates in our dataset. **Table S2.** Genome assembly statistics. **Table S3.** Accessions for Illumina-ONT hybrid assemblies. **Table S4.** Disruptions in OmpF/OmpK35 porin gene. **Table S5.** Disruptions in OmpC/OmpK36 porin gene. **Table S6.** Genomic location of resistance genes. **Table S7.** List of genes identified as ESBL or AmpC. **Table S8.** Read-based identification of carbapenemases in resistant isolates for which no resistance mechanism was found in the assemblies. **Table S9.** Size of species-specific core-genome alignments. **Table S10.** 44 multi-species geographic signatures found in a single city. **Table S11.** Functional annotation of genes present in 44 signatures. **Table S12.** Prevalence of functional categories across the 44 signatures. **Table S13.** Differences in the core genome between same-species isolates from the same patient. **Table S14.** Carbapenemase carrying isolates. **Table S15.** Cluster of related *Klebsiella pneumoniae* isolates with double-porin mutations. **Table S16.** Pairs of closely-related CSE and CRE isolates, where the CRE carried a carbapenemase not found in the CSE. **Table S17.** Identification of circular plasmids (>2 kb). **Table S18.** MOB-suite plasmid group predictions. **Table S19.** Isolates which carry a $bla_{KPC}$-containing signature. **Table S20.** Genomic background of instances of Sig5.1-CP and Sig5.6-CP.

**Additional file 2: Fig. S1.** Method for the delineation of segments shared between plasmids. **Fig. S2.** Workflow to identify geographic signatures. **Fig. S3.** Phylogenetic tree for each of the 15 species in our collection with at least five representatives. **Fig. S4.** Resistance mechanisms were diverse, with many shared across species. **Fig. S5.** Carbapenemase-carrying isolates tended to have a higher minimum inhibitory concentration than those with other resistance mechanisms. **Fig. S6.** High level of diversity and phylogenetic range among predicted plasmids. **Fig. S7.** Plasmids of diverse groups carried carbapenemases and were found in different species, and hospitals. **Fig. S8.** Limited tracing of carbapenemase localized spread using Tn*4401* isoforms and their immediate flanking sequences. **Fig. S9.** Signatures were present across diverse sequence types. **Fig. S10.** Signatures carried genes important for hospital adaptation and signature mobility. **Fig. S11.** Details of signature content. **Fig. S12.** Signatures were highly conserved and likely derived from a common ancestral sequence. **Fig. S13.** Geographic signatures with $bla_{KPC}$ can occur in multiple configurations across several species and plasmid groups. **Fig. S14.** Geographic signatures with $bla_{KPC}$ occurred in multiple configurations across several species and plasmid groups.

**Additional file 3.** Supplementary Results.

Salamzade *et al. Genome Medicine*        (2022) 14:37

Page 18 of 22

## Authors' contributions
Study design: ALM, ABO, DCH, SSH, JEK, RPB, and AME. Sample and data acquisition: SBC, WH, MLD, CEB, GMG, DK, EMP, MJF, LAC, and AME. Assays performed: TB-K, PM, LLH, JQ, and JIW. Data analysis: RS, ALM, BJW, CJW, and TPS. Consultation and supervision of analyses: ALM, TB-K, CJW, PM, SKY, SK, DCH, ESS, CAC, SSH, JEK, VMP, RPB, and AME. Prepared the original draft: RS, ALM, BJW, CJW, and AME. All authors read and approved the final manuscript.

## Availability of data and materials
The dataset analyzed during the current study is available in the National Center for Biotechnology Information (NCBI) repository under BioProject PRJNA271899 (https://www.ncbi.nlm.nih.gov/bioproject/PRJNA271899) [54]. Computer code used for the analysis of our data can be downloaded from https://github.com/broadinstitute/ConSequences [87], including a test dataset. A yaml file is provided for installation of the software and its dependencies through creation of a Conda virtual environment. ConSequences is written in Python3 and made available under the open-source license BSD3.

## Declarations

## Ethics approval and consent to participate
Institutional review board (IRB) approval was granted by the Massachusetts Institute of Technology Committee on the Use of Humans as Experimental Subjects. Samples were collected under study approvals of the IRB committees of the participating institutions: Mass General Brigham (covering both MGH and BWH); Beth Israel Deaconess Medical Center, Boston; and University of California, Irvine (IRB approval numbers 2012P001062, 2012P000205, and 2012-8957, respectively). Waivers of informed consent were granted for each institution due to the fact that bacterial isolates were collected for the study after all clinical work was completed, including after some patients were discharged, which made obtaining informed consent logistically infeasible. This study conforms to the principles of the Belmont Report and follows the ethical principles relevant to this work as outlined in the Declaration of Helsinki.

## Consent for publication
Not applicable.

## Competing interests
B.J.W. is an employee of Applied Invention (Cambridge, MA). The remaining authors declare that they have no competing interests.

## Author details
[1]Infectious Disease and Microbiome Program, Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA. [2]Present Address: Microbiology Doctoral Training Program, University of Wisconsin-Madison, Madison, WI 53706, USA. [3]Applied Invention, Cambridge, MA 02139, USA. [4]Department of Pathology, Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, MA 02215, USA. [5]Department of Ophthalmology, Department of Microbiology, Harvard Medical School and Massachusetts Eye and Ear Infirmary, 240 Charles St., Boston, MA 02114, USA. [6]Division of Infectious Disease, Brigham and Women's Hospital, Harvard Medical School, Boston, MA 02115, USA. [7]Department of Population Medicine, Harvard Medical School and Harvard Pilgrim Healthcare Institute, Boston, MA 02215, USA. [8]Department of Pathology and Laboratory Medicine, University of California Irvine School of Medicine, Orange, CA 92868, USA. [9]Division of Infectious Diseases, University of California Irvine School of Medicine, Irvine, CA 92617, USA. [10]Massachusetts General Hospital, Boston, MA 02114, USA.

## References
1. Trecarichi EM, Tumbarello M. Therapeutic options for carbapenem-resistant Enterobacteriaceae infections. Virulence. 2017;8:470–84. https://doi.org/10.1080/21505594.2017.1292196.
2. Morrill HJ, Pogue JM, Kaye KS, LaPlante KL. Treatment options for carbapenem-resistant Enterobacteriaceae infections. Open Forum Infect Dis. 2015;2. https://doi.org/10.1093/ofid/ofv050.
3. Sheu C-C, Chang Y-T, Lin S-Y, Chen Y-H, Hsueh P-R. Infections caused by carbapenem-resistant: an update on therapeutic options. Front Microbiol. 2019;10:80. https://doi.org/10.3389/fmicb.2019.00080.
4. Brennan-Krohn T, Pironti A, Kirby JE. Synergistic activity of colistin-containing combinations against colistin-resistant Enterobacteriaceae. Antimicrob Agents Chemother. 2018;62. https://doi.org/10.1128/AAC.00873-18.
5. Doi Y. Treatment options for carbapenem-resistant Gram-negative bacterial infections. Clin Infect Dis. 2019;69:S565–75. https://doi.org/10.1093/cid/ciz830.
6. Martin A, Fahrbach K, Zhao Q, Lodise T. Association between carbapenem resistance and mortality among adult, hospitalized patients with serious infections due to Enterobacteriaceae: results of a systematic literature review and meta-analysis. Open Forum Infect Dis. 2018;5. https://doi.org/10.1093/ofid/ofy150.
7. Chitnis AS, Caruthers PS, Rao AK, Lamb J, Lurvey R, De Rochars VB, et al. Outbreak of carbapenem-resistant Enterobacteriaceae at a long-term acute care hospital: sustained reductions in transmission through active surveillance and targeted interventions. Infect Control Hosp Epidemiol. 2012;33:984–92.
8. Borer A, Saidel-Odes L, Riesenberg K, Eskira S, Peled N, Nativ R, et al. Attributable mortality rate for carbapenem-resistant *Klebsiella pneumoniae* bacteremia. Infect Control Hosp Epidemiol. 2009;30:972–6. https://doi.org/10.1086/605922.
9. World Health Organization. Antibacterial agents in clinical development: an analysis of the antibacterial clinical development pipeline. 2019. https://apps.who.int/iris/bitstream/handle/10665/330420/9789240000193-eng.pdf.
10. World Health Organization. Antibacterial agents in preclinical development: an open access database. 2019. https://apps.who.int/iris/bitstream/handle/10665/330290/WHO-EMP-IAU-2019.12-eng.pdf.
11. Guidelines for the prevention and control of carbapenem-resistant Enterobacteriaceae, *Acinetobacter baumannii* and *Pseudomonas aeruginosa* in Health Care Facilities. WHO Guidelines Approved by the Guidelines Review Committee; 2017. https://apps.who.int/iris/handle/10665/259462. Accessed 8 June 2020.
12. Centers for Disease Control and Prevention (CDC). Facility guidance for control of carbapenem-resistant Enterobacteriaceae (CRE)—November 2015 update CRE toolkit. Obtenido de: https://www.Cdc.Gov/hai/pdfs/cre/cre-Guidance-508.Pdf. Acceso Día 2016;30.
13. Snitkin ES, Zelazny AM, Thomas PJ, Stock F, NISC Comparative Sequencing Program Group, Henderson DK, et al. Tracking a hospital outbreak of carbapenem-resistant *Klebsiella pneumoniae* with whole-genome sequencing. Sci Transl Med. 2012;4:148ra116. https://doi.org/10.1126/scitranslmed.3004129.
14. Roberts LW, Harris PNA, Forde BM, Ben Zakour NL, Catchpoole E, Stanton-Cook M, et al. Integrating multiple genomic technologies to investigate an outbreak of carbapenemase-producing *Enterobacter hormaechei*. Nat Commun. 2020;11:466. https://doi.org/10.1038/s41467-019-14139-5.
15. Marsh JW, Mustapha MM, Griffith MP, Evans DR, Ezeonwuka C, Pasculle AW, et al. Evolution of outbreak-causing carbapenem-resistant *Klebsiella*

*pneumoniae* ST258 at a tertiary care hospital over 8 years. MBio. 2019;10. https://doi.org/10.1128/mBio.01945-19.

16. Decraene V, Phan HTT, George R, Wyllie DH, Akinremi O, Aiken Z, et al. A large, refractory nosocomial outbreak of Klebsiella pneumoniae carbapenemase-producing *Escherichia coli* demonstrates carbapenemase gene outbreaks involving sink sites require novel approaches to infection control. Antimicrob Agents Chemother. 2018;62(12):e01689–18. https://doi.org/10.1128/AAC.01689-18.

17. Otter JA, Burgess P, Davies F, Mookerjee S, Singleton J, Gilchrist M, et al. Counting the cost of an outbreak of carbapenemase-producing Enterobacteriaceae: an economic evaluation from a hospital perspective. Clin Microbiol Infect. 2017;23:188–96. https://doi.org/10.1016/j.cmi.2016.10.005.

18. Bartsch SM, McKinnell JA, Mueller LE, Miller LG, Gohil SK, Huang SS, et al. Potential economic burden of carbapenem-resistant Enterobacteriaceae (CRE) in the United States. Clin Microbiol Infect. 2017;23:48.e9–48.e16. https://doi.org/10.1016/j.cmi.2016.09.003.

19. Adeyi OO, Baris E, Jonas OB, Irwin A, Berthe FCJ, Le Gall FG, et al. Drug-resistant infections: a threat to our economic future. Washington, DC: World Bank Group; 2017.

20. Surveillance of antimicrobial resistance in Europe 2018. European Centre for Disease Prevention and Control; 2019. https://www.ecdc.europa.eu/en/publications-data/surveillance-antimicrobial-resistance-europe-2018. Accessed 22 May 2020.

21. Centers for Disease Control and Prevention (CDC). Vital signs: carbapenem-resistant Enterobacteriaceae. MMWR Morb Mortal Wkly Rep. 2013;62:165–70 https://www.cdc.gov/mmwr/preview/mmwrhtml/mm6209a3.htm.

22. Logan LK, Weinstein RA. The epidemiology of carbapenem-resistant Enterobacteriaceae: the impact and evolution of a global menace. J Infect Dis. 2017;215:S28–36.

23. Centers for Disease Control and Prevention (U.S.). Antibiotic resistance threats in the United States, 2019: Centers for Disease Control and Prevention (U.S.); 2019. https://stacks.cdc.gov/view/cdc/82532

24. Zhang Y, Wang Q, Yin Y, Chen H, Jin L, Gu B, et al. Epidemiology of carbapenem-resistant Enterobacteriaceae infections: report from the China CRE Network. Antimicrob Agents Chemother. 2018;62. https://doi.org/10.1128/AAC.01882-17.

25. Tacconelli E, Magrini N, Kahlmeter G, Singh N. Global priority list of antibiotic-resistant bacteria to guide research, discovery, and development of new antibiotics. Geneva: World Health Organization; 2017. p. 27.

26. Woodworth KR, Walters MS, Weiner LM, Edwards J, Brown AC, Huang JY, et al. Vital signs: containment of novel multidrug-resistant organisms and resistance mechanisms - United States, 2006-2017. MMWR Morb Mortal Wkly Rep. 2018;67:396–401. https://doi.org/10.15585/mmwr.mm6713e1.

27. Jernigan JA, Hatfield KM, Wolford H, Nelson RE, Olubajo B, Reddy SC, et al. Multidrug-resistant bacterial infections in US hospitalized patients, 2012--2017. N Engl J Med. 2020;382:1309–19.

28. Tzouvelekis LS, Markogiannakis A, Psichogiou M, Tassios PT, Daikos GL. Carbapenemases in *Klebsiella pneumoniae* and other Enterobacteriaceae: an evolving crisis of global dimensions. Clin Microbiol Rev. 2012;25:682–707. https://doi.org/10.1128/CMR.05035-11.

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

30. Cerqueira GC, Earl AM, Ernst CM, Grad YH, Dekker JP, Feldgarden M, et al. Multi-institute analysis of carbapenem resistance reveals remarkable diversity, unexplained mechanisms, and limited clonal outbreaks. Proc Natl Acad Sci U S A. 2017;114:1135–40. https://doi.org/10.1073/pnas.1616248114.

31. 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 *bla*$_{KPC}$. Antimicrob Agents Chemother. 2016;60:3767–78. https://doi.org/10.1128/AAC.00464-16.

32. Paschoal RP, Campana EH, Corrêa LL, Montezzi LF, Barrueto LRL, da Silva IR, et al. Concentration and variety of carbapenemase producers in recreational coastal waters showing distinct levels of pollution. Antimicrob Agents Chemother. 2017;61. https://doi.org/10.1128/AAC.01963-17.

33. Marathe NP, Janzon A, Kotsakis SD, Flach C-F, Razavi M, Berglund F, et al. Functional metagenomics reveals a novel carbapenem-hydrolyzing mobile beta-lactamase from Indian river sediments contaminated with antibiotic production waste. Environ Int. 2018;112:279–86. https://doi.org/10.1016/j.envint.2017.12.036.

34. Gomi R, Matsuda T, Yamamoto M, Chou P-H, Tanaka M, Ichiyama S, et al. Characteristics of carbapenemase-producing Enterobacteriaceae in wastewater revealed by genomic analysis. Antimicrob Agents Chemother. 2018;62. https://doi.org/10.1128/AAC.02501-17.

35. Larsson DGJ, Andremont A, Bengtsson-Palme J, Brandt KK, de Roda Husman AM, Fagerstedt P, et al. Critical knowledge gaps and research needs related to the environmental dimensions of antibiotic resistance. Environ Int. 2018;117:132–8. https://doi.org/10.1016/j.envint.2018.04.041.

36. Gorrie CL, Mirceta M, Wick RR, Edwards DJ, Thomson NR, Strugnell RA, et al. Gastrointestinal carriage is a major reservoir of *Klebsiella pneumoniae* infection in intensive care patients. Clin Infect Dis. 2017;65:208–15. https://doi.org/10.1093/cid/cix270.

37. Kelly AM, Mathema B, Larson EL. Carbapenem-resistant Enterobacteriaceae in the community: a scoping review. Int J Antimicrob Agents. 2017;50:127–34. https://doi.org/10.1016/j.ijantimicag.2017.03.012.

38. van Loon K, Voor in 't holt AF, Vos MC. A systematic review and meta-analyses of the clinical epidemiology of carbapenem-resistant Enterobacteriaceae. Antimicrob Agents Chemother. 2017;62:1791–18.

39. Weingarten RA, Johnson RC, Conlan S, Ramsburg AM, Dekker JP, Lau AF, et al. Genomic analysis of hospital plumbing reveals diverse reservoir of bacterial plasmids conferring carbapenem resistance. mBio. 2018;9. https://doi.org/10.1128/mbio.02011-17.

40. Chng KR, Li C, Bertrand D, Ng AHQ, Kwah JS, Low HM, et al. Cartography of opportunistic pathogens and antibiotic resistance genes in a tertiary hospital environment. Nat Med. 2020;26:941–51. https://doi.org/10.1038/s41591-020-0894-4.

41. Conlan S, Thomas PJ, Deming C, Park M, Lau AF, Dekker JP, et al. Single-molecule sequencing to track plasmid diversity of hospital-associated carbapenemase-producing Enterobacteriaceae. Sci Transl Med. 2014;6:254ra126. https://doi.org/10.1126/scitranslmed.3009845.

42. Mathers AJ, Peirano G, Pitout JDD. The role of epidemic resistance plasmids and international high-risk clones in the spread of multidrug-resistant Enterobacteriaceae. Clin Microbiol Rev. 2015;28:565–91. https://doi.org/10.1128/CMR.00116-14.

43. Snitkin ES, Won S, Pirani A, Lapp Z, Weinstein RA, Lolans K, et al. Integrated genomic and interfacility patient-transfer data reveal the transmission pathways of multidrug-resistant *Klebsiella pneumoniae* in a regional outbreak. Sci Transl Med. 2017;9. https://doi.org/10.1126/scitranslmed.aan0093.

44. Han JH, Lapp Z, Bushman F, Lautenbach E, Goldstein EJC, Mattei L, et al. Whole-genome sequencing to identify drivers of carbapenem-resistant *Klebsiella pneumoniae* transmission within and between regional long-term acute-care hospitals. Antimicrob Agents Chemother. 2019;63:e01622–19.

45. Sheppard AE, Stoesser N, German-Mesner I, Vegesana K, Sarah Walker A, Crook DW, et al. TETyper: a bioinformatic pipeline for classifying variation and genetic contexts of transposable elements from short-read whole-genome sequencing data. Microb Genom. 2018;4. https://doi.org/10.1099/mgen.0.000232.

46. Göttig S, Gruber TM, Stecher B, Wichelhaus TA, Kempf VAJ. In vivo horizontal gene transfer of the carbapenemase OXA-48 during a nosocomial outbreak. Clin Infect Dis. 2015;60:1808–15. https://doi.org/10.1093/cid/civ191.

47. Mulvey MR, Haraoui L-P, Longtin Y. Multiple variants of *Klebsiella pneumoniae* producing carbapenemase in one patient. N Engl J Med. 2016;375:2408–10. https://doi.org/10.1056/NEJMc1511360.

48. Aminov RI. Horizontal gene exchange in environmental microbiota. Front Microbiol. 2011;2:158. https://doi.org/10.3389/fmicb.2011.00158.

49. Sheppard RJ, Beddis AE, Barraclough TG. The role of hosts, plasmids and environment in determining plasmid transfer rates: a meta-analysis. Plasmid. 2020;108:102489. https://doi.org/10.1016/j.plasmid.2020.102489.

50. Lerminiaux NA, Cameron ADS. Horizontal transfer of antibiotic resistance genes in clinical environments. Can J Microbiol. 2019;65:34–44. https://doi.org/10.1139/cjm-2018-0275.

Salamzade *et al. Genome Medicine*        (2022) 14:37

Page 20 of 22

51. 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. 2020;117:25043–54. https://doi.org/10.1073/pnas.2003407117.

52. CLSI. Performance standards for antimicrobial susceptibility testing. 2020.

53. Smith KP, Kirby JE. Verification of an automated, digital dispensing platform for at-will broth microdilution-based antimicrobial susceptibility testing. J Clin Microbiol. 2016;54:2288–93. https://doi.org/10.1128/JCM.00932-16.

54. Earl AM, Onderdonk AB, Kirby J, Ferraro MJ, Huang S, Spencer M, et al. Sequencing of carbapenem resistant bacteria genome sequencing: BioProject PRJNA271899, NCBI Sequence Read Archive; 2015. https://www.ncbi.nlm.nih.gov/sra/PRJNA271899

55. Wick RR, Judd LM, Holt KE. Deepbinner: demultiplexing barcoded Oxford Nanopore reads with deep convolutional neural networks. PLoS Comput Biol. 2018;14:e1006583. https://doi.org/10.1371/journal.pcbi.1006583.

56. Wick R. Porechop: adapter trimmer for Oxford Nanopore reads: GitHub; 2017. https://github.com/rrwick/Porechop

57. Krueger F. TrimGalore: a wrapper around Cutadapt and FastQC to consistently apply adapter and quality trimming to FastQ files, with extra functionality for RRBS data: GitHub; 2017. https://github.com/FelixKrueger/TrimGalore

58. Wick RR, Judd LM, Gorrie CL, Holt KE. Unicycler: resolving bacterial genome assemblies from short and long sequencing reads. PLoS Comput Biol. 2017;13:e1005595. https://doi.org/10.1371/journal.pcbi.1005595.

59. Li H. Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics. 2018;34:3094–100. https://doi.org/10.1093/bioinformatics/bty191.

60. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 2009;25:1754–60. https://doi.org/10.1093/bioinformatics/btp324.

61. Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, et al. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. PLoS One. 2014;9:e112963. https://doi.org/10.1371/journal.pone.0112963.

62. Walker BJ, Earl A. GAEMR: genome assembly evaluation, metrics and reporting: GitHub; 2018. https://github.com/broadinstitute/GAEMR

63. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, et al. BLAST+: architecture and applications. BMC Bioinformatics. 2009;10:421. https://doi.org/10.1186/1471-2105-10-421.

64. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol. 1990;215:403–10. https://doi.org/10.1016/S0022-2836(05)80360-2.

65. Zankari E, Hasman H, Cosentino S, Vestergaard M, Rasmussen S, Lund O, et al. Identification of acquired antimicrobial resistance genes. J Antimicrob Chemother. 2012;67:2640–4. https://doi.org/10.1093/jac/dks261.

66. Sayers EW, Beck J, Brister JR, Bolton EE, Canese K, Comeau DC, et al. Database resources of the National Center for Biotechnology Information. Nucleic Acids Res. 2020;48:D9–16. https://doi.org/10.1093/nar/gkz899.

67. Bradford PA. Extended-spectrum beta-lactamases in the 21st century: characterization, epidemiology, and detection of this important resistance threat. Clin Microbiol Rev. 2001;14:933–51, table of contents. https://doi.org/10.1128/CMR.14.4.933-951.2001.

68. Kurtz S, Phillippy A, Delcher AL, Smoot M, Shumway M, Antonescu C, et al. Versatile and open software for comparing large genomes. Genome Biol. 2004;5:R12. https://doi.org/10.1186/gb-2004-5-2-r12.

69. Siguier P, Perochon J, Lestrade L, Mahillon J, Chandler M. ISfinder: the reference centre for bacterial insertion sequences. Nucleic Acids Res. 2006;34:D32–6. https://doi.org/10.1093/nar/gkj014.

70. Hunt M, Mather AE, Sánchez-Busó L, Page AJ, Parkhill J, Keane JA, et al. ARIBA: rapid antimicrobial resistance genotyping directly from sequencing reads. Microb Genom. 2017;3:e000131. https://doi.org/10.1099/mgen.0.000131.

71. Robertson J, Nash JHE. MOB-suite: software tools for clustering, reconstruction and typing of plasmids from draft assemblies. Microb Genom. 2018;4. https://doi.org/10.1099/mgen.0.000206.

72. Ondov BD, Treangen TJ, Melsted P, Mallonee AB, Bergman NH, Koren S, et al. Mash: fast genome and metagenome distance estimation using MinHash. Genome Biol. 2016;17:132. https://doi.org/10.1186/s13059-016-0997-x.

73. Carattoli A, Zankari E, Garcìa-Fernandez A, Larsen MV, Lund O, Villa L, et al. PlasmidFinder and pMLST: in silico detection and typing of plasmids. Antimicrob Agents Chemother. 2014. https://doi.org/10.1128/AAC.02412-14.

74. Guglielmini J, Quintais L, Garcillán-Barcia MP, de la Cruz F, Rocha EPC. The repertoire of ICE in prokaryotes underscores the unity, diversity, and ubiquity of conjugation. PLoS Genet. 2011;7:e1002222. https://doi.org/10.1371/journal.pgen.1002222.

75. Georgescu CH, Manson AL, Griggs AD, Desjardins CA, Pironti A, Wapinski I, et al. SynerClust: a highly scalable, synteny-aware orthologue clustering tool. Microb Genom. 2018;4. https://doi.org/10.1099/mgen.0.000231.

76. Felsenstein J. Inferring phylogenies, vol. 2. Sunderland: Sinauer Associates; 2004.

77. Paradis E, Schliep K. ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. Bioinformatics. 2019;35:526–8. https://doi.org/10.1093/bioinformatics/bty633.

78. R Core Team. R: the R project for statistical computing. 2020. https://www.r-project.org/. Accessed 17 Jan 2020.

79. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 2004;32:1792–7. https://doi.org/10.1093/nar/gkh340.

80. Price MN, Dehal PS, Arkin AP. FastTree 2--approximately maximum-likelihood trees for large alignments. PLoS One. 2010;5:e9490.

81. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics. 2014;30:1312–3. https://doi.org/10.1093/bioinformatics/btu033.

82. Hodcroft E. PareTree 1.0: remove sequences, bootstraps, and branch lengths from your trees! http://emmahodcroft.com/PareTree.html. Accessed 23 Nov 2020.

83. Goris J, Konstantinidis KT, Klappenbach JA, Coenye T, Vandamme P, Tiedje JM. DNA--DNA hybridization values and their relationship to whole-genome sequence similarities. Int J Syst Evol Microbiol. 2007;57:81–91.

84. Broad Institute. Picard Toolkit: GitHub; 2014. https://github.com/broadinstitute/picard

85. Didelot X, Wilson DJ. ClonalFrameML: efficient inference of recombination in whole bacterial genomes. PLoS Comput Biol. 2015;11:e1004041. https://doi.org/10.1371/journal.pcbi.1004041.

86. Jolley KA, Bray JE, Maiden MCJ. Open-access bacterial population genomics: BIGSdb software, the PubMLST.org website and their applications. Wellcome Open Res. 2018;3:124. https://doi.org/10.12688/wellcomeopenres.14826.1.

87. Salamzade R, Manson A, Earl A. *ConSequences*: suite to delineate contiguous and conserved sequences from assemblies and search for their presence in raw sequencing data: GitHub; 2020. https://github.com/broadinstitute/ConSequences

88. Jørgensen TS, Xu Z, Hansen MA, Sørensen SJ, Hansen LH. Hundreds of circular novel plasmids and DNA elements identified in a rat cecum metamobilome. PLoS One. 2014;9:e87924. https://doi.org/10.1371/journal.pone.0087924.

89. Kothari A, Wu Y-W, Chandonia J-M, Charrier M, Rajeev L, Rocha AM, et al. Large circular plasmids from groundwater Plasmidomes span multiple incompatibility groups and are enriched in multimetal resistance genes. MBio. 2019;10. https://doi.org/10.1128/mBio.02899-18.

90. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods. 2012;9:357–9. https://doi.org/10.1038/nmeth.1923.

91. Fu L, Niu B, Zhu Z, Wu S, Li W. CD-HIT: accelerated for clustering the next-generation sequencing data. Bioinformatics. 2012;28:3150–2. https://doi.org/10.1093/bioinformatics/bts565.

92. NCBI Resource Coordinators. Database resources of the National Center for Biotechnology Information. Nucleic Acids Res. 2014;42:D7–17. https://doi.org/10.1093/nar/gkt1146.

93. Esmukov K, et al. geopy: geocoding library for Python: GitHub; 2014. https://github.com/geopy/geopy

Salamzade *et al. Genome Medicine*　　　(2022) 14:37

Page 21 of 22

94. Bradley P, den Bakker HC, Rocha EPC, McVean G, Iqbal Z. Ultrafast search of all deposited bacterial and viral genomic data. Nat Biotechnol. 2019;37:152–9. https://doi.org/10.1038/s41587-018-0010-1.

95. Toribio AL, Alako B, Amid C, Cerdeño-Tarrága A, Clarke L, Cleland I, et al. European Nucleotide Archive in 2016. Nucleic Acids Res. 2017;45:D32–6. https://doi.org/10.1093/nar/gkw1106.

96. Marçais G, Kingsford C. A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. Bioinformatics. 2011;27:764–70. https://doi.org/10.1093/bioinformatics/btr011.

97. O'Leary NA, Wright MW, Brister JR, Ciufo S, Haddad D, McVeigh R, et al. Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. Nucleic Acids Res. 2016;44:D733–45. https://doi.org/10.1093/nar/gkv1189.

98. Finn RD, Bateman A, Clements J, Coggill P, Eberhardt RY, Eddy SR, et al. Pfam: the protein families database. Nucleic Acids Res. 2014;42:D222–30. https://doi.org/10.1093/nar/gkt1223.

99. Reis-Cunha JL, Bartholomeu DC, Manson AL, Earl AM, Cerqueira GC. ProphET, prophage estimation tool: a stand-alone prophage sequence prediction tool with self-updating reference database. PLoS One. 2019;14:e0223364. https://doi.org/10.1371/journal.pone.0223364.

100. Saier MH Jr, Reddy VS, Tamang DG, Västermark A. The transporter classification database. Nucleic Acids Res. 2014;42:D251–8. https://doi.org/10.1093/nar/gkt1097.

101. Thomas CM, Thomson NR, Cerdeño-Tárraga AM, Brown CJ, Top EM, Frost LS. Annotation of plasmid genes. Plasmid. 2017;91:61–7. https://doi.org/10.1016/j.plasmid.2017.03.006.

102. Abby SS, Néron B, Ménager H, Touchon M, Rocha EPC. MacSyFinder: a program to mine genomes for molecular systems with an application to CRISPR-Cas systems. PLoS One. 2014;9:e110726. https://doi.org/10.1371/journal.pone.0110726.

103. Abby SS, Cury J, Guglielmini J, Néron B, Touchon M, Rocha EPC. Identification of protein secretion systems in bacterial genomes. Sci Rep. 2016;6:23080. https://doi.org/10.1038/srep23080.

104. Cury J, Abby SS, Doppelt-Azeroual O, Néron B, Rocha EPC. Identifying conjugative plasmids and integrative conjugative elements with CONJscan. Methods Mol Biol. 2020;2075:265–83. https://doi.org/10.1007/978-1-4939-9877-7_19.

105. Hayes F. Toxins-antitoxins: plasmid maintenance, programmed cell death, and cell cycle arrest. Science. 2003;301:1496–9. https://doi.org/10.1126/science.1088157.

106. Nekrasov SV, Agafonova OV, Belogurova NG, Delver EP, Belogurov AA. Plasmid-encoded antirestriction protein ArdA can discriminate between type I methyltransferase and complete restriction-modification system. J Mol Biol. 2007;365:284–97. https://doi.org/10.1016/j.jmb.2006.09.087.

107. Belogurov AA, Delver EP, Rodzevich OV. Plasmid pKM101 encodes two nonhomologous antirestriction proteins (ArdA and ArdB) whose expression is controlled by homologous regulatory sequences. J Bacteriol. 1993;175:4843–50. https://doi.org/10.1128/jb.175.15.4843-4850.1993.

108. Eddy SR. Accelerated Profile HMM Searches. PLoS Comput Biol. 2011;7:e1002195. https://doi.org/10.1371/journal.pcbi.1002195.

109. Akarsu H, Bordes P, Mansour M, Bigot D-J, Genevaux P, Falquet L. TASmania: a bacterial Toxin-Antitoxin Systems database. PLoS Comput Biol. 2019;15:e1006946. https://doi.org/10.1371/journal.pcbi.1006946.

110. Cock PJA, Antao T, Chang JT, Chapman BA, Cox CJ, Dalke A, et al. Biopython: freely available Python tools for computational molecular biology and bioinformatics. Bioinformatics. 2009;25:1422–3. https://doi.org/10.1093/bioinformatics/btp163.

111. Ribeiro FJ, Przybylski D, Yin S, Sharpe T, Gnerre S, Abouelleil A, et al. Finished bacterial genomes from shotgun sequence data. Genome Res. 2012;22:2270–7. https://doi.org/10.1101/gr.141515.112.

112. Stoesser N, Phan HTT, Seale AC, Aiken Z, Thomas S, Smith M, et al. Genomic epidemiology of complex, multispecies, plasmid-borne *bla*KPC carbapenemase in Enterobacterales in the United Kingdom from 2009 to 2014. Antimicrob Agents Chemother. 2020;64. https://doi.org/10.1128/AAC.02244-19.

113. Mathers AJ, Stoesser N, Sheppard AE, Pankhurst L, Giess A, Yeh AJ, et al. *Klebsiella pneumoniae* carbapenemase (KPC)-producing *K. pneumoniae* at a single institution: insights into endemicity from whole-genome sequencing. Antimicrob Agents Chemother. 2015;59:1656–63. https://doi.org/10.1128/AAC.04292-14.

114. Schürch AC, Arredondo-Alonso S, Willems RJL, Goering RV. Whole genome sequencing options for bacterial strain typing and epidemiologic analysis based on single nucleotide polymorphism versus gene-by-gene-based approaches. Clin Microbiol Infect. 2018;24:350–4. https://doi.org/10.1016/j.cmi.2017.12.016.

115. Poulou A, Voulgari E, Vrioni G, Koumaki V, Xidopoulos G, Chatzipantazi V, et al. Outbreak caused by an ertapenem-resistant, CTX-M-15-producing *Klebsiella pneumoniae* sequence type 101 clone carrying an OmpK36 porin variant. J Clin Microbiol. 2013;51:3176–82. https://doi.org/10.1128/JCM.01244-13.

116. Vandecraen J, Chandler M, Aertsen A, Van Houdt R. The impact of insertion sequences on bacterial genome plasticity and adaptability. Crit Rev Microbiol. 2017;43:709–30. https://doi.org/10.1080/1040841X.2017.1303661.

117. Ernst CM, Braxton JR, Rodriguez-Osorio CA, Zagieboylo AP, Li L, Pironti A, et al. Adaptive evolution of virulence and persistence in carbapenem-resistant *Klebsiella pneumoniae*. Nat Med. 2020;26:705–11. https://doi.org/10.1038/s41591-020-0825-4.

118. Denamur E, Matic I. Evolution of mutation rates in bacteria. Mol Microbiol. 2006;60:820–7. https://doi.org/10.1111/j.1365-2958.2006.05150.x.

119. Brandt C, Viehweger A, Singh A, Pletz MW, Wibberg D, Kalinowski J, et al. Assessing genetic diversity and similarity of 435 KPC-carrying plasmids. Sci Rep. 2019;9:11223. https://doi.org/10.1038/s41598-019-47758-5.

120. Poirel L, Bonnin RA, Nordmann P. Genetic features of the widespread plasmid coding for the carbapenemase OXA-48. Antimicrob Agents Chemother. 2012;56:559–62. https://doi.org/10.1128/AAC.05289-11.

121. Wang R, van Dorp L, Shaw LP, Bradley P, Wang Q, Wang X, et al. The global distribution and spread of the mobilized colistin resistance gene mcr-1. Nat Commun. 2018;9:1179. https://doi.org/10.1038/s41467-018-03205-z.

122. Evans DR, Griffith MP, Sundermann AJ, Shutt KA, Saul MI, Mustapha MM, et al. Systematic detection of horizontal gene transfer across genera among multidrug-resistant bacteria in a single hospital. Elife. 2020;9. https://doi.org/10.7554/eLife.53886.

123. Mahillon J, Chandler M. Insertion sequences. Microbiol Mol Biol Rev. 1998;62:725–74.

124. Grainge I, Jayaram M. The integrase family of recombinase: organization and function of the active site. Mol Microbiol. 1999;33:449–56. https://doi.org/10.1046/j.1365-2958.1999.01493.x.

125. Hegstad K, Langsrud S, Lunestad BT, Scheie AA, Sunde M, Yazdankhah SP. Does the wide use of quaternary ammonium compounds enhance the selection and spread of antimicrobial resistance and thus threaten our health? Microb Drug Resist. 2010;16:91–104. https://doi.org/10.1089/mdr.2009.0120.

126. Bennett PM. Plasmid encoded antibiotic resistance: acquisition and transfer of antibiotic resistance genes in bacteria. Br J Pharmacol. 2008;153(Suppl 1):S347–57. https://doi.org/10.1038/sj.bjp.0707607.

127. Carattoli A. Resistance plasmid families in Enterobacteriaceae. Antimicrob Agents Chemother. 2009;53:2227–38. https://doi.org/10.1128/AAC.01707-08.

128. Johnson TJ, Danzeisen JL, Youmans B, Case K, Llop K, Munoz-Aguayo J, et al. Separate F-type plasmids have shaped the evolution of the H30 subclone of *Escherichia coli* sequence type 131. mSphere. 2016;1. https://doi.org/10.1128/mSphere.00121-16.

129. Pal C, Bengtsson-Palme J, Kristiansson E, Larsson DGJ. Co-occurrence of resistance genes to antibiotics, biocides and metals reveals novel insights into their co-selection potential. BMC Genomics. 2015;16:964. https://doi.org/10.1186/s12864-015-2153-5.

130. Li L-G, Xia Y, Zhang T. Co-occurrence of antibiotic and metal resistance genes revealed in complete genome collection. ISME J. 2017;11:651–62. https://doi.org/10.1038/ismej.2016.155.

131. Popowska M, Krawczyk-Balska A. Broad-host-range IncP-1 plasmids and their resistance potential. Front Microbiol. 2013;4:44. https://doi.org/10.3389/fmicb.2013.00044.

132. Rozwandowicz M, Brouwer MSM, Fischer J, Wagenaar JA, Gonzalez-Zorn B, Guerra B, et al. Plasmids carrying antimicrobial resistance genes in Enterobacteriaceae. J Antimicrob Chemother. 2018;73:1121–37. https://doi.org/10.1093/jac/dkx488.

Salamzade *et al. Genome Medicine*     (2022) 14:37

Page 22 of 22

133. Hunte C, Screpanti E, Venturi M, Rimon A, Padan E, Michel H. Structure of a Na+/H+ antiporter and insights into mechanism of action and regulation by pH. Nature. 2005;435:1197–202. https://doi.org/10.1038/nature03692.

134. He S, Hickman AB, Varani AM, Siguier P, Chandler M, Dekker JP, et al. Insertion sequence IS26 reorganizes plasmids in clinically isolated multidrug-resistant bacteria by replicative transposition. MBio. 2015;6:e00762. https://doi.org/10.1128/mBio.00762-15.

135. Novais A, Rodrigues C, Branquinho R, Antunes P, Grosso F, Boaventura L, et al. Spread of an OmpK36-modified ST15 *Klebsiella pneumoniae* variant during an outbreak involving multiple carbapenem-resistant Enterobacteriaceae species and clones. Eur J Clin Microbiol Infect Dis. 2012;31:3057–63. https://doi.org/10.1007/s10096-012-1665-z.

136. Knopp M, Andersson DI. Amelioration of the fitness costs of antibiotic resistance due to reduced outer membrane permeability by upregulation of alternative porins. Mol Biol Evol. 2015;32:3252–63. https://doi.org/10.1093/molbev/msv195.

137. Vergalli J, Bodrenko IV, Masi M, Moynié L, Acosta-Gutiérrez S, Naismith JH, et al. Porins and small-molecule translocation across the outer membrane of Gram-negative bacteria. Nat Rev Microbiol. 2020;18:164–76. https://doi.org/10.1038/s41579-019-0294-2.

138. Cuzon G, Naas T, Nordmann P. Functional characterization of Tn4401, a Tn3-based transposon involved in $bla_{KPC}$ gene mobilization. Antimicrob Agents Chemother. 2011;55:5370–3. https://doi.org/10.1128/AAC.05202-11.

139. Naas T, Cuzon G, Villegas M-V, Lartigue M-F, Quinn JP, Nordmann P. Genetic structures at the origin of acquisition of the β-lactamase $bla_{KPC}$ gene. Antimicrob Agents Chemother. 2008;52:1257–63. https://doi.org/10.1128/AAC.01451-07.

140. Hall JPJ, Wood AJ, Harrison E, Brockhurst MA. Source-sink plasmid transfer dynamics maintain gene mobility in soil bacterial communities. Proc Natl Acad Sci U S A. 2016;113:8260–5. https://doi.org/10.1073/pnas.1600974113.

141. Lopatkin AJ, Meredith HR, Srimani JK, Pfeiffer C, Durrett R, You L. Persistence and reversal of plasmid-mediated antibiotic resistance. Nat Commun. 2017;8:1689. https://doi.org/10.1038/s41467-017-01532-1.

142. Gama JA, Zilhão R, Dionisio F. Co-resident plasmids travel together. Plasmid. 2017;93:24–9. https://doi.org/10.1016/j.plasmid.2017.08.004.

143. Gama JA, Zilhão R, Dionisio F. Conjugation efficiency depends on intra and intercellular interactions between distinct plasmids: plasmids promote the immigration of other plasmids but repress co-colonizing plasmids. Plasmid. 2017;93:6–16. https://doi.org/10.1016/j.plasmid.2017.08.003.

144. Dionisio F, Zilhão R, Gama JA. Interactions between plasmids and other mobile genetic elements affect their transmission and persistence. Plasmid. 2019;102:29–36. https://doi.org/10.1016/j.plasmid.2019.01.003.

145. Lee BY, Bartsch SM, Wong KF, Kim DS, Cao C, Mueller LE, et al. Tracking the spread of carbapenem-resistant Enterobacteriaceae (CRE) through clinical cultures alone underestimates the spread of CRE even more than anticipated. Infect Control Hosp Epidemiol. 2019;40:731–4. https://doi.org/10.1017/ice.2019.61.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.