


RESEARCH

Open Access



National genomic epidemiology investigation revealed the spread of carbapenem-resistant *Escherichia coli* in healthy populations and the impact on public health

Yan Li^{1†}, Yanyan Zhang^{2†}, Xinran Sun¹, Yuchen Wu², Zelin Yan², Xiaoyang Ju², Yonglu Huang², Hongwei Zhou², Zhiqiang Wang^{1,4}, Shaolin Wang³, Rong Zhang^{2*} and Ruichao Li^{1,4,5*} 

Abstract

Background Carbapenem-resistant *Escherichia coli* (CREC) has been considered as WHO priority pathogens, causing a great public health concern globally. While CREC from patients has been thoroughly investigated, the prevalence and underlying risks of CREC in healthy populations have been overlooked. Systematic research on the prevalence of CREC in healthy individuals was conducted here. We aimed to characterize CREC collected from healthy populations in China between 2020 and 2022 and to compare the genomes of CREC isolates isolated from healthy individuals and clinical patients.

Methods We present a nationwide investigation of CREC isolates among healthy populations in China, employing robust molecular and genomic analyses. Antimicrobial susceptibility testing, whole-genome sequencing, and bioinformatics were utilized to analyze a cohort of CREC isolates ($n = 113$) obtained from fecal samples of 5 064 healthy individuals. Representative plasmids were extracted for third-generation nanopore sequencing. We previously collected 113 non-duplicate CREC isolates (59 in 2018, 54 in 2020) collected from ICU patients in 15 provinces and municipalities in China, and these clinical isolates were used to compare with the isolates in this study. Furthermore, we employ comparative genomics approaches to elucidate molecular variations and potential correlations between clinical and non-clinical CREC isolates.

Results A total of 147 CREC isolates were identified from 5 064 samples collected across 11 provinces in China. These isolates were classified into 64 known sequence types (STs), but no dominant STs were observed. In total, seven carbapenemase genes were detected with bla_{NDM-5} ($n = 116$) being the most prevalent one. Genetic environments and plasmid backbones of bla_{NDM} were conserved in CREC isolated from healthy individuals. Furthermore, we compared clinical and healthy human-originated CRECs, revealing noteworthy distinctions in 23 resistance genes,

[†]Yan Li and Yanyan Zhang contributed equally to this work.

*Correspondence:

Rong Zhang
zhang-rong@zju.edu.cn
Ruichao Li
rchl88@yzu.edu.cn

Full list of author information is available at the end of the article



© The Author(s) 2024. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

including *bla*_{NDM-1}, *bla*_{NDM-5}, and *bla*_{KPC} (χ^2 test, $p < 0.05$). Clinical isolates contained more virulence factors associated with iron uptake, adhesion, and invasion than those obtained from healthy individuals. Notably, CREC isolates generally found in healthy people are detected in hospitalized patients.

Conclusions Our findings underscore the significance of healthy populations-derived CRECs as a crucial reservoir of antibiotic resistance genes (ARGs). This highlights the need for ongoing monitoring of CREC isolates in healthy populations to accurately assess the potential risks posed by clinical CREC isolates.

Keywords Healthy populations, Carbapenem resistance, *E. coli*, Plasmids, Diverse clones

Background

Carbapenems, which are considered the last line of defense, are frequently employed for the treatment of infections caused by multidrug-resistant pathogens [1]. However, carbapenem-resistant Enterobacteriaceae (CRE) are spreading at an alarming rate worldwide which has led to a severe public health issue [2]. Among them, *Klebsiella pneumoniae* and *E. coli* stand out as the most prevalent and deleterious [3]. CRE can induce severe infections such as urinary tract infections, bloodstream infections, and pneumonia. The emergence and spread of CRE have significantly limited therapeutic options, resulting in elevated rates of morbidity and mortality [4].

The increasing prevalence of carbapenem-resistant *Escherichia coli* (CREC) in hospitals and other healthcare facilities has raised alarms worldwide [5]. The emergence and spread of CREC primarily result from the acquisition of carbapenemase genes, such as the *bla*_{KPC}, *bla*_{NDM}, and *bla*_{OXA-48}-like resistance genes. CREC has the potential to spread rapidly within healthcare facilities, particularly in intensive care units (ICUs) [6]. The rapid increase in CREC infections in the clinical system has become a serious challenge for healthcare systems, especially in low- and middle-income countries with limited resources for infection control and surveillance [7]. Furthermore, CREC can persist in the hospital environment, specifically on contaminated surfaces and medical equipment, facilitating transmission between patients and healthcare workers [8].

Although CREC has traditionally been recognized as a significant issue in healthcare settings, recent studies have revealed their prevalence in non-clinical environments such as sewage, soils, surface waters, industrial effluents, and vegetation [9, 10]. The spread of CREC in these settings is a growing concern due to the increased risk of transmission to vulnerable populations and the potential emergence of new antibiotic resistance genes (ARGs) [11]. Furthermore, research on hospital wastewater has demonstrated that CREC from hospital environments can be transmitted to other environments, posing a significant threat to public health [12].

Although previous studies have shed light on the prevalence of CREC in hospitals, community settings,

and the environment, there remains a gap in our understanding of its prevalence among healthy populations [13]. Specifically, the diversity, persistence, transmission, and evolution of CREC within fecal microbiota in healthy individuals has not been comprehensively deciphered. To address this knowledge gap, we conducted a first national surveillance study between 2020 and 2022 to identify CREC isolates carried by healthy people in China. Using advanced genomic epidemiological analysis methods, we investigated the genomic characteristics of these isolates and highlight the potential risk they pose to human health. Our findings underscore the necessity for continuous surveillance of CREC in healthy populations and targeted control measures to mitigate this overlooked threat.

Methods

Study design

To investigate the prevalence of CREC in healthy populations, we consecutively collected 5064 non-duplicate stool specimens from healthy individuals without evidence of intestinal infection from hospitals across 11 regions in China between 2020 and 2022. These healthy individuals had come to the hospital for a routine physical examination. After incubating the samples in 5 mL Luria–Bertani broth at 37 °C for 6 h, the enriched cultures were screened for CREC isolates by streaking on MacConkey agar plates supplemented with meropenem (0.3 mg/L), and the species of isolates were identified by using MALDI-TOF MS. To avoid duplication, only one CREC isolate per stool sample was selected for further study. Ethical approval for this study was given by the Zhejiang University ethics committee (number 2023–0733). Antimicrobial susceptibility testing (AST) was performed on all non-duplicated isolates using the broth microdilution method against a panel of fifteen antimicrobials, with *E. coli* ATCC 25922 as the quality control. The resistance breakpoint of tigecycline (>2 mg/L) was interpreted based on European Committee on Antimicrobial Susceptibility Testing (EUCAST) criteria (<https://www.euca.org/>), and the Clinical and Laboratory Standards Institute (CLSI) guidelines were followed for the remaining fourteen antimicrobials [14].

DNA extraction, whole-genome sequencing, and de novo assembly

To further investigate the genetic characteristics of CREC isolates, the genomic DNA of all CREC isolates were extracted using the PureLink Genomic DNA Mini Kit (Invitrogen, USA). Then, the high-quality genomes were subjected to whole-genome sequencing by the Illumina HiSeq 2500 platform with 2×150 bp paired-end libraries. In addition, plasmids of the representative CREC isolates were extracted for third-generation nanopore sequencing to obtain complete genome sequences. The plasmids of the isolates were extracted using the Qiagen plasmid midi-kit (Qiagen, Germany) after overnight culture in 100 ml LB broth. After performing quality control and filtering by fastqc, sequencing data of Illumina and nanopore were underwent separate assembly processes using SPAdes v.3.11.0 and Flye v.2.4.2, respectively [15, 16]. The plasmid sequences were completed with a hybrid de novo assembly strategy by Unicycler v.0.4.8 [17]. Genome assemblies were further confirmed by comparing the two assembly strategies.

Genomic analysis

The ARGs, insertion sequences, and plasmid replicons were detected using the online tools with default parameters [18–20]. Virulence factors (VFs) were determined using the virulence factor database (last updated 14 October 2020) in ABRicate [21]. Multiple sequence typing (MLST) and serotyping were performed using mlst v.2.15.1 and ECTyper v.1.0 with default parameters, respectively [19, 22]. The phylotyping of *E. coli* was performed using ClermonTyping software [23]. EasyFig v.2.2.3 and Brig v.0.95 was used to generate plasmid comparison figures [24, 25]. The single nucleotide polymorphisms (SNPs) alignments were performed using Snippy v.4.0.2 [26].

Phylogenetic and genome wide association analyses

Pan-genome and core-genome analyses were performed using Roary to determine the genetic diversity and relatedness [27]. The phylogenetic tree was generated by maximum likelihood method using FastTree, and the evolutionary relationships were visualized using iTOL [28, 29]. In addition, to further explore the genomic similarities and differences of CRECs between clinical patients and healthy people, 113 CREC isolates that we had previously isolated from ICU wards were included in this study for analysis [6]. The CREC isolates were divided into healthy people group and patient group. Scoary was employed for downstream whole-genome association analysis (GWAS) of CRECs for these two groups using the presence/absence matrix output generated by Roary

as input files, to retrieve the different characteristics of both groups [30]. To annotate the functions of genomes, functional annotation of differentially expressed genes was determined based on the prokka v.1.12 alignment output [31]. The latest update database of eggNOG was used for function annotations and the annotations were further analyzed for KEGG pathway enrichment analysis [32, 33].

Statistical analysis

Statistical analysis and visualization charts were performed using R v.4.2.2 (R Foundation for Statistical Computing, Vienna, Austria). The association between ARGs and insertion sequences was analyzed based on Spearman algorithm and visualized using Gephi [34]. Chi-square analysis was used to detect differentially expressed genes between healthy individuals and patients-derived CRECs. Principal coordinate analysis (PCoA) based on Bray–Curtis distance was performed to evaluate the composition of the VFs and ARGs in the CRECs isolated from different sources.

Results

Distribution and carbapenemase genes of CREC isolates among healthy populations

In this study, we conducted a nationwide survey and isolated a total of 147 CRECs from March 2020 to September 2022. These isolates were obtained from 11 provinces in China, including Zhejiang, Guangdong, Henan, Xinjiang, Inner Mongolia, Tibet, Shandong, Hunan, Hubei, Ningxia, and Fujian (Fig. 1A, Table S1). Notably, all isolates were derived from stool samples of apparently healthy individuals, accounting for 2.90% of the total screened samples (147/5064). The geographical distribution of the samples covered a substantial portion of China, with a population of approximately 619 million individuals. This comprehensive sampling approach provides valuable insights into the prevalence and distribution of CREC isolates across different regions in China.

Among these CREC isolates collected from healthy individuals, a total of seven carbapenemase genes, including *bla*_{NDM-1}, *bla*_{NDM-4}, *bla*_{NDM-5}, *bla*_{NDM-6}, *bla*_{NDM-9}, *bla*_{NDM-39}, and *bla*_{OXA-48}, were detected (Fig. 1B–E). Notably, the *bla*_{NDM-5} gene was the most prevalent carbapenemase gene (78.91%, 116/147), followed by the *bla*_{NDM-1} gene (15.65%, 23/147). Four isolates were found to carry the *bla*_{NDM-9} gene. However, the *bla*_{NDM-6}, *bla*_{NDM-4}, *bla*_{NDM-39}, and *bla*_{OXA-48} genes were each detected in only one isolate, respectively. The number of the *bla*_{NDM} variants in the CRECs increased gradually over time, with the *bla*_{NDM-5} gene dominated among the carbapenemase genes in different years. Furthermore,

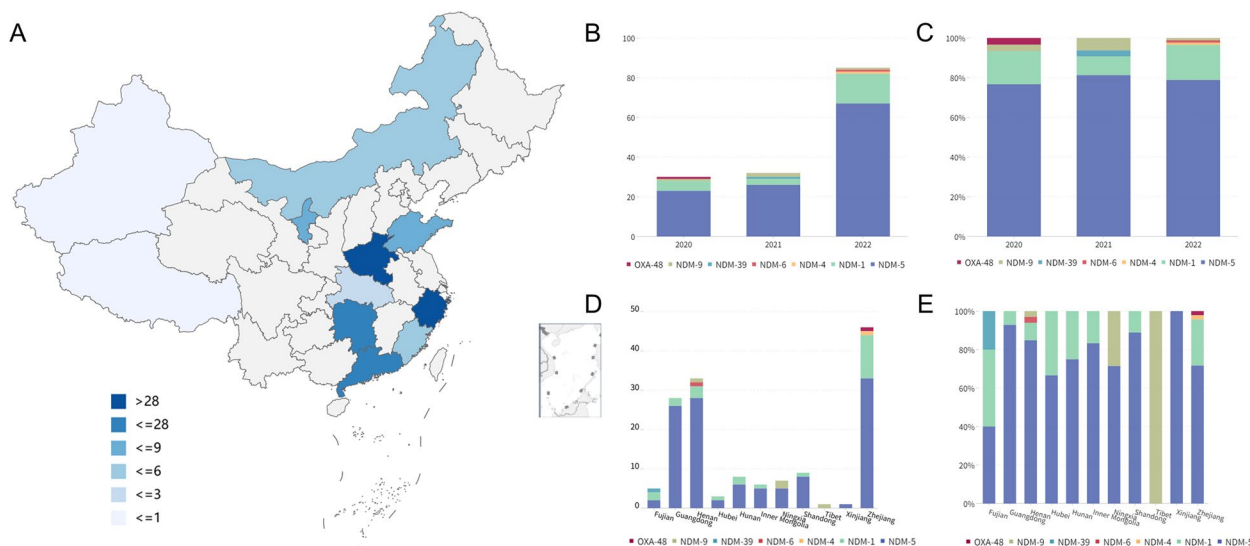


Fig. 1 Distribution and prevalence of CRECs and corresponding carbapenemases in China. **A** Prevalence of CRECs in healthy individuals in China. Different colors represent the number of CRECs. **B** Bar graph showing the number of carbapenemases by year. **C** Bar graph displaying the percentage of different carbapenemases by year. **D** Bar graph displaying the number of different carbapenemases by regions. **E** Bar graph displaying the percentage of different carbapenemases by regions

*bla*_{NDM-5} was the dominant carbapenemase gene in all regions except Fujian and Tibet.

Antimicrobial susceptibility testing and ARGs diversity of CREC isolates

Among the tested carbapenems, CREC isolates showed the highest resistance rate to ertapenem (93.20%), followed by meropenem (92.52%) and imipenem (89.12%). All CREC isolates showed high-level resistance to the third- or fourth-generation cephalosporins and their combination with β-lactam inhibitors (91.84–100%). However, a limited number of isolates exhibited resistance to colistin (3.40%), tigecycline (4.08%), and amikacin (11.56%) (Table S2). The phenotype could in most cases be explained by carriage of the corresponding ARGs. Other than carbapenemase genes, CREC isolates also carry multiple ARGs including genes conferring resistance to beta-lactams (*bla*_{CTX-M} *n* = 49, *bla*_{TEM} *n* = 87, and *bla*_{SHV} *n* = 67), sulfonamide (*sul1* *n* = 58, *sul2* *n* = 70, *sul3* *n* = 67), aminoglycoside (*aac(3)-II* *n* = 31, *aac(3)-Iva* *n* = 55, *aadA* *n* = 147, *rmtB* *n* = 15), and tetracycline (*tet(M)* *n* = 28). In addition, twelve and two CRECs were found to carry the *mcr-1* gene and the *tet(X4)* gene, respectively.

To explore the relationship between carbapenemase genes, other ARGs, and insertion sequences, a correlation network graph was constructed (Fig. S1). The network graph revealed strong associations between certain genetic elements. In particular, IS21, ISEc10, ISEc20, *bla*_{CTX-M-15}, *bla*_{EC-5}, *bla*_{OXA-1}, and ISEc24 were strongly

positively associated with *bla*_{OXA-48} (*R* > 0.3, *p* < 0.05). Similarly, IS414 was positively associated with *bla*_{NDM-1} (*R* > 0.3, *p* < 0.05). ISEc46 and IS903 were positively associated with *bla*_{NDM-4} (*R* > 0.3, *p* < 0.05). The *bla*_{NDM-5} gene was positively correlated with IS10R but negatively correlated with *bla*_{NDM-1} (*R* absolute value > 0.3, *p* < 0.05). *bla*_{CTX-M-65} was positive associated with *bla*_{NDM-6} (*R* > 0.3, *p* < 0.05). In addition, IS1326, IS1353, IS10R, *bla*_{CTX-M-65}, *aac(6′)-II*, and *aadA22* showed positive associations with *bla*_{NDM-9} (*R* > 0.3, *p* < 0.05).

Transmission of CREC multiple clones

The 147 CREC isolates belonged to 64 known STs, none of which exhibited a dominant prevalence. Our subsequent analysis revealed a significant correlation between the distribution of some STs and the geographical locations of the isolates, as demonstrated by the patterns depicted in Fig. 2A. Among the 64 STs, ST224 was the most widely dispersed ST, which was detected in five provinces and exhibited a relatively high isolation rate (10/147, 6.80%). Moreover, ST224 was also the predominant ST of the *mcr-1*-positive CRECs in this study. In terms of time distribution, only ST48 CREC was detected in all three years (Fig. 2B). The number of STs has been increasing over time. In addition, we found that some STs showed a corresponding relationship with carbapenem resistance genes. All ST410 and ST155 isolates were positive for *bla*_{NDM-5} (Fig. 2C). The *bla*_{NDM-4}, *bla*_{NDM-6}, and *bla*_{NDM-39} genes were only found in ST10, ST361, and ST746 isolates, respectively.

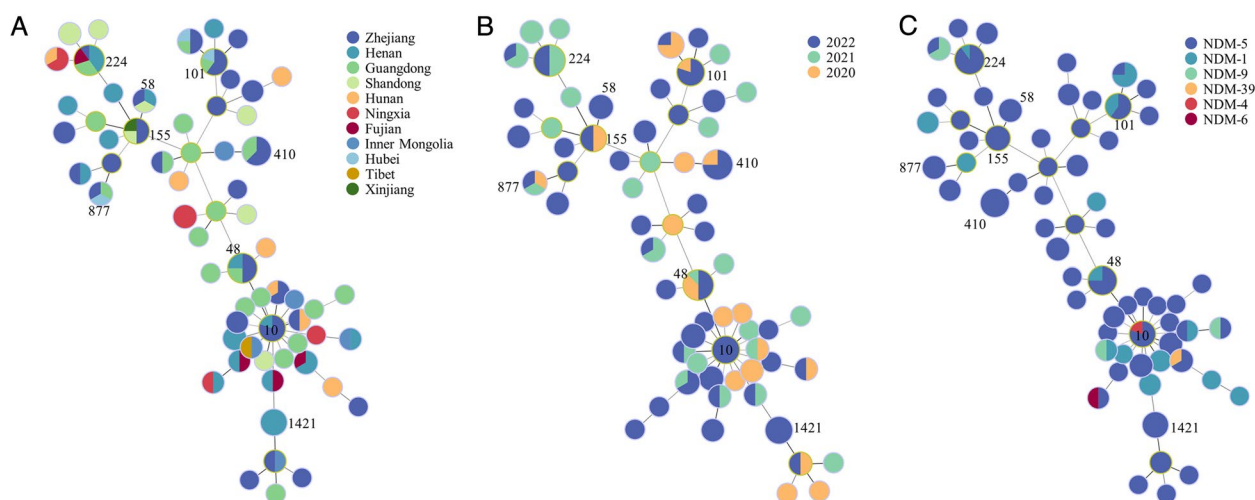


Fig. 2 Minimal spanning tree of CREC isolates. **A** STs distribution of CREC isolates in different regions. **B** STs distribution of CREC isolates in different years. **C** STs distribution of different carbapenemases

The classification of phylogenetic subgroups demonstrated that the 147 CREC isolates were distributed into 7 phylogroups (A, B1, B2, C, D, F, and G), but the majority of them were in group A (61/147, 41.50%) and B1 (60/147, 40.81%) (Fig. 3). Serotype prediction revealed that there was no dominant serotype among the CREC isolates, and the most frequently isolated serotypes were O8:H9 ($n=6$) and O88:H31 ($n=5$). Furthermore, we observed a pronounced genetic resemblance among strains situated within the same clusters in the shadow, characterized by a high degree of similarity (≤ 20 SNPs), suggestive of clonal transmission. Subsequent analyses identified a total of 19 such clusters, with strains distributed across diverse STs, all of which exhibited multidrug resistance.

Conserved genetic contexts of carbapenemase genes

The genetic contexts of the *bla*_{NDM} genes were observed to be highly conserved. A total of two major genetic contexts were identified in 116 *bla*_{NDM-5}-positive isolates, including IS3000- Δ IS*Aba125*-IS5-*bla*_{NDM-5}-*ble-trpF* and IS3000-IS*Aba125*-IS5-*bla*_{NDM-5}-*ble-trpF*. The *bla*_{NDM-5} gene was predominantly carried by either IncX3 or IncHI2A plasmids (Fig. S2, Table S3). Besides, only one genetic context IS3000-IS*Aba125*-*bla*_{NDM-1}-*ble-trpF*-IS26 were identified in *bla*_{NDM-1}-positive CREC isolates. All *bla*_{NDM-1}-positive plasmids belonged to IncX3. All *bla*_{NDM-9}-positive plasmids were of the IncHI2 type (4/4, 100%), and their conserved genetic context was Δ IS*Aba125*-*bla*_{NDM-9}-*ble-trpF*. In addition, the genetic context of only identified *bla*_{NDM-4}-positive isolate was IS3000- Δ IS*Aba125*-*bla*_{NDM-4}-*ble-trpF*.

According to the core genome diversity and plasmid diversity comparison, all *bla*_{NDM}-positive plasmids have a relatively conservative backbone structure. Besides, most plasmids showed no significant correlation with bacteria evolution (Fig. S3), implying conserved plasmids contributed to the spread of *bla*_{NDM} in CREC isolates. However, there were also some isolates with a closer evolutionary relationship and higher homology of *bla*_{NDM}-positive plasmids, indicating that clone transmission of *bla*_{NDM}-positive isolates also plays a role in the spread of CREC isolates.

Different phylogenetic evolution of CRECs from clinical and healthy populations

In order to further describe the genetic features of CRECs in healthy populations, 113 CREC isolates of clinical origins were downloaded from the NCBI database [6]. These 113 CREC isolates were collected in our previous study from ICU patients in 15 provinces and municipalities in China. Among the 260 CREC isolates, we detected 100 known STs, with a marked difference in ST distribution between clinical and non-clinical *E. coli* isolates (Fig. 4). Specifically, the CRECs isolated from healthy individuals exhibited a higher diversity of STs ($n=48$) compared to clinical CRECs ($n=20$). The two groups shared only 16 STs (16/100, 16%). ST131 was the most prevalent (34/113, 30.09%) in clinical CRECs, while ST224 being the most common in non-clinical CRECs (10/147, 6.8%). The distribution of *bla*_{KPC}-positive isolates was only observed in clinical samples, with ST131 carrying a larger number of the *bla*_{KPC} genes. The isolates located in the shaded areas of Fig. 4 are those from different sources but with high genetic similarity (≤ 20 SNPs), indicating the

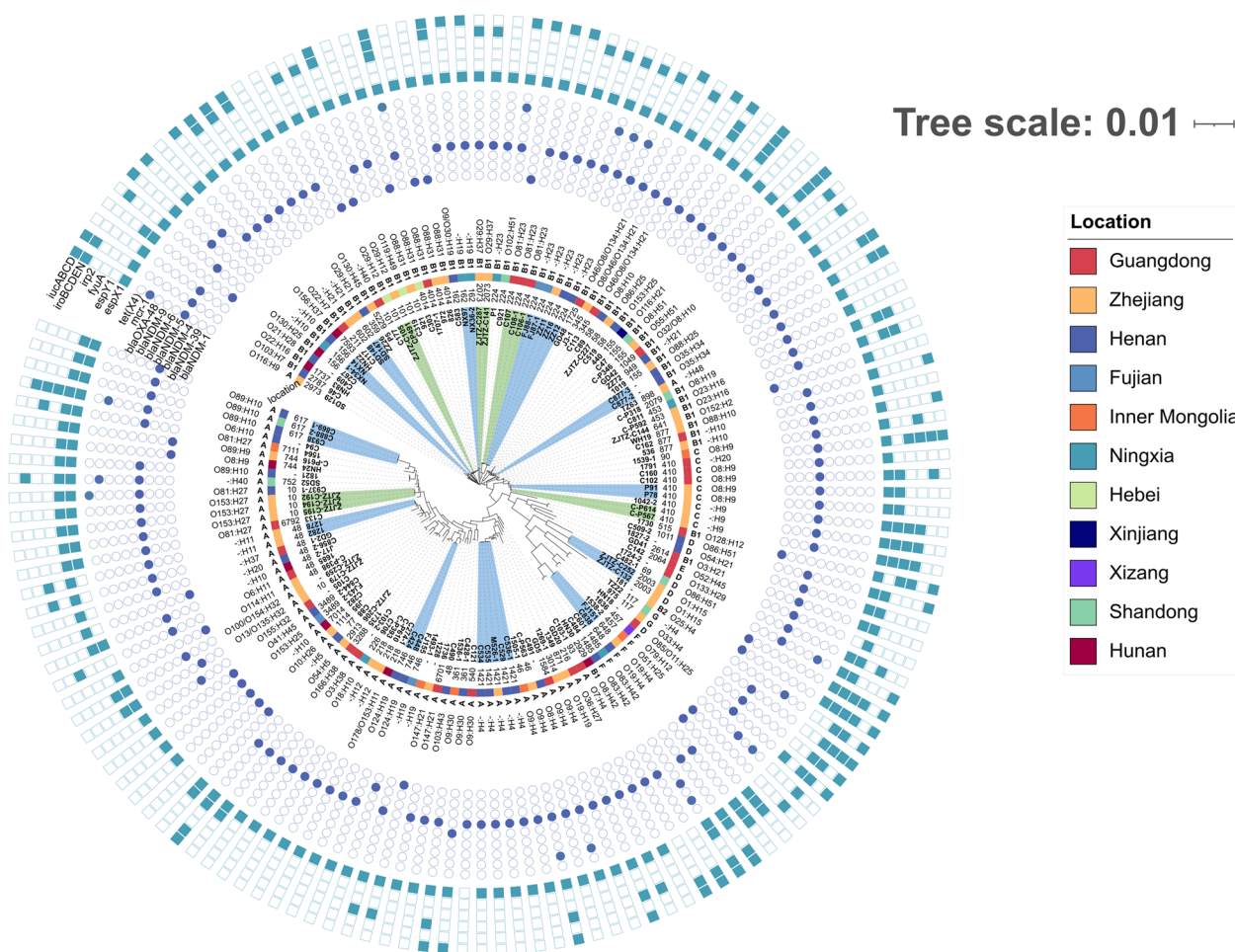


Fig. 3 Phylogenetic tree of 147 CREC isolates from China. Isolates located in the light green or light blue shaded areas contained only a few SNPs differences ($n \leq 20$). The resistance genes are indicated by square, solid graphics indicate yes, hollow no

possibility of clonal transmission between nonclinical and clinical CRECs (Table S4). The STs of these isolates were ST46 and ST410, distributed across various regions (Shanghai, Zhejiang, Inner Mongolia), forming two distinct clusters of distribution, and all of them carried carbapenemase genes.

Genomic differences in CREC isolates in clinical and non-clinical settings

Compared with 113 clinical isolates, the CREC isolates from healthy individuals carried a greater variety of insertion sequences and VFs. However, CREC from healthy individuals carries fewer types of ARGs and plasmid replicons (Fig. S4). The positive rate of the *intI1* gene in clinical isolates was 74.34% (84/113), and in isolates collected from healthy individuals, it was 75.51% (111/147). Alpha diversity analysis showed that compared with the CRECs from healthy people, the CRECs from clinical sources carried more plasmid replicons

and VFs ($p < 0.05$). However, there was no significant difference in the number of resistance genes and insertion sequences between these two sources of CRECs ($p > 0.05$) (Fig. S5). PCoA results showed that compared with the CRECs isolated from healthy people, the ARGs and VFs carried by different CREC isolates isolated from clinical samples varied greatly (Fig. 5).

Among all 120 detected ARGs, 23 ARGs exhibited significant differences between healthy and clinical population groups (χ^2 test, $p < 0.05$) (Table S5, Fig. S6A). Besides, significant differences were observed in the positive rates of carbapenemase genes between clinical and non-clinical isolates, with *bla*_{NDM-5} being more prevalent in non-clinical isolates (χ^2 test, $p < 0.01$) and *bla*_{KPC-2} being more prevalent in clinical isolates (χ^2 test, $p < 0.01$). In non-clinical CREC isolates, the enriched ARGs for aminoglycosides included *aac(3)-IVa*, *aadA*, *aph(3')-Ia*, *aph(4)-Ia*, *arr-2*, and *dfrA17*, while the positive rates of resistance genes for

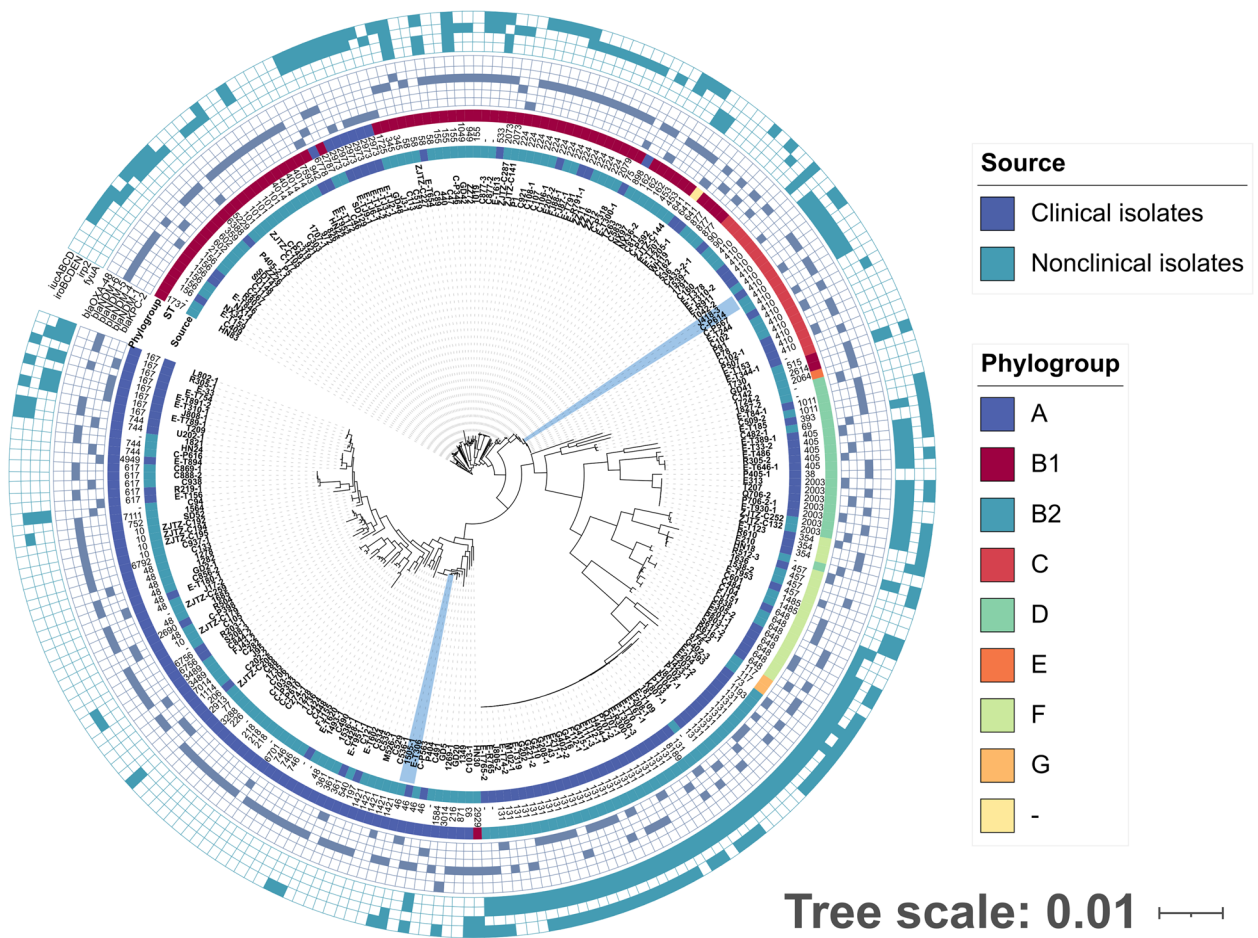


Fig. 4 Phylogenetic tree of 260 CREC isolates from clinical and non-clinical environments. Clinical and non-clinical CRECs have different genetic backgrounds, with significant differences between their STs, ARGs, and phylogroup. The resistance genes are indicated by square, solid graphics indicate yes, hollow no. The isolates in the blue-shaded areas represent isolates from different sources but with high genetic similarity (≤ 20 SNPs)

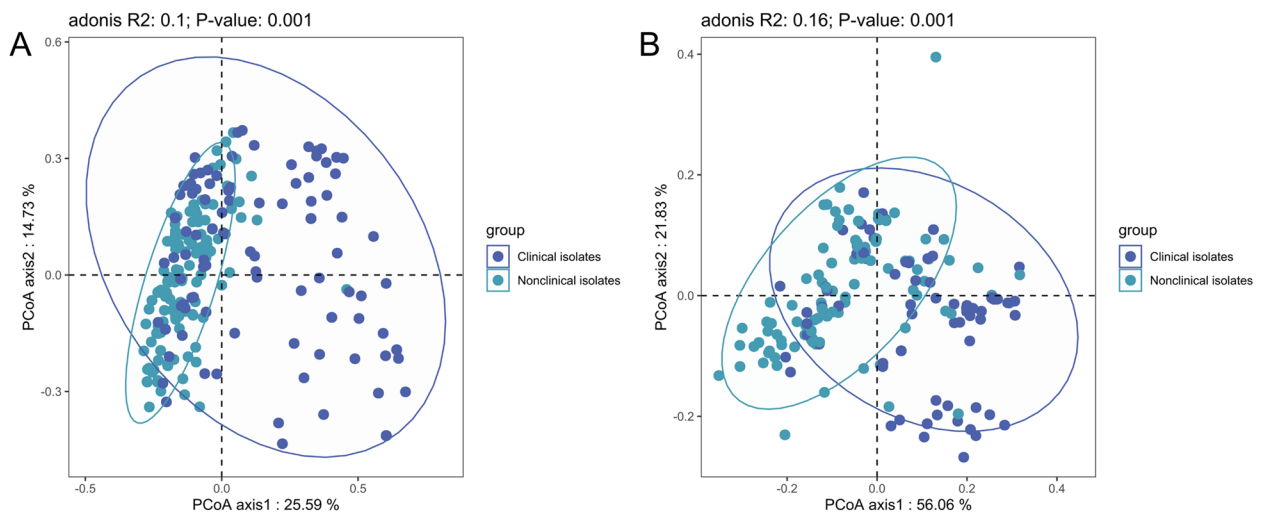


Fig. 5 Principal component analysis for ARGs and VFes. **A** Principal component analysis of ARGs in CRECs from healthy people and clinical sources. **B** Principal component analysis of VFes in CRECs from healthy people and clinical sources

β -lactams and quinolones were significantly higher than those of clinical CREC isolates.

A total of 65 VFs that showed significant differences between the two groups were significantly more abundant in clinical isolates (χ^2 test, $p < 0.01$) (Table S6, Fig. S6B). We found that the positive rates of several important virulence genes *fyuA*, *irp2*, and *iucABCD* in clinical isolates (57.52%, 57.22%, 53.10%) were higher than those in isolates isolated from healthy individuals (12.93%, 11.56%, 30.61%). All VFs with significantly higher prevalence are rich in clinical CREC, which indicated that these clinical CREC isolates have higher pathogenicity. Further looking up VFs against VFDB, these differentially abundant VFs were mainly associated with invasion (e.g., K1 capsule), iron uptake (e.g., ferrienterochelin receptor Fes), secretion (e.g., type II secretion system, type III secretion system), and adhesion (e.g., fimbrial protein). Those virulence factors are crucial in the pathogenic phase of clinical pathogenic CRECs.

Genome wide association analysis of CREC isolates

A total of 513 genes were found to be above the significance threshold (Bonferroni test, $p < 0.01$) between healthy and clinical populations. To further investigate these genes of differentially allele frequencies, including *bla*_{NDM-5} and genes related to its conserved genetic contexts, as well as *bla*_{KPC}, which were previously found to be significantly different between clinical and non-clinical isolates, these genes were subjected to KEGG pathway enrichment analysis. We found that the genetic difference of clinical CRECs were mainly enriched in biosynthesis of siderophore group nonribosomal peptides pathways and metabolic pathways ($p < 0.05$) compared with those from non-clinical isolates (Fig. S7). On the other hand, the genetic difference of non-clinical isolates was mainly enriched in microbial metabolism pathways and quorum sensing pathways ($p < 0.05$) compared with those from clinical CRECs.

Discussion

Currently, there are few studies on the prevalence of CREC in the normal flora of healthy people [13]. In this study, we conducted a systematic analysis of CREC isolates among healthy populations in different provinces of China between 2020 and 2022 and compare them with clinical CRECs with genomic analysis.

The carbapenemase genes in clinical CREC isolates that were collected from ICU patients in our previous study were mainly dominated by *bla*_{KPC-2} (53/116, 45.68%). However, unlike previous clinical CREC isolates, the most predominant carbapenemase gene was *bla*_{NDM} in healthy populations in this study (146/147, 99.32%) [6]. Asian countries, particularly China, are widely

recognized as the primary reservoirs of the *bla*_{NDM} genes [35]. An earlier study reported that 80% of the isolates in Chinese patients from 2013 to 2015 carried *bla*_{NDM-1}, with only 17.8% being *bla*_{NDM-5} [36]. However, the number of *bla*_{NDM-5}-positive *E. coli* is much higher than that of *bla*_{NDM-1}-positive *E. coli* in recent years, which may be attributed to the lower fitness cost of *bla*_{NDM-5} to host bacteria and its higher hydrolytic activity against carbapenems [37–39]. Tigecycline and polymyxin are important drugs for the clinical treatment of carbapenem-resistant Enterobacteriaceae infections [40]. However, the discovery of the *tet(X)* and *mcr-1* genes in CREC will further limit the selection of clinical drugs, potentially posing challenges to the effective management of CREC infections [41, 42].

The main ST of CRECs isolated from clinical patients was ST131 carrying the *bla*_{KPC} gene [2, 43]. The ST131 *E. coli* is a widely distributed exenteral-pathogenic *E. coli*, which has attracted extensive public attention due to its carriage of a variety of ARGs and VFs [44]. Interestingly, this particular ST was not detected in CREC isolates isolated from healthy people in the present study. The *bla*_{NDM}-positive CRECs isolated from healthy individuals exhibited a more diverse range of STs, which may further expand the range of ARGs reservoirs. However, no dominant STs were detected in these *bla*_{NDM}-positive CRECs, which is consistent with the results of previous studies [35]. This consistency underscores the significance of horizontal transmission as a pivotal route for the dissemination of carbapenemase genes in healthy individuals. This diversity in STs among healthy individuals suggests a broad dissemination of the *bla*_{NDM} genes in the normal flora, further emphasizing the importance of monitoring and controlling the spread of CREC in healthy populations.

Despite being isolated from various regions, some CRECs showed minimal genetic variation with only a few SNPs ($n \leq 20$), indicating the possibility of clonal transmission. When these healthy individuals carrying CREC move in different regions, they pose a potential threat of transmitting resistance genes to other healthy individuals. Certain clinical and non-clinical CRECs clustered together in distinct phylogenetic groups with shared genetic backgrounds and identical VFs, suggesting a common origin. Moreover, some CRECs isolated from healthy people and clinical patients also displayed high genetic similarity (SNPs ≤ 20). The occurrence of this phenomenon may result from the exposure of healthy individuals to hospital environments containing multi-drug-resistant strains. The majority of the *bla*_{NDM} genes are located on IncX3 type plasmids, consistent with previous clinical findings [6]. It has been demonstrated that IncX3 plasmids facilitate the transmission of *bla*_{NDM-5},

and the *bla*_{NDM-5} gene in IncX3 plasmid can exist stably [45, 46]. The mobile genetic elements such as insertion sequences and plasmids are highly correlated with a variety of carbapenem resistance genes, and they play important roles in mediating the spread of carbapenem resistance genes [47]. This phenomenon may contribute to the widespread dissemination of carbapenem resistance genes among healthy populations.

The presence of class I integron serves as a crucial indicator of multidrug resistance. Our findings reveal a higher positive rate of *intI1* in both healthy individuals and clinical-origin CREC isolates, with no significant difference between the two groups ($p > 0.05$). Furthermore, all CRECs isolated from healthy individuals harbored multiple resistance genes, providing additional validation to this conclusion. This underscores the strong selection pressure on healthy humans for colonization by multidrug-resistant *E. coli*. Although there were no significant differences in the number of ARGs and insertion sequences between the CREC isolates from clinical patients and those from healthy people, the former carried more diverse types of VFs and plasmid replicon types ($p < 0.05$). Moreover, the distribution of VFs in clinical isolates of CREC isolates are more diverse. The chi-square test also showed that the CREC isolates from clinical sources carried many VFs with significant differences, such as *irp2*, *fyuA*, and *iucABCD*. This increase in VFs suggests that the isolates from clinical patients may be more pathogenic compared to those isolated from healthy individuals, which aligns with the observed clinical outcomes [48, 49].

KEGG enrichment analysis results also demonstrate that clinical CREC has higher invasiveness and energy metabolism-related genes, and the specific functional differences may be further analyzed and confirmed by results from meta-transcriptome and deep metagenome sequencing to determine the differences in functionality between patient and healthy gut microbiomes. Moreover, iron is an essential nutrient for the survival of both non-pathogenic and pathogenic *E. coli*. Pathogenic *E. coli* have evolved specific mechanisms to acquire iron from host cells, as iron is often sequestered by the host as part of the innate immune response [50]. Pathogenic *E. coli* mainly use siderophores, small molecules with high affinity for iron, to acquire this essential nutrient. Enterobactin and yersiniabactin which was found that it was significantly enriched in clinical CREC are among the different types of siderophores produced by pathogenic *E. coli* to scavenge iron from host cells [51, 52]. Once acquired, iron is used by the bacteria for a range of cellular processes, including energy production and DNA replication. Iron has also been implicated in the regulation of the type III

secretion system [53], which was identified significantly enriched in clinical CREC, a major virulence factor in many pathogenic *E. coli* strains.

We acknowledge the limitations of the current study. Firstly, this was a retrospective study, and some information on the CREC isolates collected was missing, which could have affected the accuracy of our analysis. Additionally, it was not possible to determine the positive rate of CREC isolates in different regions due to the lack of complete data. Secondly, our study only focused on the molecular and genomic characteristics of CREC in the Chinese healthy populations, and thus, we lack a comprehensive understanding of CREC in healthy populations worldwide. Further research on a global scale is needed to fully comprehend the distribution and prevalence of CREC among healthy individuals in different regions.

Conclusions

This study provides a comprehensive analysis of the prevalence and distribution characteristics of CRECs in healthy individuals across the country. The CREC isolates isolated from healthy individuals were distributed across a variety of STs, and these STs were generally associated with low virulence. However, there is a possibility of transmission of these isolates to clinical patients, rendering clinical treatment ineffective. The emergence of CRECs in healthy individuals from multiple provinces of China underscores the need for stringent monitoring and appropriate measures to mitigate the future threats posed by CREC strains.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s13073-024-01310-x>.

Additional file 1: Fig. S1. The network graph describing the co-occurrence pattern of ARGs with ISs. **Fig. S2.** Circular comparison of bla_{NDM}-bearing IncH12 and IncX3 plasmids. **Fig. S3.** Phylogenetic tree generated by core genome and genetic contexts alignment. **Fig. S4.** Intersection analysis of ARGs, VFs, insertion sequences and plasmid replicons among clinical and non-clinical sources. **Fig. S5.** Alpha diversity analysis of ARGs, VFs, plasmids replicons and insertion sequences among clinical and non-clinical sources. **Fig. S6.** Manhattan plot of differential genes between clinical and non-clinical sources. **Fig. S7.** KEGG enrichment analysis of clinical and non-clinical environments.

Additional file 2: Table S1. Basic information of 147 CREC isolates collected from healthy populations. **Table S2.** MIC result of 147 CREC isolates to 15 drugs. **Table S3.** The location of the carbapenem resistance genes. **Table S4.** SNPs among 260 CREC isolates collected from healthy populations. **Table S5.** Chi-square analysis of ARGs carried by CREC in healthy people and clinical sources. **Table S6.** Chi-square analysis of VFs carried by CREC in healthy people and clinical sources.

Acknowledgements

This work was supported by the High Performance Computing Cluster of College of Veterinary Medicine, Yangzhou University.

Authors' contributions

Ruichao Li and Rong Zhang conceived and supervised the study. Yanyan Zhang, Yan Li, and Xinran Sun performed the experiments, collected the data, and prepared the draft. Yan Li, Yuchen Wu, Zelin Yan, and Xiaoyang Ju interpreted the data and conducted analysis. Yonglu Huang, Hongwei Zhou, Zhiqiang Wang, and Shaolin Wang contributed to the sample collection and manuscript editing. Rong Zhang and Ruichao Li acquired the funding. All authors read and approved the final manuscript.

Funding

This work was supported by National Key Research and Development Program of China (grant no. 2022YFD1800400), National Natural Science Foundation of China (grant no. 22193064), and the Priority Academic Program Development of Jiangsu Higher Education Institutions (PAPD).

Availability of data and materials

Whole-genome sequence data generated in this study has been deposited in the National Center for Biotechnology Information under BioProject no. PRJNA996502. Two plasmids pC519-IncHI2 (NZ_OR395176) and p1505-1-IncX3 (NZ_OR395175) with nanopore sequencing have been submitted to NCBI database. All study data is included in the article and/or supporting information.

Declarations

Ethics approval and consent to participate

Ethical approval was given by the Zhejiang University ethics committee (number 2023-0733). Informed patient consent was waived as samples were taken under a hospital surveillance framework for routine sampling. The research conformed to the principles of the Helsinki Declaration.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

¹Jiangsu Co-Innovation Center for Prevention and Control of Important Animal Infectious Diseases and Zoonoses, College of Veterinary Medicine, Yangzhou University, 48 East Wenhui Road, Yangzhou, Jiangsu 225009, P. R. China. ²Department of Clinical Laboratory, Second Affiliated Hospital of Zhejiang University, School of Medicine, Hangzhou, Zhejiang, P. R. China. ³College of Veterinary Medicine, China Agricultural University, Beijing, P. R. China. ⁴Jiangsu Key Lab of Zoonosis, Yangzhou University, Yangzhou, Jiangsu, P. R. China. ⁵Institute of Comparative Medicine, Yangzhou University, Yangzhou, Jiangsu, P. R. China.

Received: 13 October 2023 Accepted: 4 March 2024

Published online: 16 April 2024

References

- van der Bij AK, Pitout JDD. The role of international travel in the worldwide spread of multiresistant Enterobacteriaceae. *J Antimicrob Chemother.* 2012;67(9):2090–100.
- Zhang R, Liu L, Zhou H, et al. Nationwide surveillance of clinical carbapenem-resistant Enterobacteriaceae (CRE) strains in China. *EBioMedicine.* 2017;19:98–106.
- Yong D, Toleman MA, Giske CG, et al. Characterization of a new metallo-beta-lactamase gene, *bla*_{NDM-1}, and a novel erythromycin esterase gene carried on a unique genetic structure in *Klebsiella pneumoniae* sequence type 14 from India. *Antimicrob Agents Chemother.* 2009;53(12):5046–54.
- Logan LK, Weinstein RA. The epidemiology of carbapenem-resistant Enterobacteriaceae: the impact and evolution of a global menace. *J Infect Dis.* 2017;215:S28–36.
- Albiger B, Glasner C, Struelens MJ, Grundmann H, Monnet DL. European Survey of Carbapenemase-Producing Enterobacteriaceae working group. Carbapenemase-producing Enterobacteriaceae in Europe: assessment by national experts from 38 countries. *Euro Surveill.* 2015;20(45):30062.
- Zhang R, Li Y, Chen J, et al. Population genomic analysis reveals the emergence of high-risk carbapenem-resistant *Escherichia coli* among ICU patients in China. *J Infect.* 2023;86(4):316–28.
- Hassan B, Ijaz M, Khan A, et al. A role for arthropods as vectors of multidrug-resistant Enterobacteriales in surgical site infections from South Asia. *Nat Microbiol.* 2021;6(10):1259–70.
- Otter JA, Burgess P, Davies F, et al. Counting the cost of an outbreak of carbapenemase-producing Enterobacteriaceae: an economic evaluation from a hospital perspective. *Clin Microbiol Infect.* 2017;23(3):188–96.
- Berglund F, Ebmeyer S, Kristiansson E, Larsson DGJ. Evidence for wastewater as environments where mobile antibiotic resistance genes emerge. *Commun Biol.* 2023;6(1):321.
- Djenadi K, Zhang L, Murray AK, Gaze WH. Carbapenem resistance in bacteria isolated from soil and water environments in Algeria. *J Glob Antimicrob Resist.* 2018;15:262–7.
- Wang ZX, Wen Z, Jiang M, et al. Dissemination of virulence and resistance genes among *Klebsiella pneumoniae* via outer membrane vesicle: An important plasmid transfer mechanism to promote the emergence of carbapenem-resistant hypervirulent *Klebsiella pneumoniae*. *Transbound Emerg Dis.* 2022;69(5):E2661–76.
- Zurfluh K, Bagutti C, Brodmann P, et al. Wastewater is a reservoir for clinically relevant carbapenemase- and 16s rRNA methylase- producing Enterobacteriaceae. *Int J Antimicrob Ag.* 2017;50(3):436–40.
- Shen Z, Hu Y, Sun Q, et al. Emerging carriage of NDM-5 and MCR-1 in *Escherichia coli* from healthy people in multiple regions in China: a cross sectional observational study. *EClinicalMedicine.* 2018;6:11–20.
- Clinical and Laboratory Standards Institute [CLSI]. Performance standards for antimicrobial susceptibility testing. 28th ed. CLSI supplement M100. Wayne: Clinical and Laboratory Standards Institute; 2018.
- Bankevich A, Nurk S, Antipov D, et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol.* 2012;19(5):455–77.
- Kolmogorov M, Yuan J, Lin Y, Pevzner PA. Assembly of long, error-prone reads using repeat graphs. *Nat Biotechnol.* 2019;37(5):540–6.
- Wick RR, Judd LM, Gorrie CL, Holt KE. Unicycler: resolving bacterial genome assemblies from short and long sequencing reads. *PLoS Comput Biol.* 2017;13(6):e1005595.
- Siguier P, Perochon J, Lestrade L, Mahillon J, Chandler M. ISfinder: the reference centre for bacterial insertion sequences. *Nucleic Acids Res.* 2006;34(Database issue):D32–6.
- Carattoli A, Hasman H. PlasmidFinder and in silico pMLST: identification and typing of plasmid replicons in whole-genome sequencing (WGS). *Methods Mol Biol.* 2020;2075:285–94.
- Bortolaia V, Kaas RS, Ruppe E, et al. ResFinder 4.0 for predictions of phenotypes from genotypes. *J Antimicrob Chemother.* 2020;75(12):3491–500.
- Kleinheinz KA, Joensen KG, Larsen MV. Applying the ResFinder and VirulenceFinder web-services for easy identification of acquired antibiotic resistance and *E. coli* virulence genes in bacteriophage and prophage nucleotide sequences. *Bacteriophage.* 2014;4(1):e27943.
- Bessonov K, Laing C, Robertson J, et al. ECTyper: in silico *Escherichia coli* serotype and species prediction from raw and assembled whole-genome sequence data. *Microb Genom.* 2021;7(12):000728.
- Beghain J, Bridier-Nahmias A, Le Nagard H, Denamur E, Clermont O. ClermontTyping: an easy-to-use and accurate in silico method for *Escherichia* genus strain phylotyping. *Microb Genom.* 2018;4(7):e000192.
- Sullivan MJ, Petty NK, Beatson SA. Easyfig: a genome comparison visualizer. *Bioinformatics.* 2011;27(7):1009–10.
- Alikhan NF, Petty NK, Ben Zakour NL, Beatson SA. BLAST Ring Image Generator (BRIG): simple prokaryote genome comparisons. *BMC Genomics.* 2011;12:402.
- Seemann T. Snippy: rapid haploid variant calling and core genome alignment. 2020. <https://github.com/tseemann/snippy>.
- Page AJ, Cummins CA, Hunt M, et al. Roary: rapid large-scale prokaryote pan genome analysis. *Bioinformatics.* 2015;31(22):3691–3.
- Price MN, Dehal PS, Arkin AP. FastTree: computing large minimum evolution trees with profiles instead of a distance matrix. *Mol Biol Evol.* 2009;26(7):1641–50.

29. Letunic I, Bork P. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Res.* 2021;49(W1):W293–6.
30. Brynildsrud O, Bohlin J, Scheffer L, Eldholm V. Rapid scoring of genes in microbial pan-genome-wide association studies with Scoary. *Genome Biol.* 2016;17(1):238.
31. Seemann T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics.* 2014;30(14):2068–9.
32. Cantalapiedra CP, Hernandez-Plaza A, Letunic I, Bork P, Huerta-Cepas J. eggNOG-mapper v2: functional annotation, orthology assignments, and domain prediction at the metagenomic scale. *Mol Biol Evol.* 2021;38(12):5825–9.
33. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* 2000;28(1):27–30.
34. Bastian M, Heymann S, Jacomy M. Gephi: an open source software for exploring and manipulating networks. *Proceedings of the international AAAI conference on web and social media.* 2009;3(1):361–2.
35. Wu W, Feng Y, Tang G, Qiao F, McNally A, Zong Z. NDM metallo-beta-lactamases and their bacterial producers in health care settings. *Clin Microbiol Rev.* 2019;32(2):e00115–18.
36. Hu X, Xu X, Wang X, et al. Diversity of New Delhi metallo-beta-lactamase-producing bacteria in China. *Int J Infect Dis.* 2017;55:92–5.
37. Zhang Y, Xu D, He Z, Han J, Qu D. Characterization and fitness cost analysis of two plasmids carrying different subtypes of bla in aquaculture farming. *Food Microbiol.* 2023;115:104327.
38. Ali A, Gupta D, Srivastava G, Sharma A, Khan AU. Molecular and computational approaches to understand resistance of New Delhi metallo beta-lactamase variants (NDM-1, NDM-4, NDM-5, NDM-6, NDM-7)-producing strains against carbapenems. *J Biomol Struct Dyn.* 2019;37(8):2061–71.
39. Liu Z, Wang Z, Lu X, et al. Structural diversity, fitness cost, and stability of a bla_{NDM-1}-bearing cointegrate plasmid in *Klebsiella pneumoniae* and *Escherichia coli*. *Microorganisms.* 2021;9(12):2435.
40. Sun J, Chen C, Cui CY, et al. Plasmid-encoded tet(X) genes that confer high-level tigecycline resistance in *Escherichia coli*. *Nat Microbiol.* 2019;4(9):1457–64.
41. Li Y, Sun XR, Xiao X, Wang ZQ, Li RC. Global distribution and genomic characteristics of tet(X)-positive *Escherichia coli* among humans, animals, and the environment. *Sci Total Environ.* 2023;887:164148.
42. Wang Y, Tian GB, Zhang R, et al. Prevalence, risk factors, outcomes, and molecular epidemiology of mcr-1-positive Enterobacteriaceae in patients and healthy adults from China: an epidemiological and clinical study. *Lancet Infect Dis.* 2017;17(4):390–9.
43. Ding Y, Zhuang H, Zhou J, et al. Epidemiology and genetic characteristics of carbapenem-resistant *Escherichia coli* in Chinese intensive care unit analyzed by whole-genome sequencing: a prospective observational study. *Microbiol Spectr.* 2023;11(2):e0401022.
44. Gong L, Tang N, Chen D, et al. A nosocomial respiratory infection outbreak of carbapenem-resistant *Escherichia coli* ST131 with multiple transmissible bla_{KPC-2} carrying plasmids. *Front Microbiol.* 2020;11:2068.
45. Zhu WJ, Wang X, Qin JX, Liang W, Shen Z. Dissemination and stability of the bla_{NDM-5}-carrying IncX3-type plasmid among multiclinal *Klebsiella pneumoniae* isolates. *Mosphere.* 2020;5(6):e00917–20.
46. Ma T, Fu J, Xie N, et al. Fitness cost of bla_{NDM-5}-carrying p3R-IncX3 plasmids in wild-type NDM-free enterobacteriaceae. *Microorganisms.* 2020;8(3):377.
47. Acman M, Wang R, van Dorp L, et al. Role of mobile genetic elements in the global dissemination of the carbapenem resistance gene bla_{NDM}. *Nat Commun.* 2022;13(1):1131.
48. Vila J, Saez-Lopez E, Johnson JR, et al. *Escherichia coli*: an old friend with new tidings. *Fems Microbiol Rev.* 2016;40(4):437–63.
49. Bonten M, Johnson JR, van den Biggelaar AHJ, et al. Epidemiology of *Escherichia coli* bacteremia: a systematic literature review. *Clin Infect Dis.* 2021;72(7):1211–9.
50. Royer G, Clermont O, Marin J, et al. Epistatic interactions between the high pathogenicity island and other iron uptake systems shape *Escherichia coli* extra-intestinal virulence. *Nat Commun.* 2023;14(1):3667.
51. Raymond KN, Dertz EA, Kim SS. Enterobactin: an archetype for microbial iron transport. *Proc Natl Acad Sci U S A.* 2003;100(7):3584–8.
52. Perry RD, Fetherston JD. Yersiniabactin iron uptake: mechanisms and role in *Yersinia pestis* pathogenesis. *Microbes Infect.* 2011;13(10):808–17.
53. Kenny B. Mechanism of action of EPEC type III effector molecules. *Int J Med Microbiol.* 2002;291(6–7):469–77.

Publisher's Note

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