



# Genomic Features Associated with the Degree of Phenotypic Resistance to Carbapenems in Carbapenem-Resistant *Klebsiella pneumoniae*

✉ Zackery P. Bulman,<sup>a</sup> ✉ Fiorella Krapp,<sup>b\*</sup> ✉ Nathan B. Pincus,<sup>c</sup> ✉ Eric Wenzler,<sup>a</sup> Katherine R. Murphy,<sup>c\*</sup> Chao Qi,<sup>d</sup> Egon A. Ozer,<sup>b</sup> ✉ Alan R. Hauser<sup>b,c</sup>

<sup>a</sup>Department of Pharmacy Practice, University of Illinois at Chicago, Chicago, Illinois, USA

<sup>b</sup>Department of Medicine, Division of Infectious Diseases, Northwestern University Feinberg School of Medicine, Chicago, Illinois, USA

<sup>c</sup>Department of Microbiology-Immunology, Northwestern University Feinberg School of Medicine, Chicago, Illinois, USA

<sup>d</sup>Department of Pathology, Northwestern University Feinberg School of Medicine, Chicago, Illinois, USA

Zackery P. Bulman and Fiorella Krapp contributed equally to this article. Author order was determined alphabetically by last name.

Egon A. Ozer and Alan R. Hauser contributed equally to this article.

**ABSTRACT** Carbapenem-resistant *Klebsiella pneumoniae* strains cause severe infections that are difficult to treat. The production of carbapenemases such as the *K. pneumoniae* carbapenemase (KPC) is a common mechanism by which these strains resist killing by the carbapenems. However, the degree of phenotypic carbapenem resistance (MIC) may differ markedly between isolates with similar carbapenemase genes, suggesting that our understanding of the underlying mechanisms of carbapenem resistance remains incomplete. To address this problem, we determined the whole-genome sequences of 166 *K. pneumoniae* clinical isolates resistant to meropenem, imipenem, or ertapenem. Multiple linear regression analysis of this collection of largely *bla*<sub>KPC-3</sub>-containing sequence type 258 (ST258) isolates indicated that *bla*<sub>KPC</sub> copy number and some outer membrane porin gene mutations were associated with higher MICs to carbapenems. A trend toward higher MICs was also observed with those *bla*<sub>KPC</sub> genes carried by the *d* isoform of Tn4401. In contrast, *ompK37* mutations were associated with lower carbapenem MICs, and extended spectrum  $\beta$ -lactamase genes were not associated with higher or lower MICs in carbapenem-resistant *K. pneumoniae*. A machine learning approach based on the whole-genome sequences of these isolates did not result in a substantial improvement in prediction of isolates with high or low MICs. These results build upon previous findings suggesting that multiple factors influence the overall carbapenem resistance levels in carbapenem-resistant *K. pneumoniae* isolates.

**IMPORTANCE** *Klebsiella pneumoniae* can cause severe infections in the blood, urinary tract, and lungs. Resistance to carbapenems in *K. pneumoniae* is an urgent public health threat, since it can make these isolates difficult to treat. While individual contributors to carbapenem resistance in *K. pneumoniae* have been studied, few reports explore their combined effects in clinical isolates. We sequenced 166 clinical carbapenem-resistant *K. pneumoniae* isolates to evaluate the contribution of known genes to carbapenem MICs and to try to identify novel genes associated with higher carbapenem MICs. The *bla*<sub>KPC</sub> copy number and some outer membrane porin gene mutations were associated with higher carbapenem MICs. In contrast, mutations in one specific porin, *ompK37*, were associated with lower carbapenem MICs. Machine learning did not result in a substantial improvement in the prediction of carbapenem resistance nor did it identify novel genes associated with carbapenem resistance. These findings enhance our understanding of the many contributors to carbapenem resistance in *K. pneumoniae*.

**Citation** Bulman ZP, Krapp F, Pincus NB, Wenzler E, Murphy KR, Qi C, Ozer EA, Hauser AR. 2021. Genomic features associated with the degree of phenotypic resistance to carbapenems in carbapenem-resistant *Klebsiella pneumoniae*. mSystems 6:e00194-21. <https://doi.org/10.1128/mSystems.00194-21>.


**Editor** Christopher W. Marshall, Marquette University

**Copyright** © 2021 Bulman et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

Address correspondence to Zackery P. Bulman, [bulman@uic.edu](mailto:bulman@uic.edu), or Fiorella Krapp, [fiorella.krapp@upch.pe](mailto:fiorella.krapp@upch.pe).

\* Present address: Fiorella Krapp, Instituto de Medicina Tropical Alexander von Humboldt, Universidad Peruana Cayetano Heredia, Lima, Peru.

\* Present address: Katherine R. Murphy, Laboratories of Analytical Biology, National Museum of Natural History, Smithsonian Institution, Washington, DC, USA.

 Mutations in outer membrane porin genes and *bla*<sub>KPC</sub> copy number are associated with variations in carbapenem MICs in a large collection of clinical carbapenem-resistant *Klebsiella pneumoniae* isolates.

**Received** 19 February 2021

**Accepted** 19 August 2021

**Published** 14 September 2021

**KEYWORDS** carbapenem, *Klebsiella pneumoniae*, whole-genome sequencing, machine learning, antibiotic resistance

Carbapenem-resistant *Enterobacteriales* (CRE) are urgent public health threats and critical priority pathogens as defined by the World Health Organization (1). CRE are considered a leading threat in part due to their global spread and the limited availability of treatment options. Furthermore, infections caused by CRE are associated with high mortality rates, between ~20 and 50% (2–4). The most common *Enterobacteriales* species that are resistant to carbapenems include *Klebsiella pneumoniae*, *Enterobacter* spp., and *Escherichia coli*. Among these species, *K. pneumoniae* has been found to account for 57% to 74% of all CRE infections (2, 5) and 85% to 90% of bloodstream infections caused by CRE (3, 4). Carbapenem resistance can be caused by a variety of different mechanisms, and the prevalence of each varies by geographic region.

Production of a carbapenemase is the most commonly reported mechanism of carbapenem resistance among *K. pneumoniae* isolates worldwide. Molecular diagnostic approaches have taken advantage of the fact that a vast majority of carbapenem-resistant *Klebsiella pneumoniae* (CR-Kp) isolates harbor a carbapenemase gene such as the *K. pneumoniae* carbapenemase (*bla<sub>KPC</sub>*) or New-Delhi-metallo- $\beta$ -lactamase (*bla<sub>NDM</sub>*) genes. The presence of a carbapenemase gene usually predicts phenotypic carbapenem resistance, though the degree of carbapenem resistance (MIC) among these isolates greatly varies (6). This variability in the carbapenem MIC is likely driven by the presence of other genotypic factors, which also contribute to carbapenem resistance in both carbapenemase- and non-carbapenemase-producing CRE. Some genotypic variations observed in CR-Kp, such as changes in carbapenemase gene copy number (7–10) or variations in transposon isoform and promoter regions (11, 12), modify the expression level of carbapenemase genes and may explain some of the variability in carbapenem MICs. Other factors not associated with carbapenemase expression that may also increase carbapenem MICs include the expression of extended spectrum  $\beta$ -lactamases (ESBLs), which are especially problematic in conjunction with decreased carbapenem membrane permeability related to porin deficits (13, 14). These contributors to carbapenem resistance have largely been investigated in isolation, and few data exist to validate their combined influence on carbapenem MIC in clinical CR-Kp collections. We performed genomic characterization of 166 clinical *K. pneumoniae* isolates to evaluate genetic correlates with levels of carbapenem resistance.

## RESULTS

**CR-Kp isolate characteristics.** A total of 174 *K. pneumoniae* clinical isolates recovered from patients at Northwestern Memorial Hospital between October 2010 and February 2016 were identified as being resistant to at least one carbapenem by the clinical microbiology laboratory (Vitek 2). To confirm resistance, Etests were performed on each isolate. Eight of these isolates were excluded because Etest results did not confirm resistance to any of the carbapenems; thus, 166 CR-Kp isolates were included in the final analyses (see Table S1 in the supplemental material). Within this collection, rates of resistance to ertapenem, imipenem, and meropenem as defined by Etest were 97.6% (162/166), 77.1% (128/166), and 71.1% (118/166), respectively. Collection rates of CR-Kp isolates did not appreciably change between study onset and conclusion, with ~2 to 3 isolates collected per month throughout each year of the study (see Fig. S1). Urine was the most common source of the isolates ( $n = 87$ ), followed by the respiratory tract ( $n = 46$ ), blood ( $n = 13$ ), and other miscellaneous sources ( $n = 20$ ) (Table S1).

Short-read whole-genome sequencing was performed on each isolate. A majority of the CR-Kp isolates (74.1% [123/166]) belonged to sequence type 258 (ST258), while there were 10, 12, and 6 isolates in ST14, ST15, and ST16, respectively (Fig. 1). A majority of the CR-Kp isolates had the capsule locus type KL107 (62.7% [104/166]), while the next most common were categorized as KL2 (6.0% [10/166]), KL24 (3.0% [5/166]),



**FIG 1** Core genome phylogenetic tree of the 166 CR-Kp isolates included in the final analyses and corresponding genotypic/phenotypic isolate information. Sequence type and capsule locus type (if (Continued on next page)

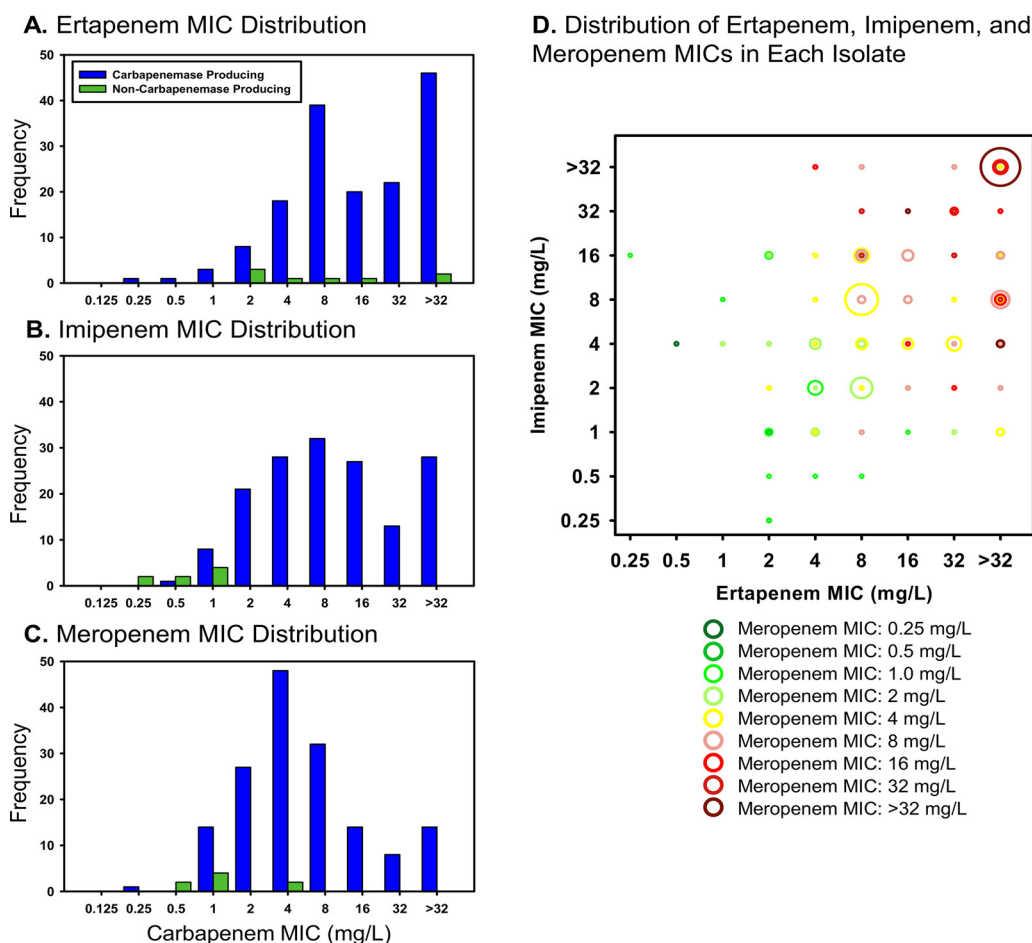
KL112 (3.0% [5/166]), or KL3 (1.8% [3/166]). Of the remaining CR-Kp, 32 (19.3%) did not have a defined KL type, and 7 isolates were categorized into 7 different KL types (Table S1). Examination of the sequences revealed that 158 (95.2%) of the 166 CR-Kp isolates carried at least one carbapenemase gene (Fig. 1). On average, the non-carbapenemase-harboring CR-Kp ( $n = 8$ ) had imipenem and meropenem MICs that were lower than those for the isolates that contained carbapenemases, while the ertapenem MIC distributions were similar for both isolates with and without a carbapenemase gene (Fig. 2A to C). Considering the inra-isolate phenotype, MICs tended to increase for all three carbapenems together (Fig. 2D). However, there were a few isolates that had a disproportionately elevated ertapenem MIC relative to the imipenem and meropenem MICs. The *bla*<sub>KPC-3</sub> gene was the most common carbapenemase gene and was present in 153 isolates (92.2% of total isolates), while the *bla*<sub>KPC-2</sub> gene was detected in 2 isolates (1.2% of total isolates). Of the remaining 3 CR-Kp isolates harboring a carbapenemase gene, 1 contained *bla*<sub>OXA-232</sub>, 1 contained both *bla*<sub>OXA-232</sub> and *bla*<sub>NDM-7</sub>, and 1 contained a novel *bla*<sub>KPC</sub> variant. This variant differed by a single nucleotide (G436A) from *bla*<sub>KPC-3</sub>, which resulted in the substitution of an arginine for a glycine residue at position 146 in the KPC-3 protein (GenBank [AEV55249.1](https://www.ncbi.nlm.nih.gov/nuclot/AEV55249.1) numbering system). This substitution lies in an alpha-helix outside the active site. In addition, each isolate carried between 2 and 5 genes coding for other  $\beta$ -lactamases of limited or extended spectrum.

Our findings indicated that carbapenem resistance in *K. pneumoniae* isolates from our collection site was largely due to *bla*<sub>KPC-3</sub>. However, despite carrying the same carbapenemase gene, these *bla*<sub>KPC-3</sub>-positive isolates differed markedly in their MICs. For each isolate, it was verified that the *bla*<sub>KPC-3</sub> gene sequence was identical. Since the carbapenem MIC could not be predicted based on the presence of a carbapenemase gene alone, additional genotypic-phenotypic relationships were explored.

**Associations between genotypes and carbapenem resistance. (i) Outer membrane porin channel alterations.** Outer membrane porins allow for the passage of  $\beta$ -lactams, including carbapenems, to their binding site in the bacterial periplasmic space. Therefore, we evaluated the association between the *ompK35*, *ompK36*, and *ompK37* genotypes and the carbapenem MICs. For *ompK35*, 35 (21.1%) CR-Kp isolates were predicted to have functional OmpK35 porin channels, whereas 129 (77.7%) had a premature stop codon or insertion sequence in the *ompK35* gene (two isolates could not definitively be associated with a phenotype and were excluded from these analyses). Presence of a premature stop codon or insertion sequence in the *ompK35* gene was associated with a 1-log<sub>2</sub> dilution median increase in the ertapenem, imipenem, and meropenem MICs ( $P < 0.005$ ) (Fig. 3A). Conversely, the *ompK36* genes for a majority of the CR-Kp isolates (135; 81.3%) were identical to wild-type genes and predicted to encode functional OmpK36 proteins, while only 8 isolates (4.8%) had a premature stop codon or were missing the *ompK36* gene. Among the 8 isolates with an absent or truncated *ompK36*, imipenem MICs were significantly lower ( $P = 0.019$ ) (Fig. 3B). However, no significant change in the ertapenem or meropenem MIC was observed ( $P > 0.05$ ). The remaining 23 CR-Kp isolates had mutations in the *ompK36* gene, but function of the porin channel could not be predicted, which excluded them from these porin channel analyses. There were only 5 (3.0%) CR-Kp isolates with both absent or truncated *ompK35* and *ompK36* genes, and their median MICs for ertapenem, imipenem, and meropenem were 16, 8, and 4 mg/liter, respec-

#### FIG 1 Legend (Continued)

predicted) are listed next to each isolate. Carbapenem MICs to meropenem, ertapenem, and imipenem are indicated by the gradient from green (low MIC) to red (high MIC). The presence of carbapenemase or other  $\beta$ -lactamase genes is depicted by blue and orange boxes, respectively. The *bla*<sub>KPC</sub> copy number is presented with a gradient from yellow (low) to red (high). The status of the *ompK35*, *ompK36*, and *ompK37* outer membrane porin channel genes are designated by the dark purple (wild-type allele), light purple (allele with mutations of unclear significance), and gray (absent or truncated allele) boxes. Transposon isoforms are indicated in blue (Tn4401a), yellow (Tn4401b), green (Tn44401d), and orange (Tn4401e). The ertapenem breakpoints for susceptibility and resistance are  $\leq 0.5$  mg/liter and  $\geq 2$  mg/liter, respectively. The imipenem and meropenem breakpoints for susceptibility and resistance are  $\leq 1$  mg/liter and  $\geq 4$  mg/liter, respectively.

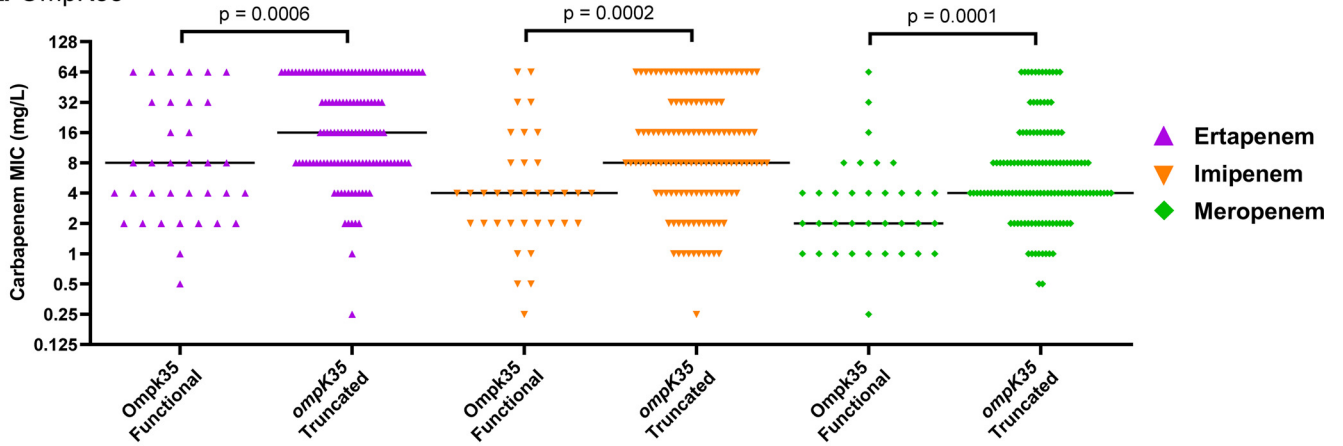


**FIG 2** Distribution of MICs obtained from Etest for ertapenem (A), imipenem (B), and meropenem (C) for CR-Kp isolates that contain carbapenemases (blue) and those without carbapenemases (green). (D) The distribution of phenotypic MICs of all three carbapenems in each CR-Kp isolate. Circles represent isolates with the corresponding ertapenem, imipenem, and meropenem MICs; the color of the circle corresponds to the meropenem MIC, and the size of the circle corresponds to the number of isolates with that phenotype (e.g., the large brown circle in the upper right corner represents 11 CR-Kp isolates that have ertapenem, imipenem, and meropenem MICs of >32, >32, and >32 mg/liter, respectively).

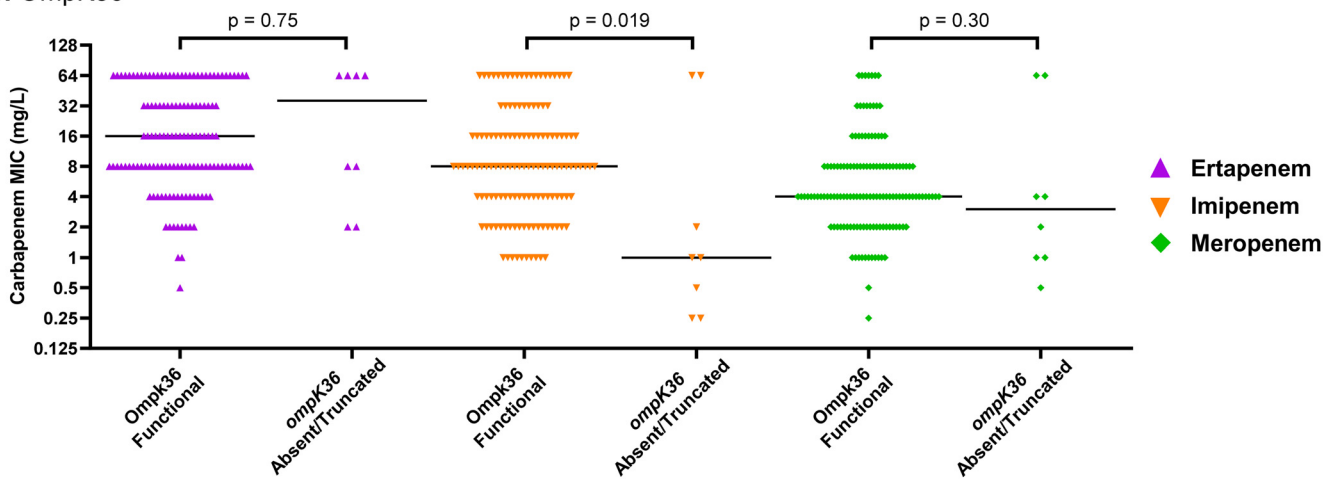
tively. Lastly, none of the isolates harbored *ompK37* genes with premature stop codons, though 37 (22.3%) contained a mutation that did not have a previously defined effect on the function of OmpK37 porin channels. Interestingly, the isolates with *ompK37* mutations had ertapenem, imipenem, and meropenem median MICs that were 2-fold less than that for the CR-Kp with wild-type *ompK37* genes, which was a significant decrease ( $P < 0.005$  for each carbapenem) (Fig. 3C).

**(ii) Presence of ESBL genes.** Although noncarbapenemase  $\beta$ -lactamases such as ESBLs have little capacity to hydrolyze carbapenems, previous data have shown they may contribute to carbapenem resistance in some isolates (15). The relationship between the presence of ESBL genes and carbapenem MICs in the CR-Kp isolates was explored. CR-Kp isolates with ( $n = 100$ ; 60.2%) and without ( $n = 66$ ; 39.8%) an ESBL gene did not have significantly different MICs for any of the carbapenems ( $P > 0.05$ ) (see Fig. S2). The subset of CR-Kp isolates with both an ESBL gene and a truncated *ompK35* allele had significantly higher carbapenem MICs ( $P < 0.005$ ) than isolates with an ESBL gene and predicted functional OmpK35; however, this appeared to be driven by the nonfunctional porin channel, not the cocarriage of the ESBL (see Fig. S3A). Among isolates with a truncated *ompK35* porin gene, the cocarriage of an ESBL did not increase the median carbapenem MICs, and mean carbapenem MICs only increased

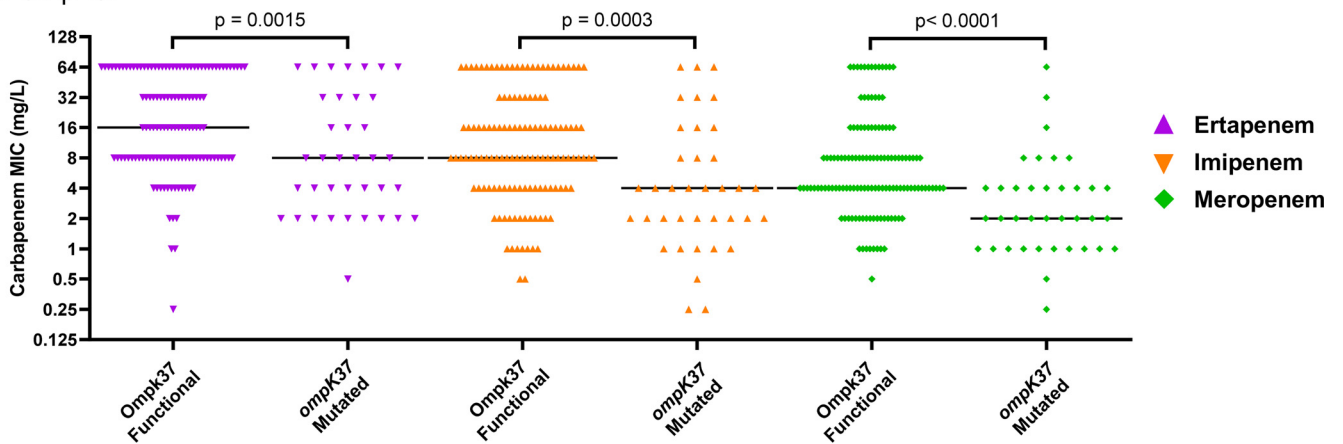
**A. OmpK35**



**B. OmpK36**

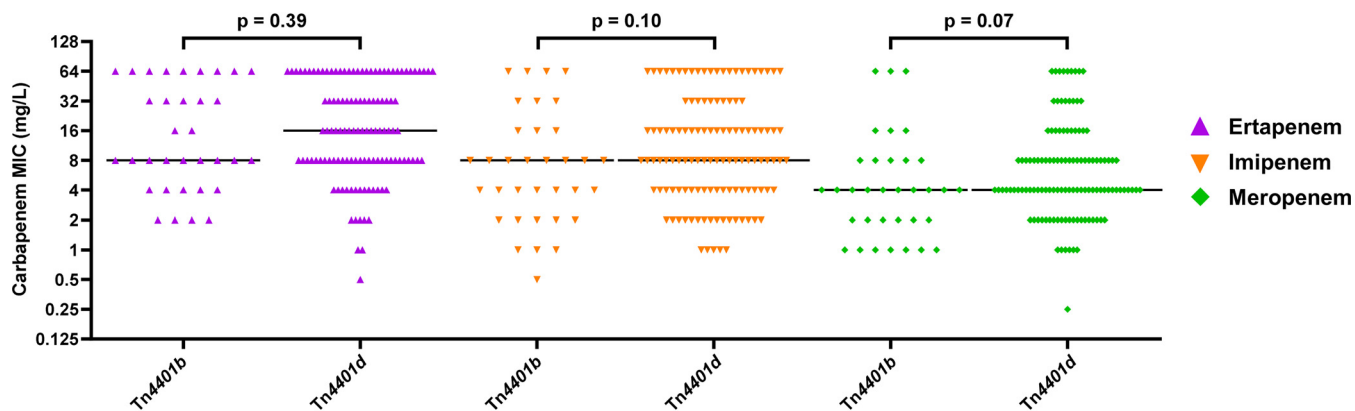


**C. OmpK37**



**FIG 3** Ertapenem (purple), imipenem (orange), and meropenem (green) MICs for each CR-Kp isolate stratified by the predicted function of OmpK35 (A), OmpK36 (B), and OmpK37 (C) outer membrane proteins. Black lines represent the median MIC for each group. MIC values of >32 mg/liter were converted to 64 mg/liter for analysis.

between 1.43 and 6.72 mg/liter. CR-Kp isolates with an ESBL gene and an absent or truncated *ompK36* gene did not have carbapenem MICs that significantly different from isolates with a wild-type *ompK36* gene ( $P > 0.05$ ) or without an ESBL gene ( $P > 0.05$ ), and no trend was noted, though there were only 5 CR-Kp isolates with this



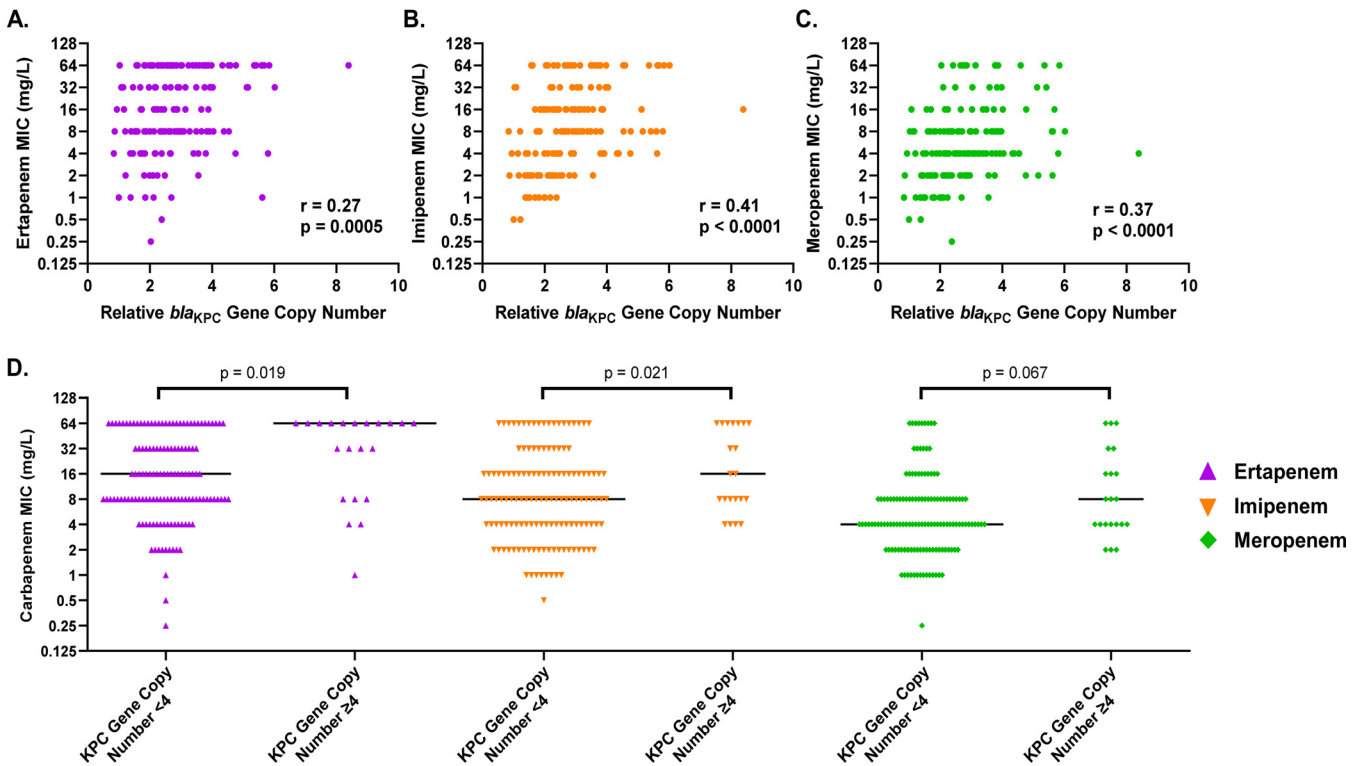
**FIG 4** Ertapenem (purple), imipenem (orange), and meropenem (green) MICs for each CR-Kp isolate stratified by the type of transposon on which the *bla*<sub>KPC</sub> gene was located. Black lines represent the median MIC for each group. MIC values of >32 mg/liter were converted to 64 mg/liter for analysis.

genotype (3.0%) (Fig. S3B). Thus, ESBLs may play at most a minor role to increase carbapenem MICs within CR-Kp isolates that coharbor carbapenemase genes.

**(iii) Transposon harboring the *bla*<sub>KPC</sub> gene.** Different Tn4401 transposon isoforms that may affect *bla*<sub>KPC</sub> expression are found in some CR-Kps (11, 12). Of the 156 CR-Kp isolates with *bla*<sub>KPC</sub>, a majority carried the *bla*<sub>KPC</sub> gene in the Tn4401d transposon isoform ( $n = 119$ ; 76.3%); one of these had one single nucleotide variation (SNV) (C6941A), while the others were identical to the Tn4401d reference sequence (7, 12). Next most common was the Tn4401b transposon isoform ( $n = 34$ ; 21.8%), though only 3 isolates had sequences identical to the reference (accession CP017937); 1 isolate had 1 SNV in the P2 promoter (G6909A), while 30 isolates had 5 SNVs, 4 of which were in the P1 promoter region (T7148G, T7152G, A7153T, C7154A, and C7158A). There was also 1 isolate with the Tn4401a transposon, 1 isolate with the Tn4401e transposon, and 1 isolate in which the transposon fell at the end of a contig and could not be typed. The isolates that contained Tn4401d had carbapenem MICs that were on average 1.9 to 5.2 mg/liter higher than isolates with Tn4401b (median MICs of 0 or 1- $\log_2$  dilution difference), which was trending toward being a significant increase for imipenem and meropenem. (Fig. 4).

**(iv) Relative copy number of the *bla*<sub>KPC</sub> gene.** Higher *bla*<sub>KPC</sub> copy number can lead to increased expression of the KPC enzyme and is associated with elevated carbapenem MICs (7). In the subgroup of CR-Kp isolates that harbored a *bla*<sub>KPC</sub> gene (*bla*<sub>KPC-2</sub> or *bla*<sub>KPC-3</sub>), a higher *bla*<sub>KPC</sub> relative copy number was significantly correlated with higher MICs for ertapenem, imipenem, and meropenem ( $P < 0.005$ ) (Fig. 5A to C). Though significant, the positive correlation was weak to moderate. The correlation was strongest for imipenem ( $r = 0.41$ ), followed by meropenem ( $r = 0.37$ ), and was the weakest for ertapenem MICs ( $r = 0.27$ ). Compared to that for CR-Kp isolates estimated to have <4 copies of the *bla*<sub>KPC</sub> gene ( $n = 135$ ; 86.5%), isolates with  $\geq 4$  copies ( $n = 21$ ; 13.5%) had 4-fold higher ertapenem ( $P = 0.019$ ) and 2-fold higher imipenem median MICs ( $P = 0.021$ ), while the meropenem median MIC was trending toward being significantly higher ( $P = 0.067$ ) (Fig. 5D). These findings demonstrate that isolates with increased copy numbers of the *bla*<sub>KPC</sub> gene tended to have elevated carbapenem MICs. In contrast to the correlation between MIC and the *bla*<sub>KPC</sub> copy number, there was less of a correlation between the *bla*<sub>ESBL</sub> copy number and the ertapenem ( $r = 0.24$ ) or imipenem ( $r = 0.23$ ) MIC; the correlation between *bla*<sub>ESBL</sub> copy number and the meropenem MIC ( $r = 0.38$ ) was similar to the correlation between meropenem MIC and *bla*<sub>KPC</sub> copy number (data not shown).

The increased copy numbers of *bla*<sub>KPC</sub> in some isolates could be the result of one of two mechanisms: duplication of the *bla*<sub>KPC</sub> gene itself, which has been reported (16, 17), or increased copy number of the entire plasmid carrying *bla*<sub>KPC</sub> (9). In nearly every isolate, the coverage ratio of the *bla*<sub>KPC</sub> gene to the other 5 genes was between 80% and 120%, with most between 90% and 110%. From this we conclude that the increased *bla*<sub>KPC</sub> copy



**FIG 5** Relationship between the relative *bla*<sub>KPC</sub> copy number in each CR-Kp isolate and its ertapenem (A), imipenem (B), and meropenem (C) MIC, where “*r*” represents the Spearman’s correlation coefficient. (D) Comparison of carbapenem MICs for isolates with fewer than 4 copies of the *bla*<sub>KPC</sub> gene and those with ≥4 copies of the *bla*<sub>KPC</sub> gene. Black lines represent the median MIC for each group. MIC values of >32 mg/liter were converted to 64 mg/liter for analysis.

number observed in most of the carbapenemase-producing isolates was likely due to increased copy number of the plasmid carrying *bla*<sub>KPC</sub>.

**(v) Multiple linear regression analysis.** Multiple linear regression analysis identified CR-Kp genes that were associated with significant changes in the MIC for each carbapenem evaluated. Final models all incorporated some established contributors to carbapenem resistance, including *bla*<sub>KPC</sub> copy number, *bla*<sub>OXA-232</sub> presence, and *ompK35* or *ompK36* alleles (Table 1). While the copy number of the ESBL genes was not associated with carbapenem MICs, each additional copy of the *bla*<sub>KPC</sub> gene was associated with a carbapenem MIC increase of 3.37 to 4.85 mg/liter. The presence of *bla*<sub>OXA-232</sub> was associated with MIC increases of 44.36 to 65.15 mg/liter. The meropenem model also showed that the presence of the wild-type *ompK37* gene was associated with significant increases in meropenem MIC (7.45 mg/liter). In general, models were only able to explain a small proportion of the carbapenem MIC variability among CR-Kp isolates (adjusted *R*<sup>2</sup> <0.3).

**(vi) Machine learning analysis.** A support vector classifier machine learning analysis was performed on the 123 ST258 isolates to predict the level of carbapenem resistance. Isolates were classified as highly resistant (MIC > 8 mg/liter) or low/intermediately resistant (MIC ≤ 8 mg/liter) to meropenem or imipenem. First, a machine learning approach was used to estimate how well high versus low resistance could be predicted based on curated features previously reported to influence carbapenem resistance: β-lactamase presence/absence, outer membrane porin phenotype, transposon isoform, and *bla*<sub>KPC</sub> copy number. This showed a mean F1 score of 0.54 for imipenem and 0.30 for meropenem in nested cross-validation, which indicates relatively low predictive performances (Fig. 6A and B). We next sought to determine whether novel genetic factors not included in our curated feature set influenced resistance to carbapenems. For this purpose, we used whole-genome sequence features, including all core genome SNV loci (3,122) and all unique accessory genome elements ≥200 bp



**TABLE 1** Multiple linear regression analysis demonstrating the influence of detected genotypic predictors on the MICs of ertapenem, imipenem, and meropenem against 166 CR-Kp isolates

Carbapenem	Molecular predictor	$\beta$	95% confidence interval	P value
Ertapenem <sup>a</sup>	<i>bla</i> <sub>KPC</sub> copy no.	4.81	2.08 to 7.54	0.001
	<i>bla</i> <sub>OXA-232</sub>	44.36	9.85 to 78.87	0.12
	Truncated <i>ompK35</i>	10.07	1.28 to 18.86	0.025
	Wild-type <i>ompK36</i>	-10.63	-20.26 to -1.00	0.031
Imipenem <sup>b</sup>	<i>bla</i> <sub>KPC</sub> copy no.	4.85	2.52-7.18	<0.001
	<i>bla</i> <sub>OXA-232</sub>	57.43	28.17 to 86.70	<0.001
	Wild-type <i>ompK35</i>	-9.09	-16.76 to -1.42	0.021
Meropenem <sup>c</sup>	<i>bla</i> <sub>KPC</sub> copy no.	3.37	1.52 to 5.21	<0.001
	<i>bla</i> <sub>OXA-232</sub>	65.15	42.75 to 87.54	<0.001
	Absent or truncated <i>ompK36</i>	14.10	2.63 to 25.58	0.016
	Wild-type <i>ompK37</i>	7.45	1.56 to 13.34	0.013

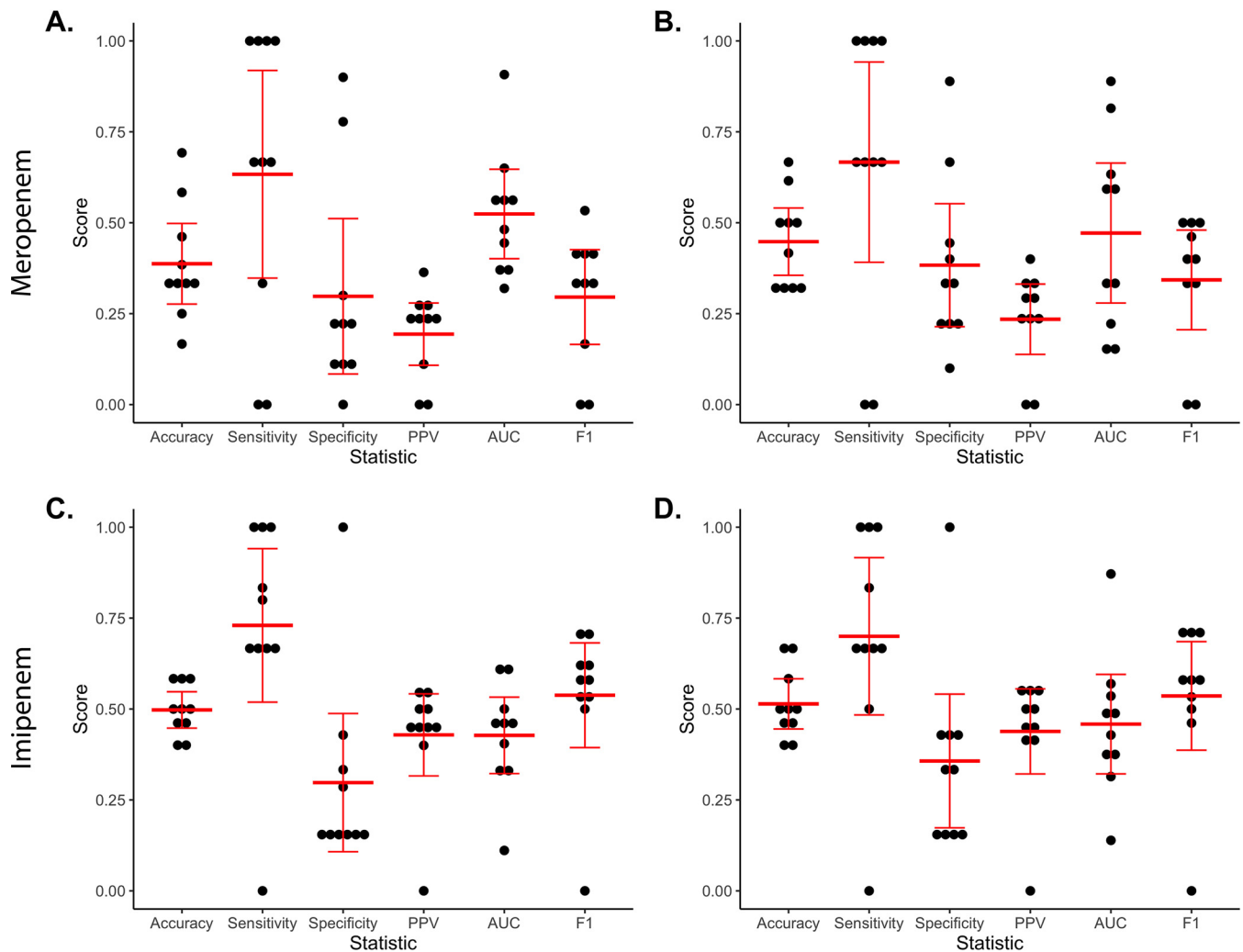
<sup>a</sup>aR<sup>2</sup> = 0.120.<sup>b</sup>aR<sup>2</sup> = 0.162.<sup>c</sup>aR<sup>2</sup> = 0.210.

(372) within the 123 ST258 isolates. The machine learning approach here showed mean F1 scores of 0.54 for imipenem and 0.34 for meropenem in nested cross-validation, which were similar to those of the curated models (Fig. 6C and D). In summary, the support vector classifier analysis, like the regression analysis, suggested that only a small proportion of the carbapenem MIC variability in the ST258 isolates was associated with previously reported factors. Broadening the feature set to include the whole genome did not improve predictions. The use of alternative machine learning algorithms (random forest, l2-regularized logistic regression, and elastic net logistic regression) yielded similar results (data not shown).

To assess whether the poor performance of the machine learning models was the result of an insufficient sample size, we conducted learning curve analyses. For both imipenem and meropenem and for both the curated and whole-genome sequence analyses, F1 scores between the training and cross-validation sets of the support vector classifier analysis had largely converged, and cross-validation scores did not substantially increase as training set sample size increased (see Fig. S4). These findings suggest that the poor performance of the support vector classifier approach is unlikely to be caused by an insufficient sample size.

## DISCUSSION

CR-Kp causes serious infections that are associated with high mortality rates (2, 3). Increasing MICs to carbapenems are associated with worse outcomes, even among carbapenemase-producing *K. pneumoniae* isolates (18, 19), but the reasons for this variability are incompletely understood. Furthermore, the ability to predict carbapenem MICs in CR-Kp may also be important in some clinical settings. For example, in low- and middle-income countries where new antibiotics with activity against CR-Kp are limited or unavailable, identification of CR-Kp isolates with lower MICs to carbapenems ( $\leq 8$  mg/liter) may enable the use of high-dose, prolonged infusion, carbapenem-containing combination regimens (20, 21). Various studies have shown how the presence of  $\beta$ -lactamases, expression levels of carbapenemase or ESBL genes, and permeability deficits, among others, can influence carbapenem activity in *K. pneumoniae*, but few studies have evaluated the combined effects of these genetic features on carbapenem MICs in clinical CR-Kp isolates. In the present study, multiple linear regression was used to identify significant predictors of ertapenem, imipenem, and meropenem MICs in a collection of 166 clinical CR-Kp isolates. Although the final models were only able to explain a portion of the phenotypic variability, many important and biologically plausible genotypes remained in



**FIG 6** Nested cross-validation performance of the support vector classifier machine learning approach to predict high versus low carbapenem resistance among ST258 *K. pneumoniae* isolates ( $n = 123$ ). Levels of meropenem and imipenem resistance were predicted using models trained on curated (A and C, respectively) and whole-genome sequence features (B and D, respectively). Accuracy, sensitivity, specificity, positive predictive value (PPV), area under the receiver operating characteristic curve (AUC), and F1 scores were determined for models built in each cross-validation fold, with mean and 95% confidence intervals displayed in red.

the models, such as mutations in the outer membrane porin genes,  $bla_{KPC}$  copy number, and the presence of  $bla_{OXA-232}$ .

Bioinformatic approaches are increasingly being used to predict antibiotic resistance from genomic content (22–26), and several studies have targeted *K. pneumoniae* and carbapenems for this purpose. Avershina and colleagues (27) applied neural network methods to  $\beta$ -lactamase  $k$ -mers in unassembled genomes to predict *K. pneumoniae* susceptibility versus resistance to carbapenems with accuracies >90%. Jaillard and colleagues used an adaptive cluster lasso algorithm to predict carbapenem susceptibility versus intermediate/resistance in *K. pneumoniae* with 91% to 93% accuracy (28). Nguyen and colleagues developed XGBoost-based models to predict carbapenem MICs for a collection of largely carbapenem-susceptible *K. pneumoniae* (29). The presence of  $bla_{KPC}$  and genes involved in recombination (recombinase and transposase genes) were most highly correlated with the carbapenem MICs in their model, which was accurately able to predict carbapenem MICs for >90% of isolates. Intriguingly, this approach yielded models with accuracies that remained as high as 63% to 71%, even when known antibiotic resistance genes were excluded and the interrogated gene set was limited to 100 randomly selected core genes (30). The authors speculated that their models were keying in on genomically widespread compensatory or epistatic

genetic changes associated with antibiotic resistance. It is difficult to compare the results of these studies to those of our study because of their differing goals. The aims of each of these studies was to predict which *K. pneumoniae* isolates were susceptible and which were resistant to carbapenems. In contrast, our study focused on only carbapenem-resistant isolates and evaluated genomic features that discriminated between isolates with very high MICs and those with more modest MICs, an arguably more difficult task that may account for the poorer performance of our models.

One genotype that was associated with significant increases in the carbapenem MICs was the presence of truncated outer membrane porin channel genes. The outer membrane porins in *K. pneumoniae*, including OmpK35 and OmpK36, permit passage of the hydrophilic carbapenems through to their periplasmic targets. Thus, loss of function of one or multiple porin channels can diminish the activity of the carbapenems and cause low-level carbapenem resistance in the absence of carbapenemase or ESBL production (31, 32). In our cohort, *ompK35* truncation or absence was common among CR-Kp isolates (>75%), while *ompK36* truncation or absence was only noted in <5% of isolates. This is consistent with other studies that have found that *ompK35* disruption is more common among clinical isolates (33, 34). Despite its previously described association with increased carbapenem MICs, disruption of *ompK36* may be less common among clinical isolates due to fitness costs (33, 35, 36). In the present study, isolates with an absent or truncated *ompK36* gene had significantly lower imipenem MICs than isolates with OmpK36 porins that were predicted to be functional; however, these results may be an artifact of the small number of isolates with an absent or truncated *ompK36* gene in our collection. Although putative nonfunctional OmpK35 is commonly identified in CR-Kp, its role in carbapenem resistance has not been completely defined; for isolates without an ESBL or carbapenemase, some studies show that OmpK35 is important for carbapenem uptake (37), while others found no significant increase in carbapenem MIC after the loss of OmpK35 (36) if OmpK36 is still intact. The impact of having a nonfunctional OmpK35 in isolates that express an ESBL or carbapenemase is also not completely defined. Some studies have shown that nonfunctional OmpK35 porins in *K. pneumoniae* can contribute to carbapenem resistance for isolates that express KPC (33) or ESBL enzymes (34). Interestingly, we found that ertapenem, imipenem, and meropenem MICs were each significantly increased in isolates with truncated *ompK35* genes, most of which contained a carbapenemase gene and an intact *ompK36* gene. This further suggests that *ompK35* mutations may in fact confer increased carbapenem resistance in isolates that express a carbapenemase.

Another outer membrane porin, OmpK37, is often expressed in isolates that are deficient in OmpK35 and OmpK36 but is usually downregulated when these porins are expressed (38). OmpK37 enables the passage of carbapenems, though it may not as readily permit other  $\beta$ -lactams to cross the membrane (31, 38, 39). Herein, we found that CR-Kp isolates with a mutation in *ompK37* of unclear functional significance had lower carbapenem MICs than isolates with a wild-type copy of *ompK37*. For example, the wild-type *ompK37* allele was found to increase the meropenem MIC in its final linear regression model. There are at least two potential explanations for this association between *ompK37* mutations and lower carbapenem MICs. First, mutations that impact the structure of the porin may enhance permeability to the  $\beta$ -lactams and could be responsible for the observed differences in carbapenem MICs (31, 40). Previous studies found that nonsynonymous mutations in the constriction region of the OmpK35 homolog OmpF increased or decreased the permeability to various cephalosporins and ampicillin (41, 42). Others have suggested that variations in the external loops of the porin channels in *K. pneumoniae* may impact permeability (31). Mutations in other regions of the porins, including the  $\beta$ -strands, have been shown to alter the electrostatic field of the pore and modify  $\beta$ -lactam passage in clinical isolates (43). In the present study, the two mutations observed in the CR-Kp isolates with lower carbapenem MICs were in  $\beta$ -strands that cross the outer membrane. Further investigation is required to determine if the mutations we observed in *ompK37* can enhance carbapenem

penetration. A second potential explanation, which is not necessarily mutually exclusive, is that isolates with these mutations are more likely to have wild-type *ompK35* and/or *ompK36* genes due to selective pressures that require CR-Kp to maintain some functional porin channels for adequate transport of nutrients. In the present study, isolates with mutations in *ompK37* were approximately 80% more likely to coharbor functional *ompK35* genes but less likely to harbor functional *ompK36* genes than isolates with wild-type *ompK37*. Thus, isolates with mutations in *ompK37* may on average also have lower carbapenem MICs due to increased drug passage through alternative porin channels. Future studies should also include functional characterization and assess expression of the porin channels to confirm or refine these proposed hypotheses.

In the present study, most CR-Kp isolates had a carbapenemase gene, and *bla*<sub>KPC</sub> was the most common. The *bla*<sub>KPC</sub> gene copy number was significantly correlated with the carbapenem MICs and was retained in all three multivariate models. However, this explained only a small portion of the variation in carbapenem MICs. Previous studies have likewise found that *bla*<sub>KPC</sub> gene expression does not always highly correlate with the gene copy number and is also influenced by differences in the upstream region of the gene and transcriptional start site (7, 44). Outer membrane porin mutations in combination with higher KPC expression may further increase the carbapenem MICs (7). The combination of outer membrane porin mutations and elevated *bla*<sub>KPC</sub> gene copy number has also been observed in KPC-Kp isolates that are resistant to meropenem-vaborbactam (45) and ceftazidime-avibactam (8). In contrast to this correlation with *bla*<sub>KPC</sub> gene copy number, the presence of a *bla*<sub>ESBL</sub> gene did not significantly increase carbapenem MICs in our collection of CR-Kp isolates. Previous studies have shown that *bla*<sub>ESBL</sub> genes such as *bla*<sub>CTX-M</sub> or *bla*<sub>SHV-5</sub> in combination with outer membrane porin deficits can cause carbapenem resistance in *K. pneumoniae* isolates that do not coharbor carbapenemase genes (13, 14). Similarly, *bla*<sub>ESBL</sub> gene duplications have been shown to increase carbapenem MICs in isolates without a carbapenemase (46). However, the present study suggests, for the first time, that neither the presence of an *bla*<sub>ESBL</sub> nor the *bla*<sub>ESBL</sub> gene copy number appreciably increases the carbapenem MICs in isolates that coharbor carbapenemase genes, which is supported by the absence of these genomic features in the final multivariate models.

The *bla*<sub>KPC</sub> gene is typically found in the Tn4401 transposon of *Enterobacteriales* isolates from the United States and Europe, while other transposons such as the Tn3-Tn4401 chimera are more widespread in China (11, 47). Isoforms of Tn4401 have also been characterized with variations in the promoters that are present upstream of the *bla*<sub>KPC</sub> gene, leading to differences in *bla*<sub>KPC</sub> gene expression and  $\beta$ -lactam resistance. For example, the presence of the *bla*<sub>KPC</sub> gene in the Tn4401a variant (P1 and P2 promoters) has been shown to lead to higher *bla*<sub>KPC</sub> expression than Tn4401b (P1, P2, and P3 promoters), thereby increasing carbapenem MICs 4- to 8-fold (11, 12). The most common Tn4401 transposon variants in the *bla*<sub>KPC</sub>-harboring CR-Kp isolates in the present study were Tn4401b and Tn4401d (P1 and P2 promoters). Naas et al. previously showed that stronger *bla*<sub>KPC</sub> expression occurs when only the P1 and P2 promoters are present, and this can subsequently increase carbapenem MICs (12). Thus, since Tn4401d has only the P1 and P2 promoters, while Tn4401b has all three promoters, isolates with Tn4401d may have higher *bla*<sub>KPC</sub> expression than those with Tn4401b. The phenotypic data from our clinical *bla*<sub>KPC</sub>-harboring isolates corresponded relatively well with the findings from this previous study, as we observed that isolates with Tn4401d had higher mean carbapenem MICs than isolates with Tn4401b. However, transposon isoforms were not found to be significantly associated with the carbapenem MICs in the multivariate models.

Regression analysis indicated that factors known to impact carbapenem MICs accounted for a relatively small portion of the variability we observed in our cohort. To determine whether as-yet-unidentified genetic features were also associated with increased or decreased carbapenem MICs, we employed an agnostic machine learning approach that interrogated the entire genomes of the ST258 isolates. Surprisingly, incorporating core genome SNVs and accessory genomic elements failed to better predict highly resistant strains

than when solely using known factors. Several explanations may account for this. Variables not included in our whole-genome feature set, such as *bla*<sub>KPC</sub> copy number, epigenetic changes, or subtle differences between homologous accessory genes, may play a large role in dictating carbapenem MICs. The majority of highly resistant isolates may each have unique MIC-altering SNV changes not shared with other isolates, making the machine learning algorithm “blind” to them. Alternatively, variability in MIC measurements (e.g., intrinsic Etest variation, instability of MICs over time, or irregularities caused by heteroresistance) may have obfuscated associations between degree of carbapenem resistance and genetic features. In future studies, these possibilities will need to be examined to gain a more thorough understanding of how CR-Kp resists killing by carbapenems.

One important limitation that may have impacted univariate and multivariate analyses of our study was that carbapenem MICs were only evaluated to a maximum concentration of 32 mg/liter. Thus, any isolate with an MIC >32 mg/liter was assigned an MIC of 64 mg/liter for statistical evaluations even though these CR-Kp isolates may in actuality have differing levels of carbapenem resistance. More precisely defined carbapenem MICs may affect the results of these analyses. Another potential limitation is that the isolates were collected from a single institution, which may impact the ability to extrapolate our findings to other medical centers with differing CR-Kp genotypes. However, the general characteristics of our CR-Kp collection appear mostly representative of resistant isolates found by other investigators, at least in the United States (2, 48). An additional limitation is that short-read sequencing was used to genomically characterize all isolates. For this reason, we were not able to define plasmid sequences, and novel plasmid insertions, deletions, and rearrangements capable of affecting *bla*<sub>KPC</sub> expression may have been missed. In future studies, long-read sequencing will be performed on these isolates to better characterize the plasmids they carry. Lastly, the carbapenem MICs were collected via Etest, which is generally reliable for determining categorical resistance in carbapenemase-producing *K. pneumoniae* but may not define the actual MIC as accurately as broth microdilution (49). Thus, our findings may not necessarily be generalizable to MIC data derived from alternative testing methods. However, since Etest MICs were used for all analyses, relative interisolate MIC differences are likely accurate at least for this testing method.

In conclusion, we collected and conducted an in-depth molecular analysis of 166 CR-Kp isolates. Multiple linear regression models identified genotypes that were associated with significant changes in carbapenem MICs. The models support the roles of several biologically plausible genotypes in contributing to carbapenem MICs among clinical CR-Kp isolates, including *bla*<sub>KPC</sub> copy number and outer membrane porins. In contrast, the presence of ESBL genes or their increased copy number were not associated with carbapenem MICs, suggesting they do not increase carbapenem resistance among isolates that already harbor a carbapenemase gene. Lastly, our data also showed a trend toward increased carbapenem MICs among isolates with Tn4401d. Machine learning analyses did not substantially improve the ability to predict isolates with high or low MICs based on their sequence. Future studies are necessary to refine the models and fully define the role of these genes in carbapenem resistance.

## MATERIALS AND METHODS

**Clinical isolate collection and carbapenem resistance testing.** A retrospective analysis was performed on 174 CR-Kp isolates that were recovered as a part of routine clinical care. The isolates were collected from consecutive patients at a tertiary care hospital in the United States (Northwestern Memorial Hospital, Chicago, IL) between October 2010 and February 2016. Carbapenem resistance was defined as displaying resistance to  $\geq 1$  carbapenem following MIC testing with a Vitek 2 system and using CLSI breakpoints (50, 51). CR-Kp isolates were archived and stored at  $-80^{\circ}\text{C}$  by the hospital microbiology laboratory as part of the routine institutional infection control policy. Isolates were subsequently recovered by streaking onto Luria-Bertani (LB) solid medium and incubating at  $37^{\circ}\text{C}$  overnight. To ensure purity, a single colony was selected from the plate for cryopreservation. Isolates were regrown on LB agar, and MICs to meropenem, ertapenem, and imipenem were measured using Etest strips (bioMérieux) according to the manufacturer's specifications to confirm carbapenem resistance. Isolates that were not resistant to at least 1 carbapenem following the Etest were excluded from the study. Observed MIC values were rounded up to the next highest log<sub>2</sub> dilution per the manufacturer's specifications. Off-scale values

(i.e., MICs >32 mg/liter) were converted to 64 mg/liter for statistical analyses. Only the first isolate per patient was included. Isolates recovered from rectal swabs were also excluded from this study. This study was approved by the Northwestern University Institutional Review Board (IRB). A waiver of informed consent was obtained due to the retrospective nature of the study. No diagnostic or treatment decisions were affected by this study.

**Whole-genome sequencing.** Isolates were grown overnight in LB broth with shaking at 37°C. DNA extraction was performed using a Maxwell 16 instrument with cell DNA purification kits (Promega Corporation, Madison, WI) according to the manufacturer's instructions. Libraries were prepared using the Illumina Nextera XT kit and sequenced using an Illumina HiSeq (150-bp paired-end reads) platform (Illumina, Inc., San Diego, CA). Sequences were trimmed using Trimmomatic v0.32 (51), and then *de novo* assembly was performed with SPAdes 3.9.1. Contigs were removed if they were shorter than 200 bp or had a mean fold coverage of  $<5\times$  per base.

**CR-Kp genome analysis.** *In silico* multilocus sequence typing (MLST) was performed by analyzing each assembled sequence with MLST 1.8 (Center for Genomic Epidemiology) (52). The capsule locus type was determined by analyzing the genome of each isolate in Kaptive, and calls with confidence levels of "good" or better were reported (53). The antimicrobial resistance genes were identified by aligning the assembled CR-Kp genomes against the Comprehensive Antibiotic Resistance Database (54) and ResFinder v3.1 (55). Whole-genome sequencing reads were aligned to the ST258 reference genome KPNIH1 (NZ\_CP008827.1) using bwa v0.7.15. This sequence was chosen as the reference as it is a publicly available closed ST258 genome. Variants relative to the reference sequence were determined using bcftools v1.9 using a minimum base quality of 25, minimum mapping quality of 30, and a haploid model as previously described (56). Only variants with a minimum SNV quality score of 200, minimum read consensus of 75%, and a minimum read depth of 5 with at least 1 read in each direction were kept. A genome alignment was created from variant positions relative to the KPNIH1 reference genome using bcftools\_filter\_and\_align ([https://github.com/egonozer/bcftools\\_filter/blob/master/bcftools\\_filter\\_and\\_align.pl](https://github.com/egonozer/bcftools_filter/blob/master/bcftools_filter_and_align.pl)). FastTree v2.1.9 was used to construct the maximum likelihood phylogenetic tree from the resulting alignment using the generalized time-reversible model with branch rescaling to optimize the Gamma20 likelihood. Phylogenetic trees were visualized using FigTree v1.4.4 and R v4.0.2 with ggtree package v2.2.4. To examine whether the *bla*<sub>SHV</sub> (ESBL), *bla*<sub>CTX-M</sub>, and *bla*<sub>KPC</sub> gene copy numbers contributed to elevated carbapenem MICs, the average number of  $\beta$ -lactamase sequencing reads were normalized to the total number of core genome sequencing reads for each isolate using sam\_to\_coverage ([https://github.com/egonozer/sam\\_to\\_coverage/blob/master/sam\\_to\\_coverage.pl](https://github.com/egonozer/sam_to_coverage/blob/master/sam_to_coverage.pl)); the ratio was used to predict the copy numbers of the  $\beta$ -lactamase genes relative to the core genome copy number. To distinguish between duplication of the *bla*<sub>KPC</sub> gene itself or increased copy number of the entire plasmid carrying *bla*<sub>KPC</sub>, we compared the average read coverage of *bla*<sub>KPC</sub> to the average read coverages of the other 5 genes on the *bla*<sub>KPC-3</sub>-carrying pRYCKPC3.1 plasmid (NC\_019151.1). The isoforms of the transposons harboring the *bla*<sub>KPC</sub> genes were analyzed by aligning the sequences upstream of the *bla*<sub>KPC</sub> genes to published reference sequences (12). Mutations in the transposons were annotated based on the Tn4401b reference sequence from pKPC\_UVA01 (accession CP017937), as previously described (57). The sequences of the outer membrane porin channel genes from each isolate were examined to predict if they encoded functional porin channels. Isolates with *ompK35*, *ompK36*, and *ompK37* alleles that were absent or had a premature stop codon or insertion sequence in the gene were considered to produce an absent or truncated porin channel, whereas isolates with genes that were identical to the *ompK35* (accession AJ011501), *ompK36* (accession JX310551, Z33506, AP006725, and JOOW00000000), and *ompK37* (accession KC534871 and NZ\_LR130541) reference sequences from isolates with a known functional OmpK protein were predicted to have functional porin channels. Isolates with nonsynonymous mutations in porin channel genes that did not lead to a premature stop codon were separately categorized as having unknown function.

**Data and regression analyses.** To determine if variations in genotype (predicted OmpK function, ESBL presence, and transposon isoform) were associated with differences in phenotype (carbapenem MIC), univariate analyses were performed and data were compared via the Mann-Whitney U test. Nonparametric continuous data are presented as the median, and a *P* value of  $\leq 0.05$  was considered statistically significant. To determine the strength and direction of the relationship between the relative *bla*<sub>KPC</sub> copy number and the carbapenem MICs, Spearman's correlation analysis was performed. The continuous variable *bla*<sub>KPC</sub> copy number was dichotomized at  $<4$  and  $\geq 4$  to compare the isolates with the highest MIC *bla*<sub>KPC</sub> copy number ( $>1$  standard deviation above mean) to the rest of the CR-Kp isolates based on the observed distribution among the sample (mean  $\pm$  standard deviation [SD]:  $2.70 \pm 1.38$ ).

To determine the combined influence of various clinically relevant genes on the ertapenem, imipenem, and meropenem MICs, multiple linear regression via a stepwise approach was utilized to evaluate the incremental change in the phenotypic carbapenem MIC associated with one or more identified genotypic markers of interest. Carbapenem MICs were modeled as observed from the Etest (i.e., 2, 4, 8, and 16 mg/liter). Any genomic element identified in  $>1$  isolate was eligible for model inclusion based on prevalence, biologic plausibility, and/or previous studies. Thus, the candidate variables (genotype) initially considered were the following: *bla*<sub>AmpC</sub> presence, *bla*<sub>CTX-M</sub> presence, *bla*<sub>CTX-M</sub> copy number, *bla*<sub>KPC</sub> presence, *bla*<sub>KPC</sub> copy number, *bla*<sub>FOX-5</sub> presence, *bla*<sub>OXA-232</sub> presence, *bla*<sub>OXA</sub> (noncarbapenemase) presence, *bla*<sub>SHV</sub> (non-ESBL) presence, *bla*<sub>SHV</sub> (ESBL) presence, *bla*<sub>SHV</sub> (ESBL) copy number, *bla*<sub>TEM</sub> presence, *bla*<sub>EB-9</sub> presence, predicted function of *ompK35/36/37*, and transposon isoform. A *P* value of  $\leq 0.05$  was considered statistically significant in the final model. Model performance was evaluated using adjusted *R*<sup>2</sup> (*aR*<sup>2</sup>). Collinearity was assessed via tolerance and variance inflation factor. Outliers, highly influential values, and leverage points were not removed since they were true observed values. Multiple linear regression analyses were performed using SPSS version 26 (IBM Corporation, Armonk, NY).

**Machine learning analysis.** Machine learning analyses were performed on the ST258 CR-Kp isolates ( $n = 123$ ) using the sci-kit learn library (v0.23.2) (<https://www.jmlr.org/papers/v12/pedregosa11a.html>) in Python v3.6.12 following the pipeline described by Pincus et al. (58) For the curated analysis,  $\beta$ -lactamase presence/absence, outer membrane porin phenotype, transposon isoform, and  $bla_{KPC}$  copy number were used as features and converted to binary variables. For example,  $bla_{KPC}$  copy number was considered “high” if  $\geq 4$  copies were present and “low” if  $< 4$  copies were present. Features present in all or none of the isolates (e.g.,  $bla_{NDM}$ ) were removed. For the whole-genome sequence analysis, all core genome (defined as sequences present in  $\geq 95\%$  of the 123 ST258 isolates) SNVs and accessory genomic elements (AGEs) of  $\geq 200$  bp were used as features. To define core genome SNVs, alignments of ST258 genomes to the KPNIH1 reference genome were created as described above, noncore loci were masked, and invariant sites were removed. AGEs in the ST258 genomes were defined using Spine v0.32, AGEnt v0.31, and ClustAGE v0.8 (59, 60), grouping together perfectly correlated accessory sequences. As part of the analysis pipeline, SNV loci were converted to binary features through one-hot-encoding.

The ability of machine learning models built using curated or whole-genome sequence features to predict whether isolates had high resistance ( $MIC > 8$  mg/liter) or low resistance ( $MIC \leq 8$  mg/liter) to imipenem or meropenem was estimated using a nested cross-validation approach. The support vector classifier algorithm was used with C, gamma, and class\_weight included as hyperparameters. F1 scores (0, lowest predictive power; 1, highest predictive power) were used as the primary scoring metric. In the outer 10-fold stratified cross-validation loop, training data for each fold were used to build a model that was evaluated against the held-out test set. Grid-search cross-validation (the inner loop) was used to select hyperparameters for each model. In addition to the F1 score, accuracy, sensitivity, specificity, positive predictive value, and area under the receiver operating characteristic curve were also calculated for each model. Learning curves were created to examine the impact of increasing training sample size on training and cross-validation F1 score. For each combination of features (curated and whole-genome sequence) and labels (meropenem and imipenem resistance), the data set was split into training and cross-validation folds through 10-fold stratified cross-validation. Subsets of examples from each training fold were used to train models with hyperparameters selected via grid search cross-validation. Training and cross-validation F1 scores were then calculated at each training sample size. Alternative machine learning algorithms (random forest, l2-regularized logistic regression, and elastic net logistic regression) were used following a similar approach. Code for the machine learning analyses are available on GitHub ([https://github.com/nathanpincus/CRKP\\_Resistance\\_Prediction](https://github.com/nathanpincus/CRKP_Resistance_Prediction)).

**Data availability.** Sequencing reads and assembled genomes have been submitted to GenBank under BioProject [PRJNA395086](https://www.ncbi.nlm.nih.gov/bioproject/PRJNA395086).

## SUPPLEMENTAL MATERIAL

Supplemental material is available online only.

**FIG S1**, PDF file, 0.1 MB.

**FIG S2**, PDF file, 0.1 MB.

**FIG S3**, PDF file, 0.2 MB.

**FIG S4**, PDF file, 0.1 MB.

**TABLE S1**, PDF file, 0.1 MB.

## ACKNOWLEDGMENTS

The project was funded in part by the Chicago Biomedical Consortium with support from the Searle Funds at The Chicago Community Trust (to Z.P.B. and A.R.H.). This project was also funded in part by the National Institutes of Health (NIH) under grants R01AI118257, U19AI135964, K24AI04831, and R21AI129167 (all to A.R.H.), R01AI148560 (to Z.P.B.), and T32GM008152 (to N.B.P.).

The content of the manuscript is solely the responsibility of the authors and does not necessarily represent the official views of the NIH.

We declare no competing interests.

Conceptualization, F.K. and A.R.H.; Formal Analysis, Z.P.B., F.K., N.B.P., E.W., E.A.O., and A.R.H.; Investigation, Z.P.B., F.K., K.R.M., and C.Q.; Resources, A.R.H.; Data Curation, Z.P.B. and F.K.; Writing—Original Draft, Z.P.B. and F.K.; Writing—Reviewing & Editing, Z.P.B., F.K., N.B.P., E.W., K.R.M., C.Q., E.A.O., and A.R.H.; Visualization, Z.P.B., F.K., N.B.P., and E.A.O.; Supervision, A.R.H.; Funding Acquisition, Z.P.B., N.B.P., and A.R.H.

## REFERENCES

- World Health Organization. 2017. Prioritization of pathogens to guide discovery, research and development of new antibiotics for drug-resistant bacterial infections, including tuberculosis. (WHO/EMP/IAU/2017.12). License CC BY-NC-SA 3.0 IGO. World Health Organization, Geneva, Switzerland.
- van Duin D, Arias CA, Komarow L, Chen L, Hanson BM, Weston G, Cober E, Garner OB, Jacob JT, Satlin MJ, Fries BC, Garcia-Diaz J, Doi Y, Dhar S, Kaye KS, Earley M, Hujer AM, Hujer KM, Domitrovic TN, Shropshire WC, Dinh A, Manca C, Luterbach CL, Wang M, Paterson DL, Banerjee R, Patel R, Evans

- S, Hill C, Arias R, Chambers HF, Fowler VG, Kreiswirth BN, Bonomo RA, Multi-Drug Resistant Organism Network Investigators. 2020. Molecular and clinical epidemiology of carbapenem-resistant *Enterobacteriales* in the USA (CRACKLE-2): a prospective cohort study. *Lancet Infect Dis* 20: 731–741. [https://doi.org/10.1016/S1473-3099\(19\)30755-8](https://doi.org/10.1016/S1473-3099(19)30755-8).
3. Gutiérrez-Gutiérrez B, Salamanca E, de Cueto M, Hsueh P-R, Viale P, Paño-Pardo JR, Venditti M, Tumbarello M, Daikos G, Cantón R, Doi Y, Tuon FF, Karaiskos I, Pérez-Nadales E, Schwaber MJ, Azap ÖK, Souli M, Roilides E, Pournaras S, Akova M, Pérez F, Bermejo J, Oliver A, Almela M, Lowman W, Almirante B, Bonomo RA, Carmeli Y, Paterson DL, Pascual A, REIPI/ESGBIS/INCREMENT Investigators. 2017. Effect of appropriate combination therapy on mortality of patients with bloodstream infections due to carbapenemase-producing *Enterobacteriaceae* (INCREMENT): a retrospective cohort study. *Lancet Infect Dis* 17:726–734. [https://doi.org/10.1016/S1473-3099\(17\)30228-1](https://doi.org/10.1016/S1473-3099(17)30228-1).
  4. Satlin MJ, Chen L, Patel G, Gomez-Simmonds A, Weston G, Kim AC, Seo SK, Rosenthal ME, Sperber SJ, Jenkins SG, Hamula CL, Uhlemann AC, Levi MH, Fries BC, Tang YW, Juretschko S, Rojzman AD, Hong T, Mathema B, Jacobs MR, Walsh TJ, Bonomo RA, Kreiswirth BN. 2017. Multicenter clinical and molecular epidemiological analysis of bacteremia due to carbapenem-resistant *Enterobacteriaceae* (CRE) in the CRE epicenter of the United States. *Antimicrob Agents Chemother* 61:e02349-16. <https://doi.org/10.1128/AAC.02349-16>.
  5. Zhang Y, Wang Q, Yin Y, Chen H, Jin L, Gu B, Xie L, Yang C, Ma X, Li H, Li W, Zhang X, Liao K, Man S, Wang S, Wen H, Li B, Guo Z, Tian J, Pei F, Liu L, Zhang L, Zou C, Hu T, Cai J, Yang H, Huang J, Jia X, Huang W, Cao B, Wang H. 2018. Epidemiology of carbapenem-resistant *Enterobacteriaceae* infections: report from the China CRE network. *Antimicrob Agents Chemother* 62:e01882-17. <https://doi.org/10.1128/AAC.01882-17>.
  6. Tamma PD, Huang Y, Opene BN, Simner PJ. 2016. Determining the optimal carbapenem MIC that distinguishes carbapenemase-producing and non-carbapenemase-producing carbapenem-resistant *Enterobacteriaceae*. *Antimicrob Agents Chemother* 60:6425–6429. <https://doi.org/10.1128/AAC.00838-16>.
  7. Kitchel B, Rasheed JK, Endimiani A, Hujer AM, Anderson KF, Bonomo RA, Patel JB. 2010. Genetic factors associated with elevated carbapenem resistance in KPC-producing *Klebsiella pneumoniae*. *Antimicrob Agents Chemother* 54:4201–4207. <https://doi.org/10.1128/AAC.00008-10>.
  8. Gaibani P, Re MC, Campoli C, Viale PL, Ambretti S. 2020. Bloodstream infection caused by KPC-producing *Klebsiella pneumoniae* resistant to ceftazidime/avibactam: epidemiology and genomic characterization. *Clin Microbiol Infect* 26:516.e1–516.e4. <https://doi.org/10.1016/j.cmi.2019.11.011>.
  9. Shen Z, Zhang H, Gao Q, Qin J, Zhang C, Zhu J, Li M. 2020. Increased plasmid copy number contributes to the elevated carbapenem resistance in OXA-232-producing *Klebsiella pneumoniae*. *Microb Drug Resist* 26: 561–568. <https://doi.org/10.1089/mdr.2018.0407>.
  10. Barwa R, Shaaban M. 2017. Molecular characterization of *Klebsiella pneumoniae* clinical isolates with elevated resistance to carbapenems. *Open Microbiol J* 11:152–159. <https://doi.org/10.2174/1874285801711010152>.
  11. Huang J, Hu X, Zhao Y, Shi Y, Ding H, Wu R, Zhao Z, Ji J. 2019. Comparative analysis of *bla*<sub>KPC</sub> expression in Tn4401 transposons and the Tn3-Tn4401 chimera. *Antimicrob Agents Chemother* 63:e02434-18. <https://doi.org/10.1128/AAC.02434-18>.
  12. Naas T, Cuzon G, Truong HV, Nordmann P. 2012. Role of ISKpn7 and deletions in *bla*<sub>KPC</sub> gene expression. *Antimicrob Agents Chemother* 56: 4753–4759. <https://doi.org/10.1128/AAC.00334-12>.
  13. Hamzaoui Z, Ocampo-Sosa A, Fernandez Martinez M, Landolsi S, Ferjani S, Maamar E, Saidani M, Slim A, Martinez-Martinez L, Boutiba-Ben Boubaker I. 2018. Role of association of *OmpK35* and *OmpK36* alteration and *bla*<sub>ESBL</sub> and/or *bla*<sub>AmpC</sub> genes in conferring carbapenem resistance among non-carbapenemase-producing *Klebsiella pneumoniae*. *Int J Antimicrob Agents* 52:898–905. <https://doi.org/10.1016/j.ijantimicag.2018.03.020>.
  14. Mena A, Plasencia V, García L, Hidalgo O, Ayestarán JI, Alberti S, Borrell N, Pérez JL, Oliver A. 2006. Characterization of a large outbreak by CTX-M-1-producing *Klebsiella pneumoniae* and mechanisms leading to *in vivo* carbapenem resistance development. *J Clin Microbiol* 44:2831–2837. <https://doi.org/10.1128/JCM.00418-06>.
  15. Paterson DL, Bonomo RA. 2005. Extended-spectrum beta-lactamases: a clinical update. *Clin Microbiol Rev* 18:657–686. <https://doi.org/10.1128/CMR.18.4.657-686.2005>.
  16. Conlan S, Thomas PJ, Deming C, Park M, Lau AF, Dekker JP, Snitkin ES, Clark TA, Luong K, Song Y, Tsai YC, Boitano M, Dayal J, Brooks SY, Schmidt B, Young AC, Thomas JW, Bouffard GG, Blakesley RW, Mullikin JC, Korlach J, Henderson DK, Frank KM, Palmore TN, Segre JA, NISC Comparative Sequencing Program. 2014. Single-molecule sequencing to track plasmid diversity of hospital-associated carbapenemase-producing *Enterobacteriaceae*. *Sci Transl Med* 6:254ra126. <https://doi.org/10.1126/scitranslmed.3009845>.
  17. Fortini D, Villa L, Feudi C, Pires J, Bonura C, Mammina C, Endimiani A, Carattoli A. 2016. Double copies of *bla*<sub>KPC-3::Tn4401a</sub> on an IncX3 Plasmid in *Klebsiella pneumoniae* Successful clone ST512 from Italy. *Antimicrob Agents Chemother* 60:646–649. <https://doi.org/10.1128/AAC.01886-15>.
  18. O'Donnell JN, Rhodes NJ, Biehle LR, Esterly JS, Patel TS, McLaughlin MM, Hirsch EB. 2020. Assessment of mortality stratified by meropenem minimum inhibitory concentration in patients with *Enterobacteriaceae* bacteraemia: a patient-level analysis of published data. *Int J Antimicrob Agents* 55:105849. <https://doi.org/10.1016/j.ijantimicag.2019.11.006>.
  19. Pea F, Della Siega P, Cojutti P, Sartor A, Crapin M, Scarpato C, Bassetti M. 2017. Might real-time pharmacokinetic/pharmacodynamic optimisation of high-dose continuous-infusion meropenem improve clinical cure in infections caused by KPC-producing *Klebsiella pneumoniae*? *Int J Antimicrob Agents* 49:255–258. <https://doi.org/10.1016/j.ijantimicag.2016.10.018>.
  20. Stewardson AJ, Marimuthu K, Sengupta S, Allignol A, El-Bouseary M, Carvalho MJ, Hassan B, Delgado-Ramirez MA, Arora A, Bagga R, Owusu-Ofori AK, Ovosi JO, Aliyu S, Saad H, Kanj SS, Khanal B, Bhattacharai B, Saha SK, Uddin K, Barman P, Sharma L, El-Banna T, Zahra R, Saleemi MA, Kaur A, Iregbu K, Uwaezuoke NS, Abi Hanna P, Feghali R, Correa AL, Munera MI, Le TAT, Tran TTN, Phukan C, Phukan C, Valderrama-Beltrán SL, Alvarez-Moreno C, Walsh TR, Harbarth S. 2019. Effect of carbapenem resistance on outcomes of bloodstream infection caused by *Enterobacteriaceae* in low-income and middle-income countries (PANORAMA): a multinational prospective cohort study. *Lancet Infect Dis* 19:601–610. [https://doi.org/10.1016/S1473-3099\(18\)30792-8](https://doi.org/10.1016/S1473-3099(18)30792-8).
  21. Sheu CC, Chang YT, Lin SY, Chen YH, Hsueh PR. 2019. Infections caused by carbapenem-resistant *Enterobacteriaceae*: an update on therapeutic options. *Front Microbiol* 10:80. <https://doi.org/10.3389/fmicb.2019.00080>.
  22. Nguyen M, Long SW, McDermott PF, Olsen RJ, Olson R, Stevens RL, Tyson GH, Zhao S, Davis JJ. 2019. Using machine learning to predict antimicrobial MICs and associated genomic features for nontyphoidal *Salmonella*. *J Clin Microbiol* 57:e01260-18. <https://doi.org/10.1128/JCM.01260-18>.
  23. Drouin A, Giguère S, Déraspe M, Marchand M, Tyers M, Loo VG, Bourgault AM, Laviolette F, Corbeil J. 2016. Predictive computational phenotyping and biomarker discovery using reference-free genome comparisons. *BMC Genomics* 17:754. <https://doi.org/10.1186/s12864-016-2889-6>.
  24. Hyun JC, Kavvas ES, Monk JM, Palsson BO. 2020. Machine learning with random subspace ensembles identifies antimicrobial resistance determinants from pan-genomes of three pathogens. *PLoS Comput Biol* 16: e1007608. <https://doi.org/10.1371/journal.pcbi.1007608>.
  25. Aytan-Aktug D, Clausen P, Bortolaia V, Aarestrup FM, Lund O. 2020. Prediction of acquired antimicrobial resistance for multiple bacterial species using neural networks. *mSystems* 5:e00774-19. <https://doi.org/10.1128/mSystems.00774-19>.
  26. Kavvas ES, Catoi E, Mih N, Yurkovich JT, Seif Y, Dillon N, Heckmann D, Anand A, Yang L, Nizet V, Monk JM, Palsson BO. 2018. Machine learning and structural analysis of *Mycobacterium tuberculosis* pan-genome identifies genetic signatures of antibiotic resistance. *Nat Commun* 9:4306. <https://doi.org/10.1038/s41467-018-06634-y>.
  27. Avershina E, Sharma P, Taxt AM, Singh H, Frye SA, Paul K, Kapil A, Naseer U, Kaur P, Ahmad R. 2021. AMR-Diag: neural network based genotype-to-phenotype prediction of resistance towards  $\beta$ -lactams in *Escherichia coli* and *Klebsiella pneumoniae*. *Comput Struct Biotechnol J* 19:1896–1906. <https://doi.org/10.1016/j.csbj.2021.03.027>.
  28. Jaillard M, Palmieri M, van Belkum A, Mahé P. 2020. Interpreting k-mer-based signatures for antibiotic resistance prediction. *Gigascience* 9: g1aa110. <https://doi.org/10.1093/gigascience/g1aa110>.
  29. Nguyen M, Brettin T, Long SW, Musser JM, Olsen RJ, Olson R, Shukla M, Stevens RL, Xia F, Yoo H, Davis JJ. 2018. Developing an *in silico* minimum inhibitory concentration panel test for *Klebsiella pneumoniae*. *Sci Rep* 8: 421. <https://doi.org/10.1038/s41598-017-18972-w>.
  30. Nguyen M, Olson R, Shukla M, VanOeffelen M, Davis JJ. 2020. Predicting antimicrobial resistance using conserved genes. *PLoS Comput Biol* 16: e1008319. <https://doi.org/10.1371/journal.pcbi.1008319>.
  31. Rocker A, Lacey JA, Belousoff MJ, Wilksch JJ, Strugnelli RA, Davies MR, Lithgow T. 2020. Global trends in proteome remodeling of the outer membrane modulate antimicrobial permeability in *Klebsiella pneumoniae*. *mBio* 11:e00603-20. <https://doi.org/10.1128/mBio.00603-20>.
  32. Doménech-Sánchez A, Martínez-Martínez L, Hernández-Allés S, del Carmen Conejo M, Pascual A, Tomás JM, Albertí S, Benedí VJ. 2003. Role



- of *Klebsiella pneumoniae* OmpK35 porin in antimicrobial resistance. *Antimicrob Agents Chemother* 47:3332–3335. <https://doi.org/10.1128/AAC.47.10.3332-3335.2003>.
33. Fajardo-Lubián A, Ben Zakour NL, Agyekum A, Qi Q, Iredell JR. 2019. Host adaptation and convergent evolution increases antibiotic resistance without loss of virulence in a major human pathogen. *PLoS Pathog* 15: e1007218. <https://doi.org/10.1371/journal.ppat.1007218>.
  34. Wise MG, Horvath E, Young K, Sahn DF, Kazmierczak KM. 2018. Global survey of *Klebsiella pneumoniae* major porins from ertapenem non-susceptible isolates lacking carbapenemases. *J Med Microbiol* 67:289–295. <https://doi.org/10.1099/jmm.0.000691>.
  35. Wong JLC, Romano M, Kerry LE, Kwong HS, Low WW, Brett SJ, Clements A, Beis K, Frankel G. 2019. OmpK36-mediated carbapenem resistance attenuates ST258 *Klebsiella pneumoniae* in vivo. *Nat Commun* 10:3957. <https://doi.org/10.1038/s41467-019-11756-y>.
  36. Tsai YK, Fung CP, Lin JC, Chen JH, Chang FY, Chen TL, Siu LK. 2011. *Klebsiella pneumoniae* outer membrane porins OmpK35 and OmpK36 play roles in both antimicrobial resistance and virulence. *Antimicrob Agents Chemother* 55:1485–1493. <https://doi.org/10.1128/AAC.01275-10>.
  37. Sugawara E, Kojima S, Nikaido H. 2016. *Klebsiella pneumoniae* major porins OmpK35 and OmpK36 allow more efficient diffusion of  $\beta$ -lactams than their *Escherichia coli* homologs OmpF and OmpC. *J Bacteriol* 198: 3200–3208. <https://doi.org/10.1128/JB.00590-16>.
  38. Doménech-Sánchez A, Hernández-Allés S, Martínez-Martínez L, Benedí VJ, Albertí S. 1999. Identification and characterization of a new porin gene of *Klebsiella pneumoniae*: its role in beta-lactam antibiotic resistance. *J Bacteriol* 181:2726–2732. <https://doi.org/10.1128/JB.181.9.2726-2732.1999>.
  39. Jacoby GA, Mills DM, Chow N. 2004. Role of beta-lactamases and porins in resistance to ertapenem and other beta-lactams in *Klebsiella pneumoniae*. *Antimicrob Agents Chemother* 48:3203–3206. <https://doi.org/10.1128/AAC.48.8.3203-3206.2004>.
  40. Acosta-Gutiérrez S, Ferrara L, Pathania M, Masi M, Wang J, Bodrenko I, Zahn M, Winterhalter M, Stavenger RA, Pagès JM, Naismith JH, van den Berg B, Page MGP, Ceccarelli M. 2018. Getting drugs into gram-negative bacteria: rational rules for permeation through general porins. *ACS Infect Dis* 4:1487–1498. <https://doi.org/10.1021/acscinfed.8b00108>.
  41. Bredin J, Saint N, Mallaéa M, Dé E, Molle G, Pagès JM, Simonet V. 2002. Alteration of pore properties of *Escherichia coli* OmpF induced by mutation of key residues in anti-loop 3 region. *Biochem J* 363:521–528. <https://doi.org/10.1042/0264-6021:3630521>.
  42. Vidal S, Bredin J, Pagès JM, Barbe J. 2005. Beta-lactam screening by specific residues of the OmpF eyelet. *J Med Chem* 48:1395–1400. <https://doi.org/10.1021/jm049652e>.
  43. Lou H, Chen M, Black SS, Bushell SR, Ceccarelli M, Mach T, Beis K, Low AS, Bamford VA, Booth IR, Bayley H, Naismith JH. 2011. Altered antibiotic transport in OmpC mutants isolated from a series of clinical strains of multi-drug resistant *E. coli*. *PLoS One* 6:e25825. <https://doi.org/10.1371/journal.pone.0025825>.
  44. Roth AL, Kurpiel PM, Lister PD, Hanson ND. 2011. *bla*<sub>KPC</sub> RNA expression correlates with two transcriptional start sites but not always with gene copy number in four genera of Gram-negative pathogens. *Antimicrob Agents Chemother* 55:3936–3938. <https://doi.org/10.1128/AAC.01509-10>.
  45. Sun D, Rubio-Aparicio D, Nelson K, Dudley MN, Lomovskaya O. 2017. Meropenem-vaborbactam resistance selection, resistance prevention, and molecular mechanisms in Mutants of KPC-producing *Klebsiella pneumoniae*. *Antimicrob Agents Chemother* 61:e01694-17. <https://doi.org/10.1128/AAC.01694-17>.
  46. Ma P, He LL, Pironti A, Laibinis HH, Ernst CM, Manson AL, Bhattacharyya RP, Earl AM, Livny J, Hung DT. 2021. Genetic determinants facilitating the evolution of resistance to carbapenem antibiotics. *Elife* 10:e067310. <https://doi.org/10.7554/eLife.67310>.
  47. Tang Y, Li G, Liang W, Shen P, Zhang Y, Jiang X. 2017. Translocation of carbapenemase gene *bla*<sub>KPC-2</sub> both internal and external to transposons occurs via novel structures of Tn7721 and exhibits distinct movement patterns. *Antimicrob Agents Chemother* 61:e01151-07. <https://doi.org/10.1128/AAC.01151-17>.
  48. Kitchel B, Rasheed JK, Patel JB, Srinivasan A, Navon-Venezia S, Carmeli Y, Brolund A, Giske CG. 2009. Molecular epidemiology of KPC-producing *Klebsiella pneumoniae* isolates in the United States: clonal expansion of multilocus sequence type 258. *Antimicrob Agents Chemother* 53:3365–3370. <https://doi.org/10.1128/AAC.00126-09>.
  49. Bulik CC, Fauntleroy KA, Jenkins SG, Abuali M, LaBombardi VJ, Nicolau DP, Kuti JL. 2010. Comparison of meropenem MICs and susceptibilities for carbapenemase-producing *Klebsiella pneumoniae* isolates by various testing methods. *J Clin Microbiol* 48:2402–2406. <https://doi.org/10.1128/JCM.00267-10>.
  50. CLSI. 2019. Performance standards for antimicrobial susceptibility testing: 29th ed; M100. CLSI, Wayne, PA.
  51. Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30:2114–2120. <https://doi.org/10.1093/bioinformatics/btu170>.
  52. Larsen MV, Cosentino S, Rasmussen S, Friis C, Hasman H, Marvig RL, Sighbak L, Sicheritz-Pontén T, Ussery DW, Aarestrup FM, Lund O. 2012. Multilocus sequence typing of total-genome-sequenced bacteria. *J Clin Microbiol* 50:1355–1361. <https://doi.org/10.1128/JCM.06094-11>.
  53. Wyres KL, Wick RR, Gorrie C, Jenney A, Follador R, Thomson NR, Holt KE. 2016. Identification of *Klebsiella* capsule synthesis loci from whole genome data. *Microb Genom* 2:e000102. <https://doi.org/10.1099/mgen.0.000102>.
  54. Jia B, Raphenya AR, Alcock B, Waglechner N, Guo P, Tsang KK, Lago BA, Dave BM, Pereira S, Sharma AN, Doshi S, Courtot M, Lo R, Williams LE, Frye JG, Elsayegh T, Sardar D, Westman EL, Pawlowski AC, Johnson TA, Brinkman FS, Wright GD, McArthur AG. 2017. CARD 2017: expansion and model-centric curation of the comprehensive antibiotic resistance database. *Nucleic Acids Res* 45:D566–D573. <https://doi.org/10.1093/nar/gkw1004>.
  55. Zankari E, Hasman H, Cosentino S, Vestergaard M, Rasmussen S, Lund O, Aarestrup FM, Larsen MV. 2012. Identification of acquired antimicrobial resistance genes. *J Antimicrob Chemother* 67:2640–2644. <https://doi.org/10.1093/jac/dks261>.
  56. Kocielek LK, Gerding DN, Espinosa RO, Patel SJ, Shulman ST, Ozer EA. 2018. *Clostridium difficile* whole genome sequencing reveals limited transmission among symptomatic children: a single-center analysis. *Clin Infect Dis* 67:229–234. <https://doi.org/10.1093/cid/ciy060>.
  57. Sheppard AE, Stoesser N, German-Mesner I, Vegesana K, Sarah Walker A, Crook DW, Mathers AJ. 2018. TETyper: a bioinformatic pipeline for classifying variation and genetic contexts of transposable elements from short-read whole-genome sequencing data. *Microb Genom* 4:e000232. <https://doi.org/10.1099/mgen.0.000232>.
  58. Pincus NB, Ozer EA, Allen JP, Nguyen M, Davis JJ, Winter DR, Chuang CH, Chiu CH, Zamorano L, Oliver A, Hauser AR. 2020. A genome-based model to predict the virulence of *Pseudomonas aeruginosa* Isolates. *mBio* 11: e01527-20. <https://doi.org/10.1128/mBio.01527-20>.
  59. Ozer EA, Allen JP, Hauser AR. 2014. Characterization of the core and accessory genomes of *Pseudomonas aeruginosa* using bioinformatic tools Spine and AGEnt. *BMC Genomics* 15:737. <https://doi.org/10.1186/1471-2164-15-737>.
  60. Ozer EA. 2018. ClustAGE: a tool for clustering and distribution analysis of bacterial accessory genomic elements. *BMC Bioinformatics* 19:150. <https://doi.org/10.1186/s12859-018-2154-x>.