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

Rauf Salamzade[1,+], Abigail L. Manson[1], Bruce J. Walker[1,2], Thea Brennan-Krohn[3], 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[4], Mary L. Delaney[5], Sanjat Kanjilal[5,6], Andrew B. Onderdonk[5], Alejandro Pironti[1], Cassiana E. Bittencourt[8], Gabrielle M. Gussin[7], Diane Kim[7], Ellena M. Peterson[8], Mary Jane Ferraro[9], David C. Hooper[9], Erica S. Shenoy[9], Christina A. Cuomo[1], Deborah T. Hung[1,9], Lisa A. Cosimi[1,5], Susan S. Huang[7], James E. Kirby[3], Virginia M. Pierce[9], Roby P. Bhattacharyya[1,9], Ashlee M. Earl[1]

[1]Infectious Disease and Microbiome Program, Broad Institute of MIT and Harvard, Cambridge, MA 02142

[2]Applied Invention, LLC, Cambridge, MA 02139

[3]Beth Israel Deaconess Medical Center, Boston, MA 02215

[4]Department of Ophthalmology, Department of Microbiology, Harvard Medical School and Massachusetts Eye and Ear Infirmary, 240 Charles St., Boston MA 02114

[5]Department of Pathology, Brigham and Women's Hospital, Harvard Medical School, Boston, MA 02115

[6]Department of Population Medicine, Harvard Medical School and Harvard Pilgrim Healthcare Institute, Boston, MA, 02215

[7]Division of Infectious Diseases, University of California Irvine School of Medicine, Irvine, CA 92617

[8]Department of Pathology and Laboratory Medicine, University of California Irvine School of Medicine, Orange, CA 92868

[9]Massachusetts General Hospital, Boston, MA 02114

[+]Present address: Microbiology Doctoral Training Program, University of Wisconsin-Madison, Madison, WI, 53706

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 U.S. hospitals over nearly five 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 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 [38–42]. CRE are readily acquired and spread from these reservoirs [13,40], including hospital-adapted *high-risk lineages* (*e.g. K. pneumoniae* sequence type (ST)-258) [43] that are associated with both intra- and inter-facility clonal transmission [13,29,44,45]. 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 [40,42].

Though bioinformatic approaches targeting the sequences surrounding carbapenemases have been used to predict the movement of carbapenemases across some CRE populations [46], 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 [39,40,42,47,48]. Furthermore, HGT rates are predicted to be high [49–51], and carbapenemase containing plasmids frequently recombine at sites of repetitive sequence, leading to mosaic plasmid structures [31,52]. 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 [40,42]. 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 three-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.

Results

Comprehensive collection of clinical specimen CREs from microbiology labs over a nearly five 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 ≥ 2μg/ml; Materials and Methods) cultured from clinical specimens between December 2013 and December 2016, regardless of species or resistance mechanism. These 146 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 (Figure 1; Table 1; Tables S1-S3). 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] (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 (Table S1).

For each isolate, we generated highly contiguous *de novo* genome assemblies (average scaffold N50 of 3.8 Mb; Table S4) using a combination of Illumina short paired-end and long-

insert mate pair libraries [53]. Echoing our previous work [30], and consistent with previous studies [29–31,54], 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 (Figure 1; Table S1). As expected by the study design, CSE isolates were similarly distributed by taxonomy (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 [43] due to their ability to cause severe and/or recurrent drug-resistant infections and rapidly spread (Materials and Methods). This revealed a striking intra-species diversity of organisms within our dataset (Figure 2; Figure S1; Table S1), 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 (Table S3). Another 14% were separated from another isolate by a maximum of 10 core SNVs, a number previously used to suggest recent common ancestry [55,56].

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 (Figure 2; Figure S1-S3; Table 2; Table S1, S5-S7; Supplementary Results). Pointing to the known role of HGT in the spread of carbapenem resistance in the *Enterobacterales* [30,42,52], species phylogenies revealed both

carriage of identical carbapenemase alleles (and their transposon Tn*4401* variants; Table S8) among distantly related isolates and heterogeneity in the carriage of resistance genes among closely related isolates (Figure 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 (Table S9), 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 $bla_{\text{CTX-M-15}}$ 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 (Table S10). Though similar to an earlier report from hospitals in Greece [57], where 19 clonally-related isolates carried $bla_{\text{CTX-M-15}}$, 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 [58–60].

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 (Table S11)), 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 (Figure S4; Table S12), 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 (Figure S5, panel B), showcasing the known genetic flexibility of plasmids [31,54,61]. Furthermore, plasmid groups were also remarkably geographically widespread [62], with nearly all (92%) of the most prevalent plasmid groups found in isolates from multiple hospitals, including all of those containing carbapenemases (Figure S5, panel C), 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 (Supplementary Results), 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} [46]. 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 (Figure S6).

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 [63], 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 (https://github.com/broadinstitute/ConSequences). 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 (≥ 99% per 10kb block) across two or more plasmid scaffolds (Figure 3, Figure S7-S8). Of the 4,605 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* [31,64]. The size of conserved segments ranged from 10 kb 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 interspecies 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 (Materials and Methods). Finally, we stringently screened these for uniqueness by searching against the ENA/SRA database using BIGSI [65].

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 (Figure S9; Table S13; Supplementary Data File). 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 kb to 311 kb), and were observed, on average, 5 times (range 2 to 19), across 2 plasmid groups (range 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 1,494 individual genes predicted within signatures, we could assign some putative function to nearly two-thirds (Figure S10-S11; Tables S14-S15). 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 [66], integrases or other recombinases [67], conjugation machinery [68], and plasmid uptake and maintenance apparatuses [69,70]. 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 (≥99%) in bacteria outside the order of *Enterobacterales* (Table S14).

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 [71](Figure 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

256 often co-located on plasmids [72–74]. Genes coding for metal resistance were also prevalent,
257 occurring in more than a third of signatures, and often co-occurring with genes for antibiotic
258 resistance, including in half of signatures containing *bla*$_{KPC}$, highlighting recent findings that
259 metal resistance and antibiotic resistance are frequently co-selected [75,76].
260

261 **Additional instances of localized carbapenemase spread were identified through**
262 **tracing of geographic signatures.**

263 Six of the 44 geographic signatures encoded *bla*$_{KPC}$. These *bla*$_{KPC}$ signatures ranged in
264 size from 18kb - 147kb (the smallest of these signatures, Sig 5.1-CP, was nested inside the
265 largest signature, Sig 5.6-CP), and were found in 21% of all *bla*$_{KPC}$-carrying CRE in our
266 collection (Figure 4; Figure S9). As expected from the filters we applied to identify geographic
267 signatures, *bla*$_{KPC}$ signatures were distributed across multiple species (2 to 4) and sequence
268 types (2 to 7), associated with multiple plasmid groups (1 to 3) and sometimes the chromosome,
269 and observed across variable time spans of at least 4 years and up to 10 years (Figure 4; Table
270 S16). Furthermore, some isolates carried more than one signature (*e.g.*, Sig1-CP and Sig4-CP),
271 sometimes on the same plasmid.
272 Though our inclusion criteria would have allowed for as many as 100 SNVs in a 10 kb
273 signature, we observed much greater identity, with most *bla*$_{KPC}$ signatures varying by only two or
274 fewer SNV differences between instances (Figure S12). This striking degree of similarity,
275 extending well beyond the edges of the Tn*4401* elements, likely indicates a recent common
276 ancestral source for these signatures, as well as their persistence and movement across
277 species boundaries within a local environment. These instances of suspected local spread of
278 carbapenemases between distantly related isolates would not have been picked up by
279 traditional epidemiology or even standard whole genome sequence analysis in most cases.
280 Furthermore, many *bla*$_{KPC}$ signature instances were harbored by isolates that were very closely

281 related (0-10 core SNVs), indicating the ability of these signatures to spread within the hospital
282 along with their bacterial hosts (Figure 4), in addition to their ability to move between bacterial
283 sequence types and species.
284

285 **A multi-genus *bla*$_{KPC}$-containing geographic signature is highly conserved**
286 **despite rapid rearrangements of its plasmid backgrounds.**

287 Although our analysis was based on highly contiguous assemblies that combined data
288 from short- and long-insert Illumina libraries, we sought to improve the assemblies further in
289 order to follow the details of signature evolution and spread across plasmids and taxonomic
290 boundaries. To do this, we re-sequenced all isolates carrying Sig5.1-CP using Oxford Nanopore
291 Technology to generate long read sequences for hybrid assembly, followed by manual curation.
292 Sig5.1-CP, a *bla*$_{KPC-3}$-carrying signature, was present in the largest number of different genera
293 and species in our prospective collection and was also found in historical isolates dating back to
294 2008 (Figure 5; Suppl. Table S17). In addition, Sig5.1-CP was almost exclusively found in a
295 single Boston hospital, including within four isolates collected over a five-week period in 2017.
296 These four isolates were initially sequenced because they were suspected to be part of a short-
297 lived *C. freundii* complex (later revealed to be *Citrobacter portucalensis*) clonal case cluster that
298 differed from one another by less than four SNVs. We found that this 2017 cluster of isolates all
299 differed by the same 12 SNVs from a 2014 isolate carrying Sig5.1-CP from the same hospital
300 that was already part of our collection.
301 The complete assembly of each individual replicon unambiguously revealed the
302 relocation of Sig5.1-CP into the chromosome and its association with three different plasmid
303 groups (Figure 5). One of these plasmids encoded both IncP and IncH replicons, which are
304 known for their broad host range [77,78], likely contributing to this signature's ability to transfer
305 to diverse hosts and persist. Although our draft assemblies were adequate to identify the

306 signature using our methodology, manual inspection of Sig5.1-CP boundaries in the completed
307 assemblies revealed that this signature could be expanded to include approximately an
308 additional 1 kb of sequence encoding a *bla*$_{TEM-1A}$ beta lactamase. Manual inspection also
309 revealed an even longer, 26kb conserved region in the 9 isolates occurring since 2012 (Figure
310 S13). This extended signature included additional cargo that could be involved in adaptation to
311 the healthcare environment, including genes predicted to encode mercury resistance, a Na+/H+
312 antiporter (possibly influencing cell viability at high pH [79]), and a multidrug efflux pump.
313 Despite Sig5.1-CP's distinct genomic contexts, it was highly conserved, maintaining perfect
314 base-level identity, with the exception of one isolate (*K. pneumoniae* MGH39 from 2012) that
315 harbored a variant of the signature with two non-synonymous SNVs, both within the same
316 predicted transposon.
317 Though possessing nearly identical chromosomes (≤3 SNVs), isolates from the *C.*
318 *portucalensis* clonal case cluster had a striking variety of plasmid and chromosomal
319 arrangements (Figure S14). While two chromosomally identical 2017 isolates also had identical
320 plasmid profiles and Sig5.1-CP location, one of the other 2017 isolates carried the signature on
321 a different plasmid, while the remaining 2017 isolate carried it on a co-integrate of both
322 plasmids. Many of the rearrangements were likely mediated by IS*26* sequences (Figure S14),
323 previously shown to drive the reorganization of plasmids through replicative transposition [80].
324 Although we could not precisely trace the divergence of the cluster of 2017 isolates from the
325 2014 isolate, multiple plasmid gains, losses, and rearrangements in the 2017 isolates appear to
326 have occurred on the same timescale as the accumulation of SNVs.
327

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 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 U.S. [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 clonal spread of CP-CRE, few studies have pointed out the contribution of porin-based CRE in nosocomial CRE spread [57,81]. 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 four-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 [82,83]. The presence of porin-based CRE clonal clusters indicate 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.

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 [31,52,54,61,84,85]. These processes are accelerated by the frequent co-occurrence of carbapenemase plasmids with other diverse plasmids within a single isolate [42]; 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 [40,42], carbapenemases are often carried by plasmids that can conjugate across species and genera [40,42]. 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 was 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 [46], 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.* [64] 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 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 adds 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 four to ten 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 [75,76] and iii) conjugation and plasmid uptake machinery, both of which can amplify bla_{KPC} and assist its persistence and spread across isolates and species [86,87]. In addition, the association of these signatures with plasmids from different incompatibility groups, particularly those adapted to different conditions such as temperature [88–90], 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 [91]. 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 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.

In conclusion, 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 U.S. hospitals. In addition to examples of 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 intermolecular 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

456 well as other organisms across these hospitals) that helped to clarify the epidemiology of
457 antibiotic resistance spread and persistence in these healthcare networks.
458

459 # Materials and Methods

460 ## Isolate collection and drug susceptibility testing

461 Our sample collection represented a continuation of our previous prospective study,
462 conducted in 2012-2013 [30]**.** Between December 2013 and November 2016, we collected
463 additional samples from symptomatic patients at Beth Israel Deaconess Medical Center
464 (BIDMC), Brigham and Women's Hospital (BWH), and Massachusetts General Hospital (MGH)
465 in Boston, MA, and the University of California Irvine (UCI) Medical Center in Orange, CA. We
466 collected all isolates from clinical samples sent to these hospitals' clinical microbiology
467 laboratories that, using the laboratories' standard operating procedures [30], were identified as
468 *Enterobacterales* with a meropenem minimum inhibitory concentration (MIC) ≥ 2 μg/mL, the
469 threshold that we used to define resistance throughout our analysis. We thus included samples
470 categorized as both *intermediate* (2 μg/mL) and *resistant* (≥ 4 μg/mL) by the Clinical and
471 Laboratory Standards Institute interpretative criteria [92]. For each resistant isolate, we also
472 collected date- and species-matched meropenem susceptible isolates. UCI submitted two
473 carbapenem susceptible isolates per resistant isolate; the other three hospitals submitted a
474 single susceptible isolate per resistant isolate. In total, we collected 347 new isolates (146
475 resistant and 201 susceptible isolates; one isolate per patient). Isolates whose meropenem
476 resistance phenotype was discordant with the genotype were retested using a previously
477 validated automated digital dispensing method [93] adapted from standard broth microdilution
478 procedures [92]. 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 - 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 274 were meropenem-resistant). 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: Mass General Brigham (covering both MGH and BWH); Beth Israel Deaconess Medical Center, Boston; and University of California, Irvine.

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. 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.

Long-read sequencing and assembly. We selected a subset of 19 isolates for further sequencing using Oxford Nanopore Technology (ONT). 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 nanograms of DNA from each sample were 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 1-4 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)[94], trimmed of any remaining adapter using Porechop (v0.2.3), and subsampled to approximately 50x depth of genome coverage. Illumina reads were trimmed of adapter using Trim Galore (v0.5.0) and subsampled to approximately 100x depth of genome coverage. Two Unicycler (v0.4.3 or v0.4.4, with default settings) [95] 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) [96]. Illumina reads were aligned to Unicycler contigs using bwa mem (v0.7.17) [97], and the resulting alignments were input to Pilon (v1.23) [98] for assembly polishing. Contigs were screened for adapter sequence and then input to GAEMR (https://github.com/broadinstitute/GAEMR) which produced chart and metric tables for use in manual assembly analysis process. Hybrid assemblies were annotated as above. Reads were submitted to SRA under 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 [99]: i) our database of carbapenem-hydrolyzing beta-lactamases [30]; ii) ResFinder [100], 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). Genes with hits to any of these databases (E-value < 10^{-10}; query coverage ≥ 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 (Tables S7 and S18). A subset of beta lactamases were designated as ESBLs by our pipeline for identifying resistance genes (Table S5), together with additional evidence from the literature [101].

In order to identify defects in porin sequences contributing to carbapenem resistance, we first identified all porin genes by sequence similarity search [99]. 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 ≥ 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 [102]. Additionally, we pinpointed porins disrupted by insertion sequences by submitting porin sequences and their promoter regions (500 bp upstream) to the ISfinder [103] BLAST facility, 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 [104], 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% (Table S19).

Plasmid Annotations. We used MOB-suite (database version from January 2, 2019) [105] 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 [106] is at least 95%. To supplement MOB-suite’s plasmid replicon database, we added 111 additional replicons from the PlasmidFinder database [107]. 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 [108]. 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 [109], 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 [106] to compare each pair of genomes in our dataset; and iii) computing a dendrogram with the neighbor joining tree algorithm [110] contained in the ape [111] package (v5.0) of R [112] (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 [113] and produced a concatenated alignment of all genes by extracting alignment columns without gaps. We then computed a phylogenetic tree using Fasttree [114] with default parameters, which we then used as input for a second iteration of Synerclust. The orthogroup output from the 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 [115] (v7.3.3), using rapid analysis of 1,000 bootstrap replicates. To generate phylogenetic trees containing only subsets of isolates, we used PareTree (http://emmahodcroft.com/PareTree.html).

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 [116]. We compared to reference assemblies obtained from the NCBI taxonomy browser (https://www.ncbi.nlm.nih.gov/taxonomy) 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 [97] and sorted the alignments with Picard SortSam (v2.20.6; http://broadinstitute.github.io/picard). Finally, we used Pilon (v1.23) [98] 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* (Table S20).

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) [117] 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 (v2.1.3) [114], 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* (Table S20). Using these alignments, we used RAxML [115] (v7.3.3) with 1,000 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 [118] 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* [43]: *E. coli* ST38, ST69, ST131, ST155, ST393, ST405, and ST648 and *K. pneumoniae* ST14, ST37, ST147, and ST258.

631 *Assessment of established genetic markers to trace local movement of bla_{KPC}.* We
632 characterized Tn*4401* structural variants and isoforms, as well as target site duplication (TSD)
633 flanking sequences of Tn*4401* using TETyper [46] for the 608 samples included in our study.
634 We analyzed 5 bp surrounding the three most common Tn*4401* isoforms (Tn*4401*a, Tn*4401*b,
635 and Tn*4401*d) in our assemblies.
636

637 **Identification of geographic signatures**

638 We developed the ConSequences software suite to identify nearly identical ≥ 10 kb
639 segments conserved between two or more plasmids (Figure S7-S8). Along with the open-
640 source code, a test dataset for running the three primary programs in ConSequences can be
641 found on its GitHub repository (https://github.com/broadinstitute/ConSequences), consisting of
642 the twelve hybrid assemblies constructed using Illumina and ONT sequencing for isolates found
643 to harbor the geographic signature Sig5.1-CP.
644 *Selection of plasmid sequences.* To construct the database of plasmid sequences that
645 we searched for geographic signatures, we included all scaffolds between 10 kb and 500 kb that
646 were not classified as chromosomal by MOB-suite [105]. We excluded scaffolds classified as
647 plasmidic and longer than 500 kb, since we found these to be often misclassified (*Plasmid*
648 *annotations*). Circular and complete representations of plasmids [119,120] were determined
649 when a plasmid scaffold showed both significant overlap between its ends (e-value < 10^{-5}) and
650 had at least five mate pair sequencing reads bridging scaffold ends.
651 *Identification of highly conserved 10 kb windows shared across pairs of plasmid*
652 *backbones*. All plasmid-predicted scaffolds were aligned in a pairwise manner using BLASTn in
653 megablast mode. In order to account for circularity of complete plasmids [119,120], we
654 duplicated 10 kb from the beginning of the scaffold and appended it to its end. For each
655 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 ≥ 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 ≥ 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 [30,46] .

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 (Figure S7). As identical segments were often obtained by using different reference scaffolds, we used CD-HIT to cluster sequences with ≥ 99% global identity and ≥ 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

682 database (nt; downloaded in December 2019), using thresholds of 98% identity and 95% query
683 coverage. Hits matching samples in NCBI sourced from the same city were retained as potential
684 geographic signatures.

685 The presence of assembly errors, including incorrect copy counts for tandem repeats,
686 could lead to the incorrect association of segments with geographies. Thus, we checked
687 whether any of the potential geographic signatures contained such tandem repeats using Pilon
688 [98]. Paired-end library sequencing reads from each isolate found to harbor a signature were
689 aligned to the signature's reference sequence using bwa mem (v0.7.17 with default settings)
690 [97]. Then Pilon (v1.23) [98] was run with the options: "--vcf --fix all,breaks --mindepth 5."
691 Instances which triggered a fix break report with the flag"TandemRepeat", indicating the
692 segment likely contained a tandem repeat motif, were identified and removed to ensure
693 geographic association was not driven by faulty estimation of the tandem repeat motif's copy
694 count.

695 We next performed a more comprehensive assessment of uniqueness for each of the
696 signatures using BIGSI [65], searching against all raw sequencing read sets in a snapshot of the
697 ENA/SRA database taken on December 2016, contemporaneous with the most recent isolation
698 dates for the 608 samples in this study. The SRA/ENA snapshot provided a broader database
699 (455,632 read sets) compared to our original screening against the nt database, which included
700 only a subset of assemblies available in the nt database and did not account for bacterial
701 samples with sequencing data but no assembly. To perform this search, we used a sliding
702 window (2 kb window; 1 kb step) across each signature to identify read sets containing at least
703 99% of all *k*-mers for each window. Matching read sets were downloaded from EBI's ENA
704 database and further searched using a *k*-mer based methodology (described below) to more
705 stringently assess whether they harbored any of the geographic signature sequences.

706 *Searching for additional instances of geographic signatures directly in raw Illumina*
707 *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 (Suppl Figure S8). 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. 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. Hits that achieved identity > 98%, 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 [121] 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 (≥ 90% identity and ≥ 80% coverage) to NCBI RefSeq's non-redundant database of bacterial proteins (BacNR); and iii) transferring protein-domain annotations from Pfam [122]. Phages were identified using ProphET [123]. 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 (Tables S14 and S15).

Antibiotic resistance genes were predicted using methods described above (*Annotation of resistance genes*). Chemical and heavy metal resistance operons, providing resistance to mercury, arsenic, tellurium, nickel, and copper, were identified by keyword searches within our combined annotations. Operons were considered when they were composed of 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 [124] of proteins to representative transporter proteins in TCDB [125] (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 [126], was flagged using MacSyFinder tool [127] with CONJscan HMMs [68,128]. 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 [69] and

anti-restriction proteins [129,130]. Toxin and antitoxin genes were predicted using HMMer v3 [131] with HMMs from the TAsmania database [70] (e-value $<10^{-5}$) and 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 [103], as described above (*Annotation of resistance genes*). 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 [124]. 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 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. Statistical significance of the differences in genome sizes and plasmid counts for CRE vs. CSE, as well as for the co-occurrence of plasmids groups, was assessed using a two-sided Wilcoxon rank sum test. The statistical significance of the differences in relaxase carriage in plasmids carrying carbapenemases vs. other plasmids were assessed using a two-sided Fisher's exact test. The

784 statistical significance for the increase in proportion of ESBLs over time was calculated using a
785 regression slope test.

786 Code availability

787 Computer code used for the analysis of our data can be downloaded from
788 https://github.com/broadinstitute/ConSequences. A yaml file is provided for installation of the
789 software and its dependencies through creation of a Conda virtual environment. ConSequences
790 is written in Python3 and made available under the open-source license BSD3.

791 Declarations

792 Ethics approval and consent to participate

793 Not applicable.

794 Consent for publication

795 Not applicable.

796 Availability of data and materials

797 The dataset analysed during the current study is available in the National Center for
798 Biotechnology Information (NCBI) repository under BioProject PRJNA271899
799 (https://www.ncbi.nlm.nih.gov/bioproject/?term=prjna271899).

800 Competing interests

801 B.J.W. is an employee of Applied Invention (Cambridge, MA). No other authors declare
802 competing interests.

Funding

This work was supported by NIH grant U19AI110818 to the Broad Institute, by NIH grant R33AI119114 to JEK. and by NIH training grants T32HD055148, T32AI007061, 1K08AI132716, and Boston Children's Hospital Office of Faculty Development Faculty Career Development fellowship to TB-K, and by Applied Invention, LLC to BJW. The contents of this publication are solely the responsibility of the authors and do not necessarily represent the official views of the funders.

Authors' contributions

Study design. ALM, ABO, DCH, DTH, LAC, SSH, JEK, RPB, and AME

Study coordination. SBC, WH, MLD, CEB, GMG, DK, EMP, MJF, and AME

Assays performed. TB-K, PM, LLH, JQ, and JIW

Data analysis. RS, ALM, BJW, CJW, TPS, and AP

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, AP, and AME

Review and approval of the final manuscript was provided by all authors.

Acknowledgements

The authors would like to gratefully acknowledge the patients from whom study isolates derived. The authors would also like to thank Thomas Abeel, Paul Cao, Gustavo Cerqueira, Christoph Ernst, Michael S. Gilmore, Zamin Iqbal, Jonathan Livny, Amy Mathers, and Noam Shoresh for helpful discussions as well as members of the Broad Bacterial Genomics group. We also thank members of the Broad Technology Labs and Microbial 'Omics Core for their assistance with data generation.

Tables

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

			Same Patient isolates		
	Total CRE	Total CSE	Unique patients	Total same-patient CRE isolates	Total same-patient CSE isolates
Boston historical collection (Cerqueira et al., 2017)	47	2	7	17	1
Initial prospective Collection (Cerqueira et al., 2017)	74	138	5	7	3
Newly sequenced prospective collection	146	201	0	0	0
Total	267	341	12	24	4

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	**145**	**3**	**98.0%**	**65.9%**
All bla_{KPC}	**124**	**2**	**98.4%**	**56.4%**
bla_{KPC-3}	**60**	**1**	**98.4%**	**27.3%**
bla_{KPC-2}	**50**	**1**	**98.0%**	**22.7%**
bla_{KPC-4}	**11**	**0**	**100.0%**	**5.0%**
bla_{KPC} (read based)[a]	**3**	**0**	**100.0%**	**1.4%**
All blaNDM	**14**	**0**	**100.0%**	**6.4%**
bla_{NDM-1}	**11**	**0**	**100.0%**	**5.0%**
bla_{NDM-5}	**3**	**0**	**100.0%**	**1.4%**
All blaSME	**5**	**1**	**83.3%**	**2.3%**
bla_{SME-1}	**2**	**1**	**66.7%**	**0.9%**

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%**	**27.7%**
Includes any ESBL	**46**	**11**	**80.7%**	**20.9%**
Includes $bla_{CTX-M-15}$	**34**	**2**	**94.4%**	**15.5%**
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%**	**14.1%**
Includes bla_{EC}	**19**	**8**	**70.4%**	**8.6%**
Includes bla_{CMY-2}	**12**	**4**	**75.0%**	**5.5%**
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 Materials and Methods).

[b]Strains may have multiple predicted AmpC and/or ESBL genes. See Tables S17 and S18 for specific resistance determinants predicted for each isolate.

Figure Legends

Figure 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].

Figure 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 grey. 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.

Figure 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 algorithm to identify geographic signatures. Conserved segments are colored. Species and hospital of isolation are indicated for each segment.

Figure 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**.

Figure 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].

Supplemental Figures

Provided as two separate multi-page PDF files:

2021_CRE_GenomeMed_SupFigures_S1.pdf and

2021_CRE_GenomeMed_SupFigures_S2-S13.pdf.

Figure S1. Phylogenetic tree for each of the 15 species in our collection with at least five representatives. Below the phylogenies, colored strips indicate resistance mechanisms and hospital of isolation for each isolate. MGH: Massachusetts General Hospital, Boston, MA; UCI: University of California, Irvine, CA; BIDMC: Beth Israel Deaconess Medical Center, Boston, MA; BWH: Brigham and Women's Hospital, Boston, MA.

Figure S2. Resistance mechanisms were diverse, with many shared across species. The numbers of resistant isolates by resistance mechanism (different colors) are depicted by species.

Figure S3. Carbapenemase-carrying isolates tended to have a higher minimum inhibitory concentration than those with other resistance mechanisms. Minimum inhibitory concentrations (MICs) of the isolates in our collection are depicted, stratified by mechanism of resistance (different colors).

Figure S4. High level of diversity and phylogenetic range among predicted plasmids. This plot displays the 215 plasmid groups (rows) contained in all 506 isolates (columns) for which

892 plasmids were predicted. Plasmids with carbapenemases are indicated in red, and plasmids
893 without carbapenemases are indicated in white. Isolates are ordered phylogenetically, while the
894 plasmid groups are ordered by the number of genera in which they occurred and clustered.
895 **Figure S5: Plasmids of diverse groups carried carbapenemases and were found in**
896 **different species, and hospitals.** The number of plasmids from groups for which we found at
897 least four instances is shown. Groups with plasmids that carry carbapenemases (CPs) are
898 depicted on a grey background, while those not observed to carry carbapenemases are shown
899 on a white background. **a,** Plasmid instances colored by carbapenemase carriage. **b**, Plasmid
900 instances colored by genus. **c**, Plasmid groups colored by hospital of origin.
901 **Figure S6: Limited tracing of carbapenemase localized spread using Tn*4401* isoforms**
902 **and their immediate flanking sequences. a,** Number of instances for Tn*4401* isoforms; and **b,**
903 combinations of the three most common Tn*4401* isoforms with their 5 bp flanking sequences
904 which were found in multiple isolates from our study. Colors indicate the proportion of instances
905 found in each of the four hospitals. The asterisk indicates forms that were found in multiple
906 species.
907 **Figure S7: Method for the delineation of segments shared between plasmids.**
908 *ConSequences* identifies the boundaries of conserved segments spanning multiple 10 kb
909 windows which can be found across multiple (> 2) isolates through assessment of conservation
910 profiles across adjacent windows along reference scaffolds. **a**, Each bar depicts a 10 kb window
911 highlighted by sliding window analysis as being conserved in multiple scaffolds. These bars are
912 ordered along the reference scaffold positionally (x-axis) and the height of bars corresponds to
913 the number of scaffolds in our isolate assemblies that have a highly similar match (≥ 99%) to the
914 10 kb sequence on the reference (colored by genus). **b**, Using a custom algorithm (*Materials*
915 *and Methods*), segments ≥ 10 kb were delineated along the reference scaffold based on
916 conservation profiles across multiple adjacent windows.

Figure S8: Workflow to identify geographic signatures. The number of plasmid segments that were retained after sequentially applying different filters to identify 44 geographic signatures is shown (*Materials and Methods*). The number of plasmid segments carrying carbapenemases (CPs) is provided in red.

Figure S9. Signatures were present across diverse sequence types. In this heatmap, each row corresponds to a unique signature, and each column corresponds to a sequence type (ST). The shading represents the percentage of signature instances belonging to different taxonomic lineages, species or ST. The bar plot to the left of the heatmap depicts the number of isolates containing each signature, highlighting their prevalence across different hospitals.

Figure S10. Signatures carried genes important for hospital adaptation and signature mobility. a, Each row in the heatmap corresponds to one of the 44 geographic signatures. Groups of signatures that nest into each other are separated by horizontal dashed lines. The predicted functions of 1,494 genes within our 44 signatures were categorized into five major functional categories, unless they fell outside of these categories (*other*) or no gene function could be predicted (*hypothetical)*. The coloring of the heatmap indicates the percentage of genes of each signature that are assigned to a particular category. The identifiers of carbapenemase-carrying signatures are shown in red type and suffixed with *-CP*. **b.** Number of genes in each signature.

Figure S11: Details of signature content. Schematics are shown for the gene content of each signature, including the five with bla_{KPC}. Genes are colored according to broad functional categorizations (*Materials and Methods*).

Figure S12: Signatures were highly conserved and likely derived from a common ancestral sequence. The number of confident, unambiguous single nucleotide variants (SNVs) differentiating signature instances was calculated for each of the 44 geographic signatures, through comparison of each instance to the signature's representative sequence using Pilon [98]. **a**, Number of isolates carrying each signature. **b**, Box plot of SNV frequencies. SNV

943 frequencies were calculated by normalizing the count of SNVs between each signature instance
944 and the reference sequence (**c**) with the signature's length.
945
946 **Figure 13. Geographic signatures with *bla*$_{KPC}$ can occur in multiple configurations across**
947 **several species and plasmid groups.** The heatmap on the left indicates the presence of
948 signatures Sig5.6-CP and Sig5.1-CP, and the alternate boundaries of the latter, across the
949 twelve isolates found to harbor the signature(s). The gene content of each signature is shown
950 on the right.
951 **Figure S14. Geographic signatures with *bla*$_{KPC}$ occurred in multiple configurations across**
952 **several species and plasmid groups. a,** The core genome single-nucleotide variants (SNVs)
953 and plasmid and geographic signature carriage of five nearly identical *Citrobacter portucalensis*
954 isolates is shown. **b,** Alignment of the MS-621 plasmids carried by all isolates **a**. Two of these
955 plasmids carry Sig5.1-CP, indicated with the bright yellow bars and triangles. The locations of
956 the *bla*$_{KPC}$ and of insertion sequence IS26 are indicated with red and grey rectangles,
957 respectively. Inversions in the alignment are indicated with orange connector lines; matching
958 regions are indicated with green connector lines. In one isolate, plasmids MS-840 and MS-621
959 cointegrated, which is indicated by blue alignment flanks.

960 ## Supplementary Tables

961 ***Provided as a single Excel spreadsheet, where each supplementary table corresponds to***
962 ***a different tab: 2021_CRE_GenomeMed_Supplementary_Tables.xlsx***
963
964 **Supplementary Table S1: Isolates in our dataset**
965 **Supplementary Table S2: Accessions for Illumina-ONT hybrid assemblies**
966 **Supplementary Table S3. Differences in the core genome between same-species isolates**
967 **from the same patient**

Supplementary Table S4. Genome assembly statistics

Supplementary Table S5: List of genes identified as ESBL or AmpC

Supplementary Table S6. Disruptions in *ompC*/*OmpK36* porin gene

Supplementary Table S7. Disruptions in o*mpF*/*OmpK35* porin gene

Supplementary Table S8: Carbapenemase carrying isolates.

Supplementary Table S9: Pairs of closely-related CSE and CRE isolates, where the CRE carried a carbapenemase not found in the CSE

Supplementary Table S10: Cluster of related Klebsiella pneumoniae isolates with double-porin mutations

Supplementary Table S11. Identification of circular plasmids (>2 kb)

Supplementary Table S12. MOB-suite plasmid group predictions

Supplementary Table S13. 44 multi-species geographic signatures specific to a single city.

Supplementary Table S14. Functional annotation of genes present in 44 signatures.

Supplementary Table S15. Prevalence of functional categories across the 44 signatures.

Supplementary Table S16: Isolates which carry a *bla*$_{KPC}$ -containing signature.

Supplementary Table S17: Genomic background of instances of Sig5.1-CP and Sig5.6-CP

Supplementary Table S18: Genomic location of the resistance genes

Supplementary Table S19: Read-based identification of carbapenemases in resistant isolates for which no resistance mechanism was found in the assemblies

Supplementary Table S20: Size of species-specific core-genome alignments

Additional Supplementary Documents

Supplementary Results: Supplementary Results.

Supplementary Data File: Representative sequences of 44 geographic signatures .

References

1. Trecarichi EM, Tumbarello M. Therapeutic options for carbapenem-resistant Enterobacteriaceae infections. Virulence. 2017;8:470–84.

2. Morrill HJ, Pogue JM, Kaye KS, LaPlante KL. Treatment Options for Carbapenem-Resistant Enterobacteriaceae Infections [Internet]. Open Forum Infectious Diseases. 2015. Available from: http://dx.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.

4. Brennan-Krohn T, Pironti A, Kirby JE. Synergistic Activity of Colistin-Containing Combinations against Colistin-Resistant Enterobacteriaceae. Antimicrob Agents Chemother [Internet]. 2018;62. Available from: http://dx.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.

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 [Internet]. Open Forum Infectious Diseases. 2018. Available from: http://dx.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. Cambridge University Press; 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.

9. World Health Organization. 2019 antibacterial agents in clinical development: an analysis of the antibacterial clinical development pipeline [Internet]. World Health Organization; 2019. Available from: 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 [Internet]. 2019. Available from: 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

1031 Care Facilities. WHO Guidelines Approved by the Guidelines Review Committee
1032 [Internet]. Geneva: World Health Organization; 2017; Available from:
1033 http://eutils.ncbi.nlm.nih.gov/entrez/eutils/elink.fcgi?dbfrom=pubmed&id=29630191&ret
1034 mode=ref&cmd=prlinks

1035 12. for Disease Control C, Prevention, Others. Facility guidance for control of
1036 carbapenem-resistant Enterobacteriaceae (CRE)—November 2015 update CRE toolkit.
1037 Obtenido de: https://www cdc gov/hai/pdfs/cre/cre-guidance-508 pdf Acceso día.
1038 2016;30.

1039 13. Snitkin ES, Zelazny AM, Thomas PJ, Stock F, NISC Comparative Sequencing
1040 Program Group, Henderson DK, et al. Tracking a hospital outbreak of carbapenem-
1041 resistant Klebsiella pneumoniae with whole-genome sequencing. Sci Transl Med.
1042 2012;4:148ra116.

1043 14. Roberts LW, Harris PNA, Forde BM, Ben Zakour NL, Catchpoole E, Stanton-Cook
1044 M, et al. Integrating multiple genomic technologies to investigate an outbreak of
1045 carbapenemase-producing Enterobacter hormaechei. Nat Commun. 2020;11:466.

1046 15. Marsh JW, Mustapha MM, Griffith MP, Evans DR, Ezeonwuka C, Pasculle AW, et
1047 al. Evolution of Outbreak-Causing Carbapenem-Resistant Klebsiella pneumoniae
1048 ST258 at a Tertiary Care Hospital over 8 Years. MBio [Internet]. 2019;10. Available
1049 from: http://dx.doi.org/10.1128/mBio.01945-19

1050 16. Decraene V, Phan HTT, George R, Wyllie DH, Akinremi O, Aiken Z, et al. A Large,
1051 Refractory Nosocomial Outbreak of Klebsiella pneumoniae Carbapenemase-Producing
1052 Escherichia coli Demonstrates Carbapenemase Gene Outbreaks Involving Sink Sites
1053 Require Novel Approaches to Infection Control. Antimicrob Agents Chemother
1054 [Internet]. 2018;62. Available from: http://dx.doi.org/10.1128/AAC.01689-18

1055 17. Otter JA, Burgess P, Davies F, Mookerjee S, Singleton J, Gilchrist M, et al. Counting
1056 the cost of an outbreak of carbapenemase-producing Enterobacteriaceae: an economic
1057 evaluation from a hospital perspective. Clin Microbiol Infect. 2017;23:188–96.

1058 18. Bartsch SM, McKinnell JA, Mueller LE, Miller LG, Gohil SK, Huang SS, et al.
1059 Potential economic burden of carbapenem-resistant Enterobacteriaceae (CRE) in the
1060 United States. Clin Microbiol Infect. 2017;23:48.e9–48.e16.

1061 19. Adeyi OO, Baris E, Jonas OB, Irwin A, Berthe FCJ, Le Gall FG, et al. Drug-resistant
1062 infections: a threat to our economic future. World Bank Group, Washington, DC. 2017;

1063 20. Surveillance of antimicrobial resistance in Europe 2018 [Internet]. European Centre
1064 for Disease Prevention and Control. 2019 [cited 2020 May 22]. Available from:
1065 https://www.ecdc.europa.eu/en/publications-data/surveillance-antimicrobial-resistance-
1066 europe-2018

1067 21. Centers for Disease Control and Prevention (CDC). Vital signs: carbapenem-
1068 resistant Enterobacteriaceae. MMWR Morb Mortal Wkly Rep. 2013;62:165–70.

1069 22. Logan LK, Weinstein RA. The Epidemiology of Carbapenem-Resistant
1070 Enterobacteriaceae: The Impact and Evolution of a Global Menace. J Infect Dis.
1071 2017;215:S28–36.

1072 23. Centers for Disease Control and Prevention (U.S.). Antibiotic resistance threats in
1073 the United States, 2019 [Internet]. Centers for Disease Control and Prevention (U.S.);
1074 2019 Nov. Available from: https://stacks.cdc.gov/view/cdc/82532

1075 24. Zhang Y, Wang Q, Yin Y, Chen H, Jin L, Gu B, et al. Epidemiology of Carbapenem-
1076 Resistant Enterobacteriaceae Infections: Report from the China CRE Network.
1077 Antimicrob Agents Chemother [Internet]. 2018;62. Available from:
1078 http://dx.doi.org/10.1128/AAC.01882-17

1079 25. Tacconelli E, Magrini N, Kahlmeter G, Singh N. Global priority list of antibiotic-
1080 resistant bacteria to guide research, discovery, and development of new antibiotics.
1081 World Health Organization. 2017;27.

1082 26. Woodworth KR, Walters MS, Weiner LM, Edwards J, Brown AC, Huang JY, et al.
1083 Vital Signs: Containment of Novel Multidrug-Resistant Organisms and Resistance
1084 Mechanisms - United States, 2006-2017. MMWR Morb Mortal Wkly Rep. 2018;67:396–
1085 401.

1086 27. Jernigan JA, Hatfield KM, Wolford H, Nelson RE, Olubajo B, Reddy SC, et al.
1087 Multidrug-Resistant Bacterial Infections in US Hospitalized Patients, 2012--2017. N Engl
1088 J Med. Mass Medical Soc; 2020;382:1309–19.

1089 28. Tzouvelekis LS, Markogiannakis A, Psichogiou M, Tassios PT, Daikos GL.
1090 Carbapenemases in Klebsiella pneumoniae and other Enterobacteriaceae: an evolving
1091 crisis of global dimensions. Clin Microbiol Rev. 2012;25:682–707.

1092 29. David S, Reuter S, Harris SR, Glasner C, Feltwell T, Argimon S, et al. Epidemic of
1093 carbapenem-resistant Klebsiella pneumoniae in Europe is driven by nosocomial spread.
1094 Nature microbiology. Nature Publishing Group; 2019;4:1919–29.

1095 30. Cerqueira GC, Earl AM, Ernst CM, Grad YH, Dekker JP, Feldgarden M, et al. Multi-
1096 institute analysis of carbapenem resistance reveals remarkable diversity, unexplained
1097 mechanisms, and limited clonal outbreaks. Proc Natl Acad Sci U S A. 2017;114:1135–
1098 40.

1099 31. Sheppard AE, Stoesser N, Wilson DJ, Sebra R, Kasarskis A, Anson LW, et al.
1100 Nested Russian Doll-Like Genetic Mobility Drives Rapid Dissemination of the
1101 Carbapenem Resistance Gene blaKPC. Antimicrob Agents Chemother. 2016;60:3767–
1102 78.

1103 32. Paschoal RP, Campana EH, Corrêa LL, Montezzi LF, Barrueto LRL, da Silva IR, et
1104 al. Concentration and Variety of Carbapenemase Producers in Recreational Coastal
1105 Waters Showing Distinct Levels of Pollution. Antimicrob Agents Chemother [Internet].
1106 2017;61. Available from: http://dx.doi.org/10.1128/AAC.01963-17

1107 33. Marathe NP, Janzon A, Kotsakis SD, Flach C-F, Razavi M, Berglund F, et al.
1108 Functional metagenomics reveals a novel carbapenem-hydrolyzing mobile beta-
1109 lactamase from Indian river sediments contaminated with antibiotic production waste.
1110 Environ Int. 2018;112:279–86.

1111 34. Gomi R, Matsuda T, Yamamoto M, Chou P-H, Tanaka M, Ichiyama S, et al.
1112 Characteristics of Carbapenemase-Producing Enterobacteriaceae in Wastewater
1113 Revealed by Genomic Analysis. Antimicrob Agents Chemother [Internet]. 2018;62.
1114 Available from: http://dx.doi.org/10.1128/AAC.02501-17

1115 35. Larsson DGJ, Andremont A, Bengtsson-Palme J, Brandt KK, de Roda Husman AM,
1116 Fagerstedt P, et al. Critical knowledge gaps and research needs related to the
1117 environmental dimensions of antibiotic resistance. Environ Int. 2018;117:132–8.

1118 36. Gorrie CL, Mirceta M, Wick RR, Edwards DJ, Thomson NR, Strugnell RA, et al.
1119 Gastrointestinal Carriage Is a Major Reservoir of Klebsiella pneumoniae Infection in
1120 Intensive Care Patients. Clin Infect Dis. 2017;65:208–15.

1121 37. Kelly AM, Mathema B, Larson EL. Carbapenem-resistant Enterobacteriaceae in the
1122 community: a scoping review. Int J Antimicrob Agents. 2017;50:127–34.

1123 38. van Loon K, Voor in t holt AF, Vos MC. A Systematic Review and Meta-analyses of
1124 the Clinical Epidemiology of Carbapenem-Resistant Enterobacteriaceae. Antimicrob
1125 Agents Chemother. 2017;62:1791–1718.

1126 39. Decraene V, Phan HTT, George R. A large, refractory nosocomial outbreak of
1127 Klebsiella pneumoniae carbapenemase-producing Escherichia coli demonstrates
1128 carbapenemase gene outbreaks …. Antimicrob Agents Chemother [Internet]. Am Soc
1129 Microbiol; 2018; Available from: https://aac.asm.org/content/62/12/e01689-18.abstract

1130 40. Weingarten RA, Johnson RC, Conlan S, Ramsburg AM, Dekker JP, Lau AF, et al.
1131 Genomic Analysis of Hospital Plumbing Reveals Diverse Reservoir of Bacterial
1132 Plasmids Conferring Carbapenem Resistance [Internet]. mBio. 2018. Available from:
1133 http://dx.doi.org/10.1128/mbio.02011-17

1134 41. Chng KR, Li C, Bertrand D, Ng AHQ, Kwah JS, Low HM, et al. Cartography of
1135 opportunistic pathogens and antibiotic resistance genes in a tertiary hospital
1136 environment. Nat Med. 2020;26:941–51.

1137 42. Conlan S, Thomas PJ, Deming C, Park M, Lau AF, Dekker JP, et al. Single-
1138 molecule sequencing to track plasmid diversity of hospital-associated carbapenemase-
1139 producing Enterobacteriaceae. Sci Transl Med. 2014;6:254ra126.

1140 43. Mathers AJ, Peirano G, Pitout JDD. The role of epidemic resistance plasmids and
1141 international high-risk clones in the spread of multidrug-resistant Enterobacteriaceae.
1142 Clin Microbiol Rev. 2015;28:565–91.

1143 44. 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 [Internet]. 2017;9. Available from: http://dx.doi.org/10.1126/scitranslmed.aan0093

45. 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. Am Soc Microbiol; 2019;63:e01622–19.

46. 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 [Internet]. 2018;4. Available from: http://dx.doi.org/10.1099/mgen.0.000232

47. 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.

48. 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.

49. Aminov RI. Horizontal gene exchange in environmental microbiota. Front Microbiol. 2011;2:158.

50. 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.

51. Lerminiaux NA, Cameron ADS. Horizontal transfer of antibiotic resistance genes in clinical environments. Can J Microbiol. 2019;65:34–44.

52. 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 [Internet]. Proceedings of the National Academy of Sciences. 2020. p. 25043–54. Available from: http://dx.doi.org/10.1073/pnas.2003407117

53. 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.

54. Stoesser N, Phan HTT, Seale AC, Aiken Z, Thomas S, Smith M, et al. Genomic Epidemiology of Complex, Multispecies, Plasmid-Borne blaKPC Carbapenemase in Enterobacterales in the United Kingdom from 2009 to 2014. Antimicrob Agents Chemother [Internet]. American Society for Microbiology Journals; 2020 [cited 2020 Jun 4];64. Available from: https://aac.asm.org/content/64/5/e02244-19.abstract

55. Mathers AJ, Stoesser N, Sheppard AE, Pankhurst L, Giess A, Yeh AJ, et al.

1181 Klebsiella pneumoniae carbapenemase (KPC)-producing K. pneumoniae at a single
1182 institution: insights into endemicity from whole-genome sequencing. Antimicrob Agents
1183 Chemother. 2015;59:1656–63.

1184 56. Schürch AC, Arredondo-Alonso S, Willems RJL, Goering RV. Whole genome
1185 sequencing options for bacterial strain typing and epidemiologic analysis based on
1186 single nucleotide polymorphism versus gene-by-gene-based approaches. Clin Microbiol
1187 Infect. 2018;24:350–4.

1188 57. Poulou A, Voulgari E, Vrioni G, Koumaki V, Xidopoulos G, Chatzipantazi V, et al.
1189 Outbreak caused by an ertapenem-resistant, CTX-M-15-producing Klebsiella
1190 pneumoniae sequence type 101 clone carrying an OmpK36 porin variant. J Clin
1191 Microbiol. 2013;51:3176–82.

1192 58. Vandecraen J, Chandler M, Aertsen A, Van Houdt R. The impact of insertion
1193 sequences on bacterial genome plasticity and adaptability. Crit Rev Microbiol.
1194 2017;43:709–30.

1195 59. Ernst CM, Braxton JR, Rodriguez-Osorio CA, Zagieboylo AP, Li L, Pironti A, et al.
1196 Adaptive evolution of virulence and persistence in carbapenem-resistant Klebsiella
1197 pneumoniae. Nat Med. 2020;26:705–11.

1198 60. Denamur E, Matic I. Evolution of mutation rates in bacteria [Internet]. Molecular
1199 Microbiology. 2006. p. 820–7. Available from: http://dx.doi.org/10.1111/j.1365-
1200 2958.2006.05150.x

1201 61. Brandt C, Viehweger A, Singh A, Pletz MW, Wibberg D, Kalinowski J, et al.
1202 Assessing genetic diversity and similarity of 435 KPC-carrying plasmids. Sci Rep.
1203 2019;9:11223.

1204 62. Poirel L, Bonnin RA, Nordmann P. Genetic features of the widespread plasmid
1205 coding for the carbapenemase OXA-48. Antimicrob Agents Chemother. 2012;56:559–
1206 62.

1207 63. Wang R, van Dorp L, Shaw LP, Bradley P, Wang Q, Wang X, et al. The global
1208 distribution and spread of the mobilized colistin resistance gene mcr-1. Nat Commun.
1209 2018;9:1179.

1210 64. Evans DR, Griffith MP, Sundermann AJ, Shutt KA, Saul MI, Mustapha MM, et al.
1211 Systematic detection of horizontal gene transfer across genera among multidrug-
1212 resistant bacteria in a single hospital. Elife [Internet]. 2020;9. Available from:
1213 http://dx.doi.org/10.7554/eLife.53886

1214 65. Bradley P, den Bakker HC, Rocha EPC, McVean G, Iqbal Z. Ultrafast search of all
1215 deposited bacterial and viral genomic data. Nat Biotechnol. 2019;37:152–9.

1216 66. Mahillon J, Chandler M. Insertion sequences. Microbiol Mol Biol Rev. 1998;62:725–
1217 74.

67. Grainge I, Jayaram M. The integrase family of recombinase: organization and function of the active site. Mol Microbiol. 1999;33:449–56.

68. 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.

69. Hayes F. Toxins-antitoxins: plasmid maintenance, programmed cell death, and cell cycle arrest. Science. 2003;301:1496–9.

70. 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.

71. 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.

72. Bennett PM. Plasmid encoded antibiotic resistance: acquisition and transfer of antibiotic resistance genes in bacteria. Br J Pharmacol. 2008;153 Suppl 1:S347–57.

73. Carattoli A. Resistance plasmid families in Enterobacteriaceae. Antimicrob Agents Chemother. 2009;53:2227–38.

74. 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 [Internet]. 2016;1. Available from: http://dx.doi.org/10.1128/mSphere.00121-16

75. 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.

76. 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.

77. Popowska M, Krawczyk-Balska A. Broad-host-range IncP-1 plasmids and their resistance potential. Front Microbiol. 2013;4:44.

78. 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.

79. 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.

80. He S, Hickman AB, Varani AM, Siguier P, Chandler M, Dekker JP, et al. Insertion

1253 Sequence IS26 Reorganizes Plasmids in Clinically Isolated Multidrug-Resistant Bacteria
1254 by Replicative Transposition. MBio. 2015;6:e00762.

1255 81. Novais A, Rodrigues C, Branquinho R, Antunes P, Grosso F, Boaventura L, et al.
1256 Spread of an OmpK36-modified ST15 Klebsiella pneumoniae variant during an outbreak
1257 involving multiple carbapenem-resistant Enterobacteriaceae species and clones. Eur J
1258 Clin Microbiol Infect Dis. 2012;31:3057–63.

1259 82. Knopp M, Andersson DI. Amelioration of the Fitness Costs of Antibiotic Resistance
1260 Due To Reduced Outer Membrane Permeability by Upregulation of Alternative Porins.
1261 Mol Biol Evol. 2015;32:3252–63.

1262 83. Vergalli J, Bodrenko IV, Masi M, Moynié L, Acosta-Gutiérrez S, Naismith JH, et al.
1263 Porins and small-molecule translocation across the outer membrane of Gram-negative
1264 bacteria. Nat Rev Microbiol. 2020;18:164–76.

1265 84. Cuzon G, Naas T, Nordmann P. Functional characterization of Tn4401, a Tn3-
1266 based transposon involved in blaKPC gene mobilization. Antimicrob Agents Chemother.
1267 2011;55:5370–3.

1268 85. Naas T, Cuzon G, Villegas M-V, Lartigue M-F, Quinn JP, Nordmann P. Genetic
1269 Structures at the Origin of Acquisition of the β-Lactamase blaKPC Gene. Antimicrob
1270 Agents Chemother. American Society for Microbiology Journals; 2008;52:1257–63.

1271 86. Hall JPJ, Wood AJ, Harrison E, Brockhurst MA. Source-sink plasmid transfer
1272 dynamics maintain gene mobility in soil bacterial communities. Proc Natl Acad Sci U S
1273 A. 2016;113:8260–5.

1274 87. Lopatkin AJ, Meredith HR, Srimani JK, Pfeiffer C, Durrett R, You L. Persistence and
1275 reversal of plasmid-mediated antibiotic resistance. Nat Commun. 2017;8:1689.

1276 88. Gama JA, Zilhão R, Dionisio F. Co-resident plasmids travel together. Plasmid.
1277 2017;93:24–9.

1278 89. Gama JA, Zilhão R, Dionisio F. Conjugation efficiency depends on intra and
1279 intercellular interactions between distinct plasmids: Plasmids promote the immigration of
1280 other plasmids but repress co-colonizing plasmids [Internet]. Plasmid. 2017. p. 6–16.
1281 Available from: http://dx.doi.org/10.1016/j.plasmid.2017.08.003

1282 90. Dionisio F, Zilhão R, Gama JA. Interactions between plasmids and other mobile
1283 genetic elements affect their transmission and persistence. Plasmid. 2019;102:29–36.

1284 91. Lee BY, Bartsch SM, Wong KF, Kim DS, Cao C, Mueller LE, et al. Tracking the
1285 spread of carbapenem-resistant Enterobacteriaceae (CRE) through clinical cultures
1286 alone underestimates the spread of CRE even more than anticipated. Infect Control
1287 Hosp Epidemiol. 2019;40:731–4.

1288 92. CLSI. Performance Standards for Antimicrobial Susceptibility Testing. 30th ed.

1289 Wayne, PA: Clinical and Laboratory Standards Institute; 2020.

1290 93. Smith KP, Kirby JE. Verification of an Automated, Digital Dispensing Platform for At-
1291 Will Broth Microdilution-Based Antimicrobial Susceptibility Testing. J Clin Microbiol.
1292 2016;54:2288–93.

1293 94. Wick RR, Judd LM, Holt KE. Deepbinner: Demultiplexing barcoded Oxford
1294 Nanopore reads with deep convolutional neural networks. PLoS Comput Biol.
1295 2018;14:e1006583.

1296 95. Wick RR, Judd LM, Gorrie CL, Holt KE. Unicycler: Resolving bacterial genome
1297 assemblies from short and long sequencing reads. PLoS Comput Biol.
1298 2017;13:e1005595.

1299 96. Li H. Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics.
1300 2018;34:3094–100.

1301 97. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler
1302 transform. Bioinformatics. 2009;25:1754–60.

1303 98. Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, et al. Pilon: an
1304 integrated tool for comprehensive microbial variant detection and genome assembly
1305 improvement. PLoS One. 2014;9:e112963.

1306 99. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search
1307 tool. J Mol Biol. 1990;215:403–10.

1308 100. Zankari E, Hasman H, Cosentino S, Vestergaard M, Rasmussen S, Lund O, et al.
1309 Identification of acquired antimicrobial resistance genes. J Antimicrob Chemother.
1310 2012;67:2640–4.

1311 101. Bradford PA. Extended-spectrum beta-lactamases in the 21st century:
1312 characterization, epidemiology, and detection of this important resistance threat. Clin
1313 Microbiol Rev. 2001;14:933–51, table of contents.

1314 102. Kurtz S, Phillippy A, Delcher AL, Smoot M, Shumway M, Antonescu C, et al.
1315 Versatile and open software for comparing large genomes. Genome Biol. 2004;5:R12.

1316 103. Siguier P, Perochon J, Lestrade L, Mahillon J, Chandler M. ISfinder: the reference
1317 centre for bacterial insertion sequences. Nucleic Acids Res. 2006;34:D32–6.

1318 104. Hunt M, Mather AE, Sánchez-Busó L, Page AJ, Parkhill J, Keane JA, et al. ARIBA:
1319 rapid antimicrobial resistance genotyping directly from sequencing reads. Microb
1320 Genom. 2017;3:e000131.

1321 105. Robertson J, Nash JHE. MOB-suite: software tools for clustering, reconstruction
1322 and typing of plasmids from draft assemblies. Microb Genom [Internet]. 2018;4.
1323 Available from: http://dx.doi.org/10.1099/mgen.0.000206

106. 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.

107. 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 [Internet]. American Society for Microbiology Journals; 2014 [cited 2020 Jun 11]; Available from: https://aac.asm.org/content/early/2014/04/22/AAC.02412-14?versioned=true

108. 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.

109. 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 [Internet]. 2018;4. Available from: http://dx.doi.org/10.1099/mgen.0.000231

110. Felsenstein J. Inferring phylogenies. Sinauer associates Sunderland, MA; 2004.

111. Paradis E, Schliep K. ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. Bioinformatics. 2019;35:526–8.

112. R Core Team. R: The R Project for Statistical Computing [Internet]. 2020 [cited 2020 Jan 17]. Available from: https://www.r-project.org/

113. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 2004;32:1792–7.

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

115. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics. 2014;30:1312–3.

116. 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. Microbiology Society; 2007;57:81–91.

117. Didelot X, Wilson DJ. ClonalFrameML: efficient inference of recombination in whole bacterial genomes. PLoS Comput Biol. 2015;11:e1004041.

118. 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.

119. 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.

120. 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 [Internet]. 2019;10. Available from: http://dx.doi.org/10.1128/mBio.02899-18

121. Li W, Godzik A. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. Bioinformatics. 2006;22:1658–9.

122. 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.

123. 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.

124. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, et al. BLAST+: architecture and applications. BMC Bioinformatics. 2009;10:421.

125. Saier MH Jr, Reddy VS, Tamang DG, Västermark A. The transporter classification database. Nucleic Acids Res. 2014;42:D251–8.

126. Thomas CM, Thomson NR, Cerdeño-Tárraga AM, Brown CJ, Top EM, Frost LS. Annotation of plasmid genes. Plasmid. 2017;91:61–7.

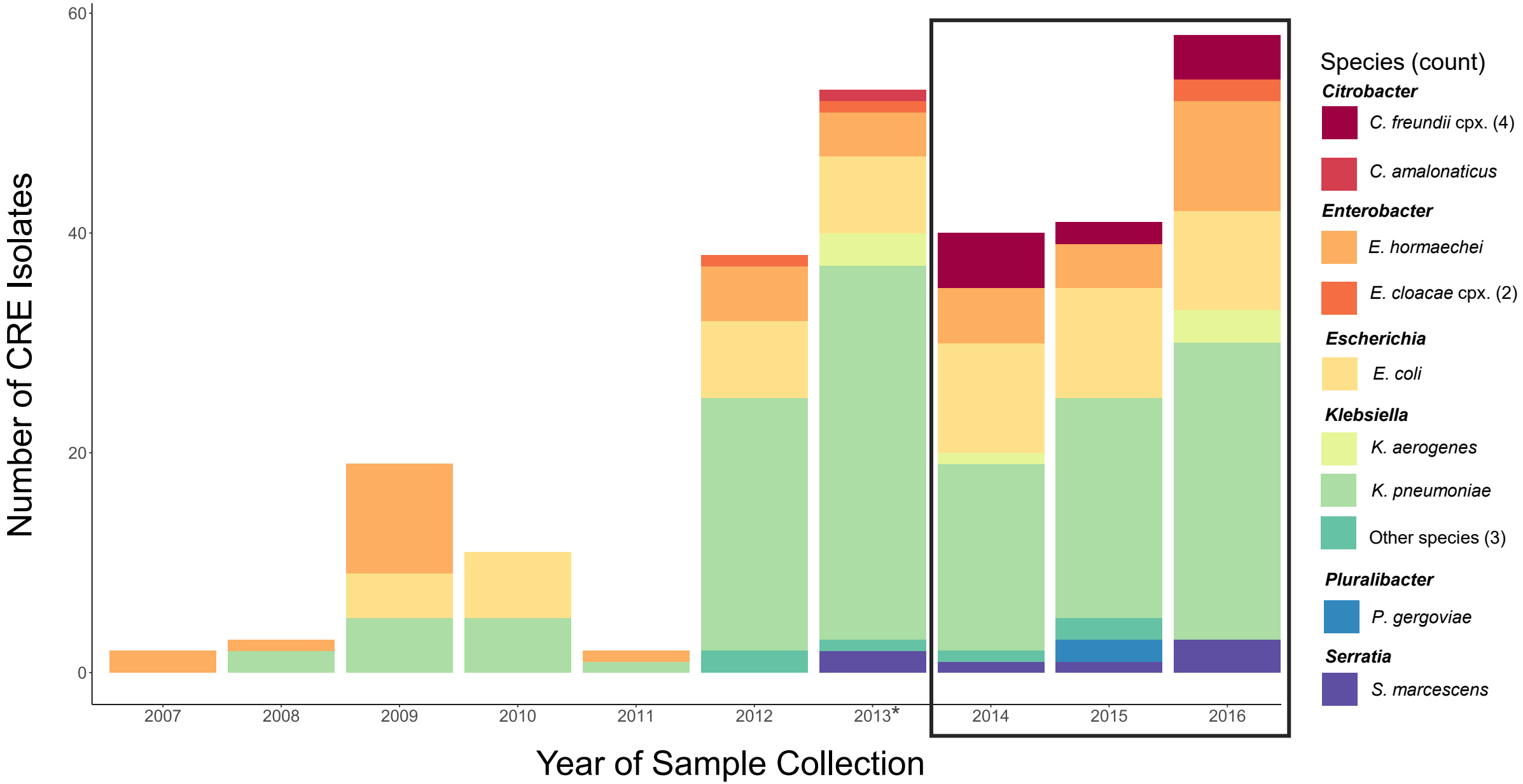
127. Abby SS, Néron B, Ménager H, Touchon M, Rocha EPC. MacSyFinder: a program to mine genomes for molecular systems with an application to CRISPR-Cas systems. PLoS One. 2014;9:e110726.

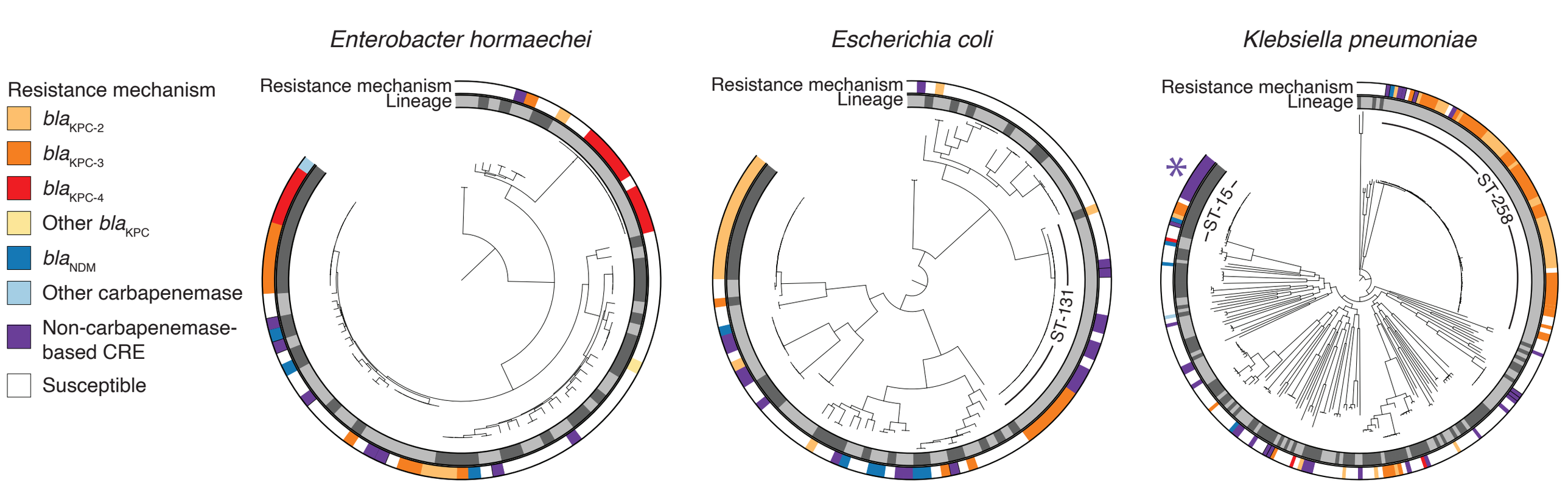
128. Abby SS, Cury J, Guglielmini J, Néron B, Touchon M, Rocha EPC. Identification of protein secretion systems in bacterial genomes. Sci Rep. 2016;6:23080.

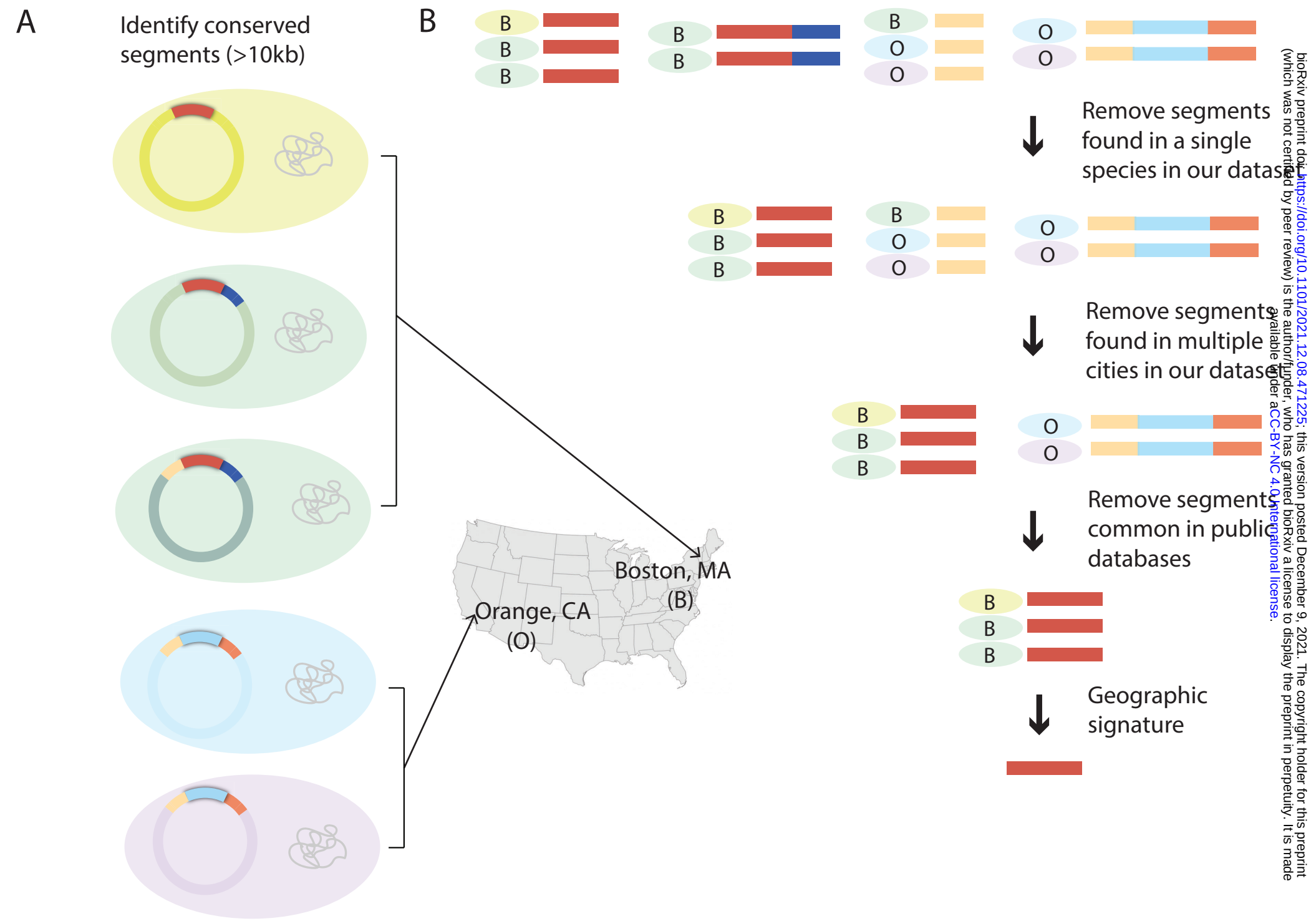
129. Nekrasov SV, Agafonova OV, Belogurova NG, Delver EP, Belogurov AA. Plasmid-encoded antirestriction protein ArdA can discriminate between type I methyltransferase and complete restriction-modification system. J Mol Biol. 2007;365:284–97.

130. 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.

131. Eddy SR. Accelerated Profile HMM Searches. PLoS Comput Biol. 2011;7:e1002195.


Number of CRE Isolates
60
40
20
0
2007
2008
2009
2010
2011
2012
2013*
2014
2015
2016
Year of Sample Collection
Species (count)
Citrobacter
C. freundii cpx. (4)
C. amalonaticus
Enterobacter
E. hormaechei
E. cloacae cpx. (2)
Escherichia
E. coli
Klebsiella
K. aerogenes
K. pneumoniae
Other species (3)
Pluralibacter
P. gergoviae
Serratia
S. marcescens


Enterobacter hormaechei
Escherichia coli
Klebsiella pneumoniae
Resistance mechanism
Lineage
Resistance mechanism
bla KPC-2
bla KPC-3
bla KPC-4
Other bla KPC
bla NDM
Other carbapenemase
Non-carbapenemase-based CRE
Susceptible
ST-131
ST-15
ST-258
*



A Species Phylogenies

B core SNVs within STs

C Signature Carriage by plasmid group

D Year of isolation

Klebsiella pneumoniae

ST-15: BIDMC 123, BIDMC 124
ST-20: MGH 235, MGH 229
ST-1138: MGH 39
ST-309: MGH 83
ST-258: UCI 117, UCI 59, UCI 13, UCI 21

0.001

Escherichia coli

ST-127: UCI 121
ST-648: BIDMC 82, BIDMC 43a, BIDMC 115, BIDMC 15, BIDMC 17A, BIDMC 17B, BIDMC 19A, BIDMC 19B, BIDMC 19C, BIDMC 2B, BIDMC 3, BIDMC 6, BIDMC 9

0.001

Citrobacter portucalensis

BIDMC 108
MGH 105, MGH279, MGH281, MGH281c, MGH283

0.001

Citrobacter freundii cpx.

MGH 103

Enterobacter hormaechei

ST-66: MGH 14, MGH 3
ST-64: BIDMC 33A, MGH 7, BWH 63, MGH 131, BIDMC 94, BWH 37, BIDMC 66, BIDMC 120

0.001

3x

2018 2009 2010 2011 2012 2013 2014 2015 2016 2017

Signatures
- Sig1-CP
- Sig4-CP
- Sig5.1-CP
- Sig5.6-CP
- Sig7-CP
- Sig14-CP
- Plasmid carriage
- Read-based

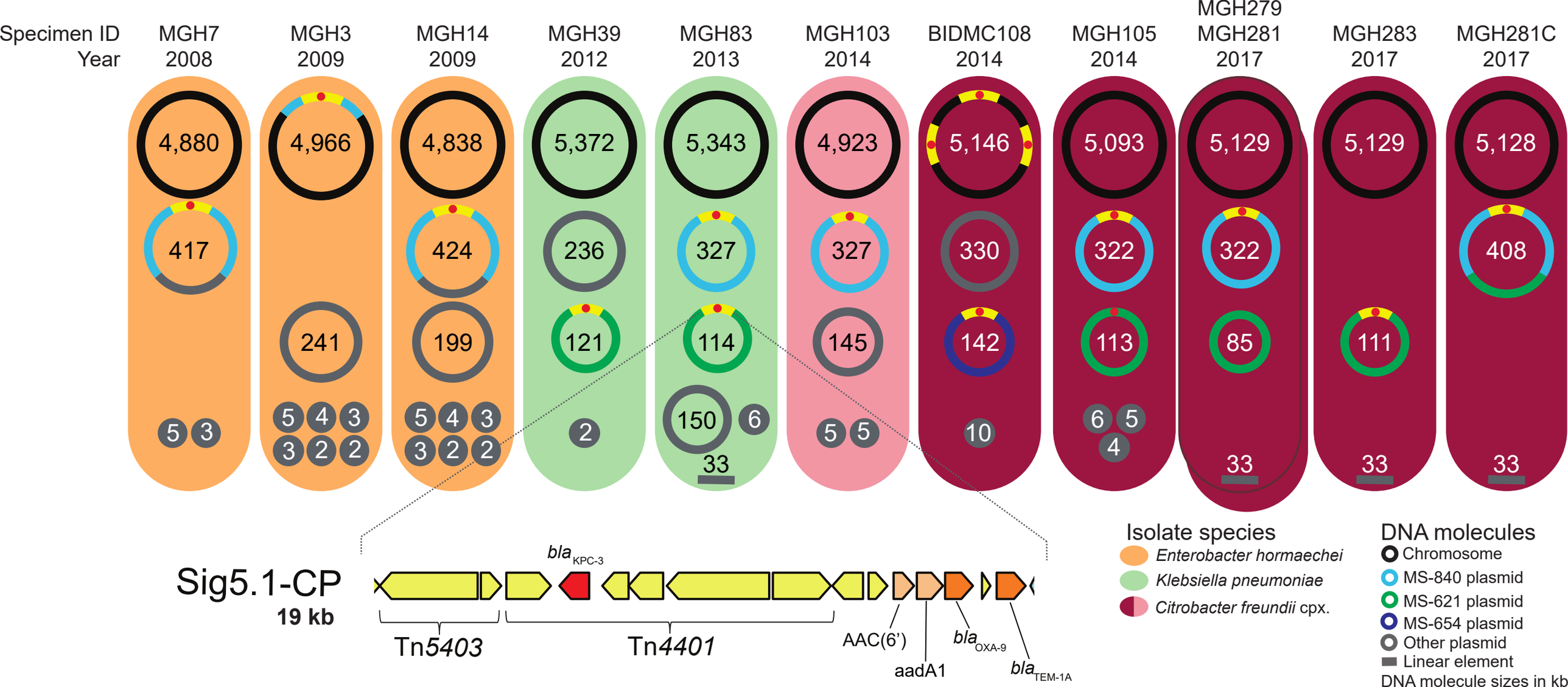
Sequence types (if available)
Same patient isolates

SNVs
>1000
100
10
1
0

1578 1233 1574 840 621 934 429 654
Plasmid Group

Read-based
Chromosomal

Year of isolation


Specimen ID
Year
MGH7
2008
MGH3
2009
MGH14
2009
MGH39
2012
MGH83
2013
MGH103
2014
BIDMC108
2014
MGH105
2014
MGH279
MGH281
2017
MGH283
2017
MGH281C
2017
4,880
417
5 3
4,966
241
5 4 3
3 2 2
4,838
424
199
5 4 3
3 2 2
5,372
236
121
2
5,343
327
114
150
6
33
4,923
327
145
5 5
5,146
330
142
10
5,093
322
113
6 5
4
5,129
322
85
33
5,129
111
33
5,128
408
33
Sig5.1-CP
19 kb
bla KPC-3
Tn5403
Tn4401
AAC(6')
aadA1
bla OXA-9
bla TEM-1A
Isolate species
Enterobacter hormaechei
Klebsiella pneumoniae
Citrobacter freundii cpx.
DNA molecules
Chromosome
MS-840 plasmid
MS-621 plasmid
MS-654 plasmid
Other plasmid
Linear element
DNA molecule sizes in kb