Genomic evolution of the globally disseminated multidrug-resistant Klebsiella pneumoniae clonal group 147

Carla Rodrigues, Siddhi Desai, Virginie Passet, Devarshi Gajjar, Sylvain Brisse

Genomic evolution of the globally disseminated multidrug-resistant *Klebsiella pneumoniae* clonal group 147

Carla Rodrigues[1], Siddhi Desai[2], Virginie Passet[1], Devarshi Gajjar[2] and Sylvain Brisse[1*]

[1] Institut Pasteur, Biodiversity and Epidemiology of Bacterial Pathogens, Paris, France.

[2] Department of Microbiology and Biotechnology Centre, Faculty of Science, The Maharaja Sayajirao University of Baroda, Vadodara, Gujarat, India.

*__Corresponding author:__

Sylvain Brisse

Institut Pasteur,

Biodiversity and Epidemiology of Bacterial Pathogens,

28 rue du Docteur Roux, F-75724 Paris, France.

E-mail: sylvain.brisse@pasteur.fr;

Phone: +33 1 45 68 83 34

Keywords: phylogenomics, pandrug resistance, last-resort antibiotics, high-risk clone, NDM-5, CRISPR/Cas systems, IncF plasmids, genomic epidemiology

Repositories: Reads and genomic sequences of the isolates analysed in this study were deposited at the European Nucleotide Archive (accession number PRJEB41234).

ABSTRACT

The rapid emergence of multidrug-resistant *Klebsiella pneumoniae* (Kp) is largely driven by the spread of specific clonal groups (CG). Of these, CG147 includes 7-gene MLST sequence types ST147, ST273 and ST392. CG147 has caused nosocomial outbreaks across the world, but its global population dynamics remain unknown. Here, we report a pandrug-resistant ST147 clinical isolate from India (strain DJ) and define the evolution and global emergence of CG147. Antimicrobial susceptibility testing (EUCAST guidelines) and genome sequencing (Illumina and Oxford Nanopore Technologies, Unicycler assembly) were performed on strain DJ. Additionally, we collated 217 publicly available CG147 genomes (NCBI, May 2019). CG147 evolution was inferred within a temporal phylogenetic framework (BEAST) based on a recombination-free sequence alignment (Roary/Gubbins). Comparative genomic analyses focused on resistance and virulence genes and other genetic elements (BIGSdb, Kleborate, PlasmidFinder, PHASTER, ICEFinder and CRISPRCasFinder).

Strain DJ had a pandrug resistance phenotype. Its genome comprised 7 plasmids and 1 linear phage-plasmid. Four carbapenemase genes were detected: $bla_{\text{NDM-5}}$ and 2 copies of $bla_{\text{OXA-181}}$ in the chromosome, and a second copy of $bla_{\text{NDM-5}}$ on an 84 kb IncFII plasmid. CG147 genomes carried a mean of 13 acquired resistance genes or mutations; 63% carried a carbapenemase gene and 83% harbored $bla_{\text{CTX-M}}$. All CG147 genomes presented GyrA and ParC mutations and a common subtype IV-E CRISPR-Cas system. ST392 and ST273 emerged in 2005 and 1995, respectively. ST147, the most represented phylogenetic branch, was itself divided into two main clades with distinct capsular loci: KL64 (74%, DJ included, emerged in 1994 and disseminated worldwide, with carbapenemases varying among world regions) and KL10 (20%, 2002, predominantly found in Asian countries, associated with carbapenemases NDM and OXA-48-like). Further, subclades within ST147-

KL64 differed at the yersiniabactin locus, OmpK35/K36 mutations, plasmid replicons and prophages. The absence of IncF plasmids in some subclades was associated with a possible activity of a CRISPR-Cas system.

K. pneumoniae clonal group CG147 comprises pandrug- or extensively-resistant isolates and carries multiple and diverse resistance genes and mobile genetic elements, including chromosomal $bla_{\text{NDM-5}}$. Its emergence is driven by the spread of several phylogenetic clades marked by their own genomic features and specific temporo-spatial dynamics. These findings highlight the need for precision surveillance strategies to limit the spread of particularly concerning CG147 subsets.

Data Summary

Sequence read files and complete genome assembly of DJ have been deposited in European Nucleotide Archive under the BioProject number PRJEB41234.

INTRODUCTION

The increasing number of antimicrobial resistant infections by *Klebsiella pneumoniae* (Kp), especially by extended-spectrum beta-lactamases (ESBL)- and carbapenemase-producing Kp, led to the declaration of Kp as an 'urgent threat' and 'priority pathogen' by public health agencies (1,2). Molecular analyses of Kp isolates has evidenced that the rapid emergence of multidrug resistant (MDR) Kp is largely driven by the geographic spread of successful clonal groups (CG; e.g., CG15, CG101, CG147, CG258, CG307) (3), some of them carrying epidemic resistance plasmids (4). In order to treat MDR Kp infections, last-resort drugs such as polymyxins (especially colistin) and tigecycline are used (3,5). Consequently, resistance is also observed to these last resort drugs, especially to colistin, and may culminate in the emergence and spread of pandrug-resistant strains (5,6). Pandrug-resistant Kp strains leave few or no therapeutic options and are associated with high mortality rates (7–11).

The 7-gene MLST sequence type (ST) ST147 has been recognized as a globally distributed antimicrobial resistance clone (12), and is closely related to ST273 and ST392, which themselves comprise MDR isolates. Based on genomic classifications, these three STs are grouped into the CG147 clonal group (13,14). The earliest studies of CG147 date from 2008-2010 in Hungary, and correspond to ciprofloxacin-resistant CTX-M-15-producing ST147 isolates, which had been disseminating in the country since 2005 (15,16). Between 2010 and 2014, CG147 (mainly ST147) was described worldwide in association with several carbapenemases (17). Most reported CG147 isolates are from clinical samples, although some were found in companion animals, river waters, chimpanzees, poultry animals and poultry environment (18–24). The above studies were locally restricted and so far, no study of the global spread and genome dynamics of this clone was performed.

Here, we report a pandrug-resistant clinical isolate from India belonging to ST147, and investigate the genomic evolution and antimicrobial resistance gene dynamics in the global CG147 population. We also analyze the phylogenetic context of virulence-associated genomic features, CRISPR loci and mobile genetic elements in this successful Kp sublineage.

MATERIAL AND METHODS

Isolation and phenotypic and genomic characterization of strain DJ

Isolate DJ was recovered from the urine of a 45-years old female patient diagnosed with a urinary tract infection (UTI) in Vadodara (Gujarat, India) in October 2016. The isolate was confirmed to be Kp by biochemical tests and 16S *rRNA* sequencing. Antibiotic susceptibility tests performed using a semi-automated commercial system (Vitek, BioMérieux) revealed resistance to all antibiotics tested. Confirmatory antimicrobial susceptibility tests were carried out for colistin and tigecycline using broth dilution; and for fosfomycin (agar supplemented with 25 mg/L glucose 6-phosphate sodium salt) using agar dilution method. Disk diffusion method was used for the other antimicrobial classes (penicillins, cephalosporins, carbapenems, monobactams, fluoroquinolones, aminoglycosides, macrolides, tetracyclines, phenicols and inhibitors of the folic acid pathway). Results were interpreted using both the Clinical and Laboratory Standards Institute (25) and the European Committee on Antimicrobial Susceptibility Testing (2018) (http://www.eucast.org/) guidelines.

DNA extraction was performed using XpressDNA Bacteria kit (MagGenome Technologies Pvt Ltd., India). Whole-genome sequencing data were generated using (i) an Illumina NextSeq-500 platform with a 2 x 150 nt paired-end protocol (Nextera XT library, Illumina, San Diego, CA); and (ii) long-read Oxford Nanopore sequencing using MinION device integrated with a FLO-MIN-106 flow cell and libraries prepared using a 1D ligation sequencing kit (SQK-LSK109) following the protocol for 1D genomic DNA (gDNA) long reads without BluePippin (Oxford Nanopore Technologies, New York, NY, USA). *De novo* assemblies of the reads were obtained using SPAdes v3.12.0 (26) for Illumina data and using Unicycler v0.4.4 (27) for hybrid assembly. Assembled sequences were annotated using

Prokka v1.12 (28). Reads and assembly were deposited at the European Nucleotide Archive database (under the BioProject PRJEB41234).

Global dataset of publicly-available genomic sequences

All publicly available CG147 Kp genomes from NCBI RefSeq repository of genome assemblies (May 2019) were downloaded. From the 245 CG147 genomes available, duplicated (n=5) and poor-quality genomes (n=6; genome size and G+C content not matching with Kp and/or more than >1000 contigs), and those without attached isolation year (n=17) were excluded. The final dataset comprised 218 genomes, including strain DJ. Sample information, accession numbers, and biological characteristics of the genomes are given in **Table S1**.

Phylogenetic analyses

For phylogenetic analyses, a core-genome alignment based on the concatenation of 4,529 core genes was obtained using Roary v3.12 (29) using a blastP identity cut-off of 90% and core genes defined as those being present in more than 90% of the isolates. Recombination events were removed from the core-genome alignment using Gubbins v2.2.0 (30). The final recombination-free alignment comprised 8,450 single-nucleotide variants (SNVs) and was used to construct a maximum-likelihood phylogenetic tree using IQ-TREE v1.6.11 (model GTR+F+ASC+G4). The tree was rooted with a Kp ST258 NJST258_2 (accession number: GCF_000597905.1) and a Kp ST37 INF042 (GCF_002752995.1) (**Figure S1**).

To evaluate the strength of the temporal signal of our molecular phylogeny, we first conducted a linear regression analysis of the root-to-tip genetic distances as a function of the sample collection year, using TempEst v1.5.3 (http://tree.bio.ed.ac.uk/software/tempest/)

(**Figure S2**). The final recombination-free alignment was then subjected to Bayesian phylogenetic analysis using BEAST v2.6.1 (run with a Markov chain Monte Carlo length of 1x 10^9, sampling every $5x10^3$ steps) (31). We used model parameters that had the best fit: GTR substitution model, lognormal relaxed clock and constant population size. Parameters estimates were computed using Tracer v1.7.1, and a maximum clade credibility tree was obtained with TreeAnnotator v2.6.0. and visualized in FigTree v1.4.4.

MLST and genomic analyses of resistance, virulence and other genetic elements

Multilocus sequence typing (MLST, 7 genes) was performed using the Institut Pasteur *Klebsiella* MLST (32) database (https://bigsdb.pasteur.fr/klebsiella/). Kleborate (33) and BIGSdb analytical tools (https://bigsdb.pasteur.fr/klebsiella/) (34) were used to define the presence of antimicrobial resistance, heavy metal tolerance, virulence genes and to characterize the capsular gene cluster . Geneious Prime 2019.1.1 software (https://www.geneious.com) was used for further manual curation of antibiotic resistance genes, and ISFinder (https://isfinder.biotoul.fr) was used to look for the insertion sequences in the resistance genes or in their genetic context. Plasmid replicons were detected using PlasmidFinder (https://cge.cbs.dtu.dk/services/PlasmidFinder/) (35), whereas prophages, integrative and conjugative elements (ICEs) and CRISPRs were identified using PHASTER (https://phaster.ca) (36), ICEfinder (https://db-mml.sjtu.edu.cn/ICEfinder/ICEfinder.html) and CRISPRCasFinder (https://crisprcas.i2bc.paris-saclay.fr/CrisprCasFinder/Index) (37), respectively. To depict co-resistance genotypes and plasmid networks, we constructed a correlation matrix for binary variables (1, presence; 0, absence) using the 'corr.test' function (Pearson method, which for a pair of binary variables compares to the Phi coefficient) from the 'corrplot' R package. Significant correlations were visualized with the 'corrplot' function from the same package. Statistical analyses to check the association of the different

categorical variables within the phylogeny groups were calculated using the χ^2 test (*P* values of <0.05 were considered statistically significant).

RESULTS

Phenotypic and genomic features of pan-drug resistant strain DJ

Strain DJ was resistant to all tested antimicrobial agents, including last-resort antimicrobials such as carbapenems, colistin, tigecycline and fosfomycin (**Table 1**), and is therefore pandrug-resistant (6). To define its genetic mechanisms of resistance, a hybrid complete genome assembly was produced (**Figure S3a**). The 5.7 Mb sequence was 56.9% G+C-rich and made up of one chromosome and 6 circularized plasmids [123kb IncFII(pKPX1); 57 kb IncR; 5.6 kb ColRNAI; 4.7 kb ColRNAI; 2.0 kb ColpVC; and 1.5 kb ColMG828]. In addition, there were two non-circularized contigs: a 84 kb IncFII plasmid and a 57 kb contig corresponding to a N15-*like* phage-plasmid (P-P) coding for a protelomerase (*telN*) responsible for the maintenance of its linear genome (38). PHASTER identified four other prophages within the chromosome. Phylogenetic analysis of strain DJ showed it belonged to *K. pneumoniae sensu stricto* (*i.e.*, phylogroup Kp1) and to the ST147-KL64 lineage previously described as endemic in India (12).

The long-read sequencing followed by hybrid assembly of strain DJ allowed the identification of multiple copies of unique antibiotic resistance genes in the different genomic elements depicted (chromosomes and plasmids). Strain DJ harbored four carbapenemase genes, corresponding to two copies of each $bla_{\text{NDM-5}}$ and $bla_{\text{OXA-181}}$. Additionally, three copies of $bla_{\text{CTX-M-15}}$ were detected (**Table 1**). The above genes were localized as follows (**Figure S3, Figure 1)**: (i) $bla_{\text{NDM-5}}$: one copy in the chromosome, and the second copy in the 84 kb IncFII plasmid; (ii) $bla_{\text{OXA-181}}$: two copies in the chromosome; and (iii) $bla_{\text{CTX-M-15}}$: two copies in the chromosome and one copy in the 57 kb IncR plasmid. In addition, a 61 kb region from the IncFII plasmid, comprising several antimicrobial resistance genes included in a class 1 integron and the replication machinery of the plasmid, was duplicated in the chromosome (**FigureS3a,c**).

Additional molecular determinants of antimicrobial resistance were observed in strain DJ. Most notably, the gene *mgrB* was disrupted by an IS*5* transposase (1057 bp), consistent with the observed colistin resistance phenotype (MIC=4 μg/ml). This strain also carried *rmtF*, *rmtB*, *strA*, *strB* and *aadA2* genes, associated with resistance to aminoglycosides including amikacin **(Table 1)**. Quinolone resistance determining region (QRDR) mutations were observed, leading to GyrA-S83I and ParC-S80I amino acid alterations. Last, there was a premature stop codon caused by a A580T substitution in RamR, a negative regulator of RamA, itself a transcriptional activator of the *acrAB* genes. Higher production of AcrAB increases the efflux of tigecycline (39), consistent with the resistance phenotype observed for this agent (MIC=16 μg/ml).

Regarding virulence genes, strain DJ harbored a complete yersiniabactin gene cluster (*ybt*10) located on an ICE*Kp*4 mobile genetic element (**Figure S3a**). Type 1 (*fimAICDFGH*) and type 3 (*mrkABCD*) fimbriae gene clusters were also observed, but none of the *rmpACD*, aerobactin and salmochelin cluster genes, typically associated with hypervirulence, were present.

208 **Table 1**. Antimicrobial susceptibility of *Klebsiella pneumoniae* strain DJ, and genes

209 potentially conferring resistance.

Class and antimicrobial agent	**Diameter (mm) or MIC (µg/ml)**	**Interpretation**[1]	**Associated resistance genes** (copy number)
Beta-lactams			$bla_{\text{NDM-5}}$ (2), $bla_{\text{OXA-181}}$ (2), $bla_{\text{CTX-M-15}}$ (3), $bla_{\text{TEM-1}}$ (2), $bla_{\text{SHV-11}}$ (1), disrupted *ompK35*, mutation in *ompK36*
Ampicillin	9	R	
Piperacillin	13	R	
Amoxycillin-clavulanic acid	11	R	
Ticarcillin-clavulanic acid	14	R	
Piperacillin-tazobactam	14	R	
Cefuroxime	11	R	
Cefotaxime	12	R	
Ceftazidime	15	R	
Cefepime	15	R	
Aztreonam	14	R	
Ertapenem	12	R	
Imipenem	14	R	
Meropenem	15	R	
Aminoglycosides			*rmtB* (1), *rmtF* (2), *aac(6')-Ib* (2), *aadA2* (3), *strAB*(2)
Amikacin	10	R	
Gentamycin	10	R	
Tobramycin	9	R	
Quinolones and Fluoroquinolones			*gyrA* and *parC* mutations
Nalidixic acid	11	R	
Ofloxacin	11	R	
Folate pathway inhibitors			*sul1* (3), *sul2* (1), *dfrA12* (3)
Trimethoprim/Sulfamethoxazole	19	R	
Trimethoprim	8	R	
Polymyxins			disrupted *mgrB*
Colistin	MIC=4	R	
Tetracyclines			*ramR* mutation[2]
Tigecycline	MIC=16	R	
Phenicols			*catA2* (1), *catB* (2)
Chloramphenicol	12	R	
Other			
Fosfomycin	MIC=128	R	*fosA*
Rifampicin	NI	-	*arr-2* (2)
Macrolides	NI	-	*mphA* (1), *ermB* (1)

210 NI, not included in the antimicrobial susceptibility testing panel; R, resistant; [1]Interpretations were based on CLSI and
211 EUCAST guidelines; [2] Nucleotide mutation resulting in a premature stop codon.

Time-scaled phylogenetic structure of the global population of sublineage CG147

We investigated the evolutionary origins of strain DJ within the global diversity of CG147 using 217 publicly available genomes of isolates collected between 2002 and 2018. CG147 genomes were mainly isolated from human samples (90%, 196/218) and hospital environment (6%, 12/218), mostly in European (40%), Southeastern Asia (17%), Southern Asia (11%) and Northern America (10%) countries (**Figure 2)**.

CG147 was deeply structured into three main branches (**Figure S1**), each corresponding to a single MLST sequence type: ST147, ST273 and ST392; the two latter are *tonB* variants of ST147. The number of genome-wide nucleotide substitutions was associated with isolation dates (root-to-tip regression analysis: R^2 =0.1123, **Figure S2**), enabling to infer a time-scaled phylogeny (**Figure 3**). The evolutionary rate within CG147 was estimated at $1.45x10^{-6}$ substitutions/site/year (95% HPD, $1.12x10^{-6}$ – $1.78\ x10^{-6}$), corresponding to 6.2 SNPs per genome per year. The last common ancestor of CG147 was estimated around year 1896, with a large uncertainty (95% highest posterior density (HPD): 1817-1962). The ST273 lineage was the first to diverge, whereas ST147 and ST392 shared a common ancestor, estimated around 1921 (95% HPD, 1868-1970) (**Figure 3**).

ST147 was the most represented (79%, 172/218 genomes) and geographically widespread lineage (**Figure 2**), and emerged around 1965 (95% HPD, 1940-1986) (**Figure 3**). The phylogenetic structure within ST147 revealed five main clades characterized by distinct capsular (KL type) and liposaccharide O antigen loci. Clades KL64-O2 (74%; 128/172) and KL10-O3a (20%, 34/172) emerged in recent years: 1994 [95% HPD, 1990-1998] and 2002 (95% HPD, 1995-2008), respectively (**Figure 3**). Whereas KL64-O2 genomes were predominantly from Europe (54%), KL10-O3a was mainly sampled from Asia (85%; **Figure 2**).

The ST392 and ST273 branches emerged in 2005 (95% HPD, 2001-2009) and 1995 (95% HPD, 1987-2001), respectively (**Figure 3**). ST392 (16% of genomes, 34/218) is distributed globally and harbors a KL27 capsular gene cluster and a O4 antigen. In contrast, ST273 (6%, 12/218) was predominantly found in Asia (64%) and carries KL74 and O3b gene clusters (**Figure 2, Figure 3**). Of note, a group of closely related genomes from the Philippines (n=5) lacked a capsular gene cluster, with only *ugd* being detected. This gene had 100% identity with the *ugd* gene from the KL74 reference strain, suggesting a recent loss of the capsular gene cluster.

Acquired antimicrobial resistance genes and their evolutionary dynamics within CG147

All genomes presented QRDR alterations in GyrA and ParC. The topoisomerase ParC S80I alteration was fully conserved, whereas the GyrA gyrase subunit S83I amino-acid change was observed in all genomes except for clade ST147-KL10, which had S83Y, caused by an ATC to TAC codon change. In addition, this clade had an D87A alteration (**Figure 3**).

The number of acquired antimicrobial resistance genes or mutations among CG147 genomes ranged from 2 to 23 (mean: 13; **Table S1**). Regarding beta-lactam resistance, in addition to the conserved chromosomal $bla_{\text{SHV-11}}$ gene, a majority of CG147 isolates carried a $bla_{\text{CTX-M}}$ (n=181, 83%), 94% of which was the $bla_{\text{CTX-M-15}}$ variant. In addition, 63% (n=137) genomes harbored at least one carbapenemase gene, with 14% (n=19) of these harboring more than one copy of the same carbapenemase gene, and/or two or more carbapenemase genes from different families (**Table S1**). Noteworthy, carbapenemase genes were significantly more frequent in ST147 (69%) compared to ST392 (35%) and ST273 (58%; p=0.0005). The two main ST147 clades were similar to this respect (ST147-KL10: 74%; ST147-KL64: 71%; p=0.78). However, their carbapenemase genes were distinct across world regions: there was a predominance of bla_{NDM} in Southeastern Asia and Northern America,

whereas the combination of *bla*$_{NDM}$ and *bla*$_{OXA-48-like}$ was almost exclusively detected in Southeastern Asia, as observed in strain DJ. In contrast in Europe, *bla*$_{OXA-48-like,}$ *bla*$_{KPC-2}$ and *bla*$_{VIM-1/-27}$ were the most frequent carbapenemases (**Figure S4; Table S1**).

Different y*bt*/ICE*Kp* subtypes (associated with hypervirulence) and OmpK35/K36 mutations (associated with multidrug resistance) were observed within the ST147-KL64 clade (**Figure 3; Table S1**). First, a group of 29 genomes (denominated subclade 1; mean number of SNPs among them: 57), emerged around 2007, was characterized by the presence of *ybt*16/ICE*Kp*12 and an altered OmpK35 protein, due to a deletion of 2 nt resulting in a premature stop codon. Within this subclade itself, a subgroup of genomes (n=19/29) harbored *bla*$_{OXA-48-like}$ genes and the OmpK36GD mutation observed previously (40). Second, a group of 22 genomes emerged around 2009 (subclade 2; mean of 60 SNPs) was defined by the presence of *ybt*10/ICE*Kp*4 and an OmpK35 gene disrupted by IS*Ecp1*-*bla*$_{CTX-M-15}$ (**Figure 1, Figure 3, Table S1**). As observed in subclade 1, all subclade 2 isolates carrying *bla*$_{OXA-181}$ shared the same OmpK36TD mutation (40). Subclades 1 and 2 also differed in plasmid replicon content: whereas the former were rich in IncR (90%) and IncHIB/IncFIB(Mar) (41%), in contrast IncFII (pKPX1) (82%), IncFII (59%) and IncR (64%) were frequent in the latter (**Table S1**). Differently, the remaining ST147-KL64 genomes (n=77) often carried IncFIB$_K$ (65%), IncFII$_K$ (49%; a common pKPN-3-derived plasmid found in Kp harboring *pco* and *sil* clusters; (4,41) and IncFIA(HI1) (38%).

Strain DJ belonged to subclade 2 and was phylogenetically closely related (<28 SNPs) to five other isolates recovered between 2014 and 2015 in different Asian countries (**Figure S3b; Figure3**) and described as extremely drug resistant or pandrug-resistant (11,42,43), with identical plasmids being observed among them but no chromosomal integration of *bla*$_{NDM-5}$ (**Figure S3**).

Convergence of antimicrobial resistance and virulence

Whereas the yersiniabactin virulence factor gene cluster was rare amongst genomes of ST392 (6%; 2/34) and ST273 (8%; 1/12), it was observed in 53% of ST147 genomes. There were two predominant variants (*ybt*16/ICE*Kp*12 and *ybt*10/ICE*Kp*4 associated with subclades 1 and 2, respectively), and four minority ones (**Figure 3; Table S1**).

Two isolates with hypervirulence genotypes, defined by the presence of *rmpA* and/or *rmpA2* or aerobactin, were observed: KpvST147L (ST147-KL14, 2016, United Kingdom) and B-8658 (ST147-KL10, 2014, Russia). The plasmid from KpvST147L was fully sequenced previously (GenBank accession number NZ_CM007852) and carries *rmpA*, *rmpA2* and aerobactin; it corresponds to a 343 kb IncFIB-IncHIB (pNDM-MAR) plasmid (**Figure S5**) (44). This plasmid was recently described in three ST147 *Kp* isolates recovered in 2018-2019 in the United Kingdom and in other sublineages (45). Here, we found that Russian strain B-8658 (ST147-KL10, 2014) also acquired the pNDM-MAR plasmid, but it lacked the *rmpA* gene (**Figure S5**).

Co-occurrence of antimicrobial resistance and heavy metal tolerance genes and plasmids

Antimicrobial resistance genes, mutations and plasmids co-occurred in a structured way (**Figure S6; Table S1**). For example, Col(BS512), ColKp3, IncFIA (HI1), Cml, EreA/B and GyrA S83Y/D87A co-occurred frequently within the ST147-KL10 clade. Associations of (i) VIM, QnrA1, AadB and IncHI2; (ii) OXA-48-like and IncL(pOXA-48) or (iii) KPC and IncFIB(pQIL) were also observed, consistent with previous descriptions of genetic elements co-carrying these genes (46–48). In addition, genes conferring tolerance to copper and silver were associated with $IncFII_K$/$IncFIB_K$ plasmids (*pco* and *sil* operons; $p<0.00001$), whereas genes for mercuric resistance were linked to IncR plasmids ($p=0.0082$). Similarly, a positive

association between the tellurite cluster and IncHIB/IncFIB (MAR) was observed (p<0.0001) (4). Last, a negative association between IncR and $IncFII_K$/$IncFIB_K$ was observed.

CRISPR-Cas systems

We investigated whether resistance gene dynamics could be influenced by CRISPR-Cas systems in CG147. There were either one (83%; 180/218) or two (17%; 37/218) CRISPR-Cas systems amongst CG147 genomes. A conserved subtype IV-E system was found in all genomes, located in the *iap – cysH* region, with a %G+C of 60.6%, and defined arbitrary as CRISPR1. Direct repeat sequences were highly conserved but the number and sequences of spacers varied across CG147 genomes. CRISPR1 variant v0 (43 spacers) was present in 56% of the genomes (122/218) and was used as reference to define CRISPR1 variants (**Figure S7; TableS2**). Four of the CRISPR1 v0 spacers (spacers 1, 28, 29 and 42) matched sequences on MDR IncF plasmids disseminated among different *Enterobacteriaceae*, including Kp. Spacer 1 matched a multicopy intergenic region, whereas spacers 28 and 29 matched a DUF3560-domain containing protein and spacer 42 a hypothetical protein located upstream of SAM-methyltransferase. Although these spacers were highly conserved among CRISPR1 variants, they were also found at high frequency (in 28% to 53% of the strains) as protospacers located in plasmid contigs. Noteworthy, in 28% (62/218) of CG147 genomes these plasmid protospacers were not detected, and the majority of these strains (65%, 40/62) belonged to subclades 1 and 2, which were also characterized by the absence of IncFII(K) and IncFIB(KpQIL) plasmid replicons. The association of the lack of IncF plasmids with IncF-targeting spacers in CRISPR1, suggests a possible activity of this CRISPR system in subclades 1 and 2. Finally, twelve of the forty-three spacers from CRISPR1 v0 targeted prophages; of these, spacers 6 and 25 were found as protospacers in some of the CG147 genomes (28% and 18%, respectively).

Six other CRISPR-Cas systems, distinct from CRISPR1, were observed in 37 genomes. CRISPR2 to CRISPR4, of type IV-A3, were strongly associated with IncHIB and/or IncFIB (pNDM-MAR) plasmids ($p<0.00001$), as previously described (49). Two variants of CRISPR2 (v1; n=12; 17 spacers and v3; n=5, 25 spacers) were prevalent (**Figure S7; TableS2**). CRISPR2 to CRISPR4 systems shared identical direct repeat sequences but showed a high diversity in the number and sequences of spacers. Their tendency to target $IncFII_K$/$IncFIB_K$ plasmids has led to suggest a role in inter-plasmid competition (49,50). However, here the presence of CRISPR2 to CRISPR4 systems was not uniformly associated with an absence of $IncFII_K$/$IncFIB_K$ plasmids.

Prophage elements

CG147 genomes harbored 0 to 7 prophages (mean of 4 prophages per genome; considering only the intact ones). Some prophages were frequent, including (i) ST147-VIM1phi7.1-*like* (GenBank accession no. NC_049451; *Myoviridae* family), which was present in 90% of the genomes (196/218); (ii) *Salmonella* phage 118970_sal3-*like* (GenBank accession no. NC_031940; *Myoviridae* family; 70% [152/218]); and (iii) *Enterobacteria* phage mEp237-*like* (GenBank accession no. NC_019704; *Siphoviridae* family; 56% [121/218]) (**Figure S7).**

The N15-*like* phage-plasmid (*Siphoviridae* family) we uncovered in the genomic assembly of strain DJ was present in 37% of CG147 genomes (81/218). It was strongly associated ($p<0.00001$) with ST147-KL64 subclades 1 and 2 and present in 92% of these genomes **(Table S1, Figure S7**). These two subclades were also enriched in other prophages (mean number of prophages: 5) compared with the remaining CG147 genomes (mean number of prophages=3).

DISCUSSION

This study was triggered by the discovery of a pandrug resistance phenotype in strain DJ from India, which prompted us to analyzed its genomic features and understand their dynamics in the context of a large genome dataset of CG147 from multiple world regions. The results revealed its deep phylogenetic structure and a capacity of CG147 members to acquire a wide range of antimicrobial resistance or virulence elements, plasmids and prophages. Of these, *Klebsiella* capsular locus (KL) switches, which occurred repeatedly, represent prominent phylogenetic markers of CG147 clades. The evolutionary dynamics of these surface structure in CG147 echo those observed in other MDR clonal groups, such as ST258 or ST307 (13,47,51). The multiple clonal expansions of sublineages with distinct KL or O types has clear implications for diagnostic, control or therapeutic strategies such as vaccination and phage therapy. Although its reliance on public genomic sequences may expose this study to a bias towards antibiotic-resistant CG147 isolates, as well as those involved in nosocomial outbreaks, this study provides new insights into the evolutionary history, epidemiology and population dynamics of this important emerging clonal group.

The genomic arsenal of antimicrobial resistance features of subclade 2, to which strain DJ belongs, suggest extensively drug and pandrug resistance is not restricted to strain DJ, but rather is a shared characteristic in this particular subclade, already disseminated among Asian countries (11,42,43,52–54). Clonal spread of this subclade between different countries was reported (52). This subclade should be closely monitored and may represent a pioneering situation ushering the worrying prospect of pan-resistance in other CG147 subclades.

In the case of strain DJ, we observed the integration of $bla_{\text{NDM-5}}$ in the chromosome. Previous studies have reported the one or multiple copies of $bla_{\text{CTX-M-15}}$ and $bla_{\text{OXA-181}}$ integrated into the chromosome of Kp isolates (55,56), including in the high-risk ST147 (52,57). However, the chromosomal integration of bla_{NDM} was, to our knowledge, only

386 reported once in Kp, in two NDM-1 producing ST14 clinical isolates from Thailand (58). In
387 that case, the chromosomal integration was mediated by IS*5* and the Tn*3* transposase. In
388 strain DJ, the integration of *bla*$_{NDM-5}$ may have been mediated by IS*Ecp*1/IS*26* (**Figure 1a**).
389 The genomic rearrangement due to the replicative transposition of IS*26* has previously been
390 shown for IncFII-*bla*$_{NDM-5}$ bearing plasmids (59). The chromosomal incorporation of
391 carbapenemase and ESBL genes is concerning, as it may stabilize these genes by promoting
392 their vertical dissemination (52).

393 The convergence of MDR and hypervirulence genotypes is being increasingly
394 observed in Kp (3). Here, this worrisome association was detected in two phylogenetically
395 distinct CG147 isolates from 2014 (Russia) and 2016 (United Kingdom), which shared an
396 MDR-Hv IncHIB/FIB plasmid (44). The recent observation of this plasmid in ST101 and
397 ST147 isolates from the United Kingdom (45), with no epidemiological link with the isolate
398 from 2016, suggests its continuous circulation and further risk of horizontal spread.

399 The ST147-KL64 lineage is globally disseminated. The evolutionary rate we
400 estimated ($1.03x10^{-6}$ substitutions/site/year) is very similar to other MDR global sublineages,
401 such as ST258 ($1.03x10^{-6}$ substitutions/site/year) and ST307 ($1.18x10^{-6}$
402 substitutions/site/year) (47,51), and slightly slower than the one estimated for ST101
403 ($2.85x10^{-6}$ substitutions/site/year) (60). It is striking that the emergence of ST147-KL64
404 lineage occurred approximately at the same time as other MDR sublineages ST258 (year
405 1995), ST307 (1994) and ST101 (1989) (47,51,60). In addition, the presence of GyrA and
406 ParC QRDR alterations is a common characteristic to these MDR high-risk sublineages
407 [CG258 and ST307: GyrA-S83I and ParC-S80I; ST101: GyrA-S83I, GyrA-D87G/N/A and
408 ParC-S80I] (47,51,60). This phenotypic and temporal conjunction points to common drivers
409 and suggests a role of the usage of fluoroquinolones, introduced into clinical practice in the
410 end of the 1980s, in the emergence of MDR Kp sublineages. This is reminiscent of the

scenarios of emergence of *Escherichia coli* ST131 and methicillin-resistant *Staphylococcus aureus* ST22 (61–63).

The drivers of genomic diversification of emerging Kp sublineages may include a combination of ecological opportunities to acquire genetic elements, local selective pressure, and molecular mechanisms that enable or restrict genetic flux. Of these, CRISPR-Cas systems may play a role (64–67). In CG258, an association was suggested between the absence of these systems and the ability to acquire IncF plasmids [such as bla_{KPC}-IncF(pKpQIL-like plasmids)] (68–70). Here we found a conserved type IV-E CRISPR-Cas system (CRISPR1) within CG147, consistent with early reports (9), with four spacers matching IncF plasmid sequences. The distribution of CRISPR1 in the broader Kp species shows a unique association with CG147, with only two exceptions (ST2746 and ST3700; based on 1001 Kp genomes representing unique STs; selected from a dataset of 4222 genomes from NCBI, November 2018; data not shown). Type IV CRISPR-Cas systems primarily target plasmids (50). However, in the majority of CG147 strains, corresponding protospacers were found in plasmid sequences, advocating that the immunity provided by this CRISPR1 system might not be fully functional. In contrast, ST147-KL64 subclades 1 and 2 were largely devoid of these protospacers and did not carry IncFIA, $IncFII_K$, $IncFIB_K$ and IncFIB (pQil) plasmids **(Figure S7)**. This observation may suggest a possible activity of the CRIPSR1 system in these recently emerged subclades. Future experimental studies are needed to explore this hypothesis. Noteworthy, it was also among these subclades that the N15-*like* phage-plasmid was prevalent. Among bacterial phyla, this phage-plasmid family was found almost exclusively in Kp (38), and we found that 58.9% of these belonged to CG147. A possible biological role of plasmid-prophages has yet to be established.

CONCLUSIONS

436 The presence of pandrug-resistant and extremely-drug resistant isolates in CG147,
437 together with the high genetic plasticity and rapid emergence dynamics of this clone,
438 represents a clear threat to public health. CG147 is globally disseminated but shows a strong
439 phylogenetic structure, with different clades being associated with specific genomic features
440 and geographical distributions. These observations underline how different variants of
441 CG147 contribute to the major public health threat posed by Kp, and call for specific
442 surveillance and directed control strategies of this clone and its particularly concerning clades
443 (12).

444 A possible link between the absence of IncF plasmids and the activation of CRISPR1
445 defense system is intriguing. This observation calls for more work on mechanistic drivers of
446 the flux of genetic elements across MDR bacterial lineages. Precise phylogenetic mapping
447 and understanding of the dynamics of antimicrobial resistance features are needed to guide
448 the development of control strategies, especially those that target specific subsets of strains
449 within pathogenic bacterial species, such as CRISPR delivery or toxic conjugation systems
450 (71,72).

DECLARATIONS

Ethics approval and consent to participate: To conduct the research, we used bacterial strain DJ, which is not considered a human sample. Accordingly, this research was not considered human research and is out of the scope of the decree n° 2016-1537 of November 16, 2016 implementing law n° 2012-300 of March 5, 2012 on research involving human subjects. Therefore, no ethics approval was needed and no informed consent was required.

Consent for publication: No personal identifying data was used; therefore, no consent was necessary.

Availability of data and materials: All data generated or analysed during this study are included in this published article and its supplementary information files.

Competing interests: The authors declare that they have no competing interests.

Funding: SD availed a UGC-BSR fellowship for research students and was supported financially by EMBO fellowship sponsored by European Molecular Biology Organization (STF_7993) for a visit in the Brisse lab. CR was supported financially by the MedVetKlebs project, a component of European Joint Programme One Health EJP, which has received funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement No 773830, and by a Roux-Cantarini grant from Institut Pasteur.

Author contributions: S.D. and D.G. coordinated the microbiological cultures of the isolate DJ and its antimicrobial susceptibility testing and biochemical characterization. V.P., C.R. and S.D. performed the genomic sequencing. C.R. designed and coordinated the comparative genomics study, with input from S.D. C.R. wrote the initial version of the manuscript. All authors provided input to the manuscript and reviewed the final version.

Acknowledgments: We thank the Plateforme de Microbiologie Mutualisée (P2M) of Institut Pasteur for Illumina sequencing.

Authors license statement: This research was funded, in whole or in part, by Institut Pasteur and by European Union's Horizon 2020 research and innovation programme. For the purpose of open access, the authors have applied a CC-BY public copyright license to any Author Manuscript version arising from this submission.

FIGURE LEGENDS

Figure 1. Genetic context of the different copies of *bla*$_{\text{NDM-5}}$ **(a)**, *bla*$_{\text{OXA-181}}$ **(b)** and *bla*$_{\text{CTX-M-15}}$ **(c)** genes identified in *K. pneumoniae* strain DJ.

Within each panel, the chromosomal and/or plasmid-encoded copies of *bla*$_{\text{NDM-5}}$, *bla*$_{\text{OXA-181}}$ and *bla*$_{\text{CTX-M-15}}$ detected in DJ are compared between them. Predicted open reading frames (ORFs) are represented on each line by colored arrows, with arrowheads indicating the direction of transcription: antimicrobial resistance genes (red), mobile genetic elements (MGEs) or gene mobilization-related genes (green); RNA-binding proteins (dark yellow); other functions (light gray). Disrupted genes are outlined with dotted lines. Dark gray blocks connecting the distinct gene regions represent homology levels, as indicated in the gradient key. The nucleotide positions of the represented regions are indicated at the left below the copy descriptors. Figures were created using the Easyfig (https://mjsull.github.io/Easyfig/).

Figure 2. Geographic **(main)** and temporal **(inset)** distribution of genomes included in this study.

The pie charts represent the frequency of each CG147 clade in each country (see size and color keys). **Inset:** Bars represent the number of isolates per year for which genome assemblies were available (NCBI RefSeq) as of May 2019, coloured by clade. Red lines represent (solid line) the number of PubMed-indexed records as of March 2020 (identified using the search criteria '*Klebsiella pneumoniae*' and 'ST147' or 'ST392' or 'ST273', resulting in a total of 99 distinct entries); and (dotted line) the cumulative number of genomes. The scale on the left Y-axis refers to total number whereas the one on the right refers to cumulative numbers.

Figure 3. Time-scaled phylogeny of 218 CG147 genomes and their epidemiological and molecular characteristics.

The phylogeny was obtained using the BEAST tool. The three main branches correspond to the three main sequence types (ST). Tree tips are colored by world region of isolation (see key). Black dots on main nodes indicate ≥95% posterior probability. The grey boxes delineate subclades 1 and 2, as indicated. Capsular (KL) and O-antigen (O) locus types and the yersiniabactin-carrying ICE*Kp* elements are colored according to their variants as shown in the legend. Antimicrobial resistance determinants are indicated by colored rectangles when present. In the antimicrobial resistance determinants columns:[1] dark pink indicates $bla_{\text{KPC-2}}$ and light pink indicates $bla_{\text{KPC-3;}}$ [2] dark pink indicates $bla_{\text{CTX-M-15}}$ and dark pink indicates other $bla_{\text{CTX-M}}$ variants; [3]dark pink indicates *mgrB* mutations and dark pink indicates *pmrB* mutations. AMGs: aminoglycosides; FPIs: folate pathway inhibitors.

Supplementary Data:

Figure S1. **Phylogenetic structure of CG147.**

The tree was obtained by maximum likelihood analysis (IQ-TREE) based on the final recombination-free alignment of concatenated nucleotide sequence alignments of 4,517 core genes. The tree was rooted with a Kp ST258 NJST258_2 (GCF_000597905.1) and a Kp ST37 INF042 (GCF_002752995.1). Branch lengths represent the number of nucleotide substitutions per site (scale, 0.001 substitution per site). The three main branches are colored by sequence type (ST) and branch tips are colored by world region of isolation (see key). Capsular (KL) and O-antigen (O) locus types and the yersiniabactin-carrying ICE*Kp* elements are colored according to their variants (see key). Subclades 1 and 2 are denoted by rectangle outlines.

Figure S2. Temporal signal in genomic sequences.

The linear correlation between year of isolation and root-to-tip distance from the maximum likelihood phylogeny for CG147 genomes collection was calculated using TempEst.

Figure S3. **Replicons of strain DJ.**

(**a**) Visual representation of the hybrid Unicycler assembly graph, obtained using Bandage (Wick et al., 2015). Nine replicons were obtained, including the circularized chromosome, 6 circularized plasmids, 1 linear phage-plasmid and 1 non-circularized plasmid contig. The size of replicons is shown.

(**b**) Sequence alignments of the chromosomes of strain DJ and the four closest genomes belonging to subclade 2 (<28 SNPs): MS6671 (GenBank Accession Number: GCA_001455995), DA48896 (GCA_003006175.1); CRKP-1215 (GCA_002786755;

=CRKP-2297) and SKGH01 (GCA_001644765.1). Strain DJ was used as reference. The outermost circle is an annotation of the reference strain showing the location of antimicrobial drug resistance genes, ytb/ICEKp, KL64 capsular locus and the IncFII chromosomal integration region. The figure was created using BRIG (Alikhan et al., 2011).

(**c**) Sequence alignments of the non-circular IncFII replicon of strain DJ with the closest plasmids of subclade 2 strains: pCRKP-2297_2 (CP024836.1); pCRKP-1215_2 (CP024840.1), MS6671_plasmidE (LN824138.1), SKGH01_p3 (CP015503.1), p48896_1 (CP024430.1) and the region integrated in the chromosome of strain DJ. The IncFII non-circularized replicon was used as reference. The outermost circle is an annotation of the reference strain showing the location of antimicrobial drug resistance genes in red. In the table the coverage, identity, size and antimicrobial resistance genes of the plasmids that were used for the comparison. In bold the resistance genes also present in DJ replicon.

(**d**) Sequence alignments of the p48896_1 (CP024430.1) plasmid with the plasmids of the closest strains: pCRKP-2297_2 (CP024836.1); pCRKP-1215_2 (CP024840.1), MS6671_plasmidE (LN824138.1), SKGH01_p3 (CP015503.1), DJ-IncFII non-circularized replicon and DJ-IncR plasmid. p48896_1 was used as reference plasmid. The outermost circle is an annotation of the reference strain showing the location of antimicrobial drug resistance genes in red.

(**e**) Sequence alignments of the 123 kb IncFII(pKPX1) plasmid of strain DJ with the closest plasmids of subclade 2 strains: pCRKP-2297_1 (CP024835.1), pCRKP-1215_1 (CP024839.1), MS6671_plasmidB (LN824135.1), SKGH01_p2 (CP015502.1), p48896_2 (CP024431.1). The IncFII(pKPX1) plasmid of strain DJ was used as reference. The outermost circle is an annotation of the reference strain showing the location of antimicrobial drug resistance genes in red. In the table the coverage, identity, size and antimicrobial

resistance genes of the plasmids that were used for the comparison. In bold the resistance genes also present in DJ replicon.

Figure S4. **Geographic origins of the genomes included in this study and their associated carbapenemase profile.**

The pie charts represent the frequency of each carbapenemase profile in each country (see key). CPKP-: no carbapenemase.

Figure S5. Sequence alignments of MDR-Hv plasmids identified in CG147 genomes.

Comparison of MDR-Hv plasmids identified in CG147 genomes (KpvST147L_NDM and B-8658) with the reference virulence plasmid pK2044 (AP006726.1) and a similar plasmid recovered in 2019 (pKpvST147B, CP040726.1), which was used as reference. The outermost circle shows annotations of the reference strain.

Figure S6. **Correlation between genotypes (Pearson method).**

Blank squares represent correlations without statistical significance ($p > 0.05$). The plot was created with the corrplot R package.

Figure S7. **Time-scaled phylogeny of CG147 genomes and their heavy metal tolerance genes, plasmid replicons, CRIPSR systems, and prophages profile.**

The phylogeny was obtained using the BEAST tool. The three main branches correspond to the three main sequence types (ST). Tree tips are colored by world region of isolation (see key). Black dots on main nodes indicate ≥95% posterior probability. The grey boxes delineate subclades 1 and 2, as indicated. Capsular (KL) and O-antigen (O) locus types and the yersiniabactin-carrying ICE*Kp* elements are colored according to their variants as shown

in the legend. The presence of heavy metal tolerance genes, plasmid replicons, CRISPR-Cas systems, protospacers and prophages is indicated. In heavy metal tolerance genes columns, lighter yellow indicates an incomplete operon. In prophages lighter orange indicates an questionable prophage. HMTG, heavy metal tolerance genes.

Table S1. Epidemiological and genomic characteristics of the 218 CG147 *K. pneumoniae* genomes analysed in this study.

Table S2. CRISPR/Cas systems and variants identified within CG147 *K. pneumoniae* genomes.

REFERENCES

1. CDC. Antibiotic Resistance Threats in the United States, 2013. 2013.

2. WHO. Antimicrobial resistance: global report on surveillance 2014. 2014.

3. Wyres KL, Lam MMC, Holt KE. Population genomics of *Klebsiella pneumoniae*. Nature Reviews Microbiology. 2020;18:344–59.

4. Navon-Venezia S, Kondratyeva K, Carattoli A. *Klebsiella pneumoniae*: A major worldwide source and shuttle for antibiotic resistance. FEMS Microbiol Rev. 2017;41:252–75.

5. Poirel L, Jayol A, Nordmann P. Polymyxins: Antibacterial Activity, Susceptibility Testing, and Resistance Mechanisms Encoded by Plasmids or Chromosomes. Clin Microbiol Rev. 2017;30:557–96.

6. Magiorakos A-P, Srinivasan A, Carey RB, Carmeli Y, Falagas ME, Giske CG, et al. Multidrug-resistant, extensively drug-resistant and pandrug-resistant bacteria: an international expert proposal for interim standard definitions for acquired resistance. Clin Microbiol Infect. 2012;18:268–81.

7. de Man TJB, Lutgring JD, Lonsway DR, Anderson KF, Kiehlbauch JA, Chen L, et al. Genomic analysis of a pan-resistant isolate of *Klebsiella pneumoniae*, United States 2016. MBio. 2018;9.

8. Lázaro-Perona F, Sotillo A, Troyano-Hernáez P, Gómez-Gil R, de la Vega-Bueno Á, Mingorance J. Genomic path to pandrug resistance in a clinical isolate of *Klebsiella pneumoniae*. Int J Antimicrob Agents. 2018;52:713–8.

9. Avgoulea K, Pilato V Di, Zarkotou O, Sennati S, Politi L, Cannatelli A, et al. Characterization of Extensively Drug-Resistant or Pandrug-Resistant Sequence Type 147 and 101 OXA-48-Producing *Klebsiella pneumoniae* Causing Bloodstream Infections in Patients in an Intensive Care Unit. Antimicrob Agents Chemother. 2018;62.

10. Guducuoglu H, Gursoy NC, Yakupogullari Y, Parlak M, Karasin G, Sunnetcioglu M, et al. Hospital outbreak of a colistin-resistant, NDM-1- A nd OXA-48-producing *Klebsiella pneumoniae:* High mortality from pandrug resistance. Microb Drug Resist. 2018;24:966–72.

11. Zowawi HM, Forde BM, Alfaresi M, Alzarouni A, Farahat Y, Chong TM, et al. Stepwise evolution of pandrug-resistance in *Klebsiella pneumoniae*. Sci Rep. 2015;5.

12. Peirano G, Chen L, Kreiswirth BN, Pitouta JDD. Emerging antimicrobial-resistant high-risk *Klebsiella pneumoniae* clones ST307 and ST147. Antimicrob Agents Chemother. 2020;64.

13. Wyres KL, Wick RR, Judd LM, Froumine R, Tokolyi A, Gorrie CL, et al. Distinct evolutionary dynamics of horizontal gene transfer in drug resistant and virulent clones of *Klebsiella pneumoniae*. PLoS Genet. 2019;15:e1008114.

14. Bialek-Davenet S, Criscuolo A, Ailloud F, Passet V, Jones L, Delannoy-Vieillard AS, et al. Genomic definition of hypervirulent and multidrug-resistant *Klebsiella pneumoniae* clonal groups. Emerg Infect Dis. 2014;20:1812–20.

15. 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;62:978–85.

16. Szilágyi E, Füzi M, Damjanova I, Böröcz K, Szonyi K, Tóth A, et al. Investigation of extended-spectrum beta-lactamase-producing *Klebsiella pneumoniae* outbreaks in Hungary between 2005 and 2008. Acta Microbiol Immunol Hung. 2010;57:43–53.

17. Munoz-Price LS, Poirel L, Bonomo RA, Schwaber MJ, Daikos GL, Cormican M, et al. Clinical epidemiology of the global expansion of *Klebsiella pneumoniae* carbapenemases. Lancet Infect Dis. 2013;13:785–96.

18. Zurfluh K, Nüesch-Inderbinen M, Morach M, Zihler Berner A, Hächler H, Stephan R. Extended-spectrum-β-lactamase-producing Enterobacteriaceae isolated from vegetables imported from the Dominican Republic, India, Thailand, and Vietnam. Björkroth J, editor. Appl Environ Microbiol. 2015;81:3115–20.

19. Ovejero CM, Escudero JA, Thomas-Lopez D, Hoefer A, Moyano G, Montero N, et al. Highly Tigecycline-resistant *Klebsiella Pneumoniae* sequence type 11 (ST11) and ST147 isolates from companion animals. Antimicrob Agents Chemother. 2017;61.

20. Sato T, Harada K, Usui M, Tsuyuki Y, Shiraishi T, Tamura Y, et al. Tigecycline Susceptibility of *Klebsiella pneumoniae* Complex and *Escherichia coli* Isolates from Companion Animals: The Prevalence of Tigecycline-Nonsusceptible *K. pneumoniae* Complex, Including Internationally Expanding Human Pathogenic Lineages. Microb Drug Resist. 2018;24:860–7.

21. Zogg AL, Simmen S, Zurfluh K, Stephan R, Schmitt SN, Nüesch-Inderbinen M. High prevalence of extended-spectrum β-Lactamase producing enterobacteriaceae among clinical isolates from cats and dogs admitted to a veterinary hospital in Switzerland. Front Vet Sci. 2018;5.

22. Zhang R, Li J, Wang Y, Shen J, Shen Z, Wang S. Presence of NDM in non-*E. coli*

667 Enterobacteriaceae in the poultry production environment. J Antimicrob Chemother.
668 2019;74:2209–13.

669 23. Suzuki Y, Nazareno PJ, Nakano R, Mondoy M, Nakano A, Bugayong MP, et al.
670 Environmental presence and genetic characteristics of carbapenemase-producing
671 Enterobacteriaceae from hospital sewage and river water in the philippines. Appl Environ
672 Microbiol. 2020;86.

673 24. Baron SA, Mediannikov O, Abdallah R, Kuete Yimagou E, Medkour H, Dubourg G,
674 et al. Multidrug-resistant *Klebsiella pneumoniae* clones from wild chimpanzees and termites
675 in Senegal. Antimicrob Agents Chemother. 2021;

676 25. CLSI. Performance Standards for Antimicrobial Susceptibility Testing. 26th ed.
677 Wayne P, editor. Clinical and Laboratory Standards Institute; 2016. CLSI Supplement M100-
678 S26.

679 26. Prjibelski A, Antipov D, Meleshko D, Lapidus A, Korobeynikov A. Using SPAdes
680 De Novo Assembler. Curr Protoc Bioinforma. 2020;70:e102.

681 27. Wick RR, Judd LM, Gorrie CL, Holt KE. Unicycler: Resolving bacterial genome
682 assemblies from short and long sequencing reads. PLoS Comput Biol. 2017;13.

683 28. Seemann T. Prokka: rapid prokaryotic genome annotation. Bioinformatics.
684 2014;30:2068–9.

685 29. Page AJ, Cummins CA, Hunt M, Wong VK, Reuter S, Holden MTG, et al. Roary:
686 Rapid large-scale prokaryote pan genome analysis. Bioinformatics. 2015;31:3691–3.

687 30. Croucher NJ, Page AJ, Connor TR, Delaney AJ, Keane JA, Bentley SD, et al. Rapid
688 phylogenetic analysis of large samples of recombinant bacterial whole genome sequences

689 using Gubbins. Nucleic Acids Res. 2015;43:e15.

690 31. Bouckaert R, Vaughan TG, Barido-Sottani J, Duchêne S, Fourment M, Gavryushkina
691 A, et al. BEAST 2.5: An advanced software platform for Bayesian evolutionary analysis.
692 PLoS Comput Biol. 2019;15.

693 32. Diancourt L, Passet V, Verhoef J, Grimont PAD, Brisse S. Multilocus Sequence
694 Typing of *Klebsiella pneumoniae* Nosocomial Isolates. J Clin Microbiol. 2005;43:4178–82.

695 33. Lam MMC, Wick RR, Watts SC, Cerdeira LT, Wyres KL, Holt KE. A genomic
696 surveillance framework and genotyping tool for *Klebsiella pneumoniae* and its related species
697 complex. Nat Commun. 2021; 12:4188.

698 34. Jolley KA, Bray JE, Maiden MCJ. Open-access bacterial population genomics:
699 BIGSdb software, the PubMLST.org website and their applications. Wellcome Open Res.
700 2018;3.

701 35. Carattoli A, Zankari E, García-Fernández A, Voldby Larsen M, Lund O, Villa L, et al.
702 In silico detection and typing of plasmids using PlasmidFinder and plasmid multilocus
703 sequence typing. Antimicrob Agents Chemother. 2014;58:3895–903.

704 36. Arndt D, Grant JR, Marcu A, Sajed T, Pon A, Liang Y, et al. PHASTER: a better,
705 faster version of the PHAST phage search tool. Nucleic Acids Res. 2016;44:W16–21.

706 37. Couvin D, Bernheim A, Toffano-Nioche C, Touchon M, Michalik J, Néron B, et al.
707 CRISPRCasFinder, an update of CRISRFinder, includes a portable version, enhanced
708 performance and integrates search for Cas proteins. Nucleic Acids Res. 2018;46:W246–51.

709 38. Pfeifer E, Moura De Sousa JA, Touchon M, Rocha EPC. Bacteria have numerous
710 distinctive groups of phage-plasmids with conserved phage and variable plasmid gene

711 repertoires. Nucleic Acids Res. 2021;49:2655–73.

712 39. Villa L, Feudi C, Fortini D, Garciá-Fernández A, Carattoli A. Genomics of KPC-
713 producing *Klebsiella pneumoniae* sequence type 512 clone highlights the role of RamR and
714 ribosomal S10 protein mutations in conferring tigecycline resistance. Antimicrob Agents
715 Chemother. 2014;58:1707–12.

716 40. Fajardo-Lubián A, Ben Zakour NL, Agyekum A, Qi Q, Iredell JR. Host adaptation
717 and convergent evolution increases antibiotic resistance without loss of virulence in a major
718 human pathogen. PLoS Pathog. 2019;15.

719 41. Rodrigues C, Machado E, Ramos H, Peixe L, Novais Â. Expansion of ESBL-
720 producing *Klebsiella pneumoniae* in hospitalized patients: a successful story of international
721 clones (ST15, ST147, ST336) and epidemic plasmids (IncR, IncFIIK). Int J Med Microbiol.
722 2014;304:1100–8.

723 42. Nahid F, Zahra R, Sandegren L. A blaOXA-181-harbouring multi-resistant ST147
724 *Klebsiella pneumoniae* isolate from Pakistan that represent an intermediate stage towards
725 pan-drug resistance. PLoS One. 2017;12.

726 43. Alfaresi M. Whole Genome Sequencing of *Klebsiella pneumoniae* Strain Unravels a
727 New Model for the Development of Extensive Drug Resistance in Enterobacteriaceae. Open
728 Microbiol J. 2018;12:195–9.

729 44. Turton JF, Payne Z, Coward A, Hopkins KL, Turton JA, Doumith M, et al. Virulence
730 genes in isolates of *Klebsiella pneumoniae* from the UK during 2016, including among
731 carbapenemase gene-positive hypervirulent K1-st23 and 'non-hypervirulent' types ST147,
732 ST15 and ST383. J Med Microbiol. 2018;67:118–28.

733 45. Turton J, Davies F, Turton J, Perry C, Payne Z, Pike R. Hybrid resistance and
734 virulence plasmids in “high-risk” clones of *Klebsiella pneumoniae*, including those carrying
735 blaNDM-5. Microorganisms. 2019;7.

736 46. Coelho A, Piedra-Carrasco N, Bartolomé R, Quintero-Zarate JN, Larrosa N, Cornejo-
737 Sánchez T, et al. Role of IncHI2 plasmids harbouring blaVIM-1, blaCTX-M-9, aac(6′)-Ib and
738 qnrA genes in the spread of multiresistant *Enterobacter cloacae* and *Klebsiella pneumoniae*
739 strains in different units at Hospital Vall d’Hebron, Barcelona, Spain. Int J Antimicrob
740 Agents. 2012;39:514–7.

741 47. Bowers JR, Kitchel B, Driebe EM, MacCannell DR, Roe C, Lemmer D, et al.
742 Genomic Analysis of the Emergence and Rapid Global Dissemination of the Clonal Group
743 258 *Klebsiella pneumoniae* Pandemic. PLoS One. 2015;10:e0133727.

744 48. van der Bij AK, Peirano G, Goessens WHF, van der Vorm ER, van Westreenen M,
745 Pitout JDD. Clinical and Molecular Characteristics of Extended-Spectrum- beta-Lactamase-
746 Producing *Escherichia coli* Causing Bacteremia in the Rotterdam Area, Netherlands.
747 Antimicrob Agents Chemother. 2011;55:3576–8.

748 49. Newire E, Aydin A, Juma S, Enne VI, Roberts AP. Identification of a Type IV-A
749 CRISPR-Cas System Located Exclusively on IncHI1B/IncFIB Plasmids in
750 Enterobacteriaceae. Front Microbiol. 2020;11.

751 50. Pinilla-Redondo R, Mayo-Muñoz D, Russel J, Garrett RA, Randau L, Sørensen SJ, et
752 al. Type IV CRISPR-Cas systems are highly diverse and involved in competition between
753 plasmids. Nucleic Acids Res. 2020;48:2000–12.

754 51. Wyres KL, Hawkey J, Hetland MAK, Fostervold A, Wick RR, Judd LM, et al.
755 Emergence and rapid global dissemination of CTX-M-15-associated *Klebsiella pneumoniae*

strain ST307. J Antimicrob Chemother. 2019;74:577–81.

52. Sonnevend Á, Ghazawi A, Hashmey R, Haidermota A, Girgis S, Alfaresi M, et al. Multihospital occurrence of pan-resistant *Klebsiella pneumoniae* sequence type 147 with an ISEcp1-directed blaOXA-181 insertion in the mgrB gene in the United Arab Emirates. Antimicrob Agents Chemother. 2017;61.

53. 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. Eurosurveillance. 2020;25.

54. Simner PJ, Antar AAR, Hao S, Gurtowski J, Tamma PD, Rock C, et al. Antibiotic pressure on the acquisition and loss of antibiotic resistance genes in *Klebsiella pneumoniae*. J Antimicrob Chemother. 2018;73:1796–803.

55. Pitout JDD, Peirano G, Kock MM, Strydom KA, Matsumura Y. The global ascendency of OXA-48-type carbapenemases. Clin Microbiol Rev. 2019;33:e00102-19.

56. Yoon E-J, Gwon B, Liu C, Kim D, Won D, Park SG, et al. Beneficial Chromosomal Integration of the Genes for CTX-M Extended-Spectrum β-Lactamase in *Klebsiella pneumoniae* for Stable Propagation . mSystems. 2020;5.

57. Rojas LJ, Hujer AM, Rudin SD, Wright MS, Domitrovic TN, Marshall SH, et al. NDM-5 and OXA-181 beta-lactamases, a significant threat continues to spread in the Americas. Antimicrob Agents Chemother. 2017;61.

58. Sakamoto N, Akeda Y, Sugawara Y, Takeuchi D, Motooka D, Yamamoto N, et al. Genomic characterization of carbapenemase-producing *Klebsiella pneumoniae* with chromosomally carried blaNDM-1. Antimicrob Agents Chemother. 2018;62.

778 59. Takeuchi D, Akeda Y, Yoshida H, Hagiya H, Yamamoto N, Nishi I, et al. Genomic
779 reorganization by IS26 in a blaNDM-5-bearing FII plasmid of *Klebsiella pneumoniae*
780 isolated from a patient in Japan. J Med Microbiol. 2018;67:1221–4.

781 60. Roe CC, Vazquez AJ, Esposito EP, Zarrilli R, Sahl JW. Diversity, virulence, and
782 antimicrobial resistance in isolates from the newly emerging *Klebsiella pneumoniae* ST101
783 Lineage. Front Microbiol. 2019;10.

784 61. Stoesser N, Sheppard AE, Pankhurst L, De Maio N, Moore CE, Sebra R, et al.
785 Evolutionary History of the Global Emergence of the *Escherichia coli* Epidemic Clone
786 ST131. MBio. 2016;7:e02162-15.

787 62. Holden MTG, Hsu LY, Kurt K, Weinert LA, Mather AE, Harris SR, et al. A genomic
788 portrait of the emergence, evolution, and global spread of a methicillin-resistant
789 *Staphylococcus aureus* pandemic. Genome Res. 2013;23:653–64.

790 63. Fuzi M, Rodriguez Baño J, Toth A. Global Evolution of Pathogenic Bacteria With
791 Extensive Use of Fluoroquinolone Agents. Front Microbiol. 2020;11.

792 64. Ostria-Hernández ML, Sánchez-Vallejo CJ, Ibarra JA, Castro-Escarpulli G. Survey of
793 clustered regularly interspaced short palindromic repeats and their associated Cas proteins
794 (CRISPR/Cas) systems in multiple sequenced strains of *Klebsiella pneumoniae.* BMC Res
795 Notes. 2015;8.

796 65. Shen J, Lv L, Wang X, Xiu Z, Chen G. Comparative analysis of CRISPR-Cas systems
797 in *Klebsiella* genomes. J Basic Microbiol. 2017;57:325–36.

798 66. Li HY, Kao CY, Lin WH, Zheng PX, Yan JJ, Wang MC, et al. Characterization of
799 CRISPR-Cas systems in clinical *Klebsiella pneumoniae* isolates uncovers its potential

800 association with antibiotic susceptibility. Front Microbiol. 2018;9.

801 67. Kamruzzaman M, Iredell JR. CRISPR-Cas System in Antibiotic Resistance Plasmids
802 in Klebsiella pneumoniae. Front Microbiol. 2020;10.

803 68. Tang Y, Fu P, Zhou Y, Xie Y, Jin J, Wang B, et al. Absence of the type I-E CRISPR-
804 Cas system in *Klebsiella pneumoniae* clonal complex 258 is associated with dissemination of
805 IncF epidemic resistance plasmids in this clonal complex. J Antimicrob Chemother.
806 2020;75:890–5.

807 69. Zhou Y, Tang Y, Fu P, Tian D, Yu L, Huang Y, et al. The type I-E CRISPR-Cas
808 system influences the acquisition of blaKPC-IncF plasmid in *Klebsiella pneumoniae*. Emerg
809 Microbes Infect. 2020;9:1011–22.

810 70. Mackow NA, Shen J, Adnan M, Khan AS, Fries BC, Diago-Navarro E. CRISPR-Cas
811 influences the acquisition of antibiotic resistance in *Klebsiella pneumoniae.* PLoS One.
812 2019;14.

813 71. López-Igual R, Bernal-Bayard J, Rodríguez-Patón A, Ghigo JM, Mazel D.
814 Engineered toxin–intein antimicrobials can selectively target and kill antibiotic-resistant
815 bacteria in mixed populations. Nat Biotechnol. 2019;37:755–60.

816 72. Bikard D, Euler CW, Jiang W, Nussenzweig PM, Goldberg GW, Duportet X, et al.
817 Exploiting CRISPR-cas nucleases to produce sequence-specific antimicrobials. Nat
818 Biotechnol. 2014;32:1146–50.

819