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[1†], Astrid V. Cienfuegos-Galletd[1,2†], Tengfei Long[1,3,4†], Gisele Peirano[5†], Tingyu Chu[1], Johann D.D. Pitout[5,6,7], Barry N Kreiswirth[1,8*], Liang Chen[1,8*]

[1]Center for Discovery and Innovation, Hackensack Meridian Health, Nutley, New Jersey, USA

[2]School of Microbiology, University of Antioquia, Medellín, Colombia

[3]Guangdong Laboratory for Lingnan Modern Agriculture, National Risk Assessment Laboratory for Antimicrobial Resistance of Animal Original Bacteria, College of Veterinary Medicine, South China Agricultural University, Guangzhou, China;

[4]Guangdong Provincial Key Laboratory of Veterinary Pharmaceutics, Development and Safety Evaluation, South China Agricultural University, Guangzhou, China;

[5]Cummings School of Medicine, University of Calgary, Calgary, Alberta, Canada

[6]Alberta Precision Laboratories, Calgary, Alberta, Canada

[7]University of Pretoria, Pretoria, Gauteng, South Africa

[8]Department of Medical Sciences, Hackensack Meridian School of Medicine, Nutley, New Jersey, USA

[†] Equal contribution

* Corresponding author: Dr. Barry Kreiswirth (Barry.Kreiswirth@hmh-cdi.org) or Liang Chen, (liangch@buffalo.edu).

Supplementary results

AMR genes in CG147

CG147 *K. pneumoniae* was characterized by carrying AMR genes confer resistance to multiple antibiotic families, including β-lactams, quinolones, aminoglycosides, rifampicin, tetracyclines, sulfonamides and trimethoprim. Other genes confer resistance to tigecycline, colistin and fosfomycin were uncommon in this clonal group. Importantly, all but two CG147 isolates had mutations related to quinolone resistance in the QRDR. GyrA-83 mutations were detected in 99.8% (n=1000) of isolates with the predominance of GyrA-83I (n=940, 92.8%) and GyrA-83Y (n=69, 6.8%). Other mutations in ParC-80I (n=1008, 99.6%) were also frequent in CG147. The double mutation GyrA-83I and ParC-80I (n=884, 87.4%), is conserved in most of the isolates, followed by the triple mutation GyrA-83Y, GyrA-87A and ParC-80I (n=68, 6.7%) (**Additional file 1: Table S1**).

Likewise, β-lactam resistance mediated by ESBL and carbapenemases acquisition was also common in CG147. We detected 67.6% (n=684) of isolates carrying at least one ESBL from the CTX-M group, being CTX-M-15 (n=647, 63.9%) the most frequent. CTX-M-15 was detected in several STs of CG147 (ST147, ST273, ST392, and ST3740). Other CTX-M groups were detected in lower proportion: CTX-M-9 (n=11, 1.1%), CTX-M-14 (n=11, 1.1%), CTX-M-55 (n=10, 1.0%) and CTX-M-3 (n=8, 1.0%). Notably CG147 carried at least one carbapenemase in 80.4% (n=814) of isolates. NDM was detected in 52.2% (n=529) of isolates. We found the predominance of NDM-1 (n=479, 47.3%), followed by NDM-5 (n=29, 2.9%) and NDM-7 (n=10, 1.0%). OXA48-like carbapenemases were identified in 17.9% (n=182) of isolates. Within this group OXA48 (n=103, 10.2%) and OXA181 (n=53, 5.2%) were the main variants. The frequency of both KPC-2 (n=58, 5.7%) and KPC-3 (n=47, 4.6%) were similar among KPC producers (n=105, 10.4%), and among VIM producers (n=46, 4.5%) we found mainly VIM-1 (n=30, 2.9%). In each continent a different β-lactamase profile was found. For instance, in Asia was frequent the presence of CTX-M-15 (42.5%) and CTX-M-15/NDM-1 (21.9%). CTX-M-15/NDM-1 was also frequent in Europe (46.1%), and

NDM-1 was predominant among isolates collected from the America region (51.8%) (**Table 1** and **Additional file 1: Table S1**).

We also detected porins mutations related to β-lactam resistance, including OmpK35 (n=531, 52.5%) and OmpK36 (n=221, 21.8%) mutations. The well-known mutations ompK36GD and ompK36TD which altered the size of the porin channel were detected in 9.6% (n=97) and 4.7% (n=48) of isolates, respectively (**Additional file 1: Table S1**). These mutations were found in ST147, ST2292, ST273, ST3740 isolates.

Ninety-five percent (n=970) of isolates harbored at least one aminoglycoside modifying enzyme with the most frequent profile *aac(3)-IId - aadA2 - aph(3')-VI, - armA* (n=112, 11.1%). In addition, 59.6% (n=603) harbored genes encoding quinolone protection proteins, such as qnrB1 (n=171, 16.9%) and qnrS1 (n=275, 27.2%). CG147 isolates also encoded resistance genes to tetracyclines (n=329, 32.5%), such as *tet(A)* (n=313, 30.9%). Importantly, only four isolates carried resistance genes to tigecycline *(tmexC1, tmexD1* and *toprJ1*). Resistance to fosfomycin was very low, with only 11 isolates carrying *fosA3* (1.1%), six of ST147 and five of ST273. Regarding colistin resistance, only 4.8% of isolates carried truncations in *mgrB* gene. We also detected plasmid genes conferring resistance to this antibiotic, including *mcr1.1* (n=7), *mcr8.1* (n=1) and *mcr9.1* (n=11). Of note, *mcr1.1* was detected in both human (n=5) and animal (n=2) isolates (**Additional file 1: Table S1**).

In addition, genes related to rifampicin resistance were present in a third of isolates (n=376, 37.2%), including the *arr-3* (n=257, 25.4%) and *arr-2* (n=119, 11.8%). Resistance to sulfonamides was mediated by *sul1, sul2* and *sul3* (n=905, 89.4%) and resistance to trimethoprim was mediated by *dfrA* and *dfrB* (n=843, 83.3%) (**Additional file 1: Table S1**).

Plasmids harbored by CG147 isolates

A total of 22 plasmid replicons types were detected in CG147 genomes. IncHI1B(pNDM-MAR) (45.2%, n=458), IncR1 (42.9%, n=435), IncFIB (pQIL) (32.4%, n=329) and IncFIA (24.1%, n=244) were the most common types. Plasmid pNDM-MAR (JN420336) was detected for the first time in *K. pneumoniae* ST15 carrying $bla_{\text{NDM-1}}$, $bla_{\text{CTX-M-15}}$, the quinolone resistance gene *qnrB1* and additional genes harboring resistance to tellurite, mercury, chloramphenicol and aminoglycosides. IncHI1B (pNDM-MAR) positive isolates also carried NDM (n=340, 74.2%) and CTX-M-15 (n=247, 53.9%). Both enzymes were also present in strains carrying IncR replicons (40.1% and 52.2%, respectively). These plasmids have been reported to carry multiple resistance genes in clinical isolates, and also have been described as part of multireplicon plasmids associated to IncA/C, IncF and IncFIIK and nontypeable backbones. KPC positive strains carried a number of plasmids from different replicons types, including IncFIA (HI1) AF250878 (28.6%, n=30), IncFIB (K) JN233704 (50.5%, n=53), IncFIB CP03223 (37.1%, n=39), IncFIB (pQIL) JN233705 (38.1%, n=40), IncHI1B (pNDM-MAR) JN420336 (24.7%, n=26), IncN AY046276 (25.7%, n=27), IncR1 DQ449578 (26.7%, n=28) (**Additional file 1: Table S1**).

Genes encoding virulence factors in CG147 isolates

We also investigated the presence of genes encoding virulence factors in CG147 isolates, such as the polysaccharide capsular type, the siderophore systems yersiniabactin, aerobactin, colibactin and salmochelin, and the hypermucoid loci (*rmpADC* and *rmpA2*). We detected 23 different capsular polysaccharide types, among which KL64 accounted for 75.8% (n=767) of isolates. This capsular type was frequent in humans (n=797, 94.8%), animals (n=6, 0.8%) and the environment (n=11, 1.4%). Additional capsular types detected were KL27 (n=88, 8.7%) and KL10 (n=68, 6.7%). Importantly, ST147 isolates carried KL64 (n=759, 86.2%), while ST273 harbored predominantly KL74 (n=25, 69.4%) and all isolates ST392 had KL27 (n=85, 100%). Yersiniabactin siderophore system was identified in 45.2% (n=458) isolates, including lineage 9 in ICEKp3 (n=209, 20.7%), lineage 16 in ICEKp12 (n=123, 12.2%) and lineage 10 in ICEKp4 (n=102, 10.1%). Additional siderophore systems found were aerobactin (n=164, 16.2%), mainly lineage *iuc* 1 (n=159, 15.7%). We only detected two isolates carrying colibactin and one carrying

salmochelin, all of ST147. Only 13.5% (n=137) of isolates from CG147 encoded the hypermucoid *rmpADC* loci and the *rmpA2* variant was truncated in the small number of isolates where it was detected (12.9%). All isolates encoding the hypermucoid locus were from human origin and of ST147 (**Additional file 1: Table S1**).

Carbapenemase

We examined the ability of ST clones within CG147 to host diverse type of carbapenemases. We found that ST147 the most frequent clone among CG147 commonly carried carbapenemases (85.0%, 749/881), while other STs such as ST273 and ST392 carried carbapenemases in 51.4%, (19/37) and 43.5% (37/85) of isolates, respectively. ST147 carbapenemase producers encoded frequently NDM-1 (n=463, 52.5%), but also KPC-2 (n=42, 4.8%), KPC-3 (n=41, 4.7%), OXA-48 (n=90, 10.2%), OXA-181 (n=52, 5.9%), and VIM-1 (n=27, 3.1%). ST147 isolates also co-harbored different carbapenemase types in low frequency, including NDM-5/OXA-181 (n=12, 1.4%), NDM-1/OXA-48 (n=7, 0.8%), and KPC-2/VIM-1 (n=6, 0.7%). Among 19 ST273 isolates carrying carbapenemases, predominated metallo-betalactamases NDM (n=10, 27.8%). We found NDM-7 (n=5, 13.9%), NDM-4 (n=4, 11.1%) and KPC-2 (n=6, 16.6%) as the most frequent genes. In addition, VIM-1 was detected in one isolate (2.8%) and the combination IMP-4/NDM-1 was also found in another isolate (2.8%). Similarly, ST392 encoded several carbapenemases types including NDM-1 (n=11, 12.9%), OXA48 (n=10, 11.8%), KPC-2 (n=8, 9.4%) and KPC-3 (n=6, 7.1%) (**Table 1, Additional file 1: Table S1**).

The I-E and IV-A3 CRISPR spacers in CG147 strains

A total of 41,240 I-E CRISPR spacers were identified among the 1,012 GC147 genomes analyzed and 96 different spacer types were identified. Dissection of the I-E CRISPR arrays detected a total of 148 unique spacer arrays with the number of spacer types varying between 6 and 86 (median, 43). A predominant array, termed the 56677-like array (the CRISPR array detected in strain 56677) was found in the majority of the collection (n=562, 55.7%). Visualization of the distribution of 96 I-E spacer types among the 1,012 GC147

strains exhibited a Z-shape profile (**Additional file 2: Fig. S2A**) where 41 of the 96 unique are present in >90% of the I-E CRISPR-Cas system harboring strains, indicating the I-E CRISPR-Cas system is conserved and an important element in the evolution of CG147.

A total of 6,062 IV-A3 spacers were identified. Compared with I-E spacers, IV-A3 spacers displayed greater diversity, with a total of 347 unique spacer types being identified. A total of 69 unique IV-A3 CRISPR arrays were identified, with the number of spacer types varying between 6 and 37 (median, 13). The SAMN18977090-like array (the CRISPR array detected in biosample SAMN18977090) was predominated, accounting for 35.2% (n=149) of isolates. The distribution of IV-A3 spacers exhibited an L-shape (**Additional file 2: Fig. S2B**) and only 11 of 347 present in >65% of the IV-A3 CRISPR-Cas harboring strains, 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. Notably, no spacers were shared between the two CRISPR-Cas systems in CG147.