BIOINFORMATICS APPROACHES FOR ANALYSING PLASMIDS……

MARÍA PÉREZ-VAZQUEZ

SPAIN (CNM-ISCIII)

ciberinfec

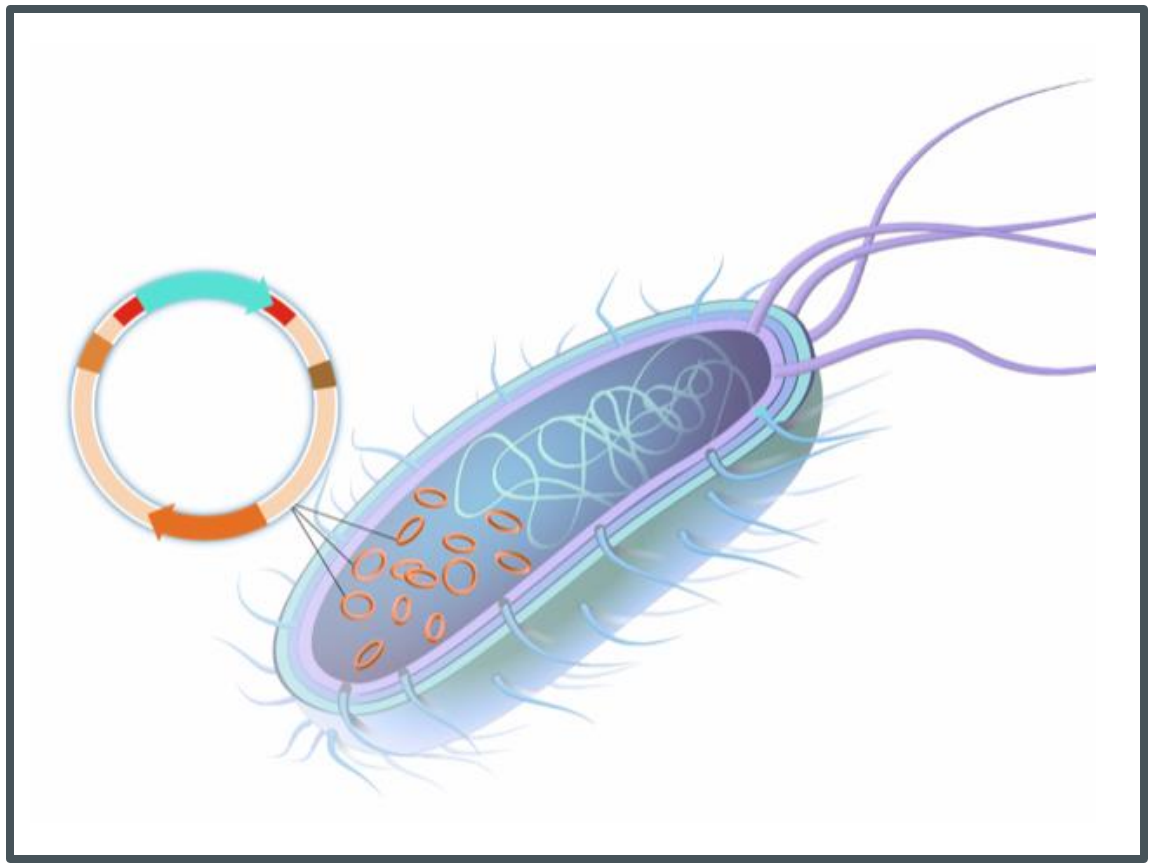
Centro de Investigación Biomédica en Red

Enfermedades infecciosas

MINISTERIO DE CIENCIA E INNOVACIÓN

Instituto de Salud Carlos III

PLASMIDS circular or linear extrachromosomal double-stranded DNA molecules capable of self-replication in a host and transfer between host cells (AMR).

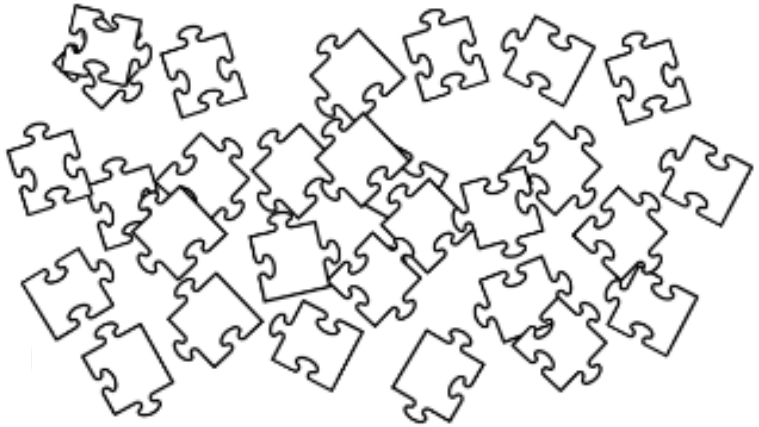


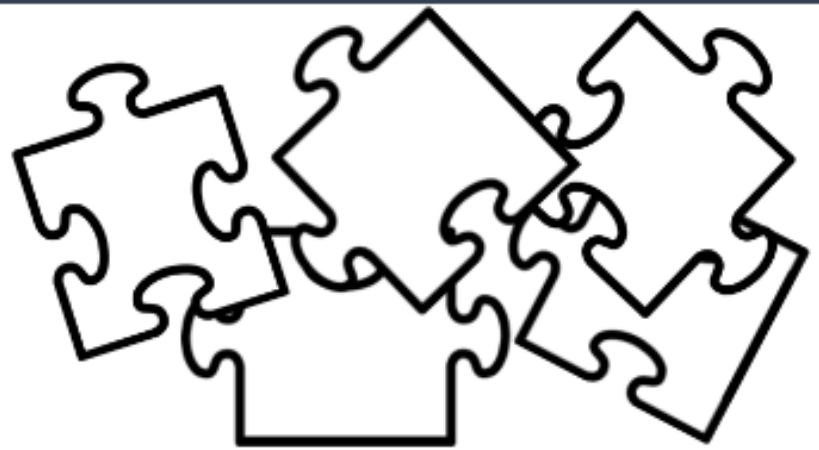
- They are highly variable both in length (from one to several hundred kilobases) and number of copies
- In Enterobacteriaceae, each strain usually has more than one plasmid.
- Plasmid usually have repetitive sequences that make difficult the assembly (small contigs. short reads).

TOOLS

Distinguish plasmid from chromosomal sequences	cBar (Zhou and Xu, 2010)	Plasmid and chromosomal sequences are distinguished based on pentamer frequencies.
	Other tools	Tools such as plasmidSPAdes and PlasmidFinder may also be used to distinguish plasmid and chromosomal sequences (Arredondo-Alonso et al., 2016).
Resolve plasmid structures from ambiguous assembly graphs	PLACNET (Lanza et al., 2014)	An input assembly graph is reconfigured according to the homology of contigs to reference sequences; the assembly graph can be visualized to allow manual pruning and correction.
	Recycler (Rozov et al., 2016)	Cycles in an assembly graph are identified and sequentially extracted from the graph, favoring cycles with minimal coverage variation across constituent contigs. Assuming different genetic units have distinct copy numbers, retrieved cycles should represent individual circular elements (plasmids, circular phages). Information from paired-end reads is used to exclude cycles that do not correspond to a single circular element, but arise from repeat elements shared across different molecules.
	plasmidSPAdes (Antipov et al., 2016)	Median coverage of longer contigs is calculated to estimate chromosomal coverage; this estimate is used as a basis for filtering putative chromosomal contigs from the assembly graph. Connected components within the filtered graph are reported as putative plasmids. This approach assumes that chromosomal contig coverage differs from plasmid contig coverage.

Short-read sequencing *vs.* Long-read sequencing





- Assembly is challenging ----- many small contigs
- Difficult to get the structure of whole plasmid

- Enable coverage of repetitive regions--- long contigs (whole plasmid)
- Take in consideration cost (vs short-read)

2018

PlasmidID: a mapping based tool for plasmid identification, annotation, typing and visualization.

Pedro J. Sola-Campoy[1,2], Sara Monzón[2], María Pérez-Vázquez[1], Adriana Ortega[1], Verónica Bautista[1], Sara Fernández-Romero[1], David Sáez[1], Noelia Lara[1], Belén Aracil[1], José Campos[1], Jesús Oteo[1] , Isabel Cuesta[2]

1 Reference and Research Laboratory for Antibiotic Resistance, CNM, ISCIII, Majadahonda, Madrid.
2 Bioinformatic Unit, ISCIII, Majadahonda, Madrid.

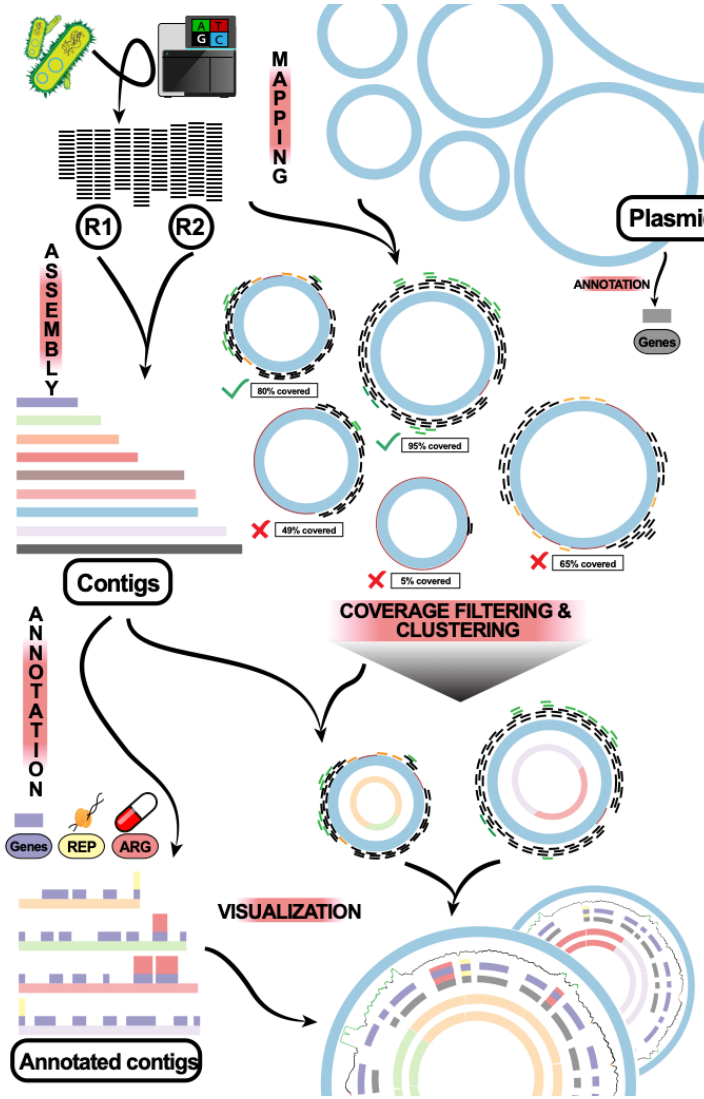
Background

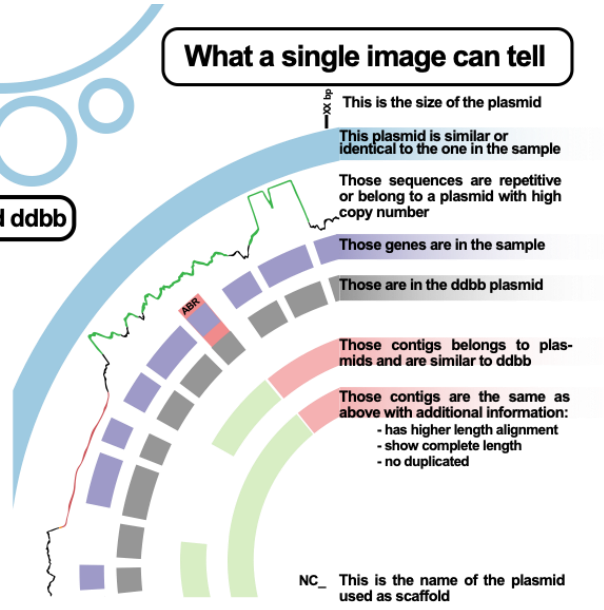
- Plasmids are key elements in ARG and virulence factors horizontal transfer
- Modular nature hinder recovery from WGS data
- Local alignment & *de novo* assembly are limited strategies

Results

PlasmidID is a bash script that maps Illumina reads over plasmid database sequences. The most covered sequences are clustered by identity to avoid redundancy and the longest are used as scaffold for plasmid reconstruction. Reads are assembled and annotated. All information generated from mapping, assembly, annotation and local alignment analyses is gathered and accurately represented in a circular image which allow user to determine plasmidic composition in any bacterial sample.

bioinformatica@isciii.es



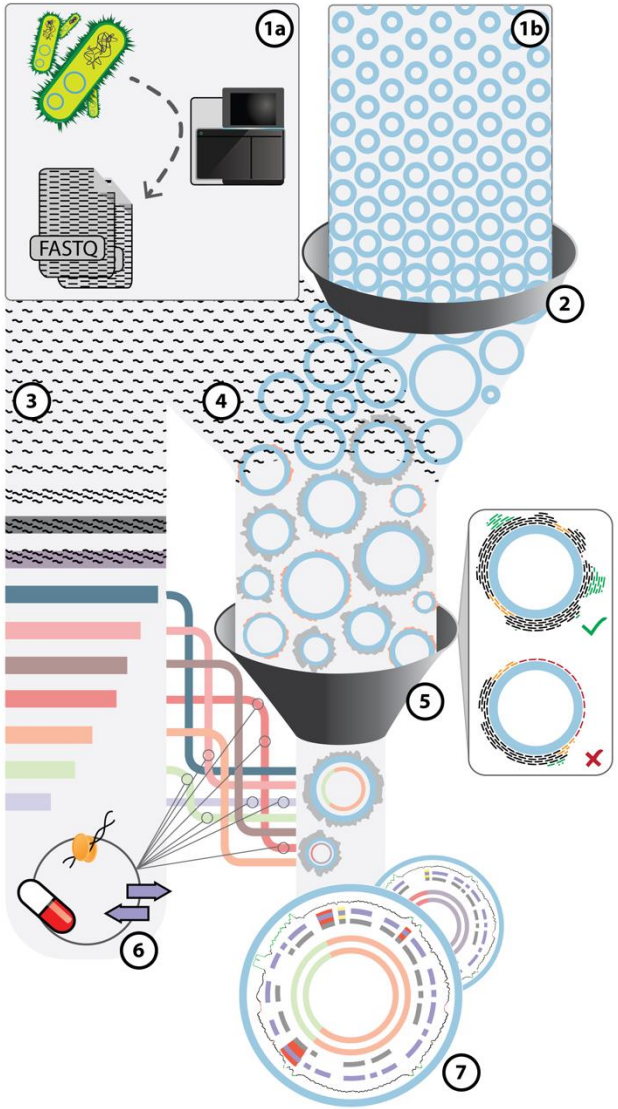


Results on high-risk *K. pneumoniae* strains

K. pneumoniae	N	Inc group	Size
ST11/OXA-48-like	3	IncL/M	~68000bp
ST147/VIM-1	2	IncN	~50000bp
ST512/KPC-3	4	IncFIB	~120000bp

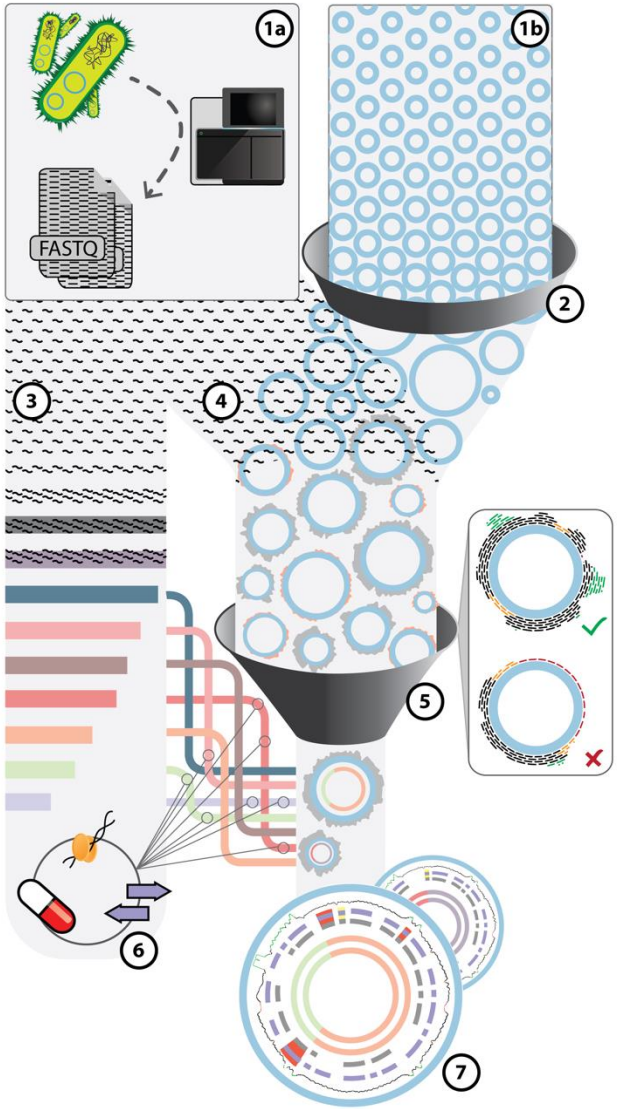
Conclusions

PlasmidID allow plasmids identification in bacterial whole genome sequences, generating images of the most likely plasmids present in a sample, easing interpretation by researchers. Since it is database dependent tool, it works for every plasmid length and species, and can be adapted to other aims such metaplasmid analysis or SMRT sequencing technologies.

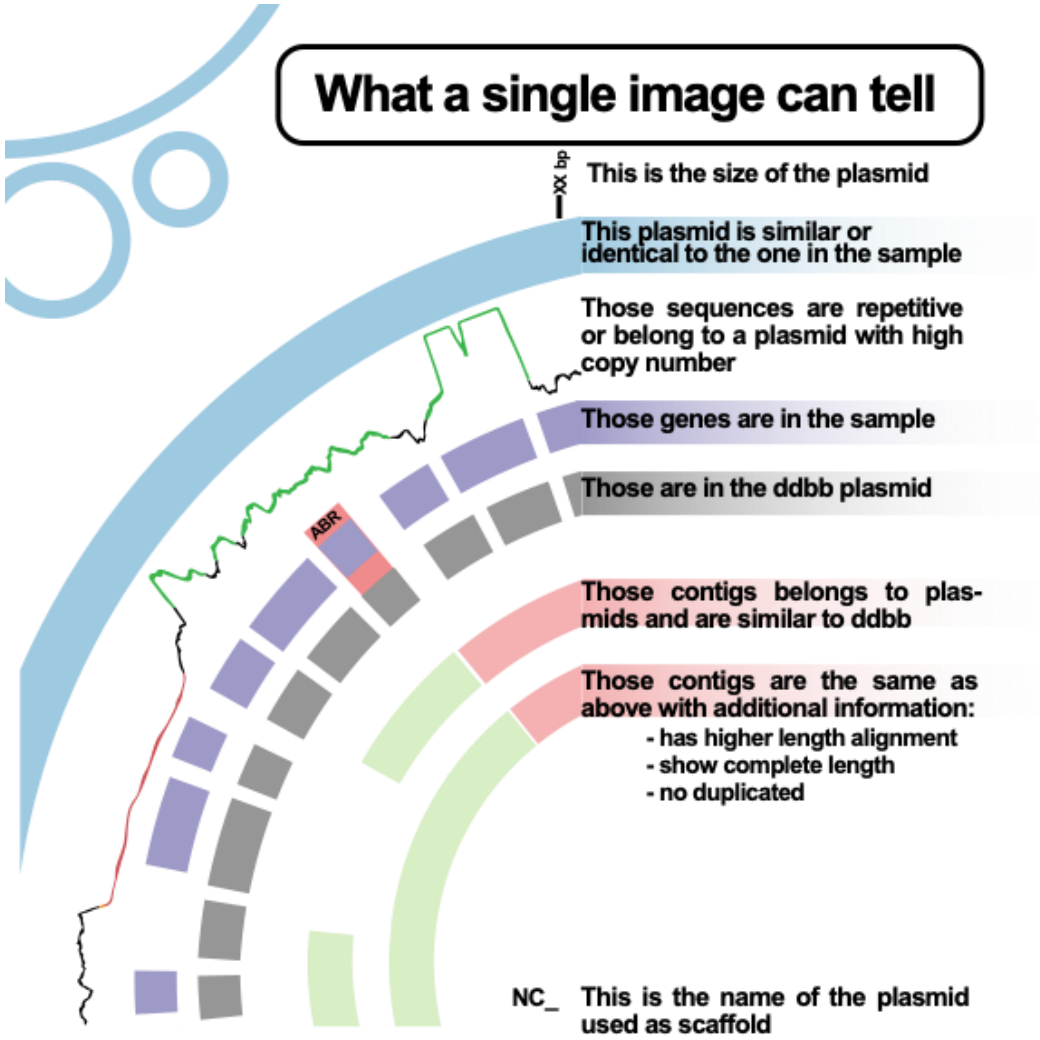


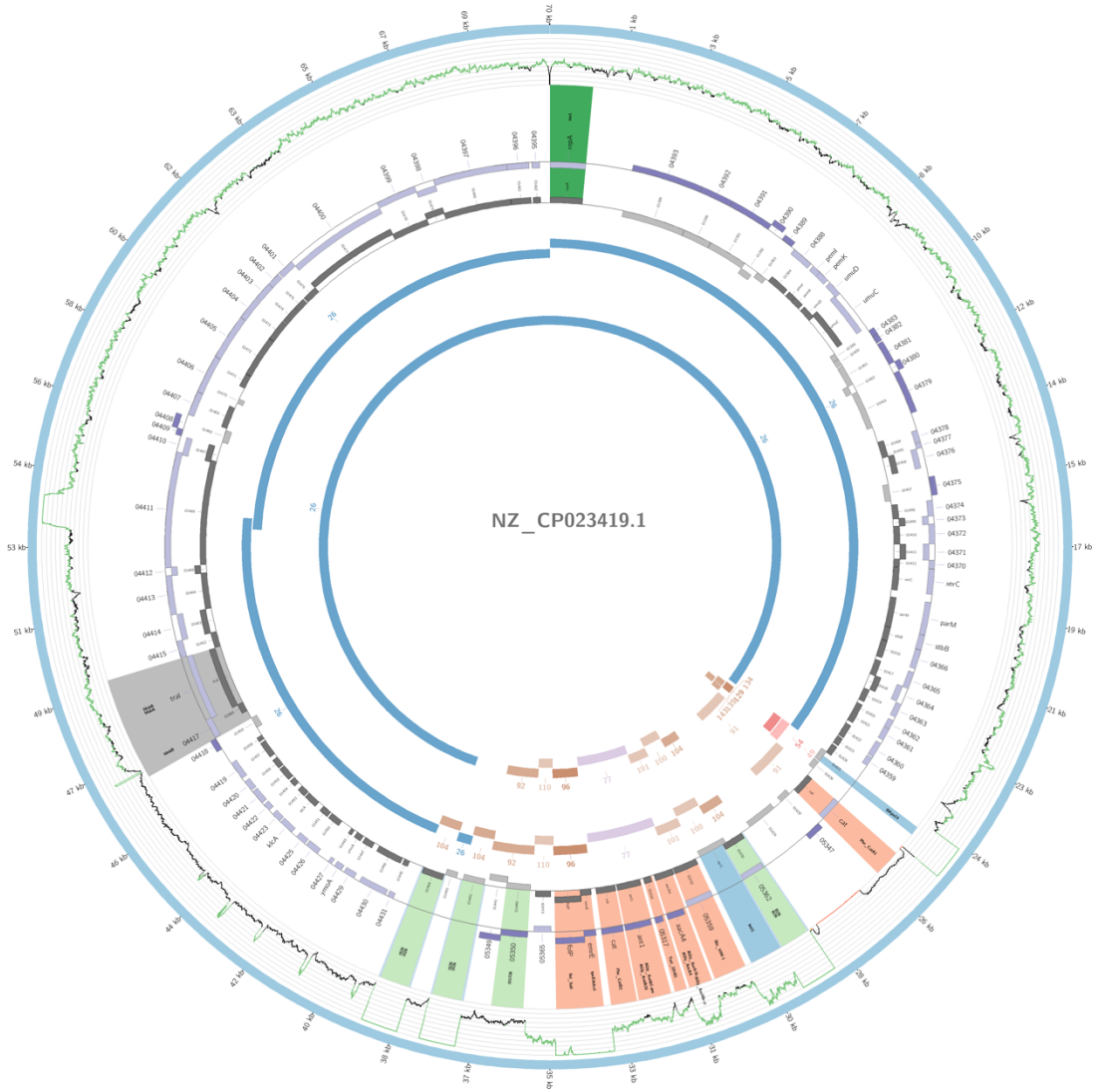
- Curated plasmid DB (NCBI)
  ftp://ftp.ncbi.nlm.nih.gov/genomes/GENOME_REPORTS/plasmids.txt
- Filtering DB (k-mers): k-mers from DB are searched in quality-filtered reads (mash), sequences with identity value higher than 0.9 are extracted and used as reference for mapping.
- Filtering DB (coverage): Plasmids with a percentage coverage < 90% are removed from successive steps.
- Clustering: Resulted plasmids are clustered (Cd-hit), longest plasmids in each cluster are selected as guide for reconstruction.

https://github.com/BU-ISCIII/plasmidID

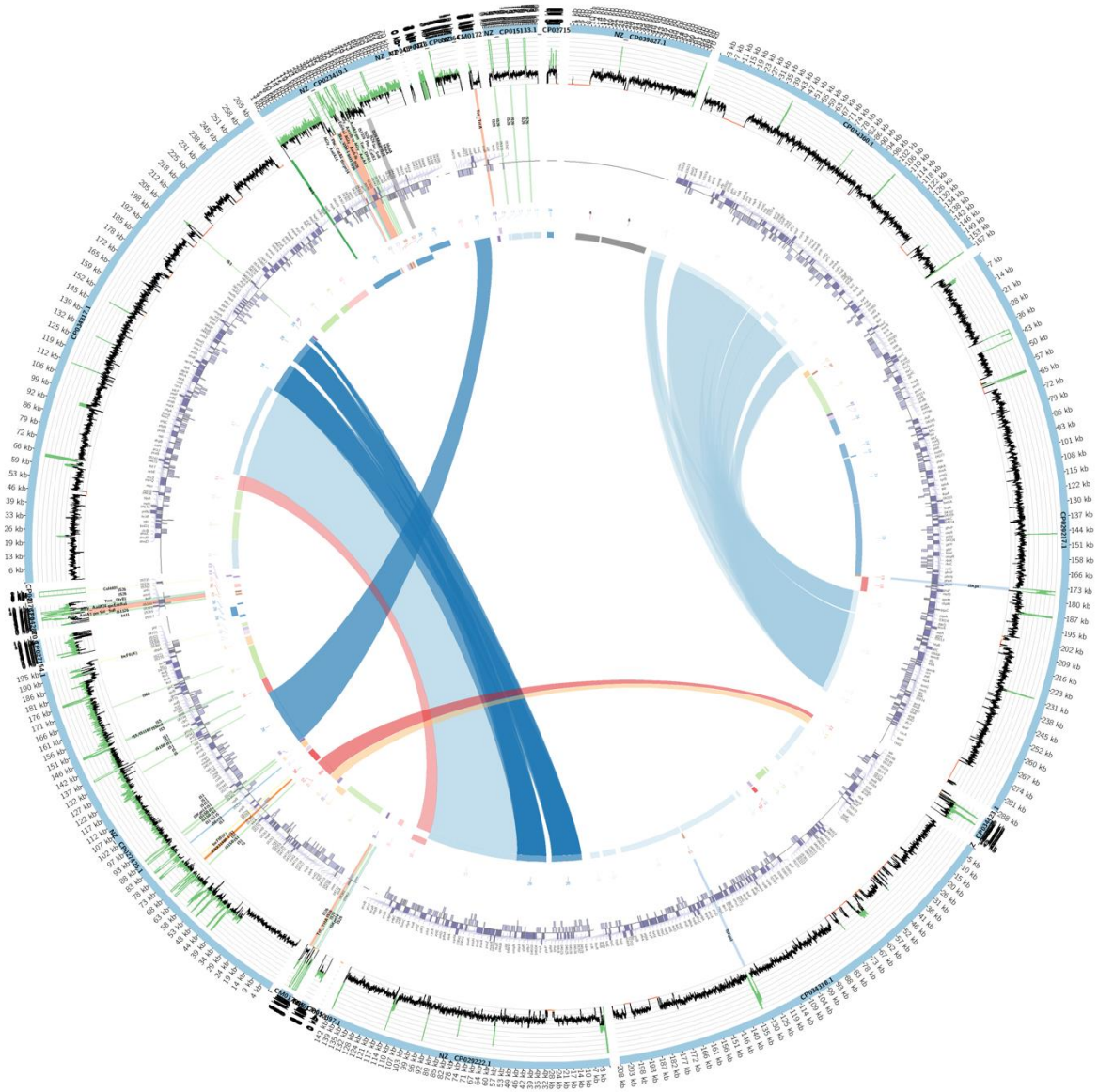


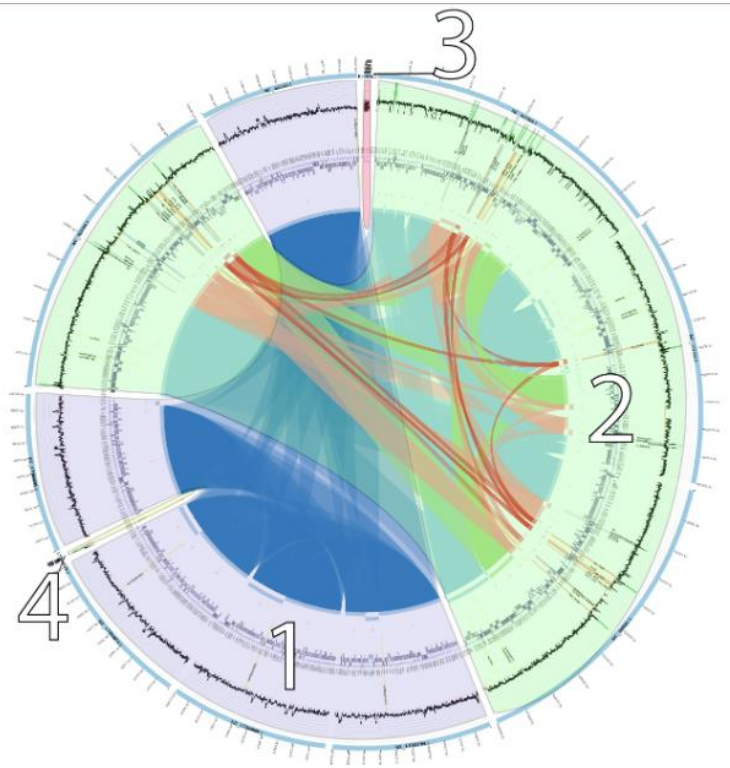
- Local alignment (BLAST+) of the contigs of problem sequence with the filtered DB. Extraction of alignments with identity higher than 90%.
- Annotation of both contigs of problem sequence and plasmids obtained from DB (prokka) .
- Additionally BLAST is used to identify target sequences, as Rep (incompatibility groups), resistance genes, virulence genes…
- Graphic representation (circus).





https://github.com/BU-ISCIII/plasmidID/wiki/Understanding-the-image%3A-track-by-track

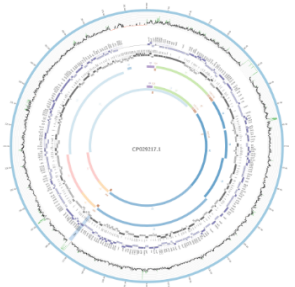
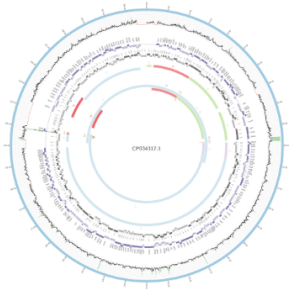




- In addition to an individual image per plasmid, PlasmidID provides a summary image that allows you to see all the detected plasmids in a single figure, very useful to detect possible redundancy in the results.

https://github.com/BU-ISCIII/plasmidID/wiki/How-to-chose-the-right-plasmids

PlasmidID REPORT

id	length	species	description	contig_name	A0209KPN
CP029217.1	288994	*Klebsiella pneumoniae*	pneumoniae strain L201 plasmid p1-L201	[15, 22, 28, 37, 42, 65, 85, 92, 109, 120, 122, 140, 164, 184, 186, 203, 204, 205, 209, 211]	 MAPPING % 96.53 ALIGN FR 1.12
CP034317.1	265109	*Klebsiella pneumoniae*	pneumoniae strain 6 plasmid pK006_1	[12, 13, 36, 52, 59, 73, 104, 152, 182, 183]	 MAPPING % 93.17 ALIGN FR 1.36

	A	B	C	D	E	F	G	H
1	AC_Number	Length	Species	Description	N	K10003	K10004	K10005
2	CP019080,1	70987	Klebsiella pneumoniae	strain DT12 plasmid pDT12	3	82,9772	83,2645	83,2969
3	K00826,1	4012	Escherichia coli	plasmid pCM959	3	94,1675	94,4168	93,7188
4	NC_005246,1	60145	Erwinia amylovora	LebB66 plasmid pEL60	3	84,1732	84,1949	84,4576
5	NC_019154,1	61881	Klebsiella pneumoniae	plasmid pOXA-48	3	97,8281	94,6704	98,586
6	NZ_CP015071,1	69471	Escherichia coli	strain Ecol_743 plasmid pEC743_OXA48	3	82,289	83,1556	82,4286
7	NZ_CP017280,1	95331	Enterobacter ludwigii	strain EN-119 plasmid pEN-119	3	84,0828	86,378	86,2364
8	NZ_CP018342,1	63588	Klebsiella pneumoniae	isolate Kp_Goe_154414 plasmid pKp_Goe_414-5	3	99,7893	97,7968	99,8097
9	NZ_CP018443.1	50611	Klebsiella pneumoniae	strain Kp_Goe_822917 plasmid pKp_Goe_917-2	3	99.7273	97.9214	99.7372

- Summary table: plasmids used as reference, the coverage percentage, the accession number and the plasmid ID (length, description)
- When more than one isolate are analysed, the table include the number of isolates that presented homology with each plasmid.

plasmidID &

- PlasmidID is used in the whole collection to identify strains that presented the most different plasmid profiles (combination)
- Long-reads-assembled plasmid sequences were introduced in the PlasmidID as database and used as a template for all Illumina-sequenced isolates

Characterization of Carbapenemase-Producing *Klebsiella oxytoca* in Spain, 2016–2017

María Pérez-Vazquez[a,b], Jesús Oteo-Iglesias[a,b], Pedro J. Sola-Campoy[a], Hugo Carrizo-Manzoni[a,c], Verónica Bautista[a,b], Noelia Lara[a,b], Belén Aracil[a,b], Almudena Alhambra[d], Luis Martínez-Martínez[b,e,f,g], José Campos[a,b], the Spanish Antibiotic Resistance Surveillance Program Collaborating Group

- Gain insight into the microbiological features and molecular epidemiology of CP *K. oxytoca*

- Emergence of CP K. oxytoca was principally due to the spread of VIM-1- and OXA-48-producing isolates in which these genes were carried by IncL, IncHI2, IncFII, and IncN plasmids.

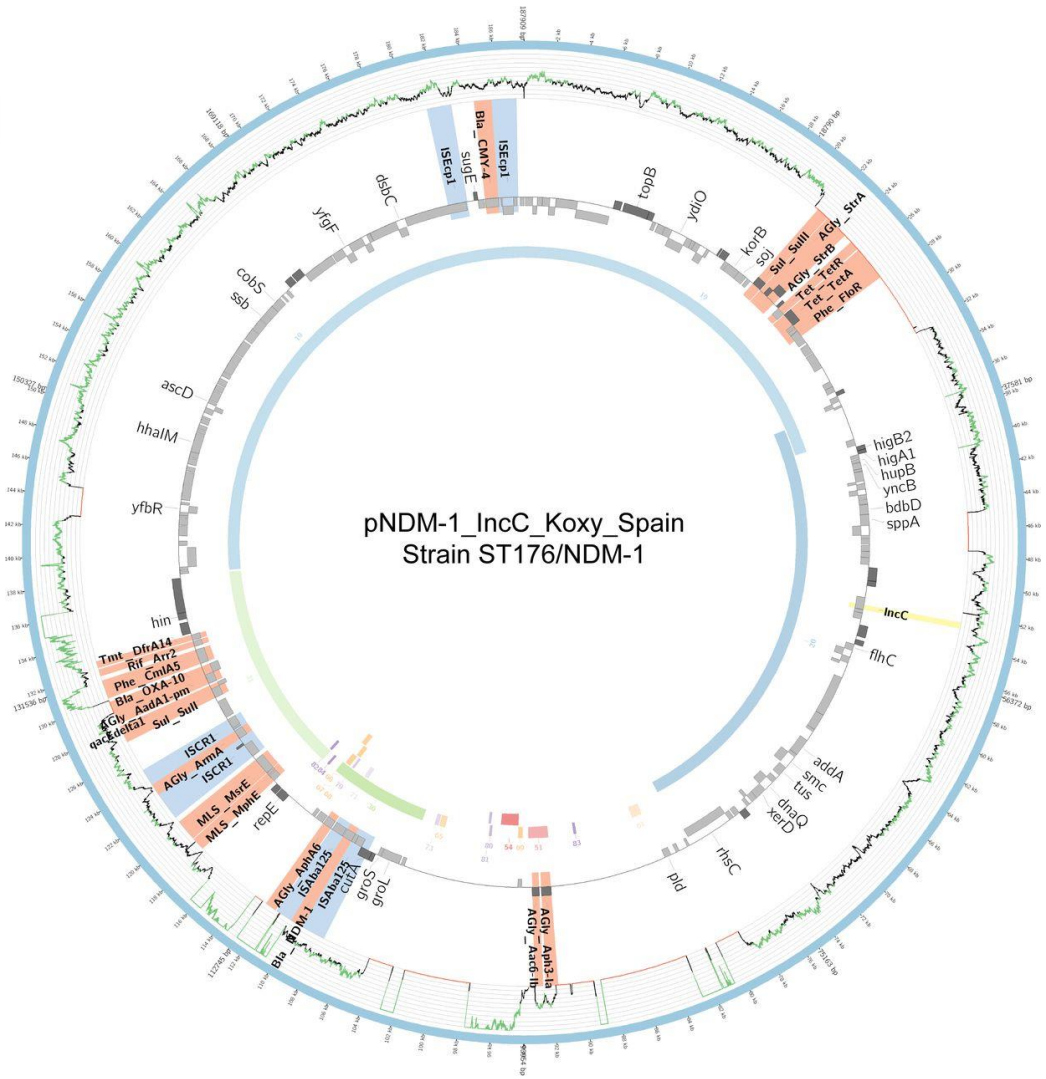
Pérez-Vázquez et al. AAC, May 2019

Characterization of Carbapenemase-Producing *Klebsiella oxytoca* in Spain, 2016–2017

TABLE 2 Main antibiotic resistance genes and other molecular markers detected by WGS of 12 representative CP *Klebsiella oxytoca* isolates[a] (Table view)

Isolate	PFGE/ST	Carbapenemase gene(s)	Plasmid(s) carrying carbapenemases genes (size [kb])	Gene(s) detected											
				bla_{OXY}	Other *bla* genes	*aadA*	*aac(6')Ib*	*arm*	*ant(2")-Ia*	*sul*	*dhfr*	*qnr*	*cat*	*tet*	*msr* and *mph*
K7533	C1/ST2	bla_{OXA-48}	IncL (~63)	$bla_{OXY-2-8}$	N	*aadA1*	ND	N	N	ND	*dfrA1*	N	N	N	N
K0100	C1/ST2	bla_{OXA-48}	IncL (~63)	$bla_{OXY-2-8}$	N	*aadA1*	ND	N	N	ND	*dfrA1*	N	N	N	N
K9103	C2/ST36	bla_{VIM-1}	IncHI2 (~250)	$bla_{OXY-2-2}$	N	*aadA1*	*aac(6')Ib*	N	N	*sul1, sul2*	*dfrA1, dfrB1*	N	*catA1, catB2*	N	*msrE, mphE*
K9635	C3/ST2	bla_{VIM-1}	IncL (~70)	$bla_{OXY-2-8}$	N	*aadA1*	*aac(6')Ib*	N	N	*sul1*	*dfrA1 dfrB1*	N	*catA1, catB2*	N	N
K9599	C4/ST36	bla_{VIM-1}	IncFII (~70)	$bla_{OXY-2-2}$	bla_{TEM-1A}	*aadA1*	*aac(6')Ib*	N	N	*sul1*	*dfrB1*	N	*catB2*	N	N
K8025	Couple/ST20	bla_{VIM-1}	IncL (~70)	$bla_{OXY-2-5}$	N	*aadA1*	*aac(6')Ib*	N	N	*sul1*	*dfrB1*	N	*catA1, catB2*	N	*msrE mphE*
K8021	Couple/ST2	bla_{VIM-1}	IncHI2 (~250)	$bla_{OXY-2-8}$	$bla_{CTX-M-9}$, bla_{SHV-12}	*aadA1, aadA2*	*aac(6')Ib*	N	*ant(2")-Ia*	*sul1*	*dfrA1, dfrA16, dfrB1*	*qnrA1*	*catA1, catB2*	tetA	N
K7929	Couple/ST141	bla_{OXA-48}	IncL (~48)	$bla_{OXY-2-2}$	N	ND	ND	N	N	N	ND	N	N	N	*msrE, mphE*
K8792	Couple/ST2	bla_{VIM-1}	ND	$bla_{OXY-2-8}$	N	*aadA1*	*aac(6')Ib*	N	N	*sul1*	*dfrA1, dfrB1*	N	*catB2*	N	N
K9534	NR/ST36	bla_{VIM-1}, bla_{KPC-3}	IncN (~50) for bla_{VIM-1}; IncFIB (~103) for bla_{KPC-3}	$bla_{OXY-2-2}$	bla_{TEM-1A}, bla_{OXA-9}	*aadA1*	*aac(6')Ib*	N	N	*sul1*	*dfrA14, dfrB1*	*qnrS1*	*catA1, catB2*	N	N
K9455	NR/ST176	bla_{NDM-1}	IncC (~154)	$bla_{OXY-2-5}$	bla_{OXA-10}, bla_{CMY-4}	*aadA1*	*aac(6')Ib*	*armA*	N	*sul1, sul2*	*dfrA14*	*qnrB32*	N	*tetA*	*msrE, mphE*
K9682	NR/ST176	bla_{VIM-1}, bla_{KPC-2}	IncL (~70) for bla_{VIM-1}; IncP6 (~39) for bla_{KPC-2}	$bla_{OXY-2-5}$	bla_{TEM-1A}	*aadA1*	*aac(6')Ib*	N	N	*sul1*	*dfrB1*	N	*catA1, catB2*	N	N

[a] C1, C2, C3, and C4, representative isolates from PFGE clusters 1, 2, 3, and 4, respectively; couple, representative isolates of PFGE clusters of only two isolates; NR, isolates not related by PFGE to any other isolate; ND, not determined; N, negative.



Pérez-Vázquez et al. AAC, May 2019

Carbapenemase-Producing *Klebsiella pneumoniae* From Transplanted Patients in Brazil: Phylogeny, Resistome, Virulome and Mobile Genetic Elements Harboring bla_{KPC-2} or bla_{NDM-1}

Otávio Hallal Ferreira Raro[1,2], Ravena Maya Cardoso da Silva[1], Edison Moraes Rodrigues Filho[3], Teresa Cristina Teixeira Sukiennik[4], Claudio Stadnik[4], Cícero Armídio Gomes Dias[1], Jesús Oteo Iglesias[2] and María Pérez-Vázquez[2]

- Describe the phylogeny, resistome, virulome, and the plasmids encoding bla_{KPC} and bla_{NDM} of CP-Kp strains isolated from transplanted inpatients.

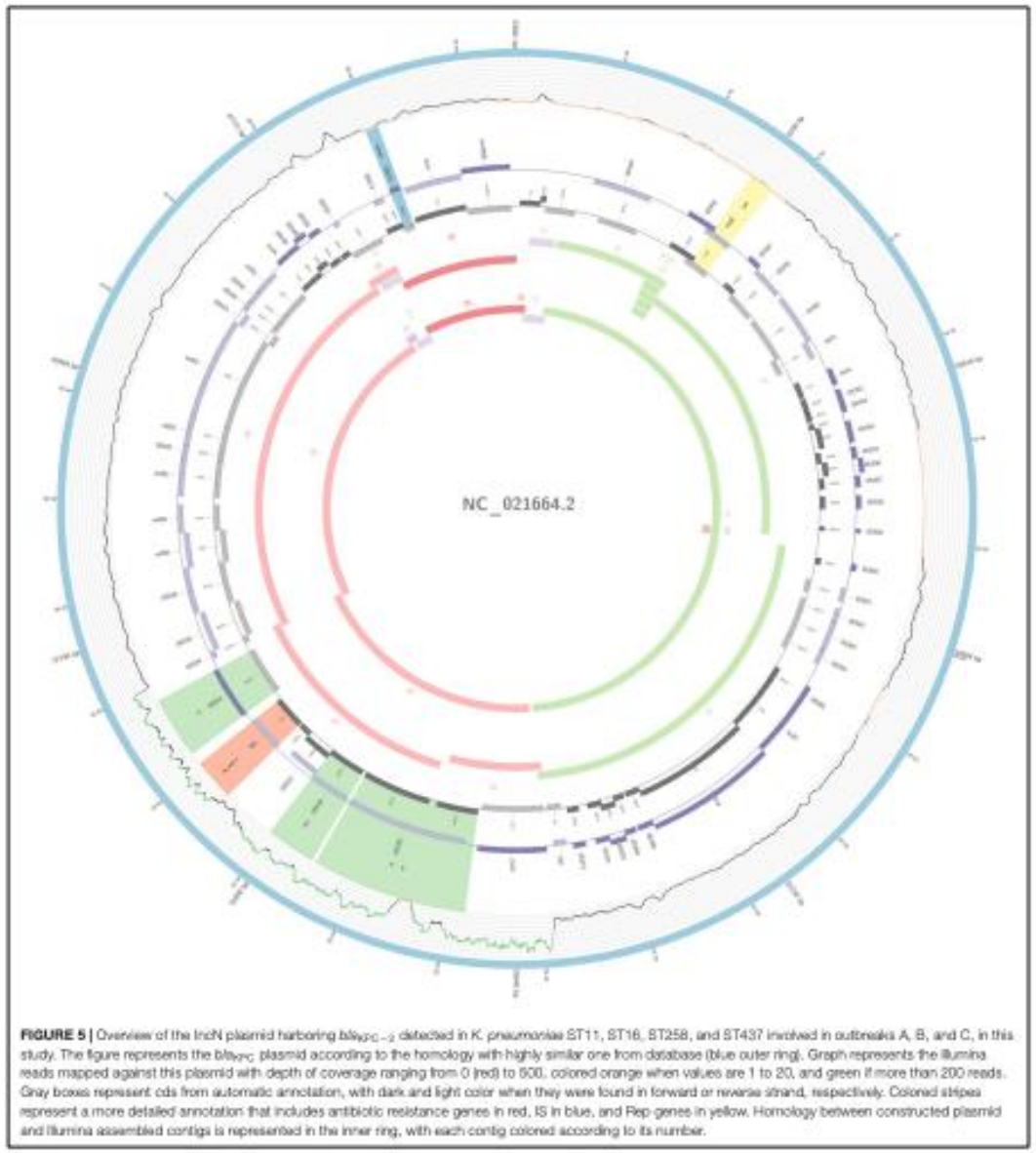


FIGURE 5 | Overview of the IncN plasmid harboring bla_{KPC-2} detected in *K. pneumoniae* ST11, ST16, ST258, and ST437 involved in outbreaks A, B, and C, in this study. The figure represents the bla_{KPC} plasmid according to the homology with highly similar one from database (blue outer ring). Graph represents the Illumina reads mapped against this plasmid with depth of coverage ranging from 0 (red) to 500, colored orange when values are 1 to 20, and green if more than 200 reads. Gray boxes represent cds from automatic annotation, with dark and light color when they were found in forward or reverse strand, respectively. Colored stripes represent a more detailed annotation that includes antibiotic resistance genes in red, IS in blue, and Rep genes in yellow. Homology between constructed plasmid and Illumina assembled contigs is represented in the inner ring, with each contig colored according to its number.

Raro et al. Frontiers, July 2020

- Plasmids containing NDM are reconstructed in the main clones of *K. pneumoniae* responsible for the dissemination of this CP in Spain
- ST437/NDM-7, ST437/NDM-1, ST147/NDM-1, ST11/NDM -1 and ST101/NDM-1.
- Both long read and short read sequencing are applied, the structure of 6 plasmids (IncFII, IncFIB, IncX3, IncR, IncN and IncC) is obtained
- IncX3 type is shared between different species.

Emergence of NDM-producing *Klebsiella pneumoniae* and *Escherichia coli* in Spain: phylogeny, resistome, virulence and plasmids encoding *bla*$_{NDM}$-like genes as determined by WGS

María Pérez-Vázquez[1,2]*, Pedro J. Sola Campoy[1], Adriana Ortega[1,2], Verónica Bautista[1,2], Sara Monzón[3], Guillermo Ruiz-Carrascoso[2,4], Jesus Mingorance[2,4], Eva M. González-Barberá[5], Concepción Gimeno[6], Belén Aracil[1,2], David Sáez[1,2], Noelia Lara[1,2], Sara Fernández[1,2], Juan José González-López[2,7], José Campos[1,2], Robert A. Kingsley[8,9], Gordon Dougan[8] and Jesús Oteo-Iglesias[1,2] on behalf of the Spanish NDM Study Group†

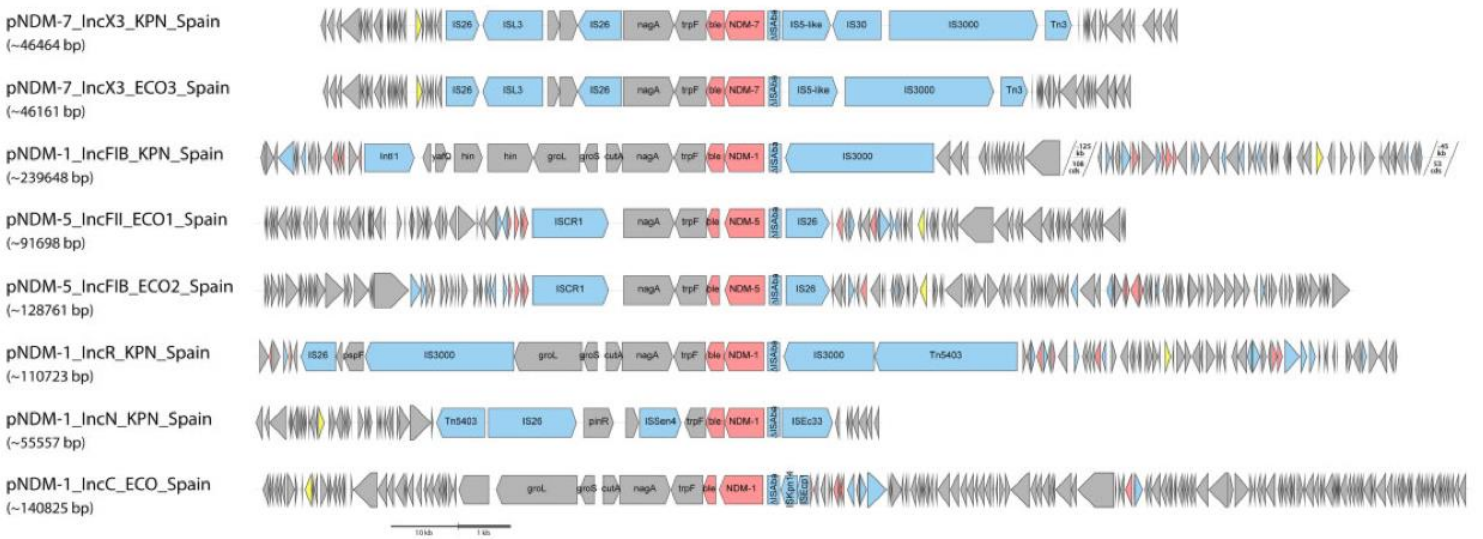


Figure 2. Major structural features of the plasmids described in this study. Predicted coding sequences are indicated with arrows oriented in the direction of transcription of each respective gene; red arrows represent resistance genes, blue arrows represent transposon-related genes and ISs, and yellow arrows represent *rep* genes.

Pérez-Vázquez et al. JAC, Sep 2019.

antibiotics

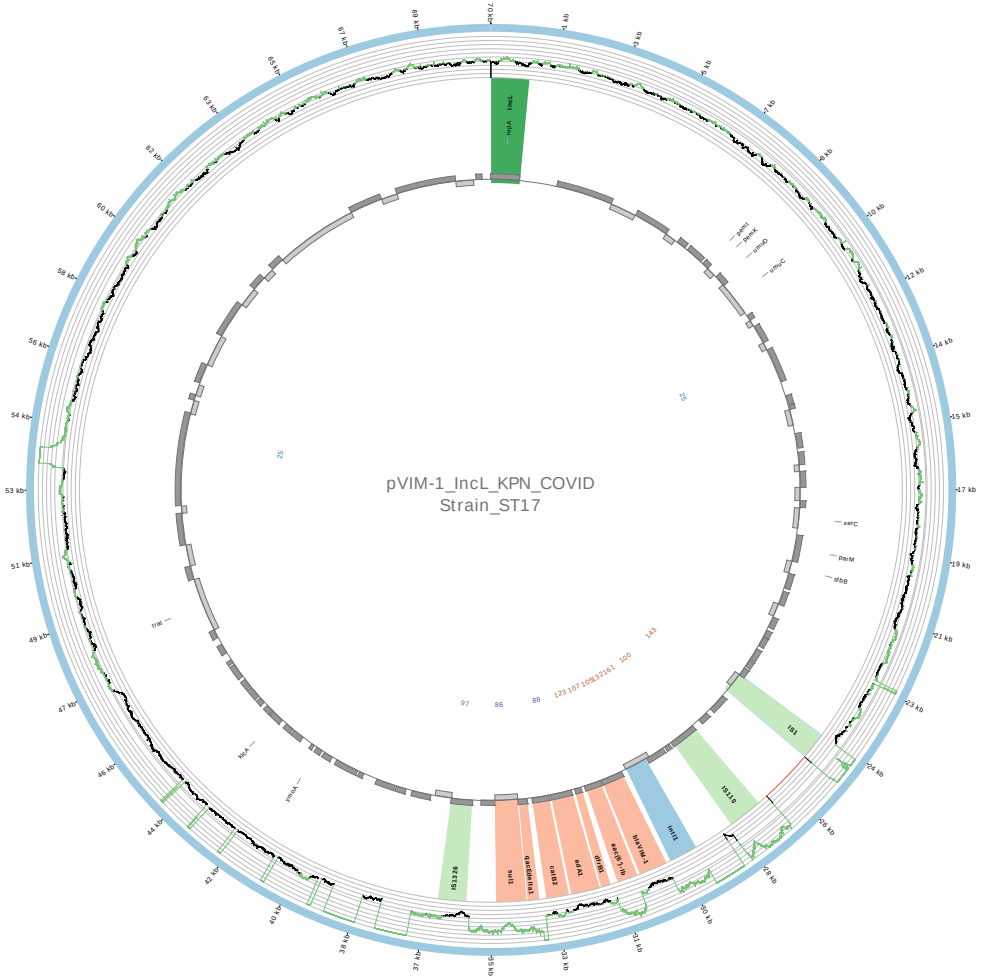
MDPI

Article

Carbapenemase-Producing *Klebsiella pneumoniae* in COVID-19 Intensive Care Patients: Identification of IncL-VIM-1 Plasmid in Previously Non-Predominant Sequence Types

Javier E. Cañada-García [1,2,†], Eva Ramírez de Arellano [1,2,†], Miguel Jiménez-Orellana [1], Esther Viedma [3], Aida Sánchez [4], Almudena Alhambra [5], Jennifer Villa [3], Alberto Delgado-Iribarren [6], Verónica Bautista [1], Noelia Lara [1], Silvia García-Cobos [1], Belén Aracil [1,2], Emilia Cercenado [7,8], María Pérez-Vázquez [1,2,*,‡] and Jesús Oteo-Iglesias [1,2,‡]

- Analyse the degree to which the COVID-19 pandemic has changed CPE status
- The aim of this research project was:
  - Characterize CP-Kpn isolates and sequence types (STs) in COVID-19 patients in the ICU
  - To be able to compare the epidemiological and microbiological characteristics of CP-Kpn isolates obtained during the COVID-19 pandemic with those obtained before the onset of the pandemic.



- Emergence of new VIM-1-producing *K. pneumoniae* clones linked to successful IncL plasmids and a class 1 integron.

SUMMARY

- Plasmid reconstruction is a tricky matter
- From our experience good results are obtained combining plasmidID (mapping tool) and log-reads data
- Not applicable (with our resources) in routine surveillance

ACKNOWLEDGMENTS

Laboratorio de RA
Unidad de Genómica, ISCIII
Unidad de Bioinformática, ISCIII
UTIC, ISCIII ……..