Carbapenemase-Encoding Gene Copy Number Estimator (CCNE): a Tool for Carbapenemase Gene Copy Number Estimation

Jianping Jiang,a Liang Chen,b,c Xin Chen,a Pei Li,a Xiaogang Xu,a Vance G. Fowler, Jr.,d David van Duin,e Minggui Wanga

aInstitute of Antibiotics, Huashan Hospital, Fudan University, Shanghai, China
bCenter for Discovery and Innovation, Hackensack Meridian Health, Nutley, New Jersey, USA
cDepartment of Medical Sciences, Hackensack Meridian School of Medicine, Nutley, New Jersey, USA
dDivision of Infectious Diseases, University of North Carolina, Chapel Hill, North Carolina, USA
eDuke Clinical Research Institute, Duke University Medical Center, Durham, North Carolina, USA

ABSTRACT Carbapenemase production is one of the leading mechanisms of carbapenem resistance in Gram-negative bacteria. An increase in carbapenemase gene (blaCarb) copies is an important mechanism of carbapenem resistance. No currently available bioinformatics tools allow for reliable detection and reporting of carbapenemase gene copy numbers. Here, we describe the carbapenemase-encoding gene copy number estimator (CCNE), a ready-to-use bioinformatics tool that was developed to estimate blaCarb copy numbers from whole-genome sequencing data. Its performance on Klebsiella pneumoniae carbapenemase gene (blaKPC) copy number estimation was evaluated by simulation and quantitative PCR (qPCR), and the results were compared with available algorithms. CCNE has two components, CCNE-acc and CCNE-fast. CCNE-acc detects blaCarb copy number in a comprehensive and high-accuracy way, while CCNE-fast rapidly screens blaCarb copy numbers. CCNE-acc achieved the best accuracy (100%) and the lowest root mean squared error (RMSE; 0.07) in simulated noise data sets, compared to the assembly-based method (23.4% accuracy, 1.697 RMSE) and the Orthologs-Based method (78.9% accuracy, 0.395 RMSE). In the qPCR validation, a high consistency was observed between the blaKPC copy number determined by qPCR and that determined with CCNE. Reverse transcription-qPCR transcriptional analysis of 40 isolates showed that blaKPC expression was positively correlated with the blaKPC copy numbers detected by CCNE (P < 0.001). An association study of 357 KPC-producing K. pneumoniae isolates and their antimicrobial susceptibility identified a significant association between the estimated blaKPC copy number and MICs of imipenem (P < 0.001) and ceftazidime-avibactam (P < 0.001). Overall, CCNE is a useful genomic tool for the analysis of antimicrobial resistance genes copy number; it is available at https://github.com/biojiang/ccne.

IMPORTANCE Globally disseminated carbapenem-resistant Enterobacteriales is an urgent threat to public health. The most common carbapenem resistance mechanism is the production of carbapenemases. Carbapenemase-producing isolates often exhibit a wide range of carbapenem MICs. Higher carbapenem MICs have been associated with treatment failure. The increase of carbapenemase gene (blaCarb) copy numbers contributes to increased carbapenem MICs. However, blaCarb gene copy number detection is not routinely conducted during a genomic analysis, in part due to the lack of optimal bioinformatics tools. In this study, we describe a ready-to-use tool we developed and designated the carbapenemase-encoding gene copy number estimator (CCNE) that can be used to estimate the blaCarb copy number directly from whole-genome sequencing data, and we extended the data to support the analysis of all known blaCarb genes and some other antimicrobial resistance genes. Furthermore, CCNE can be used to interrogate the correlations between genotypes...
and susceptibility phenotypes and to improve our understanding of antimicrobial resistance mechanisms.

**KEYWORDS** carbapenem resistance, carbapenemase-encoding gene, *Enterobacteriales*, gene copy number

The worldwide spread of carbapenem-resistant *Enterobacteriales* (CRE) is an urgent public health threat (1). The most common carbapenem resistance mechanism is the production of carbapenemases. Among them, the *Klebsiella pneumoniae* carbapenemase (KPC) is the most prevalent type in CRE (2, 3). Other carbapenemases, including New Delhi metallo-β-lactamase (NDM), some oxacillin carbapenemases (OXA), Verona integron-encoded metallo-β-lactamase (VIM), and imipenemase (IMP), are also frequently detected in *Enterobacteriales* (4). Carbapenemases may also be found in other Gram-negative bacilli, such as *Pseudomonas* and *Acinetobacter* species (5).

Carbapenemase-producing isolates often exhibit a wide range of carbapenem MICs. Higher carbapenem MICs have been associated with treatment failure (3, 6). It has been reported that the increase of carbapenemase gene (*bla*Carb) copy number contributes to increased carbapenem MICs (7), as well as to resistance to novel β-lactam and β-lactamase inhibitors, such as ceftazidime-avibactam (CAZ-AVI) (8), meropenem-vaborbactam (9), and imipenem-relebactam. However, *bla*Carb gene copy number detection is not routinely conducted during a genomic analysis, in part due to the lack of optimal bioinformatics tools. Even though some studies have examined the impact of *bla*Carb copy number variations on antimicrobial susceptibility changes, most of these studies were limited by small sample size (10–12).

Tools such as CNVnator (13), CNV-BAC (14), and CNOGpro (15) have been developed to analyze gene copy number from whole-genome sequencing (WGS) data. However, they are not optimal for *bla*Carb gene copy number detection, because *bla*Carb is often located on a plasmid with high sequence plasticity. Methods including AssemblyBased (16, 17) and OrthologsBased approaches (18) were developed to analyze the copy number of *bla*KPC in *K. pneumoniae*, *Escherichia coli*, and *Pseudomonas aeruginosa* by using WGS data. However, their accuracies on gene copy number estimation have not been fully evaluated, and no ready-to-use tools are available.

In this study, we developed a ready-to-use tool designated the carbapenemase-encoding gene copy number estimator (CCNE) to estimate the *bla*Carb copy number directly from WGS data, and we extended our data to support the analysis of all known *bla*Carbs and some other antimicrobial resistance genes. CCNE has two components, CCNE-acc and CCNE-fast. CCNE-acc detects *bla*Carb copy number in a comprehensive and high-accuracy way and uses the assembled contigs to estimate *bla*Carb copy number, without species limitations. CCNE-fast is designed for rapid screening of *bla*Carb copy number. CCNE uses a reference gene to estimate *bla*Carb copy number and it has now supported the analysis of 111 species. We evaluated the performance of CCNE-acc and CCNE-fast on *bla*KPC copy number estimation both by simulation and through analyses of real data. We compared the results with those from AssemblyBased and OrthologsBased methods. Further, we applied CCNE to analyze the *bla*KPC copy numbers of 357 KPC-producing *K. pneumoniae* clinical isolates and correlated the results with imipenem and ceftazidime-avibactam MICs.

**RESULTS**

**Implementation of CCNE.** The general procedure for CCNE in *bla*Carb copy number estimation from whole-genome sequencing data is shown in Fig. 1. After mapping the sequence reads to the assembled genome (CCNE-acc) or target gene (CCNE-fast), the mapped reads were analyzed and a genome or gene coverage profile was constructed. The coverage of each position usually follows a Poisson distribution, \( P(x; \lambda) \) (19), which has one parameter, \( \lambda \), to be estimated (see Fig. S1 in the supplemental material). CCNE-acc estimates the *bla*Carb copy number by estimating the coverages of the whole genome and the *bla*Carb-containing fragment, respectively. As GC content
has a major effect on genome coverage, the GC bias is corrected by the method proposed by Benjamini et al. (20) (see Fig. S2 in the supplemental material). The overall workflow of CCNE-acc is shown in Fig. 2A.

CCNE-fast was developed to rapidly and directly estimate the \( \text{bla}\text{Carb} \) copy number profrom WGS reads without assembly. It applied a read depth-based algorithm and calculated the ratio of the read depth of a \( \text{bla}\text{Carb} \) gene to that of a reference gene, such as \textit{rpoB} (beta subunit of RNA polymerase) in \textit{K. pneumoniae}, \textit{dinB} (DNA polymerase IV) in \textit{E. coli}, and \textit{acsA} (acetyl coenzyme A synthetase) in \textit{P. aeruginosa}. The reference genes are single-copy genes on the chromosome. CCNE-fast preprocesses WGS reads by using HTStream (quality filter and deduplication), maps them to \text{bla}\text{Carb} and reference sequences using BWA (21), and sorts the

![FIG 1](image1.png)

**FIG 1** The general procedure of determining the \text{bla}\text{Carb} copy number by read depth-based methods from WGS reads.

![FIG 2](image2.png)

**FIG 2** The workflow of CCNE in determining gene copy number using WGS reads. (A) Workflow of CCNE-acc. (B) Workflow of CCNE-fast.
reads according to nucleic acid coordinates by using SAMtools (22). It counts the base pairwise read depth by using BEDtools (23) and estimates the blaCarb copy number through a CCNE algorithm (Fig. 2B). CCNE-fast is a highly efficient tool that simply maps the WGS reads to blaCarb and reference sequences, and it can complete an analysis of a 100× K. pneumoniae WGS sample within 5 min on 4 central processing units. All the required tools have been combined in the CCNE package, which makes it a ready-to-use tool for beginners with little bioinformatics knowledge.

**CCNE supports analysis of all known blaCarbs and other antimicrobial resistance genes.** For the blaCarbs, CCNE now supports the analysis of all known blaCarbs, including blaKPC, blaIMI, blaIMP, blaOXA, etc. In addition, 377 other antimicrobial resistance genes, such as blaCTX-M, blaTEM, and vanN, are also supported (https://github.com/biojiang/ccne) (see Table S3 in the supplemental material). CCNE-acc uses assembled contigs to estimate blaCarb copy number, without species limitations. CCNE-fast uses a reference gene to estimate blaCarb copy number and now supports the analysis of 111 species (see Table S4), including all commonly isolated species published by the China Antimicrobial Surveillance Network (CHINET; http://www.chinets.com/Data/AntibioticDrugFast), such as E. coli, K. pneumoniae, Staphylococcus aureus, Acinetobacter baumannii, and P. aeruginosa. The species list will be updated and extended to cover additional antimicrobial-resistant organisms in the future.

**CCNE achieved high accuracy in simulation compared with available algorithms.** To evaluate the performance of CCNE, simulation reads were utilized to compare CCNE with available algorithms, including AssemblyBased (16) and OrthologsBased (18) methods. The pseudogenomes with different numbers of copies of blaKPC (from 1 to 10) were obtained by mixing simulated reads from sequences of chromosome and blaKPC-carrying plasmid in different ratios (1:1 to 1:10). In non-noise data sets, all methods acquired compatible high accuracies (>96%) and a low root mean squared error (RMSE; <0.22) (Fig. 3A). In noise data sets, CCNE-acc achieved the best accuracy (100%) and the lowest RMSE (0.07), compared to the AssemblyBased (23.4% accuracy and 1.697 RMSE) and the OrthologsBased (78.9% accuracy and 0.395 RMSE) methods (Fig. 3B). CCNE-fast also achieved a high accuracy (97.1% accuracy and 0.21 RMSE) in noise data sets. In tandem repeat (Fig. 3C) and mixed (Fig. 3D) data sets, CCNE-acc, CCNE-fast, and OrthologsBased methods acquired high accuracies and low RMSEs, while the AssemblyBased method largely underestimated gene copy numbers. Detailed results are shown in Table S5 in the supplemental material. Simulation results suggested that CCNE-acc and CCNE-fast are more appropriate for real-world sequencing.
data, which contain mixed or repeated reads from chromosome and mobile genetic elements with different sequencing lengths and depths.

**CCNE revealed higher consistency with qPCR than other algorithms.** To confirm the reliability of *bla*Carb copy number estimation from whole-genome sequencing data, qPCR was employed to determine the *bla*Carb copy numbers. Three isolates with approximately 1, 3, and 9 copies of *bla*KPC (based on CCNE) were selected from 357 CRACKLE-2 China isolates (Fig. 4A). The *bla*KPC copy numbers determined by qPCR were 0.46 ± 0.02, 2.35 ± 0.15, and 8.81 ± 0.76 (means ± standard errors), while copy numbers estimated with CCNE-fast were 0.45 ± 0.05, 2.22 ± 0.13, and 9.05 ± 0.21, respectively (Fig. 4B). The *bla*KPC copy numbers were 0.38 ± 0.03, 1.75 ± 0.08, and 8.36 ± 0.05 when determined by CCNE-acc, 0.37 ± 0.04, 1.58 ± 0.07, and 7.60 ± 0.22 when estimated with the AssemblyBased method, 0.39 ± 0.05, 1.82 ± 0.07, and 8.36 ± 0.25 when estimated with the OrthologsBased method. The consistency between the results from qPCR and whole-genome sequencing-based methods demonstrated that *bla*Carb copy number estimations from whole-genome sequencing are reliable.
bla<sub>KPC</sub> copy numbers estimated by CCNE positively correlated with bla<sub>KPC</sub> gene expression levels and antimicrobial resistance levels. To determine the association between CCNE estimates of bla<sub>KPC</sub> copy number and gene expression, the transcriptional level of bla<sub>KPC</sub> in 20 isolates with a CCNE-estimated single bla<sub>KPC</sub> copy and 20 isolates with CCNE-estimated multiple bla<sub>KPC</sub> copies (3 to 7 copies) was determined by qRT-PCR. The MIC ranges (and MIC<sub>50</sub> in parentheses) of meropenem, imipenem, and CAZ-AVI for the 20 CRKP isolates with low numbers of copies of bla<sub>KPC</sub> were 32 to 128 µg/mL (128 µg/mL), 16 to 512 µg/mL (256 µg/mL), and 1 to 4 µg/mL (2 µg/mL), respectively. For the 20 CRKP isolates with high numbers of copies of bla<sub>KPC</sub>, these values were 64 to 512 µg/mL (256 µg/mL), 256 to 1,024 µg/mL (512 µg/mL), and 2 to 8 µg/mL (4 µg/mL), respectively. The transcriptional levels of bla<sub>KPC</sub> in isolates estimated by CCNE to carry a single bla<sub>KPC</sub> copy were significantly lower than levels in isolates estimated by CCNE to carry multiple bla<sub>KPC</sub> copies (1.15 ± 0.46 versus 5.08 ± 2.28; \( P < 0.001 \)) (Fig. 5). Estimated bla<sub>KPC</sub> copy numbers were positively correlated with bla<sub>KPC</sub> expression levels (Pearson coefficient, 0.76; 95% confidence interval [CI], 0.68 to 0.83; \( P < 0.001 \)).

To study the association between the bla<sub>KPC</sub> copy numbers estimated by CCNE and antimicrobial resistance, 357 sequenced ST11 bla<sub>KPC-2</sub>-harboring K. pneumoniae isolates with same ompK35 and ompK36 genotypes were selected. The bla<sub>KPC</sub> copy numbers estimated by CCNE-acc showed that 42.23% of the isolates harbored more than one copy of bla<sub>KPC</sub>, and the average copy number was 1.6 (range, 0.4 to 7) (see Fig. S4 in the supplemental material), indicating that isolates harboring multiple copies of bla<sub>KPC</sub> are common in clinical bla<sub>KPC-2</sub>-harboring ST11 K. pneumoniae. Then, the association between the copy numbers and the imipenem and ceftazidime-avibactam MICs were analyzed. The Kruskal-Wallis test showed that CCNE-estimated copy numbers were significantly associated (\( P < 0.05 \)) with MICs of imipenem (Fig. 6A) and CAZ-AVI (Fig. 6B). The estimated bla<sub>KPC-2</sub> copy numbers were positively correlated with the imipenem MIC (Pearson coefficient, 0.278; 95% CI, 0.181 to 0.370; \( P < 0.001 \)) and CAZ-AVI MIC (Pearson coefficient, 0.164; 95% CI, 0.060 to 0.264; \( P < 0.001 \)). A recent study in China found that increased bla<sub>KPC-2</sub> copy number changes were associated with decreased CAZ-AVI susceptibility in carbapenem-resistant P. aeruginosa (18). We applied CCNE-fast to determine the bla<sub>KPC-2</sub> copy numbers and assessed the correlation with the CAZ-AVI MICs using the data from the above study. Similarly, positive associations between bla<sub>KPC-2</sub> copy numbers and CAZ-AVI MIC values were detected, and our CCNE-fast method showed consistent results with their findings (see the supplemental material for more details).
**blaKPC-2 copy number variations among K. pneumoniae, E. coli, and P. aeruginosa genomes.** To investigate the blaKPC-2 copy number profiles among different species, publicly available genome assemblies of *K. pneumoniae*, *E. coli*, and *P. aeruginosa* were obtained from the NCBI GenBank database. Genomic analysis showed that 26.56% (2,840/10,692) of *K. pneumoniae*, 0.54% (127/23,447) of *E. coli*, and 0.58% (34/5,850) of *P. aeruginosa* isolates harbored blaKPC-2 (see Table S6 in the supplemental material). Among them, 808 *K. pneumoniae*, 46 *E. coli*, and 11 *P. aeruginosa* strains had matched SRA short read data available for CCNE analysis. In blaKPC-2-harboring strains, 51.86% of *K. pneumoniae* (see Fig. S4A), 76.09% of *E. coli* (see Fig. S4B), and 54.55% of *P. aeruginosa* (see Fig. S4C) harbored more than one copy of blaKPC-2, and the average copy numbers were 2.0, 4.1, and 1.7, respectively, demonstrating that isolates with multiple copies of blaKPC-2 are widely distributed in these species. The average blaKPC-2 copy number for *E. coli* strains appeared to be significantly higher than those for *K. pneumoniae* and *P. aeruginosa* (see Fig. S4D), while no significant difference was found between *K. pneumoniae* strains from NCBI and isolates from CRACKLE-2 China cohort.

Among the 808 blaKPC-2-harboring *K. pneumoniae* isolates for which there are raw data in NCBI, 67 of them have been completely sequenced. A total of 65 completed isolates harbor plasmid-borne blaKPC-2, of which four isolates harbor blaKPC-2 on 2 plasmids (with 2 to 4 estimated copies), and 2 harbor chromosomal blaKPC-2 (with 1 estimated copy). By replicon analysis with PlasmidFinder, we found that among the 61 blaKPC-2 carrying plasmids, 18 plasmids belong to IncFII(pHN7A8)/IncR, 9 belong to IncFIB(pQil)/IncFII_pKP91, 5 belong to IncA/C2, 5 belong to repA_pKPC-2, and the remaining 14 belong to other replicon families. As shown in Fig. S5, the Wilcoxon test revealed that the estimated copy number of blaKPC-2 in isolates with IncFII(pHN7A8)/IncR plasmids was significantly higher than that of blaKPC-2 in

---

**FIG 6** Correlation between estimated blaKPC-2 copy numbers and MICs of imipenem and CAZ-AVI in 357 blaKPC-2-harboring ST11 *K. pneumoniae* isolates. (A) Distribution of estimated blaKPC-2 copy numbers in imipenem groups with different MICs. (B) Distribution of estimated blaKPC-2 copy numbers in CAZ-AVI groups with different MICs.

![Correlation between estimated blaKPC-2 copy numbers and MICs](image-url)
isolates with IncFIB(pQil)/IncFII_pKP91. No significant difference was observed between other groups. For the 46 blaKPC-2-harboring E. coli isolates, 20 of them have been analyzed, and all the blaKPC-2 genes were located on plasmids. Among them, 6 plasmids belonged to IncN, 4 belonged to IncA/C2, and the remaining 10 belonged to other replicon families. For 11 blaKPC-2-harboring P. aeruginosa isolates, 2 of them have been completed and both the blaKPC-2 genes were located on untypeable plasmids.

**DISCUSSION**

In this study, we developed a robust and ready-to-use toolset, CCNE, including CCNE-acc and CCNE-fast, to estimate blaCarb copy numbers from WGS data, and we evaluated its performance on blaKPC copy number estimation through simulation, qPCR, and comparison with available algorithms. Our discovery is an important contribution to the field in several ways.

The blaCarb copy number plays an important role in the resistance to carbapenems and novel β-lactam and β-lactamase inhibitor combinations in carbapenemase-producing isolates. The routine technique for the detection of blaCarb copy number is qPCR. However, the need for optimizing efficient primers and establishing standard curves limits use of qPCR for large-scale studies. Given the increasing application of WGS in antimicrobial studies, WGS-based copy number determination is a more straightforward and scalable approach. Some WGS analysis methods, including AssemblyBased and OrthologsBased methods, have been applied to analyze the blaKPC copy numbers in K. pneumoniae, E. coli, and P. aeruginosa (16). However, they all require assembling the WGS reads into contigs and/or predicting the single-copy genes from the contigs, making these analyses time-intensive and computing resource-intensive. More importantly, their accuracies for copy number estimation have not been validated, and no ready-to-use tools are available. The current study overcomes these challenges by estimating the blaCarb copy number directly from WGS data, validating the estimation algorithm with simulation and qPCR, and implementing the algorithm into a ready-to-use tool provided to the community.

CCNE was initially developed to analyze the blaCarb copy number in carbapenemase-producing isolates, and it supports the analysis of all known blaCarbs. Given that gene copy number increases are also commonly found in other antimicrobial resistance genes, for example, the multiple blaTEM18 copy-mediated piperacillin-tazobactam resistance in E. coli (24) and tandem vanM repeats that drive vancomycin resistance in Enterococcus faecium (25), CCNE has been extended to support an additional 377 antimicrobial resistance genes, including blaCTXM, blaTEM, and vanM. CCNE-fast uses predefined reference genes for common clinical species, such as E. coli, K. pneumoniae, S. aureus, A. baumannii, and P. aeruginosa. CCNE-fast will be updated to support additional antimicrobial-resistant organisms in the future. CCNE-acc uses the assembled contigs to estimate blaCarb copy numbers, without species limitations.

In simulations, CCNE achieved the highest accuracy and the lowest RMSE compared with AssemblyBased and OrthologsBased methods. In real data, CCNE determinations of gene numbers were also consistent with qPCR findings. In a validation study, the blaKPC copy numbers estimated with CCNE were positively correlated with blaKPC gene expression levels and elevated imipenem and ceftazidime-avibactam MICs, which are consistent with the results from previous qPCR-based studies (7, 12, 26). As both CCNE-fast and qPCR use the reads count or CT values of blaKPC-2 and rpoB to estimate the copy number and the GC bias was not corrected, the results of CCNE-fast showed higher consistency with qPCR results in this study.

Several isolates in this study harbored (estimates of) less than one copy of blaKPC, which may have been caused by two factors. First of all, the sequencing bias introduced by library preparation can lead to the gene copy number being over- or underestimated (27). Second, as CCNE determines the average gene copy number of a clone community or population, it will produce an estimate of less than one gene copy if the bacterial cells in the community lost the gene, especially when the gene is located on a transposon or plasmid (28).

While augmenting the blaCarb copy number can lead to carbapenem resistance, enhanced production of other multiple carbapenemases, such as the combination of
KPC and NDM, can also result in this phenotype (29). In addition, other mechanisms, including porin mutations (30) and efflux pump overexpression (31), may also contribute to resistance phenotypes.

In conclusion, CCNE is a ready-to-use tool to determine the gene copy number of all known blaCarbs and most other antimicrobial resistance genes from 111 species. The evaluation for bla_KPC copy number estimation in KPC-producing K. pneumoniae by simulation, qPCR, and comparison with available algorithms revealed that CCNE is a robust and useful tool for blaCarb copy number estimation using WGS data. CCNE can be further used to interrogate the correlations between genotypes and susceptibility phenotypes and to improve our understanding of antimicrobial resistance mechanisms.

MATERIALS AND METHODS

Overview of CCNE. CCNE has two components, CCNE-acc and CCNE-fast, which are designed for different application scenarios. CCNE-acc estimates blaCarb copy numbers by utilizing whole-genome reads mapping coverage information. CCNE-fast is designed for rapid screening of blaCarb copy numbers when the target gene is known. CCNE-acc assumes that read coverage for each position of the whole genome and the blaCarb-containing fragment follow a Poisson distribution, and it estimates the blaCarb copy number by fitting two Poisson models. The blaCarb copy number in CCNE-acc is estimated according to this equation: copy number(CCNE-acc)_{blaCarb} = (\lambda_{blaCarb}^2)/(\lambda_{Genome}^2), where \lambda_{blaCarb} and \lambda_{Genome} are the parameters of two Poisson distributions and are approximated by the medians of read coverages of blaCarb-containing fragments and the whole genome.

CCNE-fast is a reference-based method. It only counts the reads mapped to the target blaCarb and the single-copy reference gene. The blaCarb copy number in CCNE-fast is estimated according to this equation:

\[
\text{Copy number (CCNE - fast)}_{blaCarb} = \frac{1}{\sum_{i=1}^{f} (\frac{\lambda_{Carb} - f}{\lambda_{ref} - f} \frac{ci}{ci+1})} \sum_{i=1}^{f} (\frac{\lambda_{Carb} - f}{\lambda_{ref} - f} \frac{ci}{ci+1})
\]

where L_{blaCarb} is the blaCarb gene length, L_{ref} is the reference gene length, f is the length of flanking sequence excluded in the calculation, and ci is the read depth of the ith base in the gene. Here, f is applied for controlling the bias introduced by the reads mapped to the gene boundaries.

Simulation of genomes with different copies of bla_KPC. To evaluate the performance of CCNE, we used the bla_KPC gene as an example of blaCarb. bbmap (https://sourceforge.net/projects/bbmap/) was used to generate the simulation reads based on the complete genome sequence of K. pneumoniae clinical isolate HS01777, which contains one bla_KPC-carrying plasmid and 4 other plasmids. The non-noise data sets contain the chromosome and bla_KPC-carrying plasmid reads, mixed from 1:1 to 1:10 ratio (chromosome:bla_KPC-carrying plasmid ratio), with an average chromosome read coverage of 100. The noise data sets were generated by adding all 4 other plasmids for simulated reads from HS01777 to non-noise data sets on a random coverage (from 100, 200, or 300 to 1,100). The tandem repeat data sets were generated by repeating the bla_KPC-carrying plasmid reads, mixed from 1:1 to 1:5. All the data sets in each ratio group were generated 100 times.

Comparison of CCNE with other bla_KPC copy number estimation methods. The AssemblyBased method was implemented according to the methods of Stoesser et al. (16). In brief, the raw sequencing reads of HS01777 were assembled into contigs by SPAdes v3.11.1 (32), and the bla_KPC-carrying contig was identified by BLAST using the bla_KPC sequence as the query sequence. The bla_KPC copy number was estimated by dividing the read depth of the contig containing bla_KPC by the average read depth for the entire genome assembly (weighted by contig length). The OrthologsBased method was implemented according to the methods of Zhu et al. (18). The bla_KPC copy number in the orthologs-based method was estimated by calculating the ratio of read depth of bla_KPC to the average depth of single-copy genes. The single-copy genes in HS01777 were identified by using BUSCO with the single-copy genes derived from the BUSCO v4.0.0 Enterobacterales ortholog database, 10th edition (33). The scripts and results have been placed on https://github.com/biojiang/ccne_simulation. Boxplots were generated in R, and the accuracies were calculated based on the read depth ratios, according to the following equation: accuracy = (true positives)/true positives + false negatives, and the true positives value was calculated by comparing the rounded number of read depth ratios with the true copy numbers.

The root mean squared error (RMSE) was calculated according to the following equation:

\[
\text{RMSE} = \sqrt{\frac{1}{N} \sum_{k=1}^{N} (C_k - \bar{C}_k)^2}
\]

where N is the sample size, k is the sample index, C_k is the simulated copy number of bla_KPC, and \bar{C}_k is the estimated bla_KPC copy number.
Whole-genome sequencing data of 357 KPC-producing K. pneumoniae isolates. Whole-genome sequencing data of 357 ST11 KPC-producing K. pneumoniae isolates were acquired from the Second Consortium on Resistance Against Carbapenems in Klebsiella and other Enterobacteria (CRACKLE-2) project of the MDRO Network China cohort (34). All isolates harbored the same ompK35 (ΔOmpK35, OmpK35 truncation) and ompK36 (OmpK36GD, OmpK36 134-135 GD insertion) mutant genotypes, to avoid effects of the porin mutation on the resistance phenotypes (see Table S1 in the supplemental material) (30).

Determining blaKPC copy numbers by qPCR. The blaKPC copy numbers were measured relative to rpoB by using qPCR as previously described (7). The blaKPC and rpoB genes were cloned into pMD-18T to generate pMD-18T-blaKPC-2 and pMD-18T-rpoB, respectively. Plasmids pMD-18T-blaKPC-2 and pMD-18T-rpoB were used as the templates, and primers KPC-qRT-F/KPC-qRT-R and RPOB-qRT-F/RPOB-qRT-R were used to establish the corresponding standard curves for the gene copy number determinations. Three isolates (ARLG-7036, ARLG-7038, and ARLG-6599) with 1, 3, and 9 copies of blaKPC (based on CCNE estimation) were selected from 357 CRACKLE-2 China isolates for qPCR testing. The isolates were grown on Luria-Bertani agar, and 5 single colonies of each isolate (15 samples in total) were randomly selected for culture and DNA extraction. The total 15 DNA samples were tested for blaKPC copy number by qPCR. All amplifications were carried out in triplicate using three different DNA preparations. Primers are shown in Table S2 of the supplemental material. The 15 DNA samples used in qPCR detection were also subject to whole-genome sequencing. The sequencing was performed on the Illumina NovaSeq 6000 platform (Illumina, San Diego, CA) with 150-bp paired-end reads, as previously described (34).

Transcriptional analysis of blaKPC. Quantitative reverse transcription-PCR (RT-qPCR) was employed to assess the transcriptional expression levels of blaKPC as previously described (35). The K. pneumoniae clinical isolate HS02244 with a single copy of blaKPC-2 was used as a reference for every batch of the qRT-PCR analysis. In this study, 20 K. pneumoniae clinical isolates with a CCNE-acc-estimated single copy of blaKPC and 20 isolates with multiple copies from among the 357 CRACKLE-2 China isolates were randomly selected for blaKPC transcriptional analysis. All amplifications were carried out in triplicate from three different RNA preparations. Primers are shown in Table S2 of the supplemental material.

Association between CCNE-determined blaKPC copy numbers and MICs. MICs of imipenem and CAZ-AVI were determined using the CLSI reference method for K. pneumoniae clinical strains (36). The boxplots of the CCNE-acc-determined copy numbers of each of the imipenem and CAZ-AVI groups with different MICs were generated by using R 4.0.0. The overall P values were calculated by using the Kruskal-Wallis test. The correlation coefficients, P values, and confidence intervals were calculated by using the Pearson correlation test. All tests were performed in R.

blaKPC copy number detection in the NCBI database using CCNE. The genome assemblies of 10,692 K. pneumoniae, 23,447 E. coli, and 5,850 P. aeruginosa isolates were downloaded from NCBI GenBank (accessed June 2021) and screened for blaKPC-2 by using BLASTN. The biosamples harboring blaKPC-2 were searched against the NCBI SRA database to query and download SRA data, followed by conversion to fastq format using SRA tools. The copy numbers of blaKPC-2 from public data were determined by using CCNE-fast.

Data availability. The data have been deposited with links to BioProject accession number PRJNA658369 in the NCBI BioProject database.

SUPPLEMENTAL MATERIAL

Supplemental material is available online only.

SUPPLEMENTAL FILE 1, PDF file, 0.9 MB.

ACKNOWLEDGMENTS

This work was supported by the National Natural Science Foundation of China (81991531, 81773785, and 81903673). Research reported in this publication was supported in part by the National Institute of Allergy and Infectious Diseases of the National Institutes of Health under award number UM1AI104681. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. The computing power in this work was provided by the National Clinical Research Center for Geriatric Diseases (Huashan Hospital).

We thank Keri R. Baum (Duke Clinical Research Institute) and Carol Bova Hill (Duke Clinical Research Institute) for their kind help in the study.

REFERENCES

1. Grundmann H, Glasner C, Albiger B, Aanensen DM, Tomlinson CT, Andrasevic AT, Canton R, Carmeli Y, Friedrich AW, Giske CG, Glupczynski Y, Gniadkowski M, Livermore DM, Nordmann P, Poirel L, Rossolini GM, Seifert H, Vatopoulos A, Walsh T, Woodford N, Monnet DL. 2017. Occurrence of carbapenemase-producing Klebsiella pneumoniae and Escherichia coli in the European survey of carbapenemase-producing Enterobacteriaceae (EuSCAPE): a prospective, multinational study. Lancet Infect Dis 17:153–163. https://doi.org/10.1016/S1473-3099(16)30325-2.

2. Nordmann P, Poirel L. 2019. Epidemiology and diagnostics of carbapenem resistance in Gram-negative bacteria. Clin Infect Dis 69:S521–S528. https://doi.org/10.1093/cid/ciz824.
3. van Duin D, Arias CA, Komarow L, Chen L, Hanson BM, Weston G, Cober E, Garner OB, Jacob JT, Satlin MJ, Fries BC, García-Diaz J, Doi Y, Dhar S, Kaye KS, Earley M, Hujer AM, Hujer KM, Domitrz TN, Shrophore 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 Enterobacteriaceae in the USA (CRACKLE-2): a prospective cohort study. Lancet Infect Dis 20:731–741. https://doi.org/10.1016/S1473-3099(20)30755-8.

4. Queenan AM, Bush K. 2007. Carbapenemases: the versatile beta-lactamases. Clin Microbiol Rev 20:440–458. https://doi.org/10.1128/CMR.00001-07.

5. Patel G, Bonomo RA. 2013. Stormy waters ahead: global emergence of carbapenemases. Front Microbiol 4:48. https://doi.org/10.3389/fmicb.2013.00048.

6. David S, Reuter S, Harris SR, Gla snear C, Feltwell T, Argimon S, Abudahab K, Goater R, Giani T, Enrico G, Aspbury M, Sjunnebo S, Eu SWG, Group ES, Feil EJ, Rossolini GM, Aanensen DM, Grundmann H, ESGEM Study Group. 2019. Epidemi c of carbapenem-resistant Klebsiella pneumoniae in Europe is driven by nosocomial spread. Nat Microbiol 4:1919–1929. https://doi.org/10.1038/s41564-019-0492-8.

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. Coppi M, Di Pilato V, Monaco F, Giani T, Conaldi PG, Rossolini GM. 2020. ceftazidime-avibactam resistance associated with increased blaKPC gene copy number mediated by pfpQIL plasmid derivatives in sequence type 258 Klebsiella pneumoniae. Antimicrob Agents Chemother 64. https://doi.org/10.1128/AAC.01816-19.

9. Sun D, Rubio-Aparicio D, Nelson K, Dudley MN, Nelson K. 2017. Meropenem-vaborbactam resistance selection, resistance prevention, and molecular mechanisms in mutants of KPC-producing Klebsiella pneumoniae. Antimicrob Agents Chemother 61. https://doi.org/10.1128/AAC.01694-17.

10. Sun D, Chen Y, Hua X, Chen Y, Kong L, Wu X, Jiang Y, van Schaik W, Qu T, Yu Y. 2020. Tandem amplification of the vanM gene cluster drives vancomycin resistance in vancomycin-variable Enterococci. J Antimicrob Chemother 75:283–291. https://doi.org/10.1093/jac/dkz461.

11. Girlich D, Bonnin RA, Jousset A, Naas T. 2017. Promoter characterization and expression of the blaKPC gene in Escherichia coli, Pseudomonas aeruginosa and Acinetobacter baumannii. J Antimicrob Chemother 72:1597–1601. https://doi.org/10.1093/jac/dkw422.

12. Aird D, Ross MG, Chen WS, Danielsson M, Fennell T, Russ C, Jaffe DB, Nusbaum C, Grinke A. 2011. Analyzing and minimizing PCR amplification bias in illumina sequencing libraries. Genome Biol 12:R18. https://doi.org/10.1186/gb-2011-12-2-r18.

13. Wei DW, Wong NK, Song Y, Zhang G, Wang G, Li J, Feng J. 2020. Is256 veers genotypic plasticity and genetic rearrangement toward carbapenem hyperresistance under sublethal antibiotics. mBio 13. https://doi.org/10.1093/mboic/nbaa018.

14. Ham DC, Mahon G, Bhaurla SK, Horwich-Scholefeld S, Klein L, Dotson N, Rasheed JK, McAllister G, Stanton RA, Karlsson M, Lonsway D, Huang JY, Brown AC, Walters MS. 2021. Gram-negative bacteria harboring multiple carbapenemases, United States, 2012–2019. Emerg Infect Dis 27:516.e1–516.e4. https://doi.org/10.3201/eid2712.200111.

15. Fajardo-Lubian A, Ben Zakour NL, Agyekum A, Qi Q, Iredell JR. 2019. Host genomic plasticity and genetic rearrangement toward carbapenem hyperresistance under sublethal antibiotics. mBio 13. https://doi.org/10.1093/mboic/mbz111.

16. Manni M, Berkeley MR, Seppey M, Simao FA, Zdobnov EM. 2021. BUSCO–flury pumps and AmpC, and metallo-beta-lactamases in carbapenem-resistant Pseudomonas aeruginosa isolates from South Korea. Int J Antimicrob Agents 40:168–172. https://doi.org/10.1016/j.ijantimicag.2020.04.004.

17. Prijbelski A, Antipov D, Meleshko D, Lapidus A, Korobeynikov A. 2020. Using SPAdes de novo assembler. Curr Protoc Bioinformatics 70:e102. https://doi.org/10.1002/cpbi.102.

18. Mannoni M, Berger MF, Pompelli F, Miosso-DA, Zafirovski S, Karadi R, Bonomo RA, Rutala WA, Rodríguez-Baño JG, Rodriguez-Valera F, Noone TC, Sahl WH, Vazquez-Eroles J, Villegas MV, Evans S, Hill C, Arai R, Baumann D, Fournier V, Lascols C, Tchuem Tchuenté L, Corbel M, Westh K, Gaze MS, Sonenshein L, Iredell R, Latif A, Shin D, Jafari L, Gaudy A, Marcais G, Bonomo RA. 2018. Accurate genotyping of Klebsiella pneumoniae using the near full-genome coverage of a global collection of clinical isolates. Mol Microbiol 107:1645–1660. https://doi.org/10.1111/mmi.14111.

19. Lee JY, Ko KS. 2012. OprD mutations and inactivation, expression of efflux pumps and AmpC, and metallo-beta-lactamases in carbapenem-resistant Pseudomonas aeruginosa isolates from South Korea. Int J Antimicrob Agents 40:168–172. https://doi.org/10.1016/j.ijantimicag.2012.04.004.

20. Wang M, Haufler S, Epicentrum D, Meleshko D, Lapidus A, Korobeynikov A. 2020. Using SPAdes de novo assembler. Curr Protoc Bioinformatics 70.e102. https://doi.org/10.1002/cpbi.102.

21. Manni M, Berger MF, Pompelli F, Miosso-DA, Zafirovski S, Karadi R, Bonomo RA, Rutala WA, Rodríguez-Baño JG, Rodriguez-Valera F, Noone TC, Sahl WH, Vazquez-Eroles J, Villegas MV, Evans S, Hill C, Arai R, Baumann D, Fournier V, Lascols C, Tchuem Tchuenté L, Corbel M, Westh K, Gaze MS, Sonenshein L, Iredell R, Latif A, Shin D, Jafari L, Gaudy A, Marcais G, Bonomo RA.
outcomes and bacterial characteristics of carbapenem-resistant *Klebsiella pneumoniae* complex among patients from different global regions (CRACKLE-2): a prospective, multicentre, cohort study. Lancet Infect Dis 22:401–412. https://doi.org/10.1016/S1473-3099(21)00399-6.

35. Ye M, Ding B, Qian H, Xu Q, Jiang J, Huang J, Ou H, Hu F, Wang M. 2017. *In vivo* development of tigecycline resistance in *Klebsiella pneumoniae* owing to deletion of the *ramR* ribosomal binding site. Int J Antimicrob Agents 50:523–528. https://doi.org/10.1016/j.ijantimicag.2017.04.024.

36. Clinical Laboratory Standards Institute. 2015. Performance standards for antimicrobial susceptibility testing: 25th informational supplement. CLSI M100-S25. CLSI, Wayne, PA.