

RESEARCH

Open Access



# Intricate interplay of CRISPR-Cas systems, anti-CRISPR proteins, and antimicrobial resistance genes in a globally successful multi-drug resistant *Klebsiella pneumoniae* clone

Jianping Jiang<sup>1†</sup>, Astrid V. Cienfuegos-Gallet<sup>1,2†</sup>, Tengfei Long<sup>1,3,4†</sup>, Gisele Peirano<sup>5†</sup>, Tingyu Chu<sup>1</sup>, Johann D. D. Pitout<sup>5,6,7</sup>, Barry N. Kreiswirth<sup>1,8\*</sup> and Liang Chen<sup>1,8,9\*</sup> 

## Abstract

**Background** *Klebsiella pneumoniae* is one of the most prevalent pathogens responsible for multiple infections in healthcare settings and the community. *K. pneumoniae* CG147, primarily including ST147 (the founder ST), ST273, and ST392, is one of the most globally successful MDR clone linked to various carbapenemases.

**Methods** One hundred and one CG147 strains were sequenced and additional 911 publicly available CG147 genome sequences were included for analysis. The molecular epidemiology, population structure, and time phylogeny were investigated. The virulome, resistome, and mobilome were analyzed, and the recombination in the capsular region was studied. The CRISPR-Cas and anti-CRISPR were identified. The interplay between CRISPR-Cas, anti-CRISPR, and carbapenemase-encoding plasmids was analyzed and experimentally validated.

**Results** We analyzed 1012 global CG147 genomes, with 80.4% encoding at least one carbapenemase (NDM [529/1012, 52.3%], OXA-48-like [182/1012, 17.7%], and KPC [105/1012, 10.4%]). Surprisingly, almost all CG147 strains (99.7%, 1009/1,012) harbor a chromosomal type I-E CRISPR-Cas system, with 41.8% (423/1012) containing an additional plasmid-borne type IV-A3 CRISPR-Cas system, and both target IncF plasmids, e.g., the most prevalent KPC-encoding pKpQIL-like plasmids. We found the presence of IV-A3 CRISPR-Cas system showed a negative correlation with the presence of KPC. Interestingly, a prophage-encoding anti-CRISPR AcrIE8.1 and a plasmid-borne anti-CRISPR AcrIE9.2 were detected in 40.1% (406/1012) and 54.2% (548/1012) of strains, respectively, which displayed positive correlations with the presence of a carbapenemase. Plasmid transfer experiments confirmed that the I-E and IV-A3 CRISPR-Cas systems significantly decreased ( $p < 0.001$ ) KPC-encoding pKpQIL plasmid conjugation frequencies, while the AcrIE8.1 and AcrIE9.2 significantly increased ( $p < 0.001$ ) pKpQIL conjugation frequencies and protected plasmids from elimination by CRISPR-Cas I-E system.

<sup>†</sup>Jianping Jiang, Astrid V. Cienfuegos-Gallet, Tengfei Long and Gisele Peirano contributed equally to this work.

\*Correspondence:

Barry N. Kreiswirth  
barry.kreiswirth@hnh-cdi.org  
Liang Chen  
liangch@buffalo.edu

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



**Conclusions** Our results indicated a complex interplay between CRISPR-Cas, anti-CRISPR, and mobile genetic elements that shape the evolution of CG147. Our findings advance the understanding of multi-drug resistance mechanisms and will aid in preventing the emergence of future MDR clones.

**Keywords** *Klebsiella pneumoniae*, CG147, CRISPR-Cas, Anti-CRISPR, Genomic epidemiology

## Background

*Klebsiella pneumoniae* is one of the most prevalent pathogens, responsible for a large range of infections, such as bloodstream-associated, intra-abdominal, respiratory tract, and urinary tract infections. In addition, *K. pneumoniae* can harbor a substantial number of antimicrobial resistance (AMR) genes, making it a significant AMR pathogen in global health [1].

Several globally successful and epidemiologically predominant clones that disproportionately contribute to the global disease burden have been identified in *K. pneumoniae* [2]. The clonal group (CG) 258, which frequently carries *K. pneumoniae* carbapenemases (KPCs), is a clinically significant example that has disseminated globally [3]. Recently, another globally successful MDR clone, CG147, primarily including sequence type (ST) 147 (the founder ST), ST273, ST392, and other minor variants, has been increasingly highlighted [4, 5]. CG147 was initially reported in Hungary and Spain in 2008–2009, characterized by the production of CTX-M-15 extended-spectrum  $\beta$ -lactamase (ESBL), and the harboring of *gyrA* S83I and *parC* S80I mutations in the quinolone resistance-determining region (QRDR) [6, 7]. Since 2009, CG147 strains encoding KPC, VIM, NDM, and OXA-48-like carbapenemases have emerged and disseminated across European countries, subsequently extending to Asia and Northern and Southern American countries within a short period [8–13], and caused multiple outbreaks [11, 14, 15].

Within the *K. pneumoniae* population, horizontal gene transfer (HGT) of mobile genetic elements (MGEs), specifically AMR plasmids, plays a substantial role in both the development and spread of globally successful clones, including CG147. In contrast to gene movement and to prevent exogenous nucleic acid invasion, bacterial hosts have developed various defense mechanisms against HGT, among which the CRISPR-Cas system is the most important [16]. The CRISPR-Cas system is an adaptive immune system consisting of clustered regularly inter-spaced short palindromic repeats (CRISPRs) and CRISPR-associated (Cas) proteins [17]. In *K. pneumoniae*, the chromosome-borne type I-E CRISPR-Cas system has been reported to restrict the transfer and maintenance of *bla*<sub>KPC</sub>-harboring IncF plasmids [18], and predictably, their presence correlate with the low detection of IncF plasmids in specific clones, such as in ST23 and ST15 [18–20].

Conversely, to counteract the CRISPR-Cas system, MGEs have evolved various strategies to evade immunity [16]. One prevalent mechanism is the production of CRISPR-Cas inhibitors known as anti-CRISPRs (Acrcs) [21]. Notable examples of Acrcs such as prophage-borne AcrIE8.1 and plasmid-borne AcrIE9.2 that block the I-E CRISPR-Cas system have been reported in certain *K. pneumoniae* clones and other species [22, 23].

Previous genomic studies have shown that CG147 strains, similar to ST23, universally harbor a chromosome-based I-E CRISPR-Cas system [24]. However, unlike ST23, CG147 has successfully evolved into a global MDR successful clone, despite harboring the CRISPR-Cas system. The factors contributing to the epidemiological success of *K. pneumoniae* CG147 are unclear and therefore warrants further investigation. In this study, we investigate 1012 global *K. pneumoniae* CG147 strains, reconstruct their molecular evolution, and explore the interplay between the CRISPR-Cas systems, anti-CRISPR proteins, and MGEs. Our study provides an in-depth understanding of this unique globally successful MDR clone.

## Methods

### *K. pneumoniae* clonal group 147 genomes

CG147 contains ST147, ST273, ST392, and other minor variants including ST1880, ST2292, ST2358, and ST3740. A total of 911 CG147 isolates collected worldwide were retrieved from the NCBI whole genome sequence database or short reads archive (dated as 12/2022), comprising all the genomes publicly available at the time of the study. Additional 101 CG147 isolates from our collection were also included in this study. The accession number, host, isolation source, geographical origin, and genotype data are listed in the Additional file 1: Table S1.

### Whole-genome sequencing and analysis

Whole-genome sequencing was performed as previously described using Illumina NextSeq 500 platform with 2×150 bp paired-end reads [25]. Draft genomes were assembled using SPAdes v3.13.0 [26]. Additionally, two strains, Kp46564 and Kp46596, were sequenced using Oxford Nanopore and assembled to closure via Unicycler 0.4.9 [27] leveraging both short and long sequencing reads. Kleborate v2.3 [28] was employed to screen the genome assemblies of *K. pneumoniae* CG147 for the

prediction of sequence type, capsular type and the detection of virulence loci, including hypermucoity genes *rmpADC*, *rmpA2* and siderophore systems, such as yersiniabactin (*ybt*), colibactin (*clb*), aerobactin (*iuc*), salmochelin (*iro*), as well as their associated mobile elements (ICEKp). Resistance genes were called by AMRFinderPlus v3.10.20 [29] and ARIBA v2.14.6 [30]. CRISPR-Cas was identified using cctyper v1.8.0 [31]. Redundant CRISPR arrays and spacers were removed by CD-HIT v4.8.1 [32]. The self-targeting spacer sequence, defined as an additional copy of spacer sequences located in the genomes but that are not in the CRISPR array, was identified by BLASTn v2.12.0 [33]. Anti-CRISPR proteins were detected by BLAST against the AcrDB [34] database. The prophages were identified by PHASTEST [35]. The linear alignment of prophages were visualized by Easyfig v2.1 [36]. Plasmid replicons were identified by ABRicate v1.01 [37] using the PlasmidFinder [38] database with 90% identity and 80% query coverage cutoffs.

#### ***K. pneumoniae* CG147 strains phylogeny**

The paired-end raw fastq reads were mapped to the reference genome Kp46564 (Genbank: CP059317.1) using Snippy v4.6.0 [39]. If raw sequencing data of genomes from public database are not available, genome assemblies were used to simulate 10 M 150 bp paired-end reads using the wgsim algorithm from Samtools v1.16.1 [40], followed by mapping to reference as describe above. Repeated and recombination regions were detected using MUMmer v4.0.0 [41] and Gubbins v3.0.0 [42], respectively, followed by filtering in Snippy analysis. The maximum likelihood phylogenetic tree was constructed using RAxML v1.0.1 [43].

To estimate node dates of CG147 strains, a subset of 306 genomes was selected and analyzed using the BactDating v1.1 R package [44]. The 306 genomes were selected according following priority: (a) covering all the major clades on the maximum likelihood phylogenetic tree; (b) having available meta-data, including time and geographical information; (c) having available raw sequencing data; and (d) having a high N50 if raw sequence data were not available. The isolation time and the recombination-corrected tree from Gubbins output were used as the inputs in BactDating. We implemented a mixed model with  $10^8$  iterations to ensure that the Markov chain Monte Carlo (MCMC) was run for long enough to converge (the effective sample size of the inferred parameters  $\alpha$ ,  $\mu$ , and  $\sigma$  were  $>200$ ). Three BactDating replicates and one with a randomized tip date were conducted, and the convergence were evaluated by Gelman diagnostic using the coda v0.19 R package. The temporal signal significance was determined by comparing the first replicate model to the model with

randomized tip date using the model compare function of the BactDating package. The maximum likelihood phylogenetic tree and resulting BactDating tree were annotated using iTOL v5 [45].

For CRISPR-Cas systems and self-targeting spacer analysis, the genomes were dereplicated using Assembly Dereplicator v0.3.2 [46] with a cutoff of mash distance 0.0001 to avoid interference from duplicated strains.

#### **CPS switch analysis**

The recombination regions of different subclones corresponding to the reference genome Kp46564 were extracted from the Gubbins output. The potential donor of the CPS recombination region was analyzed according to the previous methods [47, 48]. The SNP densities of KP131\_H (Biosample: SAMN13301701), 47,226 (Biosample: SAMN37917565), 5085 (Biosample: SAMN20338550), 50,490 (Biosample: SAMN37917589), 44,383 (Biosample: SAMN37917502), 4300STDY6542362 (Biosample: SAMEA4394732), 53,323 (Biosample: SAMN37917624), EuSCAPE\_TR302 (Biosample: SAMEA3729896), 49,189 (Biosample: SAMN44379006), 53,330 (Biosample: SAMN37917631), and AR\_0097-R3 (Biosample: SAMN29227743) were analyzed in comparison to the reference strain Kp46564. The SNP density plot was generated by ggplot2 v3.4.4 R package.

#### **Prediction of the co-existence of self-targeting sequence and carbapenemas-encoding genes**

The self-targeting sequence and carbapenemas-encoding gene were considered to co-exist on the same fragment if one of the following criteria was fulfilled: (a) they co-existed on the same assembled plasmid or contig and (b) the self-targeting sequence and carbapenemas-encoding gene carrying contigs had the same average read coverages, with evidence supporting their potential co-existence based on the NCBI *Klebsiella* plasmid database. To estimate the average reads coverages, the raw reads were mapped to the contig sequences by BWA v0.7.17 [49]. The reads coverage for each site (non-coding, repeat and IS regions were excluded) was calculated by bedtools v2.30.0 [50]. Subsequently, 100 sites were randomly selected, and coverages were compared by T test.  $p < 0.05$  was considered as statistically significant.

#### **Construction of the *acrIE9.2* and IETarget mutant**

The *acrIE9.2* knockout mutant in clinical plasmid pKp-QIL-03 [51] was constructed using the two-plasmid CRISPR-Cas9 system, as previously described [52, 53]. In brief, the pCasKp-Apr was electroporated into pKpQIL-03-harboring *E. coli* DH10B and selected on LB agar plates containing 30  $\mu\text{g/ml}$  apramycin (APR) and 1  $\mu\text{g/ml}$  meropenem (MEM) at 30°C. The 20-nt base pairing

region (N20) of a sgRNA was designed using Geneious Prime 2023 and inserted into the plasmid pSgKp-Rif to generate pSgKp-TacrIE9.2 via Golden Gate Cloning. A 90 bp ssDNA oligonucleotide repair template was synthesized from Azenta Life Sciences (South Plainfield, NJ). pSgKp-TacrIE9.2 and ssDNA repair template were mixed and electroporated into arabinose-induced pKpQIL-03 and pCasKp-Apr harboring *E. coli* DH10B competent cell. Electroporants were selected on LB plates supplemented with rifampicin (RIF, 100 µg/ml), 30 µg/ml APR, and 1 µg/ml MEM, followed by PCR confirmation. Single colonies were transferred to LB agar plates containing 5% sucrose and incubated at 37 °C to remove pSgKp-TacrIE9.2 and pCasKp-Apr vectors to generate pKpQIL-03ΔacrIE9.2-harboring *E. coli* DH10B. Subsequently, a 1315 bp region, containing the four common type I-E protospacers on pKpQIL-03ΔacrIE9.2 was knocked out using the same method, resulting in *E. coli* DH10B harboring the pKpQIL-03ΔacrIE9.2ΔIETarget plasmid.

#### Type I-E CRISPR-Cas spacers, *acrIE8.1*, *acrIE9.2*, and type IV-A3 CRIPSR-Cas system cloning

The sequences of *acrIE8.1* from Kp46596, *acrIE9.2* from pKpQIL-03, and IV-A3 CRIPSR-Cas systems, including CasIVa3-1 from BK42810 with 19 spacers and CasIVa3-2 from BK54145 with 7 spacers, were amplified using Q5 Hot Start DNA polymerase (NEB) with pKpQIL-03 plasmid DNA or the genomic DNA from BK42810 and BK54145 as the templates. Purified PCR products were then integrated into the pClone-Apr plasmid following manufacturer instructions for NEBuilder HiFi DNA assembly. Correct constructs were identified through PCR and confirmed by Sanger sequencing, resulting in the creation of plasmid pClone-*acrIE8.1*, pClone-*acrIE9.2*, pClone-CasIVa3-1, and pClone-CasIVa3-2.

#### Experimental validation the function of CRISPR-Cas systems and anti-CRISPRs

To validate the immunity of the I-E CRISPR-Cas system, we used pKpQIL-03ΔacrIE9.2-harboring and pKpQIL-03ΔacrIE9.2ΔIETarget-harboring DH10B as the donor strains, respectively, and *K. pneumoniae* strain BK56682 (containing an intact type I-E CRISPR-Cas system with the corresponding spacers against pKpQIL-03) as the recipient strain and conducted a mating conjugation assay. In brief, donor and recipient strains were grown at log phase before mixing in a 1:1 ratio and placed onto a 0.22-µm filter paper on LB agar and incubated at 37°C. After 6-h mating, the filter paper was washed, and the mixtures were series diluted and plated onto APR and MEM plates or APR only plates. The conjugation efficiency was calculated as the number of transconjugants divided by the number of recipient strains.

To validate the anti-CRISPR function of the *acrIE8.1* and *acrIE9.2*, we use pClone-*acrIE8.1* and pClone-*acrIE9.2*-harboring *K. pneumoniae* BK56682 as the recipient strains, respectively, and pKpQIL-03ΔacrIE9.2-harboring DH10B as the donor strain and conducted a mating conjugation assay as above.

For the function of the IV-A3 CRISPR-Cas system, we used two IV-A3 CRISPR-Cas system-harboring DH5a constructs (pClone-CasIVa3-1 with the IV-A3 CRIPSR-Cas system from BK42810 and pClone-CasIVa3-2 with the IV-A3 CRIPSR-Cas system from BK54145) as the recipient strains and pKpQIL-03ΔacrIE9.2-harboring DH10B as the donor strain and conducted a mating conjugation assay as above. An empty pClone-Apr vector was used as control. The conjugation efficiency was calculated as above.

#### Stability of newly formed transconjugants

To examine the stability of pKpQIL-03 within an I-E CRISPR-Cas bacterial host, and the impact of anti-CRISPR proteins (*AcrIE8.1* and *AcrIE9.2*), we performed serial passages of the transconjugants of pKpQIL-03ΔacrIE9.2-harboring BK56682 with and without *acrIE8.1* and *acrIE9.2* genes for 3 days for a total of 60 generations in fresh LB. After every 10 passages, the culture was diluted (1:100) into fresh LB and incubated with 200 rpm shaking. To measure the percentage of pKpQIL-03ΔacrIE9.2-positive strains, the plasmid stability was calculated as the percentage of colonies grown on LB agar with 1 µg/ml MEM to that on LB agar without MEM. Strains, vectors, and primers used in this study were listed in Additional file 1: Table S2 and Table S3.

#### Statistical analysis

We performed a descriptive analysis to describe CG147 isolates. Categorical data was summarized using absolute and relative frequencies, and the distribution of quantitative data was summarized using the median and interquartile range. The transfer frequencies were compared by T test, and  $p < 0.05$  was considered as statistically significant. KPC, NDM, OXA-48-like carbapenemase, or CTX-M ESBL was analyzed respectively against STs, regions of isolation, IV-A3 CRISPR-Cas system, *AcrIE8.1*, *AcrIE9.2*, and *AcrIF11* with logistic regression model. All the variables were treated as categories. The logistic regression was performed by using R and Wald tests were used for each variable with  $p < 0.05$  considered as statistically significant.

## Results

### General characteristics of CG147

The 1012 strains originated from 59 different countries and spanned the period between 2002 and 2020.

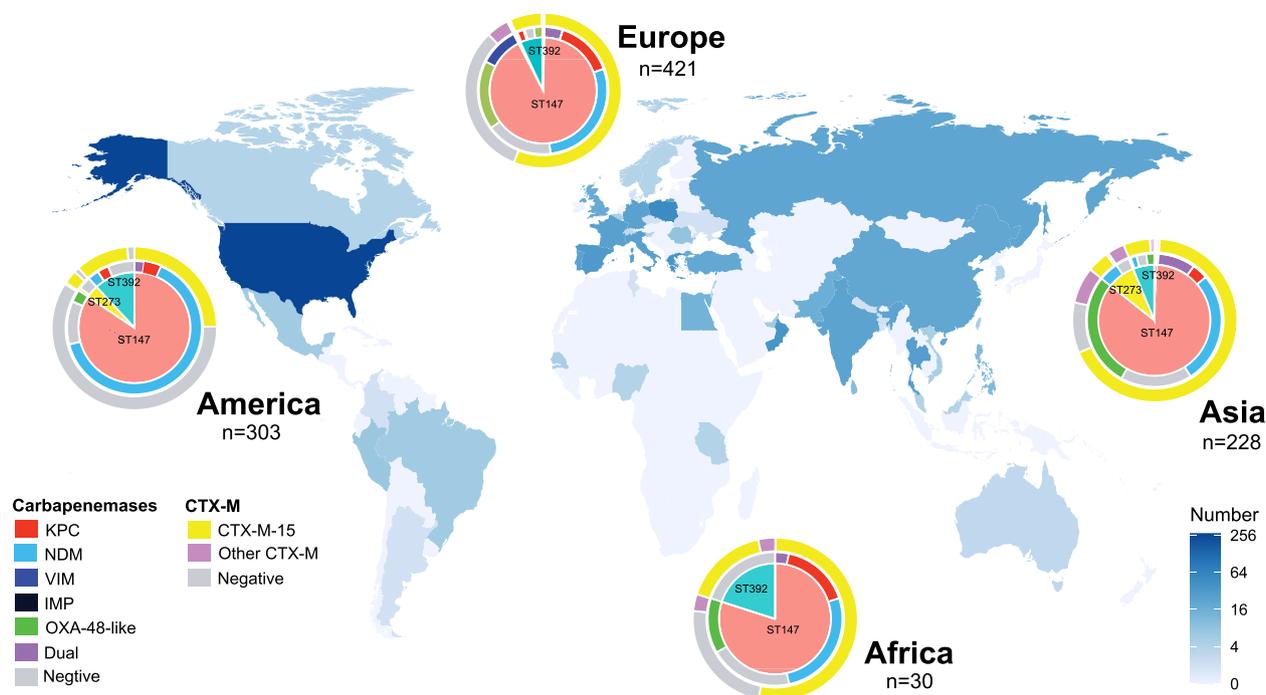
Isolates were collected from humans ( $n=954$ , 94.3%), animals ( $n=10$ , 1%), the environment ( $n=14$ , 1.4%), and unknown ( $n=34$ , 3.4%) sources. Isolates of human origin were obtained frequently from rectal swabs or stools ( $n=302$ , 31.6%), followed by urine ( $n=176$ , 18.5%), blood ( $n=99$ , 10.4%), and the upper respiratory tract ( $n=65$ , 6.8%). Most isolates were collected in North America ( $n=278$ , 27.5%) and countries in Southern Europe ( $n=241$ , 23.8%), Eastern Europe ( $n=86$ , 8.5%), and Western Asia ( $n=76$ , 7.5%) (Fig. 1). The major sequence types include ST147 ( $n=881$ , 87.1%), ST392 ( $n=85$ , 8.4%), and ST273 ( $n=36$ , 3.6%) (Table 1).

*K. pneumoniae* CG147 strains harbor AMR genes conferring resistance to multiple antibiotic families, including  $\beta$ -lactams, fluoroquinolones, aminoglycosides, rifampicin, tetracyclines, sulfonamides, and trimethoprim. Genes conferring resistance to tigecycline, colistin, and fosfomycin were less commonly observed (Additional file 1: Table S1). In this collection, we detected that 67.6% ( $n=684$ ) of isolates carried at least one CTX-M ESBL gene, with CTX-M-15 ( $n=647$ , 94.6%) being the most frequent. Notably, 80.4% ( $n=814$ ) of isolates carried at least one carbapenemase gene. *bla*<sub>NDM</sub> was detected in 52.2% ( $n=529$ ) of isolates, *bla*<sub>OXA-48-like</sub> in 17.9% ( $n=182$ ) of isolates, *bla*<sub>KPC</sub> in 10.4% ( $n=105$ ), and *bla*<sub>VIM</sub> in 4.5% ( $n=46$ ) of isolates. Interestingly, in 2010, the frequency of KPC and NDM was similar; however, after 2014, NDM

predominated among CG147 strains (Additional file 2: Fig. S1). The details of AMR genes, plasmids, capsular polysaccharide (CPS) types, and virulence factors are summarized in Additional file 3: Supplementary results.

### The population structure and phylogenetic dating of CG147

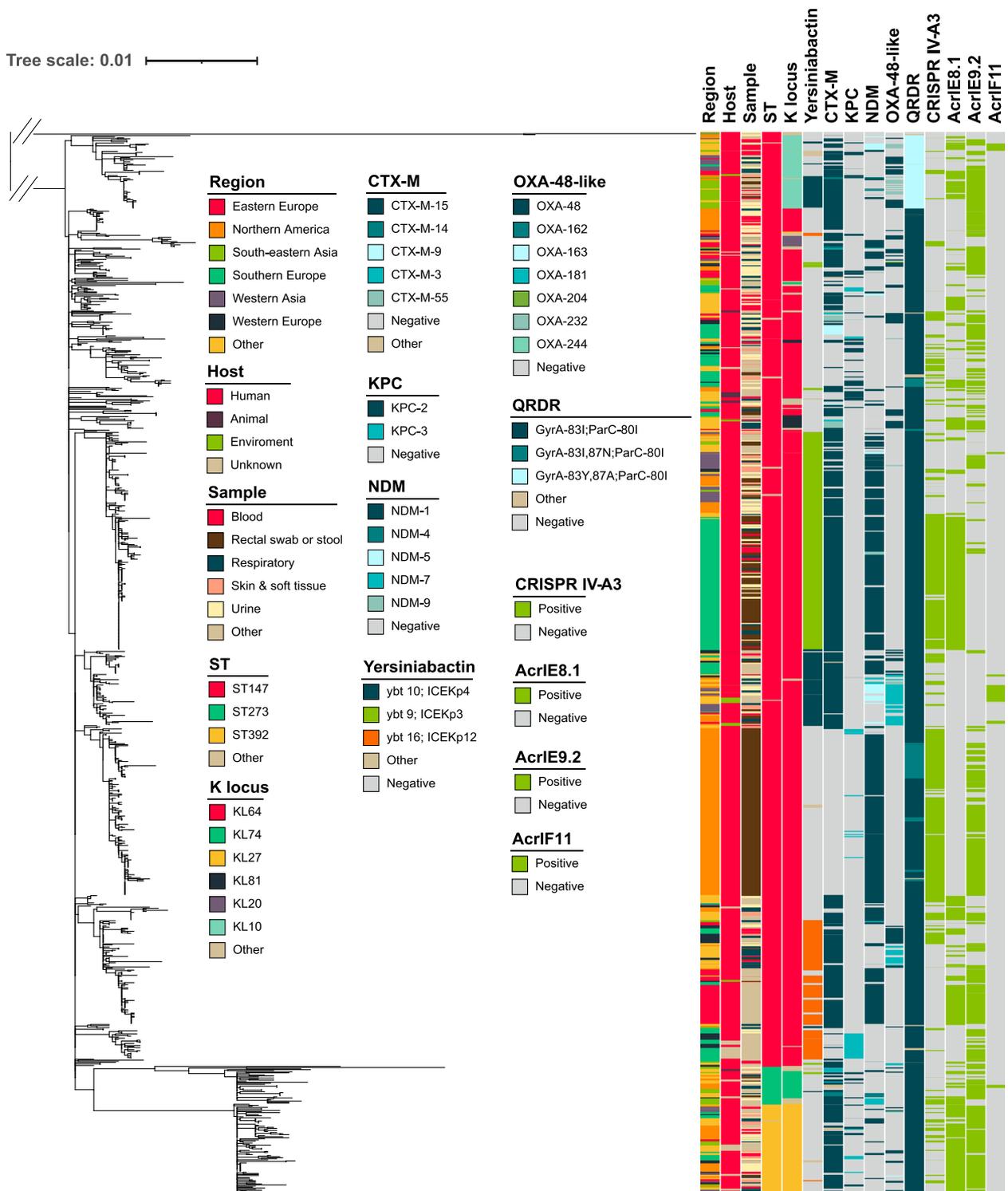
The CG147 phylogeny revealed a basal clade containing two quinolone susceptible isolates of ST147-KL55 and a larger clade containing the remaining isolates which carried QRDR mutations. In this larger clade, three primary branches emerged, corresponding to ST147 ( $n=879$ , 86.9%), ST273 ( $n=36$ , 3.6%), and ST392 ( $n=85$ , 8.4%). Within the ST147 clade, two additional branches were identified. The first branch comprised 70 isolates, predominantly ST147 ( $n=69$ , 98.6%) with one ST1880. Most of ST147 isolates within this branch encoded the capsular loci KL10 (66/69, 95.7%). These strains were collected in Southeastern Asia (27/69, 39.1%) and Western Asia (12/69, 17.4%), primarily from human hosts (66/69, 95.7%) and isolated from blood (17/69, 24.6%) and urine (15/69, 21.7%) specimens. The QRDR mutation GyrA-83Y-GyrA-87A-Par80I (67/69, 97.1%) was frequently found in this clade. Additionally, a small subclade within ST147-KL10 encoded yersiniabactin (*ybt10*-ICEKp4) (29/66, 43.3%) (Fig. 2).



**Fig. 1** Isolate map depicting the collection sites of 1012 CG147 isolates over the world. Pie graphs show ST distribution in each continent, and donut graphs show carbapenemase (inner ring) and CTX-M (outer ring) distribution within each ST

**Table 1** General characteristics of CG147 genomes publicly available from 2002 to 2020

Characteristics	ST147 (n=886)	ST273 (n=36)	ST392 (n=86)	Other ST (n=4)	Total (n=1012)
Continent (n, %)					
Africa	24 (2.71)	0 (0)	6 (6.98)	0 (0)	30 (2.96)
Asia	192 (21.67)	19 (52.78)	15 (17.44)	2 (50)	228 (22.53)
Europe	397 (44.81)	2 (5.56)	21 (24.42)	1 (25)	421 (41.6)
America	253 (28.56)	13 (36.11)	36 (41.86)	1 (25)	303 (29.94)
Other	2 (0.23)	0 (0)	1 (1.16)	0 (0)	3 (0.3)
K locus (n, %)					
Other KL	27 (3.05)	11 (30.56)	0 (0)	0 (0)	38 (3.75)
KL10	67 (7.56)	0 (0)	0 (0)	1 (25)	68 (6.72)
KL20	10 (1.13)	0 (0)	0 (0)	0 (0)	10 (0.99)
KL27	2 (0.23)	0 (0)	86 (100)	0 (0)	88 (8.7)
KL64	764 (86.23)	0 (0)	0 (0)	3 (75)	767 (75.79)
KL74	2 (0.23)	25 (69.44)	0 (0)	0 (0)	27 (2.67)
KL81	14 (1.58)	0 (0)	0 (0)	0 (0)	14 (1.38)
CTX-M (n, %)					
Negative	304 (34.31)	8 (22.22)	8 (9.30)	3 (75)	323 (31.92)
Other	4 (0.45)	3 (8.33)	3 (3.49)	0 (0)	10 (0.99)
CTX-M-3	5 (0.56)	3 (8.33)	0 (0)	0 (0)	8 (0.79)
CTX-M-9	11 (1.24)	0 (0)	0 (0)	0 (0)	11 (1.09)
CTX-M-14	3 (0.34)	4 (11.11)	0 (0)	0 (0)	7 (0.69)
CTX-M-15	549 (61.96)	18 (50.0)	75 (87.21)	1 (25)	643 (63.54)
CTX-M-55	10 (1.13)	0 (0)	0 (0)	0 (0)	10 (0.99)
KPC (n, %)					
Negative	803 (90.63)	30 (83.33)	72 (83.72)	2 (50.00)	907 (89.62)
KPC-2	42 (4.74)	6 (16.67)	8 (9.30)	2 (50.00)	58 (5.73)
KPC-3	41 (4.63)	0 (0)	6 (6.98)	0 (0)	47 (4.64)
NDM (n, %)					
Negative	380 (42.89)	26 (72.22)	74 (86.05)	3 (75)	483 (47.73)
NDM-1	466 (52.6)	1 (2.78)	12 (13.95)	0 (0)	479 (47.33)
NDM-4	4 (0.45)	4 (11.11)	0 (0)	0 (0)	8 (0.79)
NDM-5	28 (3.16)	0 (0)	0 (0)	1 (25)	29 (2.87)
NDM-7	5 (0.56)	5 (13.89)	0 (0)	0 (0)	10 (0.99)
NDM-9	3 (0.34)	0 (0)	0 (0)	0 (0)	3 (0.3)
OXA-48-like (n, %)					
Negative	720 (81.26)	34 (94.44)	74 (86.05)	2 (50.00)	830 (82.02)
OXA-48	91 (10.27)	2 (5.56)	10 (11.63)	0 (0)	103 (10.18)
OXA-162	3 (0.34)	0 (0)	0 (0)	0 (0)	3 (0.3)
OXA-163	0 (0)	0 (0)	1 (1.16)	0 (0)	1 (0.10)
OXA-181	52 (5.87)	0 (0)	0 (0)	1 (25)	53 (5.24)
OXA-204	1 (0.11)	0 (0)	0 (0)	0 (0)	1 (0.10)
OXA-232	19 (2.14)	0 (0)	0 (0)	1 (25)	20 (1.98)
OXA-244	0 (0)	0 (0)	1 (1.16)	0 (0)	1 (0.10)
Yersiniabactin (n, %)					
Other	10 (1.13)	2 (5.56)	2 (2.33)	0 (0)	14 (1.38)
Negative	439 (49.55)	31 (86.11)	82 (95.35)	2 (50.00)	554 (54.74)
ybt10-ICEKp4	101 (11.40)	1 (2.78)	0 (0)	2 (50.00)	104 (10.28)
ybt16-ICEKp12	121 (13.66)	0 (0)	2 (2.33)	0 (0)	123 (12.15)
ybt9-ICEKp3	215 (24.27)	2 (5.56)	0 (0)	0 (0)	217 (21.44)



**Fig. 2** Phylogenetic analysis of 1012 genomes of CG147 strains. Colors in columns illustrate regions of origin, host, specimen, sequence type (ST), K locus, yersiniabactin, CTX-M, KPC, NDM, OXA-48-like, QRDR substitutions (GyrA and ParC), CRISPR-Cas type IV-A3 system, and anti-CRISPR proteins

The second branch consisted of ST147 isolates ( $n=818$ ) predominantly encoding the KL64 capsular locus (767/818, 93.8%). The most common QRDR mutation pattern among ST147-KL64 isolates was GyrA-83I-ParC-80I (711/767, 92.7%), followed by GyrA-83I-GyrA-87N-ParC-80I (48/767, 6.3%), with a small proportion carrying other mutations (8/767, 1%). Within this branch, four subclades were identified, correlated with different yersiniabactin lineages. The first subclade comprised 207 isolates encoding *ybt9* in the ICEKp3 element (207/207, 100%). These isolates were exclusively recovered from human hosts (207/207, 100%), predominantly from Southern Europe (126/207, 60.9%), with rectal swabs or stool specimens being the most common sources (107/207, 51.7%). This subclade frequently carried  $\beta$ -lactamases such as CTX-M-15 (196/207, 94.7%) and NDM-1 (188/207, 90.8%), with no isolates encoding KPC carbapenemases (Fig. 2).

The second ST147-KL64 subclade carried *ybt10* in the ICEKp4 element (69/72, 95.8%). These isolates were collected primarily in North America (14/72, 19.4%), Southern Europe (12/72, 16.7%), and Western Asia (12/72, 16.7%). Most were recovered from human hosts (61/72, 84.7%), with urine (16/61, 26.2%), and respiratory specimens (12/61, 19.7%) being the most common sample types. This subclade commonly encoded ESBL CTX-M-15 (69/72, 95.8%) and OXA-48-like carbapenemases (50/72, 69.4%), primarily OXA-181 (37/50, 51.4%). NDM variants, mainly NDM-5 (16/72, 22.2%) and NDM-1 (15/72, 20.8%), were also present, though only one isolate carried KPC-2 (Fig. 2).

The third ST147-KL64 subclade ( $n=99$ ) primarily encoded *ybt16* in the ICEKp12 element (91/99, 91.9%). These strains were collected mostly in Eastern Europe (50/99, 50.1%), Southern Europe (8/99, 8.1%), and Western Europe (18/99, 18.2%) and were primarily of human origin (98/99, 98.9%). This subclade was frequently associated with CTX-M-15 (92/99, 92.3%) and NDM-1 (57/99, 57.6%), with some isolates also encoding OXA-48-like carbapenemases (32/99, 32.3%). None carried KPC carbapenemases, and CRISPR-IV-A3 was present in 17.2% (17/99) of isolates. A smaller subclade ( $n=27$ ) also carried *ybt16* in ICEKp12 (26/27, 96.3%), with 12/27 isolates from human sources. This subclade encoded KPC-3 (24/27, 88.9%) but no other carbapenemases (NDM or OXA-48-like). Only three isolates (11.1%) encoded CTX-M-15 (Fig. 2).

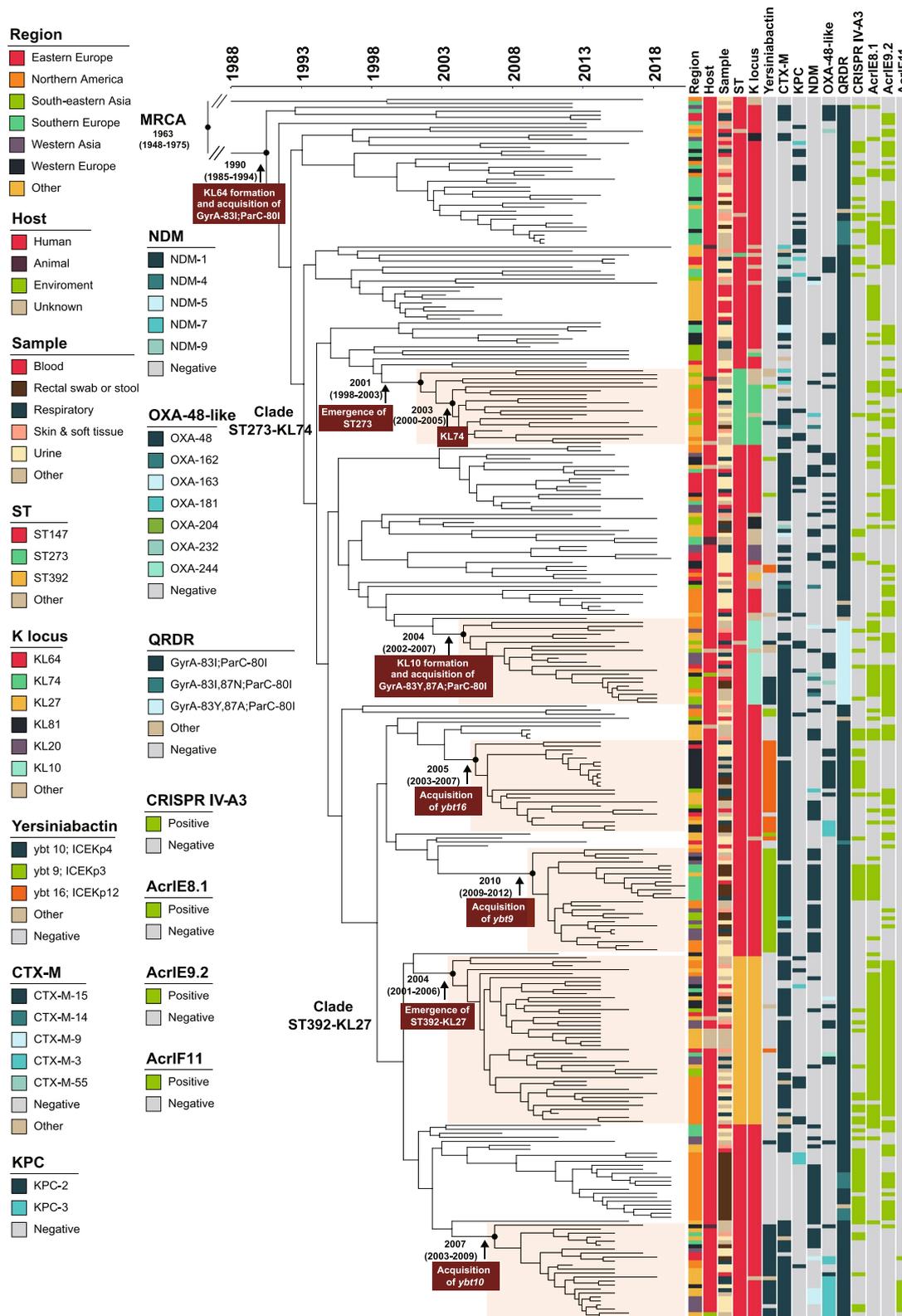
The second main branch included all ST273 isolates ( $n=36$ ), frequently encoding KL74 (25/36, 69.4%). These strains were mostly collected in Southeastern Asia (10/36, 27.8%) and North America (9/36, 25.0%). CTX-M-type ESBLs were common (28/36, 77.8%), mainly CTX-M-15 (18/36, 50.0%). Yersiniabactin was

rare among ST273 isolates, with 86.1% (31/36) testing negative. Only ten isolates carried NDM, two carried KPC-2, and another two had OXA-48 carbapenemases (Fig. 2).

The last main branch consisted of all ST392 isolates ( $n=86$ ), exclusively encoding KL27 (86/86, 100%). Most strains were from human sources (79/86, 91.9%), with specimen types including the urine (22/86, 25.6%), skin and soft tissue (13/86, 15.2%), respiratory (11/86, 12.8%), and blood (11/86, 12.8%). Yersiniabactin was rare in this branch, with 95.4% (82/86) testing negative; only two isolates carried *ybt16*-ICEKp12. CTX-M-15 was frequent (75/86, 87.2%), while carbapenemases were less common, with KPC detected in 17.3% (14/86), NDM-1 in 13.9% (12/86), and OXA-48-like in 13.9% (12/86) of isolates.

Three ST147-KL64 subclades from different regions independently acquired distinct yersiniabactin loci. In 2005 (95% CI, 2003–2007), a subclade primarily from Western Europe acquired *ybt16*; in 2009 (95% CI, 2009–2012), a Southern European subclade acquired *ybt9*; and in 2007 (95% CI, 2003–2009), a subclade from Western Asia and Eastern Europe acquired *ybt10*, suggesting geographically associated acquisition of yersiniabactin in CG147. The *rmpADC* and *rmpA2* loci were estimated to have been acquired in 2002 (95% CI, 2000–2004) and 1998 (95% CI, 1994–2002), respectively.

We then implemented a Bayesian phylogeny analysis to reconstruct the global evolutionary history of CG147 using a subset of 306 genomes. The BactDating model estimates that the most recent common ancestor of CG147 was around 1963 (95% confidence interval [CI], 1948–1975) (Fig. 3). Our phylogenetic analysis identified an ancestral branch corresponding to the two ST147 isolates harboring KL55, which were predicted to be susceptible to most antibiotics, except for ampicillin, due to the chromosomal *bla*<sub>SHV-11</sub> gene. Around 1990 (95% CI, 1985–1994), ST147 acquired two QRDR mutations (GyrA-83I and ParC-80I) and K locus (KL) 64, followed by acquisition of plasmids harboring resistance determinants, such as ESBLs and carbapenemases, and became an epidemic MDR cluster. Among this cluster, most strains were ST147-KL64; however, three major subclades evolved, including the emergence of ST273-KL74 in ~2003 (95% CI, 2000–2005), ST147-KL10 in ~2004 (95% CI, 2002–2007), and ST392-KL27 in ~2004 (95% CI, 2001–2006). Most of the CG147 strains ( $n=884$ ) harbored GyrA-83I and ParC-80I mutations; whereas, 48 strains harbored an additional GyrA-87N mutation. Interestingly, besides ParC-80I, the majority of ST147-KL10 strains (97.1%, 67/69) harbored the GyrA-83Y and GyrA-87A mutations, which appears to have originated by a genomic recombination event (see below). The detailed population structure of



**Fig. 3** Phylogenetic dating of 306 genomes of CG147 strains. Colors in columns illustrate regions of origin, host, specimen, sequence type (ST), K locus, yersiniabactin, CTX-M, KPC, NDM, OXA-48-like, QRDR substitutions (GyrA and ParC), type IV-A3 CRISPR-Cas system, AcrIE8.1, AcrIE9.2, and AcrIF11. Relevant evolutionary events are displayed in red boxes, and selected divergence time and 95% CIs are shown at the nodes

CG147 and phylogenetic dating of other genetic factors are described in Additional file 3: Supplementary results.

#### Capsular variation of CG147

Our phylogenetic dating showed ST147-KL64 was likely descendant from ST147-KL55, while ST392-KL27, ST273-KL74, and ST147-KL10 originated from ST147-KL64 (Fig. 3). To unravel the recombination events associated with KL switching, we evaluated four major subclones with different ST, KL, and O locus combinations: i.e., ST147-KL64-O2v1, ST147-KL10-O3/O3a, ST392-KL27-O4, and ST273-KL74-O3b (Fig. 4A). The recombination analysis showed that all the KL-associated recombination was located in a ~793 Kb region (Fig. 4B) that spanned from a UbiX family flavin prenyltransferase gene (locus ID H1D29\_06765 in Kp46564) to the MFS transporter gene (H1D29\_12740) and encompassed the MLST locus of *tonB*, the fluoroquinolone-resistant-associated gene *gyrA*, and K and O antigen encoding operons. In detail, ST147-KL64-O2v1 resulted from a ~370 Kb recombination event that spanned from *gltX* (H1D29\_08410) to *hisA* (H1D29\_10050) between a ST147-KL55-O3a and a ST392-KL64-O2v1-like strain; ST392-KL27-O4 resulted from a ~558 Kb recombination event that spanned from *bglX* (H1D29\_09590) and iron ABC transporter permease (H1D29\_12350) between a ST147-KL64-O2v1 and a ST1758-KL27-O4-like strain; ST147-KL10-O3/O3a resulted from a ~305 Kb recombination event that spanned from UbiX family flavin prenyltransferase (H1D29\_08925) and alpha/beta fold hydrolase (H1D29\_10260) between a ST147-KL64-O2v1 and a ST3603-KL10-O3/O3a strain, and ST273-KL74-O3b resulted from a ~636 Kb recombination event that spanned from the ABC transporter substrate-binding protein (H1D29\_09595) and MFS transporter (H1D29\_12740) between ST147-KL64-O2v1 and a ST147-KL74-O3b strain (Fig. 4A). In addition, minor subclones including ST147-KL107-O2v1, ST147-KL81-OL101, and ST147-KL20-O2v1 were investigated, and they too resulted from recombination within the ~793 Kb region (Fig. 4B). An SNP density

plot illustrated in Fig. 4C depicts the recombination regions.

#### CRISPR-Cas systems in CG147

We next analyzed the distribution of CRISPR-Cas systems in 1012 CG147 strains. We identified two distinct CRISPR-Cas systems: a chromosome-borne type I-E system in 99.7% ( $n=1009$ ) and a plasmid-borne type IV-A3 system in 41.8% ( $n=423$ ) of the genomes.

A total of 96 different CRISPR spacers (short DNA fragments that are stored in CRISPR arrays for target recognition) were identified and 41 spacers were present in >90% of the I-E CRISPR-Cas system harboring strains (Additional file 3: Supplementary results), indicating the I-E CRISPR-Cas system is conserved and an important element in the evolution of CG147.

To investigate the potential impact of I-E CRISPR-Cas system on the distribution of carbapenemase-encoding genes, we analyzed the presence of I-E protospacers (a complementary DNA sequence that is recognized by CRISPR-Cas system) in 1084 completely sequenced carbapenemase-encoding plasmids obtained from *K. pneumoniae* genomes in the NCBI database (accessed in 02/2023). Our investigation revealed the presence of 11 I-E protospacers on 423 carbapenemase-encoding plasmids. Notably, four protospacer sequences (IESP-2, [99.2%, 1001/1009], IESP-4, [99.1%, 1000/1009], IESP-30, [93%, 938/1009] and IESP-31, [92.6%, 934/1009], Additional file 1: Table S4) were identified highly prevalent in CG147 strains. The four highly prevalent protospacers were located within an approximate 830 bp region, including *traC*, a gene of DUF3560 domain, and the intergenic region between single-stranded DNA-binding protein (*ssb*) and SAM-methyltransferase gene on IncF plasmids. Notably, these protospacers-harbored IncF plasmids, such as IncFII(pHN7A8) (e.g., NZ\_CP116905.1), IncFIB(pQil) (e.g., NZ\_CP103551.1), IncFIB(Kpn3) (e.g., NZ\_CP098346.1), IncFII(pKP91) (e.g., NZ\_OW970445.1), or FII(pBK30683) (e.g., NZ\_CP087613.1) plasmids, are frequently associated with KPC.

The IV-A3 CRISPR-Cas system was identified on IncHI1B/IncFIB plasmids in CG147 and most IV-A3 CRISPR-Cas system harboring isolates were collected

(See figure on next page.)

**Fig. 4** **A** Hypothesized evolutionary history in *K. pneumoniae* CG147 strains. **B** MLST allele locations on Kp46564 genome (GenBank CP059317). The blue arrow denotes the genome of Kp46564, and the green region shows the ~793-Kbp putative recombination region between the ST147 and other genomes. **C** Core genome SNP distributions of strains ST392-KL27, ST1758-KL27, ST147-KL10, ST3603-KL10, ST147-KL107, ST147-KL81, ST147-KL20, and ST147-KL55 compared to reference strain Kp46564. The number of SNPs (y axis) per 1000 nt is plotted according to the position on the Kp46564 genome (x axis). *cps* and *lps* are illustrated by small vertical bars in blue and green respectively. The box indicates the recombination region

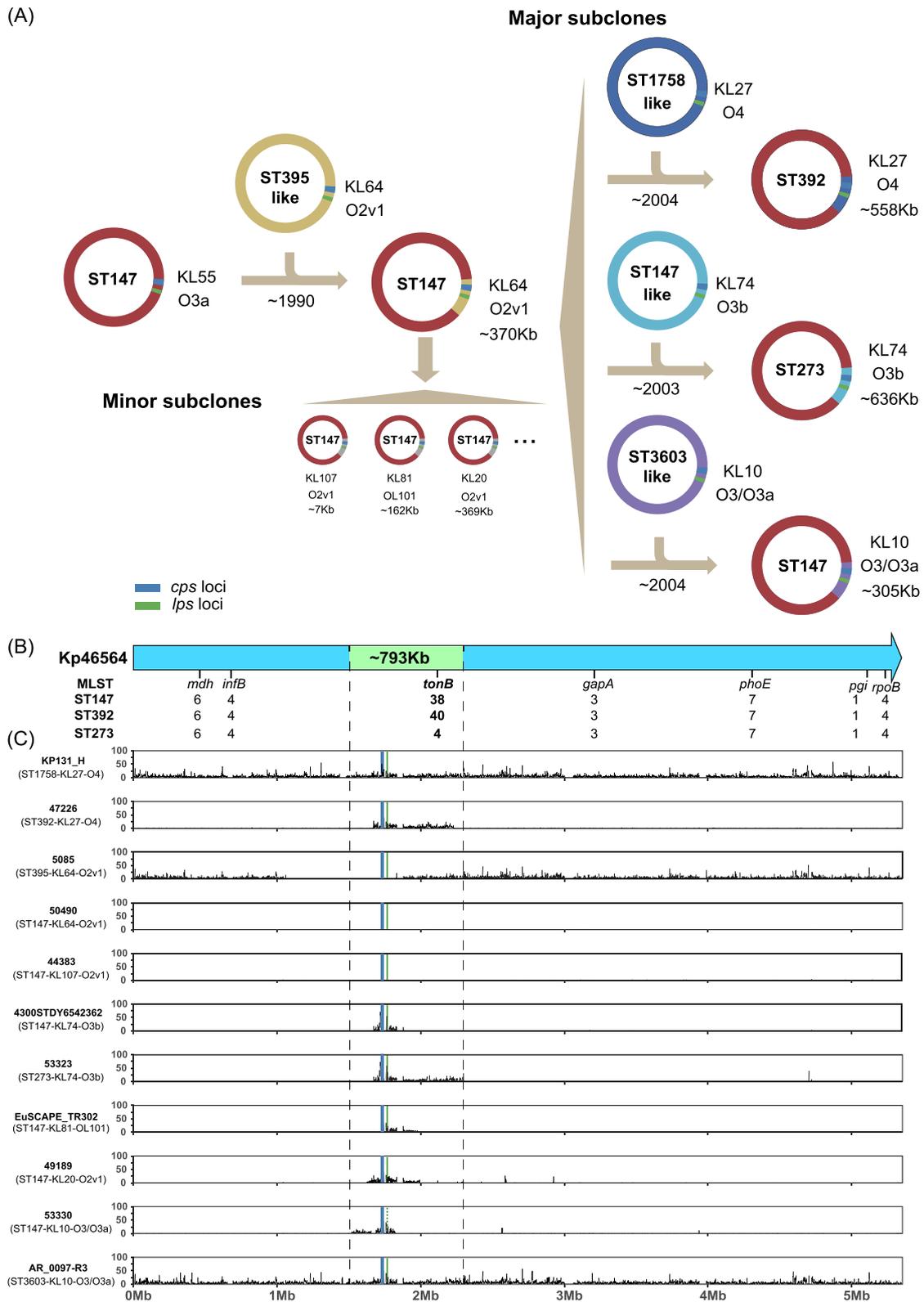


Fig. 4 (See legend on previous page.)

from North America ( $n=167$ , 39.5%) and Southern Europe ( $n=146$ , 33.8%). These isolates clustered into two major subclades on the phylogenetic tree (Fig. 2). A total of 347 unique spacers were identified and only 11 spacers present in >65% of the IV-A3 CRISPR-Cas harboring strains (Additional file 3: Supplementary results), suggesting that the acquisition of IV-A3 CRISPR-Cas system was a relatively recent event in comparison to the acquisition of I-E CRISPR-Cas system.

Among the 347 unique IV-A3 spacers, 16 were found to target 402 carbapenemase-encoding plasmids from the NCBI database. Among these, only one spacer (IVA3SP-10, [65.2%, 276/423], Additional file 1: Table S5) was present in >65% strains, 13 spacers were harbored by >2 strains, while the remaining were identified in individual strains. The mostly frequent encountered protospacer (IVA3SP-10) was located in *traL* on KPC- or CTX-M-harboring IncF plasmids, such as plasmids belonging to IncFII(pHN7A8) (e.g., NZ\_CP107017.1), IncFIB(pQil) (e.g., NZ\_CP027056.1), or FII(pBK30683) (e.g., NZ\_CP059890.1) replicon types.

Among the 402 carbapenemase-encoding plasmids targeted by IV-A3 CRISPR-Cas system (from NCBI database), 332 were also targeted by I-E CRISPR-Cas system, indicating a redundant anti-IncF plasmid strategy by two different CRISPR-Cas systems in CG147.

#### Anti-CRISPR proteins in CG147

We then mined the distribution of anti-CRISPR proteins in 1012 CG147 genomes. We identified anti-type I-E CRISPR protein AcrIE8.1 in 40.1% ( $n=406$ ), AcrIE9.2 in 54.2% ( $n=548$ ), and anti-type I-F CRISPR protein AcrIF11 in 3.2% ( $n=32$ ) CG147 strains (20.7% [ $n=209$ ] strains harbored both AcrIE8.1 and AcrIE9.2).

The gene *acrIE8.1* was located on different prophages that size around 46 Kb. Two major prophage types were characterized: Phage-1 was in the arginine-tRNA ligase site on both the ST147 and ST273 chromosomes and Phage-2 was identified in ST392 where it inserts in the threonine-tRNA ligase site (Additional file 2: Fig. S3A). Phage-1 and Phage-2 are both intact prophages with defined attachment (*att*) sites, have high sequence similarity with each other, and likely evolved from a common ancestor (Additional file 2: Fig. S3A). The *acrIE8.1* harboring strains were clustered into three major subclades on the phylogenetic tree, which corresponds to the presence of the two prophages (Additional file 2: Fig. S3B).

The genomic analysis showed that the *acrIE9.2* was predominantly associated with a group of IncF plasmids, such as the epidemic *bla*<sub>KPC</sub>-harboring pKpQIL plasmids (e.g., CP059315, CP023928, and CP059315) in CG147 strains. Furthermore, the *acrIF11* was identified on different IncF plasmids that lacked both I-E and

IV-A CRISPR-Cas systems targeted sequences. These plasmids were more commonly harbored by *E. coli* (e.g. CP046417.1). The strains harboring *acrIF11* are concentrated into one major cluster, all of which belonged to ST147 (Fig. 2).

The phylogenetic dating indicates that the chromosome-borne I-E CRISPR-Cas system emerged in CG147 as early as 1963 (95% CI, 1948–1975). The plasmid-borne IV-A3 CRISPR-Cas system was acquired around 1992 (95% CI, 1987–1996), which aligns with the emergence of anti-CRISPRs, including the phage-borne *acrIE8.1* (1990 [95% CI, 1985–1994]) and the plasmid-borne *acrIE9.2* (1992 [95% CI, 1987–1996]) in CG147. In contrast, the acquisition of *acrIF11* appears to be a more recent event (2006 [95% CI, 2003–2009]).

#### Self-targeting CRISPR spacers in CG147

We next analyzed the presence of self-targeting CRISPR spacers (additional copies of spacer sequences found elsewhere in the genomes but not in the CRISPR array) among CG147 strains. The existence of self-targeting spacers primarily indicates the functionality of anti-CRISPR mechanisms [54]. To avoid duplications, we included 646 non-redundant strains in this study, all of which possessed I-E CRISPR-Cas systems (Additional file 1: Table S6).

Interestingly, we found that a significant proportion (71.2%, 460/646) of strains contain I-E CRISPR-Cas self-targeting spacers. Among these, 218 harbored self-targeting spacers on plasmids, 87 on chromosomes, and 155 on both plasmids and chromosomes. Of the 242 strains harboring chromosomal self-targeting spacers, 87.6% (212/242) had self-targeting spacers located in prophage regions. This finding is unexpected because self-targeting spacers were expected to be eliminated by the endogenous I-E CRISPR-Cas system. However, further analysis revealed that nearly all the strains with I-E self-targeting spacers (442/460, 96.1%) also harbored at least one anti-type I-E CRISPR gene to overcome the presumed I-E immunity. This included 224 (48.7%) strains with *acrIE8.1*, 359 (78.0%) strains with *acrIE9.2*, and 141 (30.1%) strains possessing both. In contrast, the occurrence of anti-type I-E CRISPR proteins among strains lacking type I-E self-targeting spacers was considerably lower (2 strains with *acrIE8.1* and 3 strains with *acrIE9.2*, 2.9%, 5/186) ( $p < 0.001$ ). This significant difference strongly suggests that AcrIE8.1 and AcrIE9.2 provide protection against I-E CRISPR-Cas system, consistent with previous experimental tests [22, 23].

To further understand other factors that affect the presence of I-E self-targeting spacers, we conducted a logistic regression analysis of the presence of self-targets with STs, isolation regions, and the presence of anti-CRISPR

proteins AcrIE8.1, AcrIE9.2, and AcrIF11. The analysis showed that only AcrIE8.1 (odds ratio [OR]=537.98,  $p < 0.001$ ) and AcrIE9.2 (OR=977.55,  $p < 0.001$ ) significantly associated with the presence of I-E self-targeting spacers (Table 2), providing additional credence to the hypothesis that AcrIE8.1 and AcrIE9.2 are countermeasures against the I-E CRISPR-Cas system.

We also examined the factors associated with the presence of IV-A3 CRISPR-Cas system. Among the 646 strains, 195 harbored the IV-A3 CRISPR-Cas system and ~60% isolates (116/195) possessed IV-A3 self-targeting spacers. Among them, 40 strains harbored self-targeting spacers on plasmids, 61 on chromosomes, and 15 on both plasmids and chromosomes, of which 46.1% (35/76) had self-targeting spacers located within prophage regions. However, the logistic regression analysis failed to detect any significant associations between the presence of anti-CRISPR proteins (AcrIE8.1, AcrIE9.2, and AcrIF11) and the occurrence of IV-A3 self-targeting spacers (Table 2), suggesting that these anti-CRISPR proteins may not have evolved as a countermeasure against the IV-A3 CRISPR-Cas system.

**Associations between CRISPR-Cas systems, anti-CRISPR proteins, and AMR genes in CG147**

We next investigated the relationships between CRISPR-Cas systems, anti-CRISPR proteins, and AMR plasmids in CG147 at a population level (Fig. 5A). We employed a

logistic regression analysis to examine the associations of AMR genes with STs, isolation regions, and the presence of IV-A3 CRISPR-Cas system, as well as anti-CRISPR proteins AcrIE8.1, AcrIE9.2, and AcrIF11.

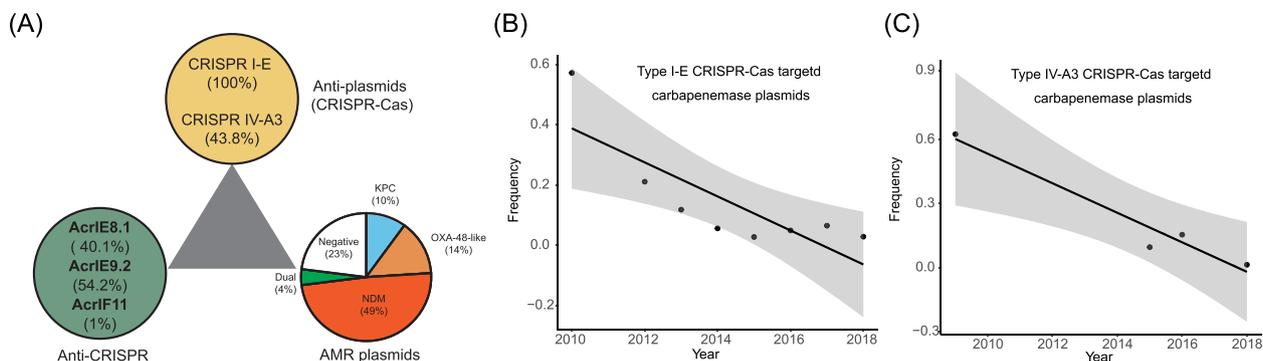
When both STs and geographic regions were controlled, we observed significant differences in the detection of the IV-A3 CRISPR-Cas system among various AMR strains. IV-A3 CRISPR-Cas system was commonly found in strains harboring NDM (OR=4.16,  $p < 0.001$ ), but less frequently associated with KPC-(OR=0.52,  $p = 0.02$ ) and CTX-M-positive strains (OR=0.42,  $p < 0.001$ ) (Table 3). To investigate this observation in more detail, we analyzed 50 completely assembled CG147 genomes including both chromosomes and plasmids. The result showed that the primary target of the IV-A3 CRISPR-Cas system, *traL*, was consistently present in plasmids associated with KPC and CTX-M production (e.g., CP023916, CP075261.1, and CP023928), but absent in plasmids linked to NDM or OXA-48-like genes. Furthermore, we noted instances where *bla*<sub>NDM-1</sub> co-existed with the IV-A3 CRISPR-Cas system on the same plasmids (e.g., CP066856). This co-occurrence may contribute to the increased association between NDM and the IV-A3 CRISPR-Cas system.

The result analyzing the prevalence of the anti-CRISPR proteins showed a positive association between AcrIE8.1 and NDM (OR=4.16,  $p < 0.001$ ) and CTX-M (OR=2.33,  $p < 0.001$ ), but a negative correlation with KPC (OR=0.48,

**Table 2** Logistic regression of type I-E and type IV-A3 self-targeting spacers with STs, regions, type IV-A3 CRISPR-Cas system, and anti-CRISPR proteins in 646 non-redundant strains

	Type I-E self-targeting spacer (n = 460)		Type IV-A3 self-targeting spacer (n = 116)	
	Odds ratio (95% CI)	p value	Odds ratio (95% CI)	p value
ST				
ST273 vs. ST147	4.33 (0.52–33.61)	0.160	0.35 (0.06–1.25)	0.168
ST392 vs. ST147	1.24 (0.03–100.14)	0.929	1.02 (0.48–2.08)	0.956
Region				
Eastern Europe vs. Northern America	2.04 (0.3–11.15)	0.429	1.52 (0.61–3.58)	0.348
Western Europe vs. Northern America	1.32 (0.34–5.21)	0.683	1.82 (0.96–3.48)	0.066
Southern Europe vs. Northern America	0.29 (0.04–1.77)	0.199	1.34 (0.59–2.92)	0.47
South-eastern Asia vs. Northern America	0.12 (0.01–1.24)	0.064	0.79 (0.27–2.02)	0.642
Western Asia vs. Northern America	0.59 (0.12–2.64)	0.495	0.56 (0.21–1.36)	0.224
Other vs. Northern America	0.68 (0.17–2.64)	0.578	1.34 (0.71–2.53)	0.368
CRISPR-Cas				
Type IV-A3 CRISPR-Cas	1.08 (0.38–2.92)	0.883	NA	NA
Anti-CRISPR				
AcrIE8.1	537.98 (134.69–3856.23)	<0.001***	1.16 (0.72–1.86)	0.533
AcrIE9.2	977.55 (283.88–5329.1)	<0.001***	0.76 (0.49–1.19)	0.232
AcrIF11	0.25 (0.02–2.14)	0.275	0.38 (0.06–1.36)	0.2

\*\*\* $p < 0.001$



**Fig. 5** **A** Linking the interactions between CRISPR-Cas systems, anti-CRISPR protein, and AMR plasmids in CG147. **B** The frequencies of type I-E CRISPR-Cas targeted carbapenemase plasmids in CG147 during 2010–2018. **C** The frequencies of type IV-A3 CRISPR-Cas targeted carbapenemase plasmids in CG147 during 2009–2018

**Table 3** Logistic regression of KPC, NDM, OXA-48-like and CTX-M with STs, regions, type IV-A3 CRISPR-Cas system, and anti-CRISPR proteins in 646 non-redundant strains

	KPC (n = 94)		NDM (n = 235)		OXA-48-like (n = 138)		CTX-M (n = 465)	
	Odds ratio (95% CI)	p value	Odds ratio (95% CI)	p value	Odds ratio (95% CI)	p value	Odds ratio (95% CI)	p value
ST								
ST273 vs. ST147	2.22 (0.71–6.29)	0.146	0.29 (0.1–0.73)	0.013*	0.28 (0.04–1.01)	0.096	0.82 (0.34–2.12)	0.67
ST392 vs. ST147	1.51 (0.65–3.48)	0.332	0.15 (0.07–0.32)	<0.001***	0.61 (0.25–1.36)	0.247	2.1 (0.94–5.09)	0.081
Region								
Eastern Europe vs. Northern America	1.02 (0.34–2.7)	0.967	0.37 (0.16–0.79)	0.012*	7.74 (3–21.01)	<0.001***	2.9 (1.3–7.06)	0.013*
Western Europe vs. Northern America	2.8 (1.27–6.15)	0.01*	0.09 (0.03–0.22)	<0.001*	9.9 (4.21–25.22)	<0.001***	1.23 (0.64–2.4)	0.531
Southern Europe vs. Northern America	2.84 (1.46–5.62)	0.002**	0.22 (0.12–0.39)	<0.001*	2.07 (0.83–5.39)	0.122	1.22 (0.7–2.11)	0.485
South-eastern Asia vs. Northern America	0.1 (0.01–0.53)	0.03*	0.64 (0.31–1.29)	0.216	4.38 (1.55–12.5)	0.005**	3.34 (1.46–8.47)	0.007**
Western Asia vs. Northern America	0.1 (0.01–0.5)	0.026*	0.85 (0.45–1.61)	0.622	10.42 (4.49–26.37)	<0.001***	5.2 (2.35–12.81)	<0.001***
Other vs. Northern America	0.95 (0.46–1.93)	0.877	0.47 (0.27–0.79)	0.005**	4.21 (1.93–10.06)	<0.001***	2.8 (1.61–4.92)	<0.001***
CRISPR-Cas								
Type IV-A3 CRISPR-Cas	0.52 (0.29–0.9)	0.023*	4.16 (2.81–6.22)	<0.001***	1.2 (0.74–1.93)	0.451	0.42 (0.28–0.63)	<0.001***
Anti-CRISPR								
AcrIE8.1	0.48 (0.25–0.87)	0.02*	1.85 (1.21–2.84)	0.005**	0.94 (0.58–1.53)	0.816	2.33 (1.46–3.8)	<0.001***
AcrIE9.2	3.18 (1.85–5.65)	<0.001***	0.91 (0.62–1.34)	0.637	0.9 (0.57–1.41)	0.633	0.79 (0.53–1.18)	0.253
AcrIF11	NA	0.984	7.02 (2.67–21)	<0.001***	16.83 (5.81–58.76)	<0.001***	2.73 (0.75–17.65)	0.19

\*  $P < 0.05$ ; \*\*  $P < 0.01$ ; \*\*\*  $P < 0.001$

$p = 0.02$ ). A more in-depth analysis of the association between KPC subtypes and AcrIE8.1 (Additional file 1: Table S7) revealed that the negative correlation was

primarily driven by the over-representation of KPC-3 in AcrIE8.1-negative CG147 strains (OR = 0.04,  $p = 0.004$ ). Notably, common  $bla_{KPC-3}$ -harboring IncN plasmids

(similar to LT838197) were detected in most KPC-3-producing isolates from Portugal and France. Of significance, these plasmids do not possess the I-E protospacers.

Additionally, our results unveiled a significantly positive association between AcrIE9.2 and KPC (OR=3.18,  $p < 0.001$ ), particularly with KPC-2 (OR=4.77,  $p < 0.001$ ), while no such association was observed with NDM, OXA-48-like, or CTX-M. Through the analysis of complete genome sequences, we have identified the co-existence of *acrIE9.2* with *bla*<sub>KPC-2</sub> on certain plasmids, notably the epidemic pKpQIL plasmids (e.g., CP059315, CP023928, and CP059315). Interestingly, as aforementioned, pKpQIL-like plasmids also harbor I-E and IV-A3 CRISPR-Cas systems target sequences, illustrating a unique example of the co-occurrence of AMR genes, the CRISPR-Cas targets, and anti-CRISPR genes in the same epidemic plasmid vector.

Lastly, we evaluated the proportions of CRISPR-Cas system-targeted carbapenemase plasmids over time. Our findings indicate that the frequencies of I-E CRISPR-Cas (Fig. 5B) and IV-A3 CRISPR-Cas systems that targeted (Fig. 5C) carbapenemase plasmids declined year by year from 2009 to 2019.

#### Experimental verification of the function of CRISPR-Cas systems and anti-CRISPR proteins from CG147

To assess the anti-plasmid capability of the I-E CRISPR-Cas system, we conducted a conjugation assay comparing the transfer frequency of the pKpQIL plasmids, both with and without type I-E CRISPR-Cas targets, into strain BK56682. BK56682 is an ST147 strain, possessing the I-E CRISPR-Cas system but it lacks the IV-A3 CRISPR-Cas system and *acr* genes. For this investigation, the KPC-encoding pKpQIL plasmid was selected as the test plasmid as it is an epidemic *bla*<sub>KPC</sub>-harboring vector in various Enterobacteriales species, and it contains targets for both I-E and IV-A3 CRISPR-Cas systems. Specifically, we tested plasmid pKpQIL-03, which carries the four highly frequent I-E protospacers. It is important to note that the native pKpQIL-03 also contains *acrIE9.2*, and therefore, to avoid any potential confounding effects from this anti-CRISPR, we knocked out *acrIE9.2* on pKpQIL-03 (designated as pKpQIL-03Δ*acrIE9.2*) in the test plasmid. Subsequently, we knocked out the 1315 bp region containing the four common type I-E protospacers on pKpQIL-03Δ*acrIE9.2* (designated as pKpQIL-03Δ*acrIE9.2*ΔIETarget). The results demonstrated that the conjugation frequency of pKpQIL-03Δ*acrIE9.2* reduced by approximately 7000 times ( $p < 0.001$ ) compared to pKpQIL-03Δ*acrIE9.2*ΔIETarget (Fig. 6A), confirming the anti-plasmid functionality of the I-E CRISPR-Cas system in CG147.

We then evaluated the anti-CRISPR efficacy of AcrIE8.1 against the I-E CRISPR-Cas system in CG147. The gene *acrIE8.1* was cloned into plasmid vector, pClone, and transferred to BK56682. This was followed by a conjugation assay comparing the transfer frequency of the pKpQIL plasmid into BK56682, both with (pClone-*acrIE8.1*) and without (pClone) *acrIE8.1*. The conjugation assays revealed that the transfer frequency of pKpQIL-03Δ*acrIE9.2* to BK56682 with pClone-*acrIE8.1* was 4.2 times higher ( $p < 0.05$ ) than to BK56682 with the control plasmid (Fig. 6B), thus confirming the anti-CRISPR effect of AcrIE8.1.

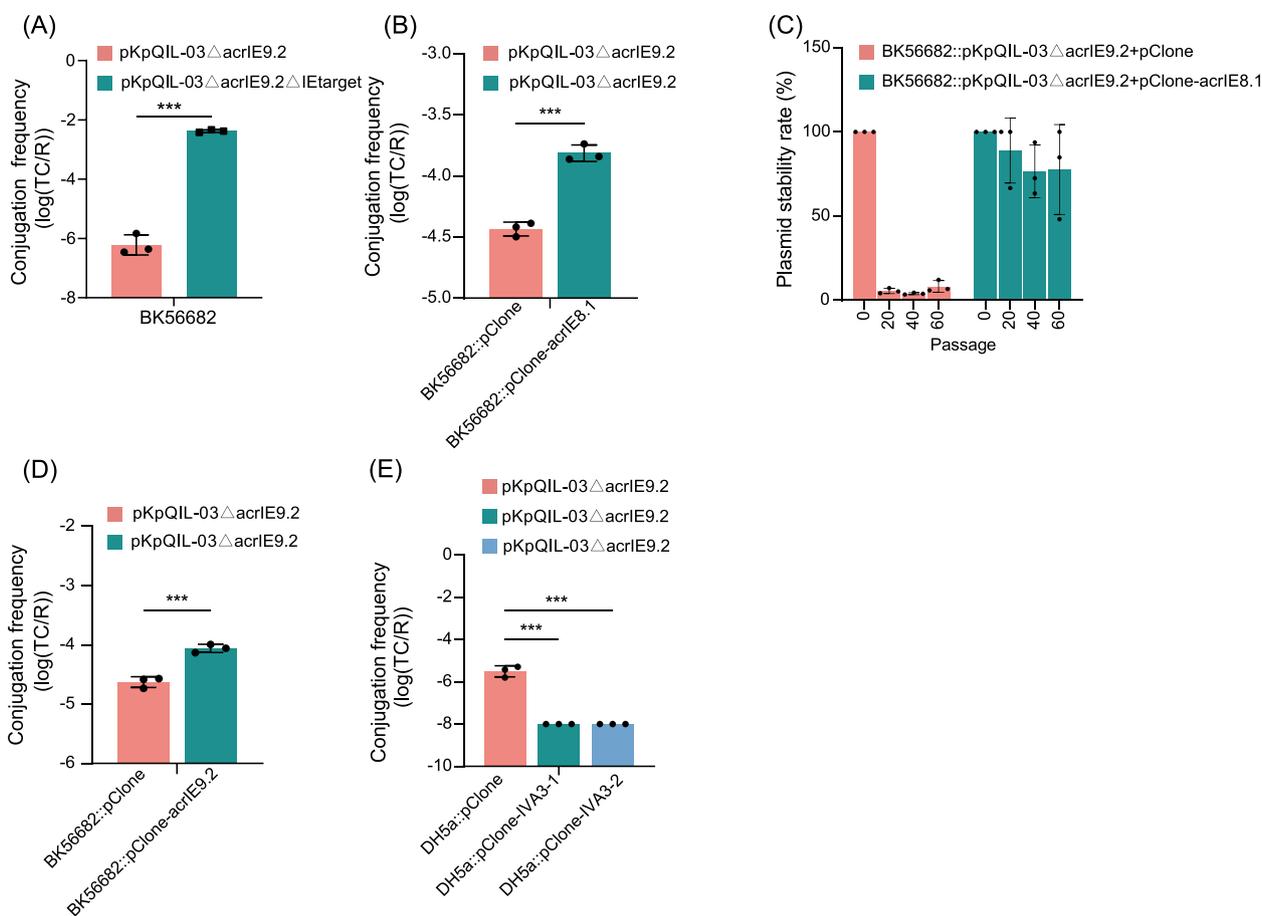
To assess the dynamics of the I-E CRISPR-Cas system over time, we conducted plasmid stability experiments and compared the results of serial passaging the strains. The findings showed that the carriage rate of pKpQIL-03Δ*acrIE9.2* remained around 80% after 60 passages in strains with a repressed I-E CRISPR-Cas system, while the carriage rate dropped to approximately 10% in negative control strains after 20 passages (Fig. 6C), confirming the plasmid immunity of the I-E CRISPR-Cas system.

We also explored the anti-CRISPR effect of AcrIE9.2 by conjugating pKpQIL-03Δ*acrIE9.2* to BK56682 with (pClone-*acrIE9.2*) and without *acrIE9.2* (pClone control). The conjugation frequency of pKpQIL-03Δ*acrIE9.2* in BK56682::pClone-*acrIE9.2* was 3.7 times higher ( $p < 0.001$ ) compared to the transfer in BK56682::pClone, indicating that AcrIE9.2 also possesses an anti-CRISPR effect (Fig. 6D).

To verify the functionality of the plasmid-borne IV-A3 CRISPR-Cas system, we cloned two different IV-A3 CRISPR-Cas arrays (both with spacers targeting pKpQIL-03) into pClone (pClone-IVA3-1 and pClone-IVA3-2) and transferred them to *E. coli* DH5a. Through conjugation assays, we observed that the conjugation frequencies of pKpQIL-03 transferred to DH5a::pClone-IVA3-1 and DH5a::pClone-IVA3-2 were approximately 900 times lower ( $p < 0.001$ ) than that of pKpQIL-03 transferred to DH5a with the control plasmid (Fig. 6E), indicating the robust anti-plasmid function of the IV-A3 CRISPR-Cas system. Lastly, despite the detection of AcrIF11 in our experiments, no significant anti-CRISPR effects on the I-E and IV-A3 CRISPR-Cas systems were observed (data not shown).

#### Discussion

CG147 is a unique MDR clone that conservatively harbors (found in 99.7% of analyzed genomes) a chromosomal I-E CRISPR-Cas system. This type of CRISPR-Cas system usually targets IncF plasmids [18], which are common AMR vectors in Enterobacteriales. In comparison to the CG23 clone, which also possesses the I-E CRISPR-Cas system and has few acquired AMR plasmids, CG147

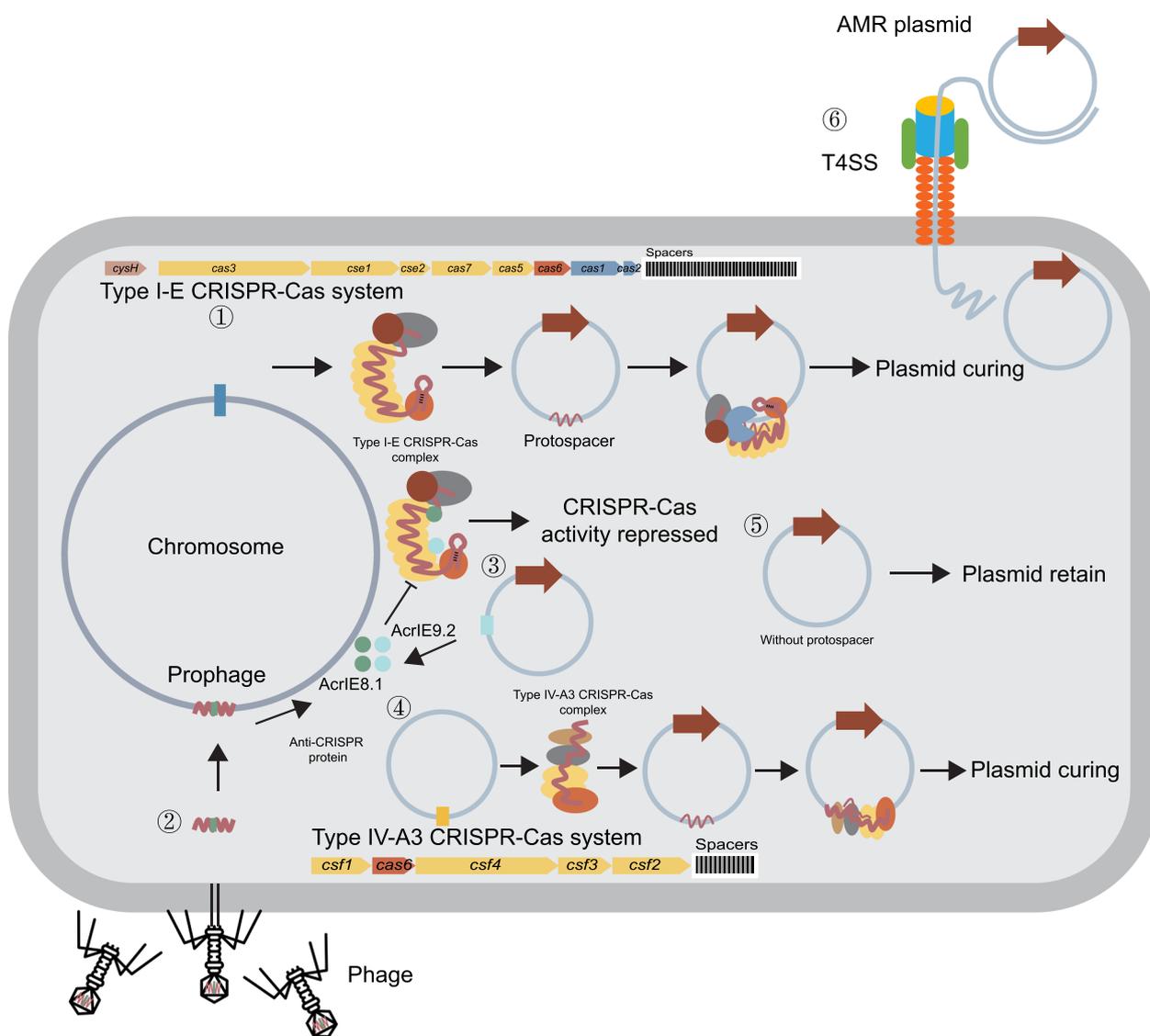


**Fig. 6** **A** Conjugation frequencies of pKpQIL-03ΔacrIE9.2 and pKpQIL-03ΔacrIE9.2ΔIEtarget to BK56682. **B** Conjugation frequencies of pKpQIL-03ΔacrIE9.2 transferred to BK56682::pClone and BK56682::pClone-acrIE8.1. **C** Plasmid stabilities of pKpQIL-03ΔacrIE9.2 in BK56682::pClone and BK56682::pClone-acrIE8.1. **D** Conjugation frequencies of pKpQIL-03ΔacrIE9.2 transferred to BK56682::pClone and BK56682::pClone-acrIE9.2. **E** Conjugation frequencies of pKpQIL-03ΔacrIE9.2 transferred to DH5a::pClone, DH5a::pClone-IVA3-1 (no transconjugant observed) and DH5a::pClone-IVA3-2 (no transconjugant observed). TC transconjugants, R recipient strains. \* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$

uniquely evolved into a highly successful MDR clone that harbors diverse AMR plasmids while retaining the I-E CRISPR-Cas system. CG147 also differs from other successful clones, such as CG258 and CG307, which lack the chromosomally encoded I-E CRISPR-Cas system. Previous studies suggested that the absence of this immunity mechanism contributes to epidemiological success of CG258 [55]. Our results shed light on the intricate interplay between CRISPR-Cas, anti-CRISPR, and MGEs in contributing to the success of CG147 (see Fig. 7).

Through the examination of the spacers within the I-E CRISPR-Cas systems of CG147, we identified prevalent spacers targeting a conserved region on IncF plasmids. Experimental investigations into the functionality of the I-E CRISPR-Cas system in CG147 confirmed its ability to eliminate plasmids with protospacers (Fig. 6). Initially, this finding appears to be contradictory to the MDR

feature of CG147. However, further analysis revealed a significant proportion of IncF plasmids in CG147 strains harboring self-targeting spacers, indicating potential dysfunction or inhibition of I-E CRISPR-Cas system. Subsequent mining of CG147 genomes unveiled that an overwhelming 96.1% of strains with I-E self-targeting spacers harbored at least one anti-type I-E CRISPR gene, including *acrIE8.1* [22] and *acrIE9.2* [23]. In contrast, the occurrence of anti-type I-E CRISPR proteins among strains lacking I-E self-targeting spacers was considerably lower. Moreover, our experimental verification demonstrated the inhibitory effects of AcrIE8.1 and AcrIE9.2 on the I-E CRISPR-Cas system. At the population level, our analysis revealed a positive correlation between AcrIE8.1 and KPC-2, NDM, and CTX-M ESBL, while AcrIE9.2 exhibited a positive correlation with KPC. These findings suggest that anti-type I-E CRISPR proteins, specifically



**Fig. 7** A proposed model depicting the interactions between CRISPR-Cas systems, anti-CRISPR protein, and AMR plasmids in CG147. Chromosomal type I-E CRISPR-Cas confers immunity to AMR plasmid harboring the protospacer; prophage or plasmid borne anti-CRISPR protein hinder the function of CRISPR-Cas system; plasmid borne type IV-A3 CRISPR-Cas provides additional immunity against AMR plasmid; and AMR plasmid lacking CRISPR-Cas protospacer remain unaffected and can persist within the bacterial host

AcrIE8.1 and AcrIE9.2, play a pivotal role in facilitating the acquisition of IncF plasmids carrying AMR genes into CG147 and overcoming the immunity effect of the I-E CRISPR-Cas system. This mechanism contributes significantly to the molecular epidemics of this globally successful MDR clone.

We also detected a plasmid-borne IV-A3 CRISPR-Cas system in 41.8% of the 1012 analyzed CG147 strains. The IV-A3 CRISPR-Cas system has been predominantly found on IncHI1B/IncFIB plasmids and reported to target the *tra* genes on IncF plasmids [56, 57]. IV-A3

CRISPR-Cas system lacks the Cas3-like nuclease gene [58], and a recent study revealed that this system acted similar to the CRISPRi-like mechanism which transcriptionally repress the target genes [59, 60]. It is worth highlighting that instances of self-targeting by the IV-A3 CRISPR-Cas system have been documented in clinical isolates, even in the absence of known anti-type IV-A3 CRISPR mechanisms [61]. In this study, our conjugation experiment confirmed that the anti-plasmid function of the IV-A3 CRISPR-Cas system in CG147 (Fig. 6). In addition, our investigation unveiled a positive correlation

between the presence of the IV-A3 CRISPR-Cas system and NDM, while revealing negative correlation with KPC and CTX-M ESBL plasmids. The negative correlation is likely attributable to the IV-A3 CRISPR-Cas system's targeting of *traL* on IncF type plasmids, which are common carriers of KPC and CTX-M ESBL genes [62]. Conversely, the positive correlation may arise from the co-existence of IV-A3 CRISPR-Cas system and NDM on the same plasmid. We observed that a large proportion of strains harboring the IV-A3 CRISPR-Cas system also contained self-targeting spacers. Further gene expression analysis found no significant difference in IV-A3 CRISPR-Cas system activity between isolates with and without self-targeting spacers (data not shown). One can hypothesize that the high prevalence of IV-A3 self-targeting spacers may be linked to its distinct CRISPR interference like function [59, 61]. Additionally, the presence of unknown anti-type IV-A3 CRISPR protein or regulator cannot be ruled out.

Our phylogenetic dating analysis revealed that the I-E CRISPR-Cas system integrated into the chromosome of CG147 strains around the 1960s. However, the emergence of acquired antimicrobial resistance (AMR) genes or chromosomal alterations, such as QRDR mutations, took place in the 1990s. This coincided with the increased clinical use of third-generation cephalosporins, carbapenems, and fluoroquinolones since the 1980s [63, 64]. Given that CTX-M-15 and KPC genes are commonly located on IncF plasmids, the CG147 strains exhibited a remarkable ability to adapt and circumvent the directed immunity. To overcome the antibiotic pressure, they adopted a novel anti-type I-E CRISPR mechanism through the horizontal acquisition of MGEs, including phages and plasmids. These MGEs encoded *acr* genes, allowing the strains to overcome the endogenous anti-plasmid I-E CRISPR-Cas system. This adaptation rendered the CG147 isolates resistant to cephalosporins and carbapenems. Concurrently, CG147 strains developed an additional MGE-mediated anti-plasmid IV-A3 CRISPR-Cas system. This system, utilizing a mechanism distinct from the I-E CRISPR-Cas system, provides strains additional biological flexibility during interactions with AMR plasmids containing protospacers, ultimately offering additional IncF plasmid immunity. On the other side of this dynamic interplay, the AMR gene ingeniously integrated into the plasmids harboring anti-CRISPR genes, exemplified by the *bla*<sub>KPC</sub>-harboring pKpQIL plasmids. pKpQIL descended from the KPC-negative pKPN4-like plasmid through the integration of the *bla*<sub>KPC</sub>-harboring transposon Tn4401 [65]. This finding also contributes to our understanding of the molecular basis underlying the success of this epidemic KPC vector. Additionally, it is noteworthy that the Acrs (AcrIE8.1 and AcrIE9.2) only provide a low level of anti-CRISPR efficiency. Over time, our results indicate a gradual decrease in the

prevalence of KPC-harboring IncF plasmids (Additional file 2: Fig. S1).

It is plausible that other non-MGE encoding features could contribute to the epidemiological success of CG147. In *K. pneumoniae*, capsular polysaccharide (CPS) and lipopolysaccharide (LPS) are two main virulence and antigen-encoding factors. Our investigation unveiled a dynamic landscape characterized by multiple capsular switches driven by recombination events within CG147 and leading to the emergence of major subclones. Notably, ST147-KL64-O2v1 stands out the most predominant subclone in CG147, globally disseminated and sharing KL similarities with ST11 [66], another globally successful clone commonly found in China. Previous research suggested that KL64 may have superior selection advantage over KL47 in the ST11 background [66]. Our results also showed the ST147-KL64 has remained the most dominant subclone in CG147 since the 2000s, prompting the need for further studies to determine whether the success of CG147 is linked to KL64 specially.

## Conclusions

In summary, the success of CG147 as a MDR clone appears to be driven by an intricate interplay between CRISPR-Cas immunity, anti-CRISPR mechanisms, MGEs, and adaptations to host-pathogen interactions. While the chromosomally encoded I-E CRISPR-Cas system in CG147 typically restricts plasmid acquisition, CG147 has circumvented this immunity through the acquisition of anti-type I-E CRISPR genes (e.g., AcrIE8.1 and AcrIE9.2), allowing the strain to retain diverse AMR plasmids. This contrasts with other clones like CG258, where the absence of CRISPR-Cas immunity may facilitate plasmid accumulation. The plasmid-mediated IV-A3 CRISPR-Cas system also contributes a unique layer of regulation, selectively targeting IncF plasmids associated with specific AMR genes while being counterbalanced by possible unknown anti-type IV or regulation mechanisms. This unique dual CRISPR strategy, alongside plasmid and phage-borne anti-CRISPR genes, exemplifies an evolved mechanism enabling CG147 to adapt to increasing antibiotic pressures. The timing of CG147's AMR evolution aligns with shifts in clinical antibiotic use, suggesting that environmental pressures drove the selection of both CRISPR inhibition and AMR acquisition. Additionally, capsular switching events within CG147, particularly the dominance of the ST147-KL64-O2v1 subclone, underscore the significance of recombination in generating epidemiologically successful lineages. The association of CG147 with virulence-related features, such as specific capsule and yersiniabactin loci, highlights the contributions of both

genetic adaptability and immunity-related mechanisms to CG147's success as a global MDR pathogen. Our study offers valuable insights into the genomic epidemiology and resistance mechanisms that have driven the widespread dissemination of the CG147 clone.

#### Abbreviations

CG	Clonal group
ST	Sequence type
CRISPR	Clustered regularly interspaced short palindromic repeats
KPC	<i>Klebsiella pneumoniae</i> Carbapenemase
AMR	Antimicrobial resistance
QRDR	Quinolone resistance-determining region
ESBL	Extended-spectrum beta-lactamase
MDR	Multidrug resistance
HGT	Horizontal gene transfer
MGE	Mobile genetic element
MCMC	Markov chain Monte Carlo

#### Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s13073-025-01428-6>.

Additional file 1: Supplementary Table S1–S7. Table S1. The information of CG147 strains in this study. Table S2. Primers used in this study. Table S3. Plasmids and strains used in this study. Table S4. Type I-E CRISPR-Cas spacer sequences. Table S5. Type IV-A3 CRISPR-Cas spacer sequences. Table S6. STs, regions, type IV-A3 CRISPR-Cas system and anti-CRISPR proteins of 646 non-redundant strains by carbapenemases and ESBL CTX-M in 646 non-redundant strains. Table S7. Logistic regression of KPC-2 and KPC-3 with STs, regions, type IV-A3 CRISPR-Cas system and anti-CRISPR proteins in 646 non-redundant strains.

Additional file 2: Supplementary Fig. S1–S3. Fig. S1. Distribution of NDM and KPC positive CG147 isolates between 2000 and 2020. Verdigris dot: NDM; red: KPC. Figure S2. Distribution of spacers of type I-E CRISPR-Cas arrays and type IV-A3 CRISPR-Cas arrays in CG147. Fig. S3. Prophages in CG147 strains. Alignment of two major types of *acrIE8.1* carrying prophages. The distribution of *acrIE8.1* and the containing prophages in 1,012 CG147 strains with different STs and K locus.

Additional file 3: Supplementary results. The supplementary description of the results

#### Acknowledgements

The study was in part supported by NIAID grant R01AI090155.

#### Authors' contributions

JJ., A.V.C., and G.P. collected the genomic data. JJ. and L.C. performed bioinformatics analysis. T. L. and T. C. performed gene knockout and experimental verification. JJ. and A.V.C. performed statistical analysis and wrote the initial draft of the manuscript. L. C., B.N.K., and J.D.D.P. critically reviewed and revised the manuscript. L.C. was responsible for overall study design, data analysis, and the final manuscript review. All authors read and approved the final manuscript.

#### Data availability

All raw sequencing data have been deposited in GenBank under the BioProject accession number PRJNA549322 (<https://www.ncbi.nlm.nih.gov/bioproject/PRJNA549322>). The custom code used in this study is freely available at GitHub [67].

#### Declarations

#### Ethics approval and consent to participate

Not applicable.

#### Consent for publication

Not applicable.

#### Competing interests

The authors declare no competing interests.

#### Author details

<sup>1</sup>Center for Discovery and Innovation, Hackensack Meridian Health, Nutley, NJ, USA. <sup>2</sup>School of Microbiology, University of Antioquia, Medellín, Colombia. <sup>3</sup>Guangdong Laboratory for Lingnan Modern Agriculture, National Risk Assessment Laboratory for Antimicrobial Resistance of Animal Origin Bacteria, College of Veterinary Medicine, South China Agricultural University, Guangzhou, China. <sup>4</sup>Guangdong Provincial Key Laboratory of Veterinary Pharmaceuticals, Development and Safety Evaluation, South China Agricultural University, Guangzhou, China. <sup>5</sup>Cummings School of Medicine, University of Calgary, Calgary, AB, Canada. <sup>6</sup>Alberta Precision Laboratories, Calgary, AB, Canada. <sup>7</sup>University of Pretoria, Pretoria, Gauteng, South Africa. <sup>8</sup>Department of Medical Sciences, Hackensack Meridian School of Medicine, Nutley, NJ, USA. <sup>9</sup>Present address: School of Pharmacy and Pharmaceutical Sciences, University at Buffalo, Buffalo, NY, USA.

Received: 29 April 2024 Accepted: 2 January 2025

Published: 30 January 2025

#### References

- Holt KE, Wertheim H, Zadoks RN, Baker S, Whitehouse CA, Dance D, et al. Genomic analysis of diversity, population structure, virulence, and antimicrobial resistance in *Klebsiella pneumoniae*, an urgent threat to public health. *Proc Natl Acad Sci U S A*. 2015;112(27):E3574–81.
- Wyres KL, Lam MMC, Holt KE. Population genomics of *Klebsiella pneumoniae*. *Nat Rev Microbiol*. 2020;18(6):344–59.
- Deleo FR, Chen L, Porcella SF, Martens CA, Kobayashi SD, Porter AR, et al. Molecular dissection of the evolution of carbapenem-resistant multilocus sequence type 258 *Klebsiella pneumoniae*. *Proc Natl Acad Sci U S A*. 2014;111(13):4988–93.
- Peirano G, Chen L, Kreiswirth BN, Pitout JDD. Emerging antimicrobial-resistant high-risk *Klebsiella pneumoniae* clones ST307 and ST147. *Antimicrob Agents Chemother*. 2020;64(10):e01148–20.
- Falcone M, Giordano C, Barnini S, Tiseo G, Leonildi A, Malacarne P, et al. Extremely drug-resistant NDM-9-producing ST147 *Klebsiella pneumoniae* causing infections in Italy, May 2020. *Euro Surveill*. 2020;25(48):2001779.
- Coelho A, Mirelis B, Alonso-Tarres C, Nieves Larrosa M, Miro E, Cliville Abad R, et al. Detection of three stable genetic clones of CTX-M-15-producing *Klebsiella pneumoniae* in the Barcelona metropolitan area. *Spain J Antimicrob Chemother*. 2009;64(4):862–4.
- Damjanova I, Toth A, Paszti J, Hajbel-Vekony G, Jakab M, Berta J, et al. Expansion and countrywide dissemination of ST11, ST15 and ST147 ciprofloxacin-resistant CTX-M-15-type beta-lactamase-producing *Klebsiella pneumoniae* epidemic clones in Hungary in 2005—the new “MRSAs”? *J Antimicrob Chemother*. 2008;62(5):978–85.
- Richter SN, Frasson I, Franchin E, Bergo C, Lavezzo E, Barzon L, et al. KPC-mediated resistance in *Klebsiella pneumoniae* in two hospitals in Padua, Italy, June 2009–December 2011: massive spreading of a KPC-3-encoding plasmid and involvement of non-intensive care units. *Gut Pathog*. 2012;4(1):7.
- Giakkoupi P, Papagiannitsis CC, Miriagou V, Pappa O, Polemis M, Tryfinopoulou K, et al. An update of the evolving epidemic of *bla*<sub>KPC-2</sub>-carrying *Klebsiella pneumoniae* in Greece (2009–10). *J Antimicrob Chemother*. 2011;66(7):1510–3.
- Peirano G, Pillai DR, Pitondo-Silva A, Richardson D, Pitout JD. The characteristics of NDM-producing *Klebsiella pneumoniae* from Canada. *Diagn Microbiol Infect Dis*. 2011;71(2):106–9.
- Wang X, Xu X, Li Z, Chen H, Wang Q, Yang P, et al. An outbreak of a nosocomial NDM-1-producing *Klebsiella pneumoniae* ST147 at a teaching hospital in mainland China. *Microb Drug Resist*. 2014;20(2):144–9.
- Lee CS, Vasoo S, Hu F, Patel R, Doi Y. *Klebsiella pneumoniae* ST147 coproducing NDM-7 carbapenemase and RmtF 16S rRNA methyltransferase in Minnesota. *J Clin Microbiol*. 2014;52(11):4109–10.

13. Rocha LK, dos Santos Neto RL, da Costa Guimaraes AC, Almeida AC, Vilela MA, de Moraes MM. Plasmid-mediated *qnrA1* in *Klebsiella pneumoniae* ST147 in Recife. *Brazil Int J Infect Dis*. 2014;26:49–50.
14. Di Pilato V, Henrici De Angelis L, Aiezza N, Baccani I, Niccolai C, Parisio EM, et al. Resistome and virulome accretion in an NDM-1-producing ST147 sublineage of *Klebsiella pneumoniae* associated with an outbreak in Tuscany, Italy: a genotypic and phenotypic characterisation. *Lancet Microbe*. 2022;3(3):e224–e34.
15. Lapp Z, Crawford R, Miles-Jay A, Pirani A, Trick WE, Weinstein RA, et al. Regional spread of *bla*<sub>NDM-1</sub>-containing *Klebsiella pneumoniae* ST147 in post-acute care facilities. *Clin Infect Dis*. 2021;73(8):1431–9.
16. Smith WPJ, Wucher BR, Nadell CD, Foster KR. Bacterial defences: mechanisms, evolution and antimicrobial resistance. *Nat Rev Microbiol*. 2023;21(8):519–34.
17. Horvath P, Barrangou R. CRISPR/Cas, the immune system of bacteria and archaea. *Science*. 2010;327(5962):167–70.
18. Zhou Y, Tang Y, Fu P, Tian D, Yu L, Huang Y, et al. The type I-E CRISPR-Cas system influences the acquisition of *bla*<sub>KPC</sub>-IncF plasmid in *Klebsiella pneumoniae*. *Emerg Microbes Infect*. 2020;9(1):1011–22.
19. Lam MMC, Wyres KL, Duchene S, Wick RR, Judd LM, Gan YH, et al. Population genomics of hypervirulent *Klebsiella pneumoniae* clonal-group 23 reveals early emergence and rapid global dissemination. *Nat Commun*. 2018;9(1):2703.
20. Tang Y, Fu P, Zhou Y, Xie Y, Jin J, Wang B, et al. Absence of the type I-E CRISPR-Cas system in *Klebsiella pneumoniae* clonal complex 258 is associated with dissemination of IncF epidemic resistance plasmids in this clonal complex. *J Antimicrob Chemother*. 2020;75(4):890–5.
21. Pawluk A, Davidson AR, Maxwell KL. Anti-CRISPR: discovery, mechanism and function. *Nat Rev Microbiol*. 2018;16(1):12–7.
22. Pinilla-Redondo R, Shehreen S, Marino ND, Fagerlund RD, Brown CM, Sorensen SJ, et al. Discovery of multiple anti-CRISPRs highlights anti-defense gene clustering in mobile genetic elements. *Nat Commun*. 2020;11(1):5652.
23. Wang C, Sun Z, Hu Y, Li D, Guo Q, Wang M. A Novel Anti-CRISPR AcrIE9.2 Is associated with dissemination of *bla*<sub>KPC</sub> plasmids in *Klebsiella pneumoniae* sequence type 15. *Antimicrob Agents Chemother*. 2023;67(4):e0154722.
24. Rodrigues C, Desai S, Passet V, Gajjar D, Brisse S. Genomic evolution of the globally disseminated multidrug-resistant *Klebsiella pneumoniae* clonal group 147. *Microb Genom*. 2022;8(1):000737.
25. Wang M, Earley M, Chen L, Hanson BM, Yu Y, Liu Z, et al. Clinical 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*. 2022;22(3):401–12.
26. Pribelski A, Antipov D, Meleshko D, Lapidus A, Korobeynikov A. Using SPAdes de novo assembler. *Curr Protoc Bioinformatics*. 2020;70(1):e102.
27. 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.
28. Lam MMC, Wick RR, Watts SC, Cerdeira LT, Wyres KL, Holt KE. A genomic surveillance framework and genotyping tool for *Klebsiella pneumoniae* and its related species complex. *Nat Commun*. 2021;12(1):4188.
29. Feldgarden M, Brover V, Gonzalez-Escalona N, Frye JG, Haendiges J, Haft DH, et al. AMRFinderPlus and the Reference Gene Catalog facilitate examination of the genomic links among antimicrobial resistance, stress response, and virulence. *Sci Rep*. 2021;11(1):12728.
30. Hunt M, Mather AE, Sanchez-Buso L, Page AJ, Parkhill J, Keane JA, et al. ARIBA: rapid antimicrobial resistance genotyping directly from sequencing reads. *Microb Genom*. 2017;3(10):e000131.
31. Russel J, Pinilla-Redondo R, Mayo-Munoz D, Shah SA, Sorensen SJ. CRISPR-CasTyper: automated identification, annotation, and classification of CRISPR-Cas loci. *CRISPR J*. 2020;3(6):462–9.
32. Li W, Godzik A. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics*. 2006;22(13):1658–9.
33. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, et al. BLAST+: architecture and applications. *BMC Bioinformatics*. 2009;10:421.
34. Huang L, Yang B, Yi H, Asif A, Wang J, Lithgow T, et al. AcrDB: a database of anti-CRISPR operons in prokaryotes and viruses. *Nucleic Acids Res*. 2021;49(D1):D622–9.
35. Wishart DS, Han S, Saha S, Oler E, Peters H, Grant JR, et al. PHAST-EST: faster than PHASTER, better than PHAST. *Nucleic Acids Res*. 2023;51(W1):W443–50.
36. Sullivan MJ, Petty NK, Beatson SA. Easyfig: a genome comparison visualizer. *Bioinformatics*. 2011;27(7):1009–10.
37. Seemann T. Abricate. <https://github.com/tseemann/abricate>. GitHub. 2020.
38. Carattoli A, Zankari E, Garcia-Fernandez A, Voldby Larsen M, Lund O, Villa L, et al. In silico detection and typing of plasmids using PlasmidFinder and plasmid multilocus sequence typing. *Antimicrob Agents Chemother*. 2014;58(7):3895–903.
39. Seemann T. Rapid haploid variant calling and core genome alignment. <https://github.com/tseemann/snippy>. GitHub. 2020.
40. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The sequence alignment/map format and SAMtools. *Bioinformatics*. 2009;25(16):2078–9.
41. Marçais G, Delcher AL, Phillippy AM, Coston R, Salzberg SL, Zimin A. MUMmer4: a fast and versatile genome alignment system. *PLoS Comput Biol*. 2018;14(1):e1005944.
42. Croucher NJ, Page AJ, Connor TR, Delaney AJ, Keane JA, Bentley SD, et al. Rapid phylogenetic analysis of large samples of recombinant bacterial whole genome sequences using Gubbins. *Nucleic Acids Res*. 2015;43(3):e15.
43. Stamatakis A. RAXML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*. 2014;30(9):1312–3.
44. Didelot X, Croucher NJ, Bentley SD, Harris SR, Wilson DJ. Bayesian inference of ancestral dates on bacterial phylogenetic trees. *Nucleic Acids Res*. 2018;46(22):e134.
45. Letunic I, Bork P. Interactive Tree Of Life (iTOL) v4: recent updates and new developments. *Nucleic Acids Res*. 2019;47(W1):W256–9.
46. Wick RR, Holt KE. Assembly Dereplicator. <https://github.com/rrwick/Assembly-dereplicator>. GitHub. 2022.
47. Chen T, Wang Y, Zhou Y, Zhou W, Chi X, Shen P, et al. Recombination drives evolution of carbapenem-resistant *Klebsiella pneumoniae* sequence type 11 KL47 to KL64 in China. *Microbiol Spectr*. 2023;11(1):e0110722.
48. Chen L, Peirano G, Kreiswirth BN, Deviney R, Pitout JDD. Acquisition of genomic elements were pivotal for the success of *Escherichia coli* ST410. *J Antimicrob Chemother*. 2022;77(12):3399–407.
49. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*. 2009;25(14):1754–60.
50. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*. 2010;26(6):841–2.
51. Chen L, Chavda KD, Melano RG, Jacobs MR, Koll B, Hong T, et al. Comparative genomic analysis of KPC-encoding pKpQIL-like plasmids and their distribution in New Jersey and New York Hospitals. *Antimicrob Agents Chemother*. 2014;58(5):2871–7.
52. Hao M, He Y, Zhang H, Liao XP, Liu YH, Sun J, et al. CRISPR-Cas9-mediated carbapenemase gene and plasmid curing in carbapenem-resistant Enterobacteriaceae. *Antimicrob Agents Chemother*. 2020;64(9):e00843–20.
53. Wang Y, Wang S, Chen W, Song L, Zhang Y, Shen Z, et al. CRISPR-Cas9 and CRISPR-assisted cytidine deaminase enable precise and efficient genome editing in *Klebsiella pneumoniae*. *Appl Environ Microbiol*. 2018;84(23):e01834–18.
54. Yin Y, Yang B, Entwistle S. Bioinformatics identification of anti-CRISPR loci by using homology, guilt-by-association, and CRISPR self-targeting spacer approaches. *mSystems*. 2019;4(5):e00455–19.
55. Shen J, Zhou J, Xu Y, Xiu Z. Prophages contribute to genome plasticity of *Klebsiella pneumoniae* and may involve the chromosomal integration of ARGs in CG258. *Genomics*. 2020;112(1):998–1010.
56. Newire E, Aydin A, Juma S, Enne VI, Roberts AP. Identification of a type IV-A CRISPR-Cas system located exclusively on IncHI1B/IncFIB plasmids in Enterobacteriaceae. *Front Microbiol*. 2020;11:1937.
57. Long J, Zhang J, Xi Y, Zhao J, Jin Y, Yang H, et al. Genomic insights into CRISPR-harboring plasmids in the *Klebsiella* genus: distribution, backbone structures, antibiotic resistance, and virulence determinant profiles. *Antimicrob Agents Chemother*. 2023;67(3):e0118922.
58. Pinilla-Redondo R, Mayo-Munoz D, Russel J, Garrett RA, Randau L, Sorensen SJ, et al. Type IV CRISPR-Cas systems are highly diverse and involved in competition between plasmids. *Nucleic Acids Res*. 2020;48(4):2000–12.

59. Benz F, Camara-Wilpert S, Russel J, Wandera KG, Cepaite R, Ares-Arroyo M, et al. Type IV-A3 CRISPR-Cas systems drive inter-plasmid conflicts by acquiring spacers in trans. *Cell Host Microbe*. 2024;32(6):875-886.e9.
60. Cepaite R, Klein N, Miksys A, Camara-Wilpert S, Ragozius V, Benz F, et al. Structural variation of types IV-A1- and IV-A3-mediated CRISPR interference. *Nat Commun*. 2024;15(1):9306.
61. Guo X, Sanchez-Londono M, Gomes-Filho JV, Hernandez-Tamayo R, Rust S, Immelmann LM, et al. Characterization of the self-targeting type IV CRISPR interference system in *Pseudomonas oleovorans*. *Nat Microbiol*. 2022;7(11):1870-8.
62. Rozwandowicz M, Brouwer MSM, Fischer J, Wagenaar JA, Gonzalez-Zorn B, Guerra B, et al. Plasmids carrying antimicrobial resistance genes in Enterobacteriaceae. *J Antimicrob Chemother*. 2018;73(5):1121-37.
63. Klein NC, Cunha BA. Third-generation cephalosporins. *Med Clin North Am*. 1995;79(4):705-19.
64. Papp-Wallace KM, Endimiani A, Taracila MA, Bonomo RA. Carbapenems: past, present, and future. *Antimicrob Agents Chemother*. 2011;55(11):4943-60.
65. Leavitt A, Chmelnitsky I, Carmeli Y, Navon-Venezia S. Complete nucleotide sequence of KPC-3-encoding plasmid pKpQIL in the epidemic *Klebsiella pneumoniae* sequence type 258. *Antimicrob Agents Chemother*. 2010;54(10):4493-6.
66. Zhou K, Xiao T, David S, Wang Q, Zhou Y, Guo L, et al. Novel subclone of carbapenem-resistant *Klebsiella pneumoniae* sequence type 11 with enhanced virulence and transmissibility. *China Emerg Infect Dis*. 2020;26(2):289-97.
67. Jiang J. Custom code and workflow. [https://github.com/biojiang/ST147\\_CRKP\\_paper](https://github.com/biojiang/ST147_CRKP_paper). GitHub. 2023.

### Publisher's Note

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