

PlasmidID:

a mapping based tool for plasmid identification, annotation, classification and representation. Application as high-risk strains plasmid detection

Pedro J. Sola-Campoy^[1,2]

, Sara Monzón^[2], María Pérez-Vázquez^[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.



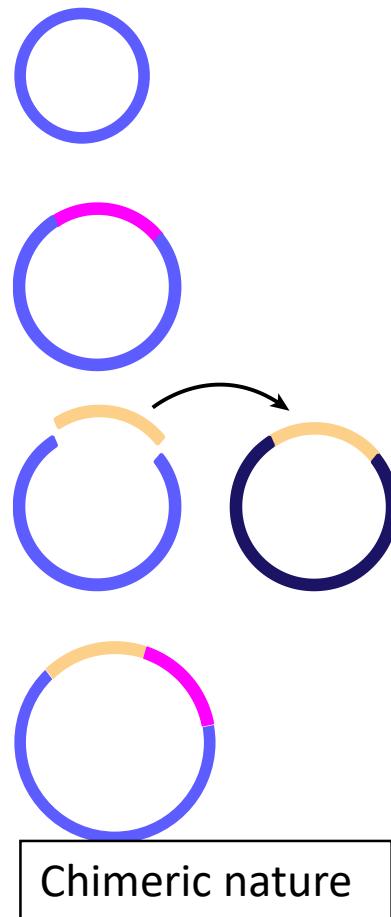
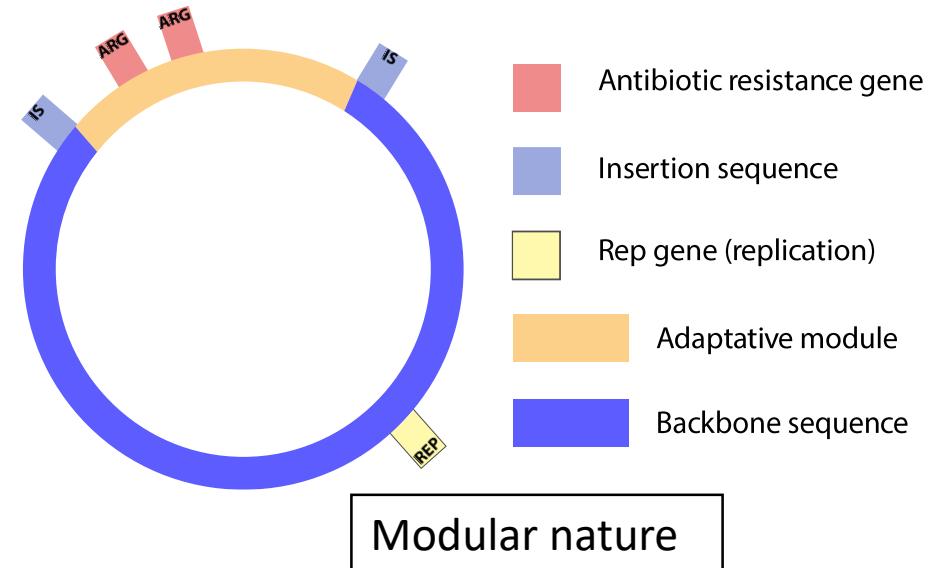
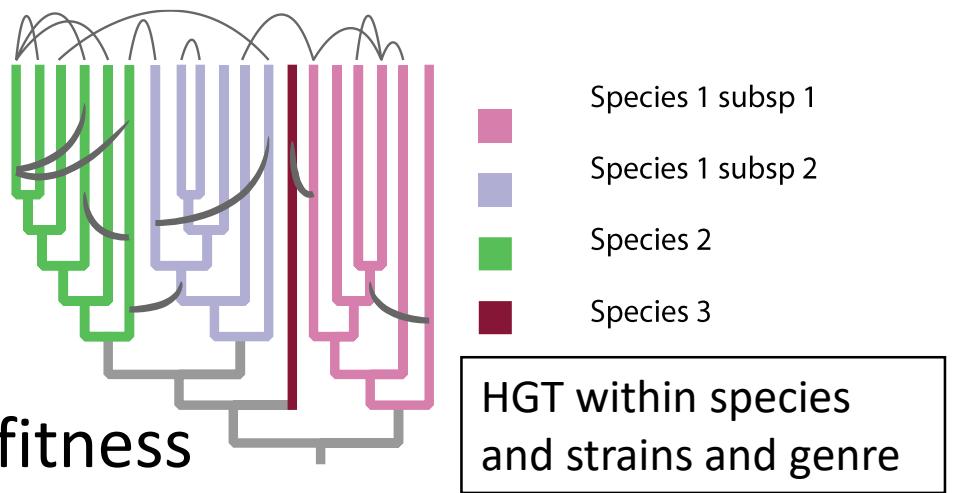
> BU-ISCIII

bioinformatica@isciii.es



BACKGROUND Plasmids

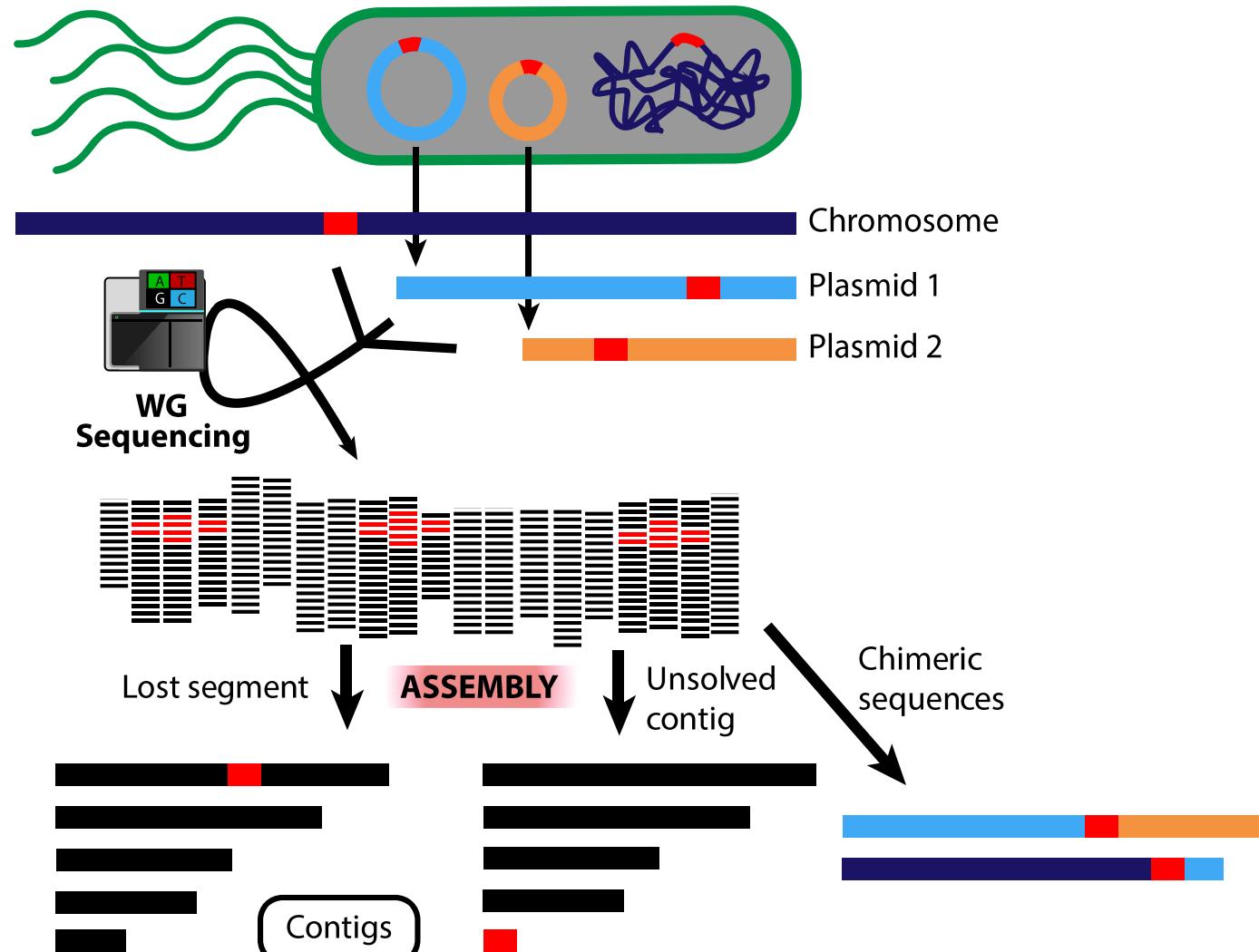
- Spread through HGT
 - ARG
 - Virulence factors
 - Accessory genes
- Improve bacterial survival and fitness
- Modular and Chimeric nature
 - Backbone
 - Replication and plasmid survival
 - Adaptative
 - Host survival (ABR)
 - Flanked by IS and/or transposons
 - Transferable / able to combine
 - Chimeric



BACKGROUND

Mapping

- *de novo* assembly
 - [PlasmidSPAdes / Recycler](#)
 - Fail with plasmids with the same copy number as chromosome
 - Assembly algorithm (image)
- BLASTed DDBB
 - [PLACNET](#)
 - Homologous sequences
 - Plasmid vs Chromosome
 - Plasmid vs Plasmid (same or different)
 - Small plasmids underrepresented
 - Assembly algorithm (image)

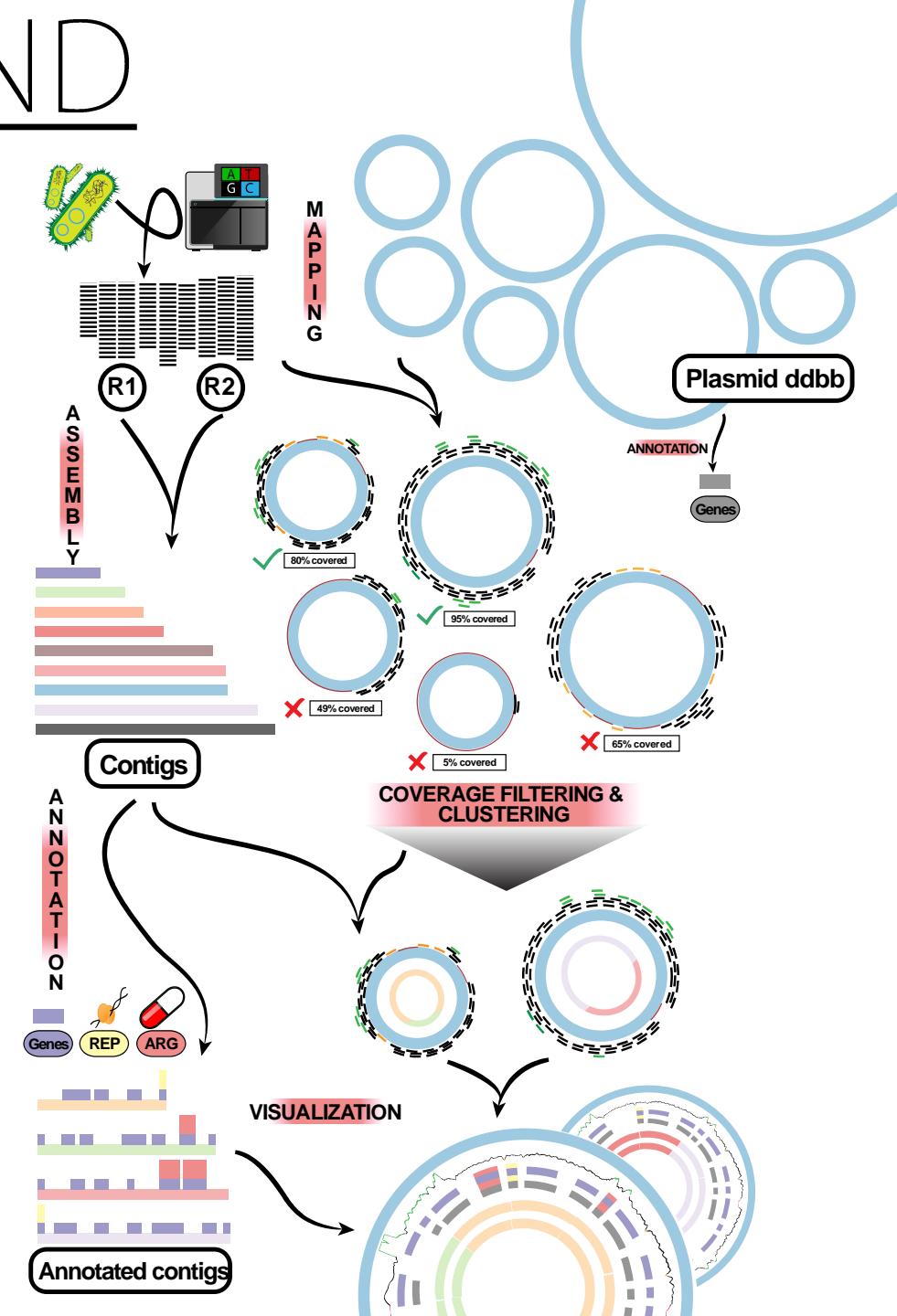




BACKGROUND

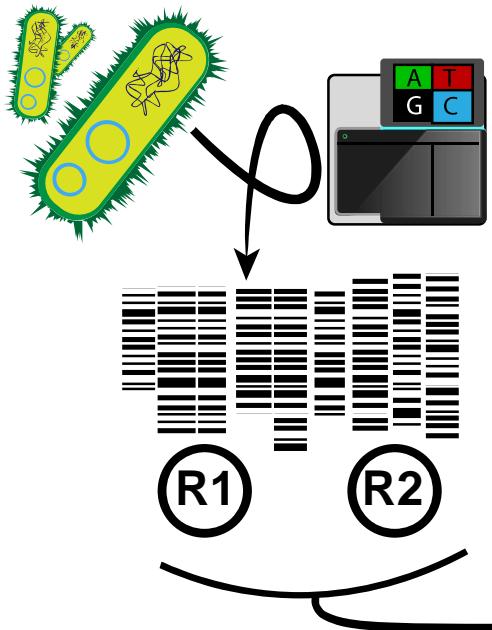
Proposal

- BASH implemented tool
- Rely on mapping to identify putative plasmids
 - Mapping based
- Assembly assisted
 - Positional information (reconstruction)
 - Annotation
 - Typing
- Pipeline
 - Mapping (bowtie2)
 - Coverage (bedtools)
 - Cluster (cd-hit)
 - Assembly (SPAdes)
- Annotation
 - Prokka/BLAST
 - Visualization (circus)



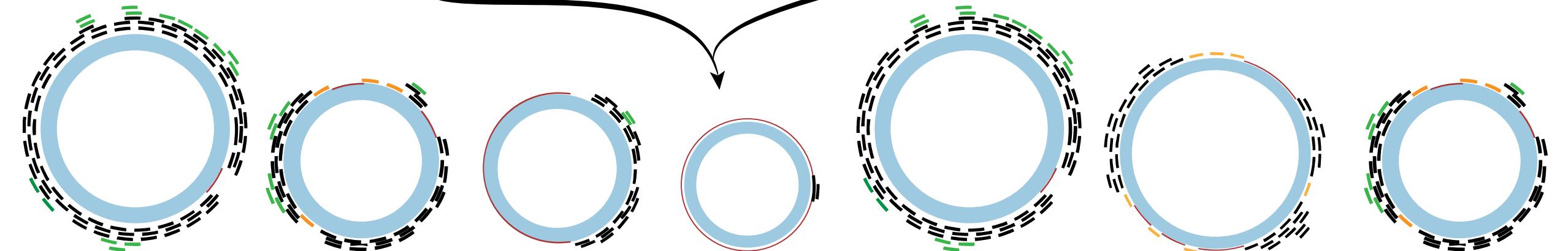
PlasmidID pipeline

Mapping



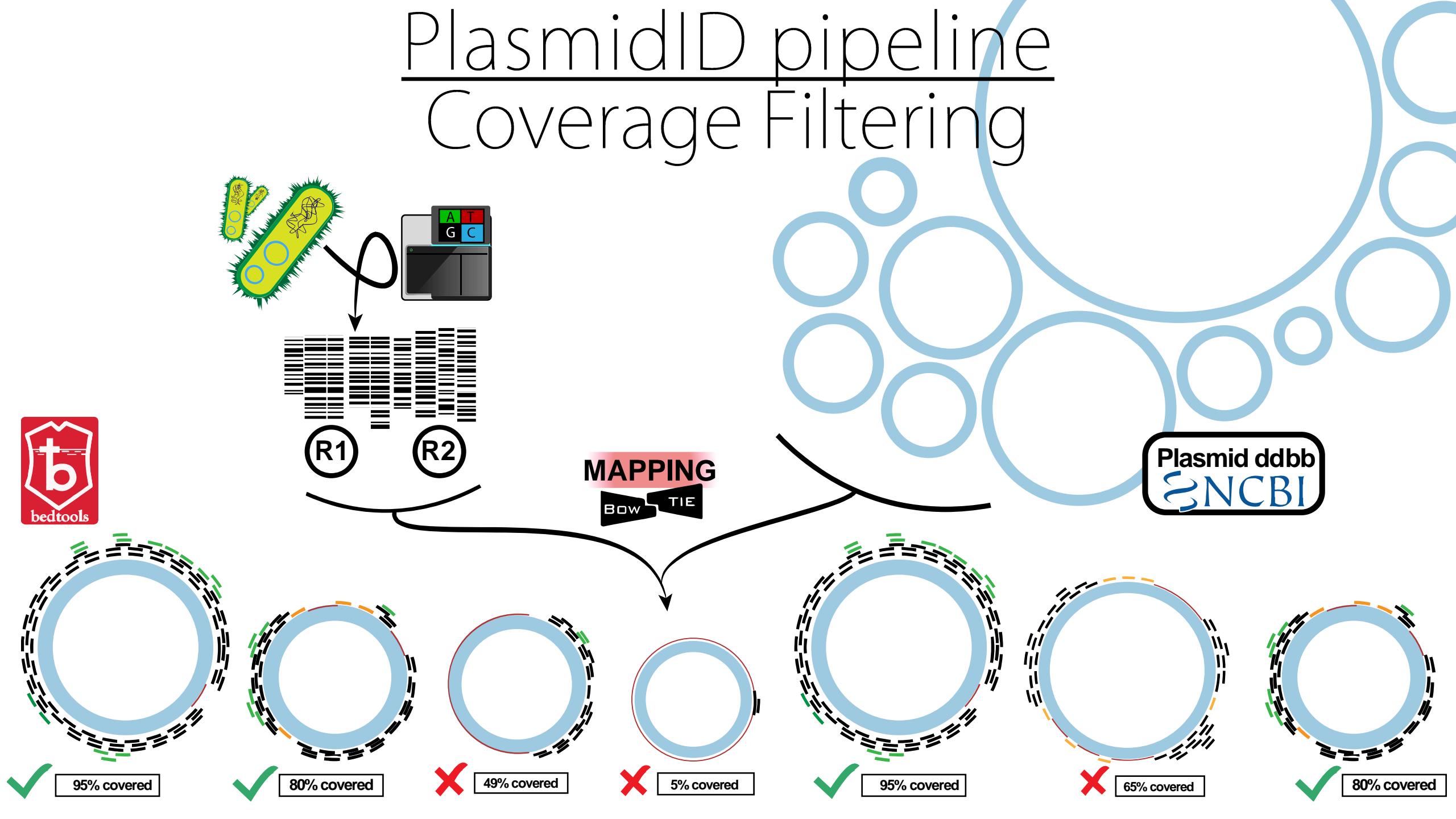
MAPPING
BOW TIE

Plasmid ddbb
NCBI



PlasmidID pipeline

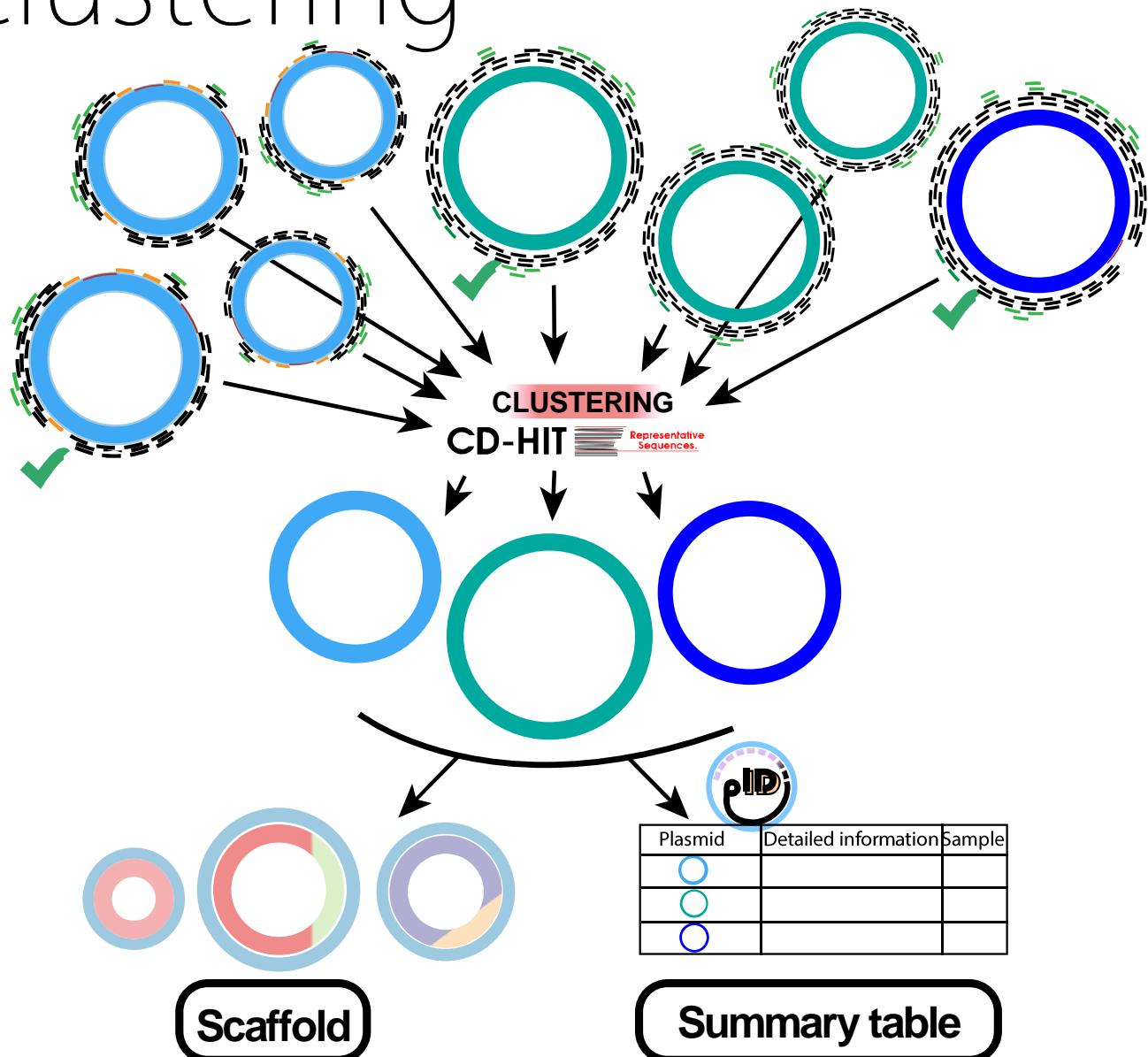
Coverage Filtering



PlasmidID pipeline

Clustering

- DDBB redundancy
 - Refseq
- Biased database
 - Enterobacteriaceae
 - ABR bearing plasmids
- Clustering (CD-HIT)
 - Longest representative
 - Default homology percentage 80%



PlasmidID results

Summary table

AC_Number	Length	Species
NC_019049.1	4012	<i>E. coli</i>
NZ_CP015856.1	115432	<i>E. coli</i>
NZ_CP018246.1	95341	<i>E. coli</i>
NZ_CP015022.1	95170	<i>E. coli</i>
NZ_CP018253.1	95229	<i>E. coli</i>

AC_Number	Length	Species	Description	N	<i>E. coli</i>	<i>K. pneumoniae</i>	<i>S. enterica</i>
NC_019049.1	4012	<i>E. coli</i>	plasmid pCM959	1	99.9999		
NZ_CP015856.1	115432	<i>E. coli</i>	strain EDL933-1 plasmid pO157	1	99.9983		
NZ_CP018246.1	95341	<i>E. coli</i>	strain 472 plasmid pO157	1	99.9979		
NZ_CP015022.1	95170	<i>E. coli</i>	strain SRCC 1675 plasmid pSRCC 1675	1	99.9979		
NZ_CP018253.1	95229	<i>E. coli</i>	strain 9000 plasmid pO157	1	99.9979		
NC_007414.1	92077	<i>E. coli</i>	O157:H7 str. EDL933 plasmid pO157	1	99.9978		
NZ_CP017670.1	92755	<i>E. coli</i>	strain PA20 plasmid pO157	1	99.9978		
NZ_CP017435.1	92624	<i>E. coli</i>	O157:H7 strain 1130 plasmid pO157	1	99.9978		
NZ_CP018244.1	92522	<i>E. coli</i>	strain 350 plasmid pO157	1	99.9978		
NZ_CP018242.1	92495	<i>E. coli</i>	strain 319 plasmid pO157	1	99.9978		
NZ_CP018251.1	91789	<i>E. coli</i>	strain 10671 plasmid pO157	1	99.9978		
NZ_CP018238.1	91420	<i>E. coli</i>	strain 155 plasmid pO157	1	99.9978		
NZ_D5999999.1	92381	<i>E. coli</i>	O157:H7 str. TW14588 plasmid pTW14588 scf_1117778227960	1	99.9935		
NZ_CP017252.1	92691	<i>E. coli</i>	strain NADC 5570/86-24/6564 plasmid pO157	1	99.9935		
NZ_CP018240.1	94170	<i>E. coli</i>	strain 272 plasmid pO157	1	99.9289		
NZ_CP015021.1	81401	<i>E. coli</i>	strain 28RC1 plasmid p28RC1	1	99.9177		
NZ_CP006633.1	3319	<i>E. coli</i>	PCN033 plasmid p1PCN033	1	99.7288		
NZ_AVCD01000003.1	98062	<i>E. coli</i>	O157:H7 str. F8092B plasmid pO157	1	99.5085		
NZ_CP015847.1	95598	<i>E. coli</i>	O157:H7 strain FRIK2069 plasmid pO157	1	99.5		
NZ_CP015816.1	95910	<i>E. coli</i>	O157:H7 strain JEONG-1266 plasmid pO157	1	99.4974		
NZ_CP010305.1	94730	<i>E. coli</i>	O157:H7 str. SS52 plasmid pO157	1	99.4954		
NC_013010.1	94601	<i>E. coli</i>	O157:H7 str. TW14359 plasmid pO157	1	99.4884		
NZ_AETX01000217.1	89762	<i>E. coli</i>	O157:H7 str. G5101 plasmid pO157	1	86.9533		
NZ_CP010135.1	91607	<i>E. coli</i>	strain D1 plasmid A complete genome	1	81.3104		
NC_016846.1	111195	<i>K. pneumoniae</i>	subsp. pneumoniae HS11286 plasmid pKPHS2	1	99.9991		
NC_016839.1	105974	<i>K. pneumoniae</i>	subsp. pneumoniae HS11286 plasmid pKPHS3	1	99.9981		
NC_016838.1	122799	<i>K. pneumoniae</i>	subsp. pneumoniae HS11286 plasmid pKPHS1	1	99.9976		
NZ_CP015133.1	26450	<i>K. pneumoniae</i>	strain Kpn555 plasmid pKPN-d6b	1	99.9962		
NC_016840.1	3751	<i>K. pneumoniae</i>	subsp. pneumoniae HS11286 plasmid pKPHS4	1	99.9733		
NC_016847.1	3353	<i>K. pneumoniae</i>	subsp. pneumoniae HS11286 plasmid pKPHS5	1	99.9702		
NC_016841.1	1308	<i>K. pneumoniae</i>	subsp. pneumoniae HS11286 plasmid pKPHS6	1	99.9235		
NZ_CP011334.1	2954	<i>E. coli</i>	O104:H4 str. C227-11 plasmid	1	90.5213		
NZ_CP012565.1	112059	<i>K. pneumoniae</i>	strain UCLAOXA232K plasmid pUCLAOXA232-4	1	89.0683		
NZ_CP016924.1	111539	<i>K. pneumoniae</i>	isolate 23 plasmid unnamed1	1	89.0433		
CM003767.1	3652	<i>P. aeruginosa</i>	strain BH6 plasmid pBH6 whole genome shotgun sequence	1	87.2125		
NZ_CP012427.1	126466	<i>K. pneumoniae</i>	strain KP5 plasmid pSg1-1	1	86.5837		
NZ_CP008931.1	111693	<i>K. pneumoniae</i>	strain PMK1 plasmid pPMK1-B	1	85.077		
NZ_CP015755.1	109349	<i>K. pneumoniae</i>	strain W14 plasmid unnamed2	1	82.6135		
NZ_CP007137.1	64562	<i>E. coli</i>	O145:H2 str. RM12581 plasmid pRM12581	1	81.6661		
NC_003385.1	106516	<i>S. enterica</i>	subsp. enterica serovar Typhi str. CT18 plasmid pHCM2	1	99.9999		
NC_003384.1	218160	<i>S. enterica</i>	subsp. enterica serovar Typhi str. CT18 plasmid pHCM1	1	99.9977		
NZ_LN866946.1	141119	<i>S. enterica</i>	subsp. enterica serovar Serftenberg genome assembly NCTC10384 plasmid: 4	1	98.852		
NC_013365.1	204604	<i>E. coli</i>	O111:H str. 11128 plasmid pO111_1 DNA	1	96.7591		
NZ_LN866945.1	147787	<i>S. enterica</i>	subsp. enterica serovar Serftenberg genome assembly NCTC10384 plasmid: 3	1	96.3096		
NC_009981.1	208409	<i>S. enterica</i>	subsp. enterica serovar Choleraesuis plasmid pMak1	1	95.2819		
NC_002305.1	180461	<i>S. typhi</i>	plasmid R27	1	93.7034		
NZ_LN866944.1	727905	<i>S. enterica</i>	subsp. enterica serovar Serftenberg genome assembly NCTC10384 plasmid: 2	1	92.0865		
NZ_CP017724.1	106449	<i>S. enterica</i>	subsp. enterica serovar Stanleyville plasmid pSARB26_01	1	90.9647		
NZ_CP011655.1	109688	<i>C. freundii</i>	strain CAV1741 plasmid pCAV1741-110	1	89.5403		
NC_016825.1	181431	<i>S. enterica</i>	subsp. enterica serovar Typhi str. P-stx-12 plasmid	1	85.6904		
NC_023289.2	239151	<i>E. coli</i>	strain T23 plasmid pEQ1	1	83.9859		

<i>E. coli</i>	<i>K. pneumoniae</i>	<i>S. enterica</i>
99.9999		
99.9983		
99.9979		
99.9979		
99.9979		

99.4884	
86.9533	
81.3104	
	99.9991
	99.9981
	99.9976
	99.9962

●

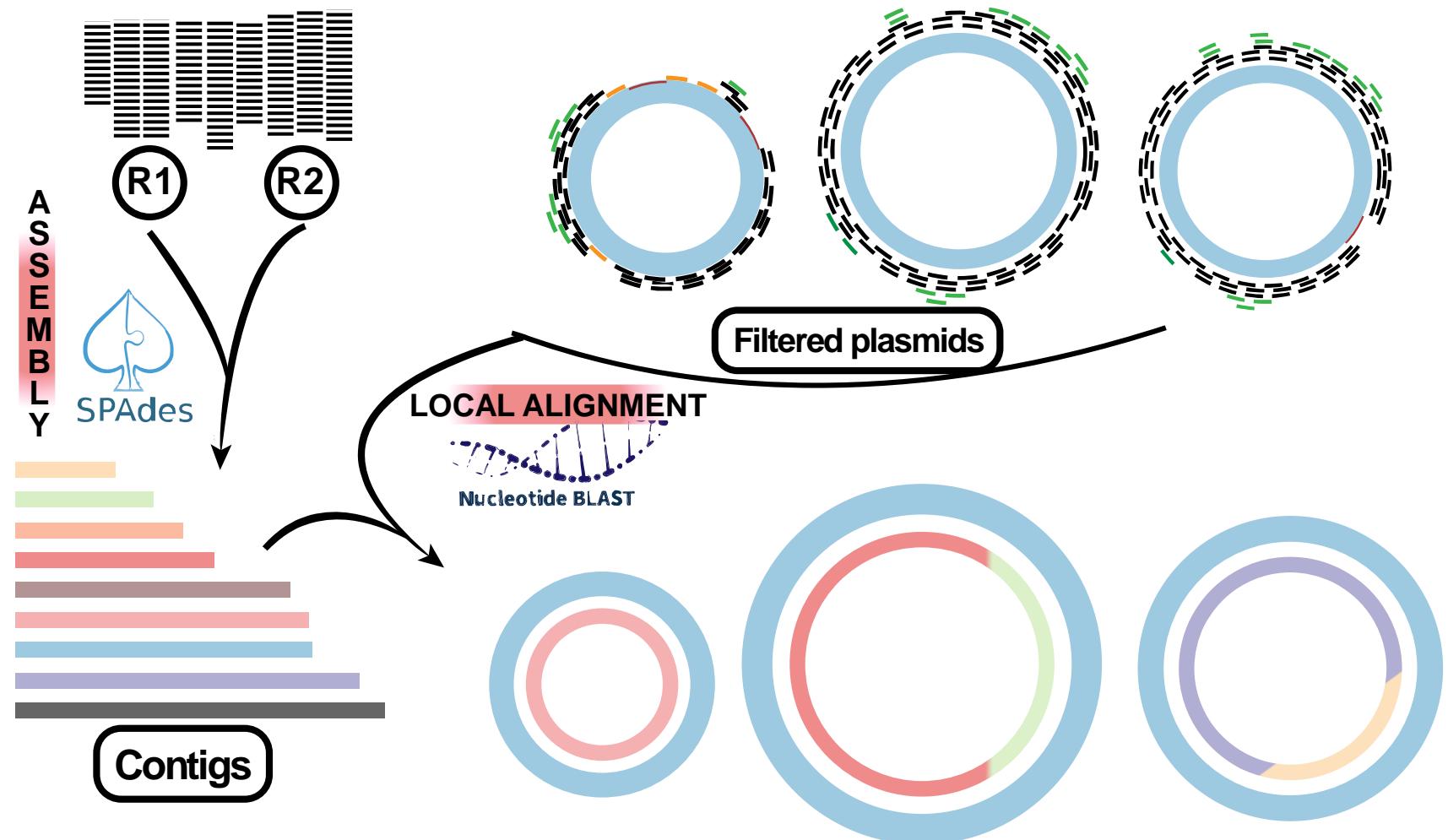
●

●

PlasmidID pipeline

Alignment

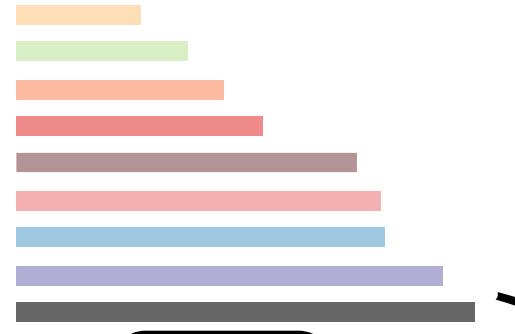
- Assembly
 - All sequences
 - Chromosome
 - Plasmids
- Local alignment
 - BLAST+
 - Filtered only by identity
 - Only relevant alignment
- Clustered plasmids are used as scaffold



PlasmidID pipeline

Annotation

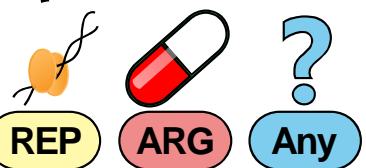
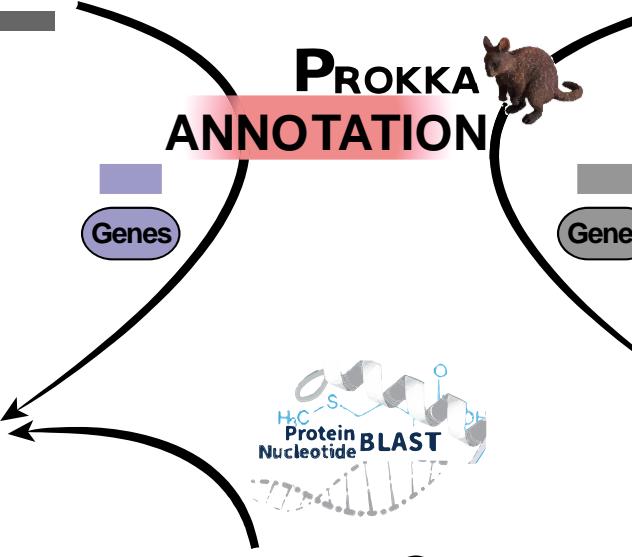
- Automatic annotation
 - Prokka
 - DDBB plasmid
 - Contigs
 - Gff to bed
- Specific annotation
 - BLAST+
 - ABR & REP
 - User input FASTA



Contigs



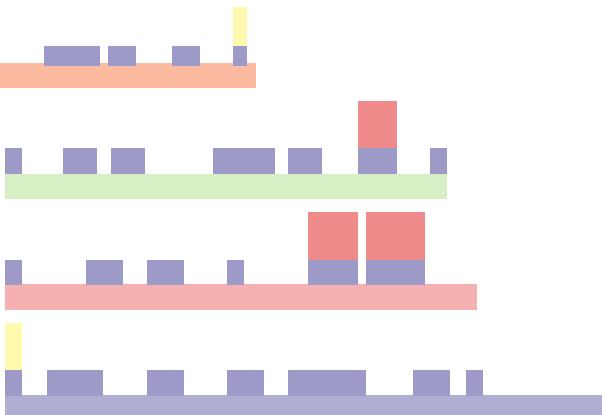
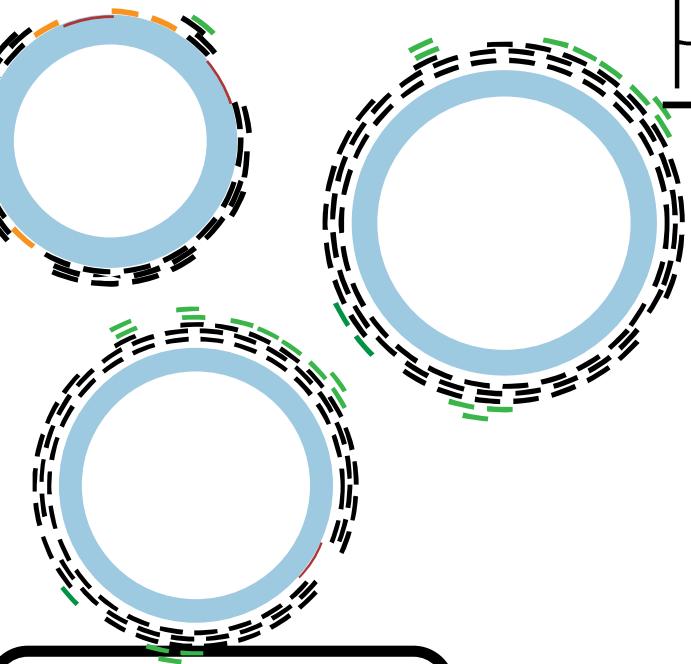
Annotated contigs



Plasmid ddbb
NCBI

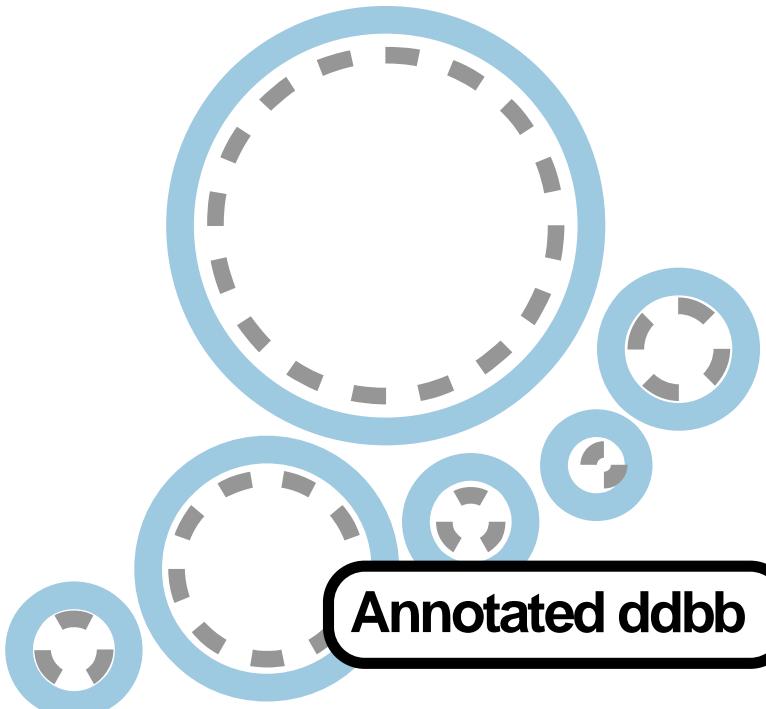
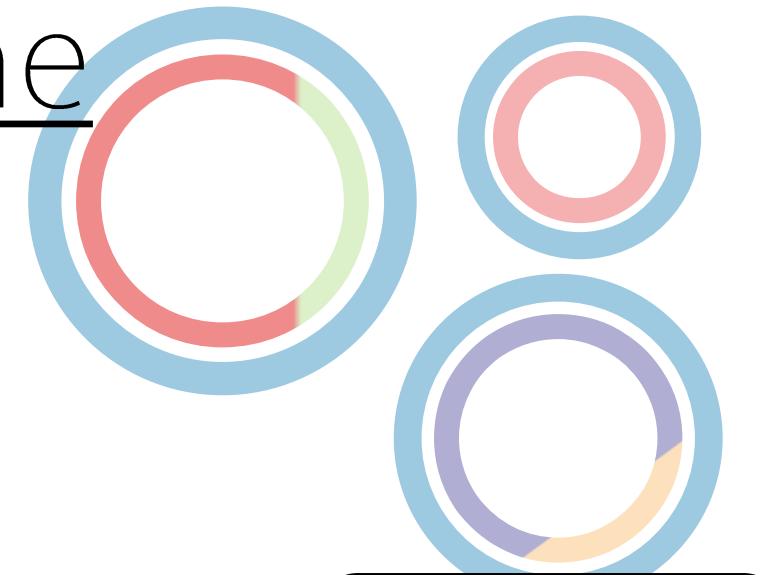
Annotated ddbb

PlasmidID pipeline Visualization



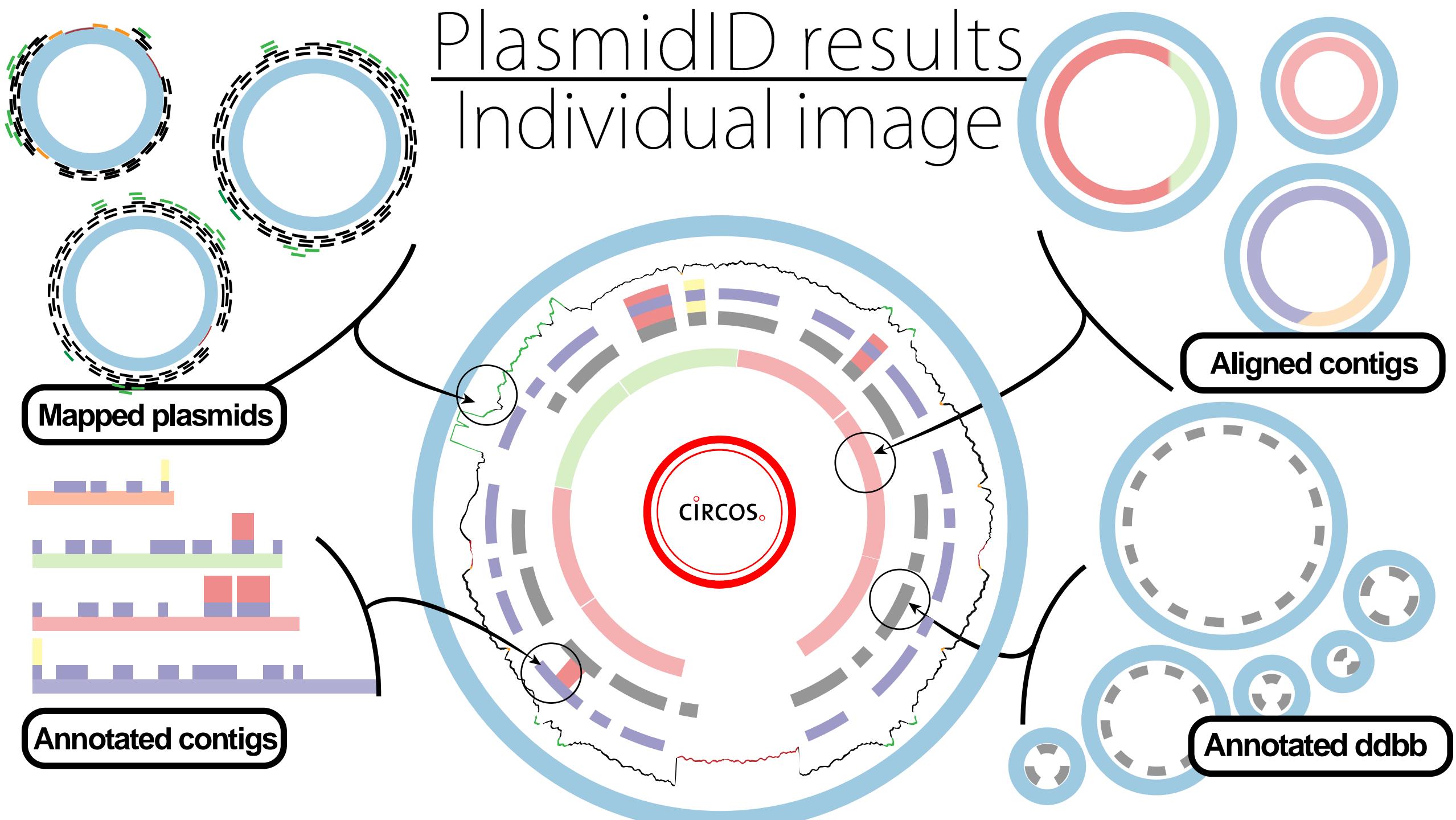
Annotated contigs

- Independently gathered information
 - Mapping
 - Local alignment
 - Specific annotation
 - Automatic annotation
 - Plasmid DDBB
 - Contigs
- All adapted to scaffold plasmid



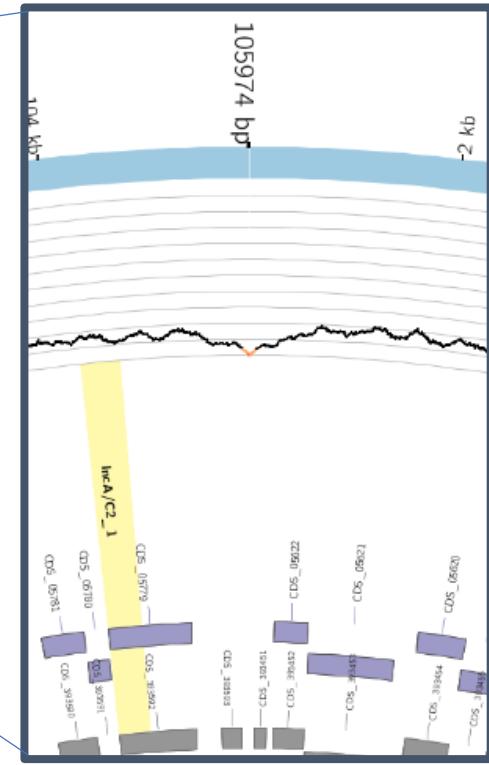
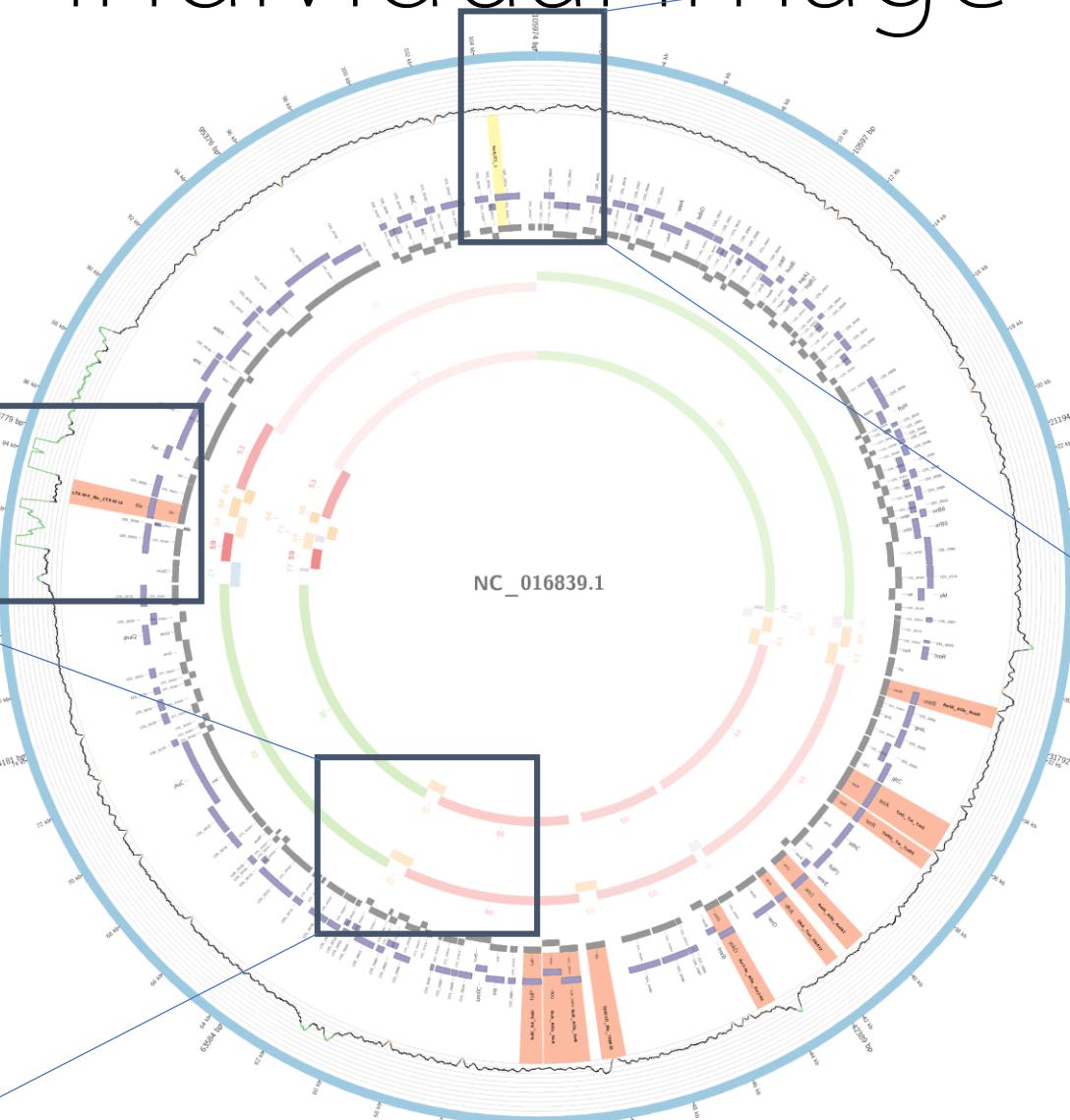
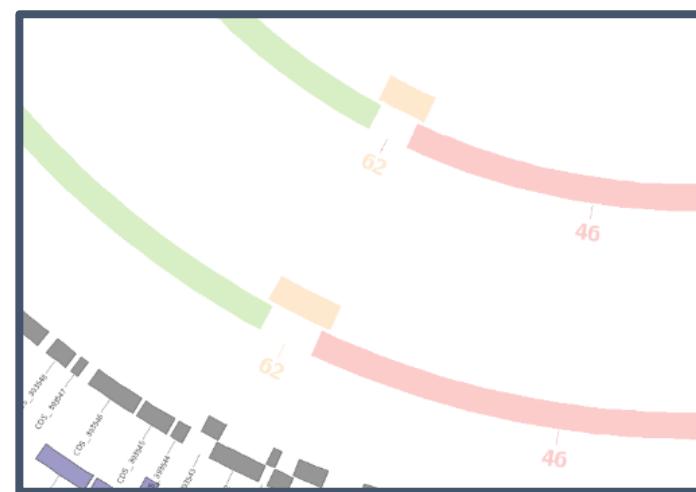
PlasmidID results

Individual image



PlasmidID results

Individual image



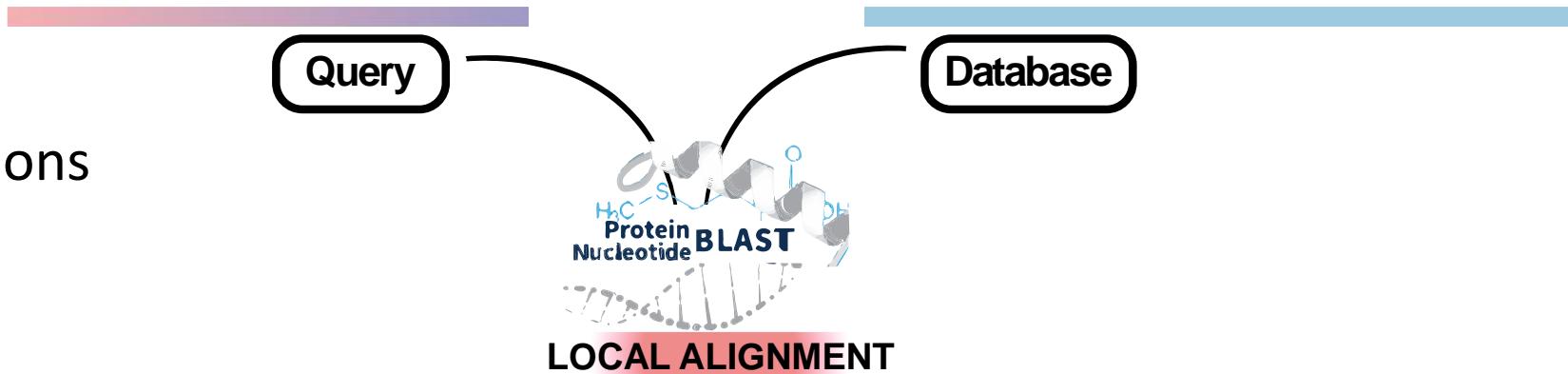
- *In silico* data
 - *K. pneumoniae*

PlasmidID pipeline

Local alignment vs Complete

- Local alignment

- Filtered by homology
- Detect sequence duplications
 - Insertion sequences
- Distant homology



- Complete Contig Track

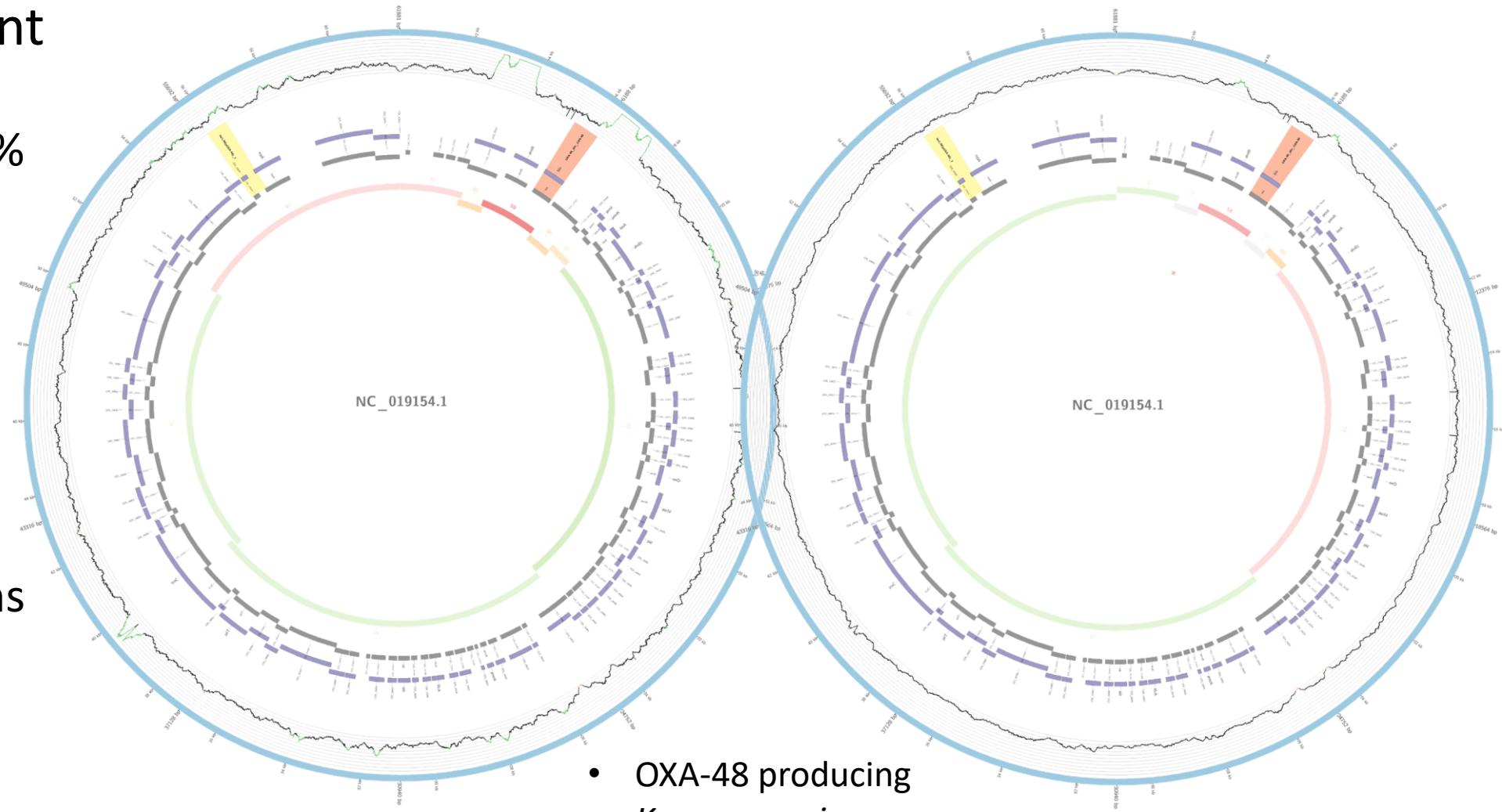
- Filtered by alignment length
- Use previous positional information
- Representative of the sample
 - No reference
- Know the total length of the contig



PlasmidID pipeline

Complete contig track

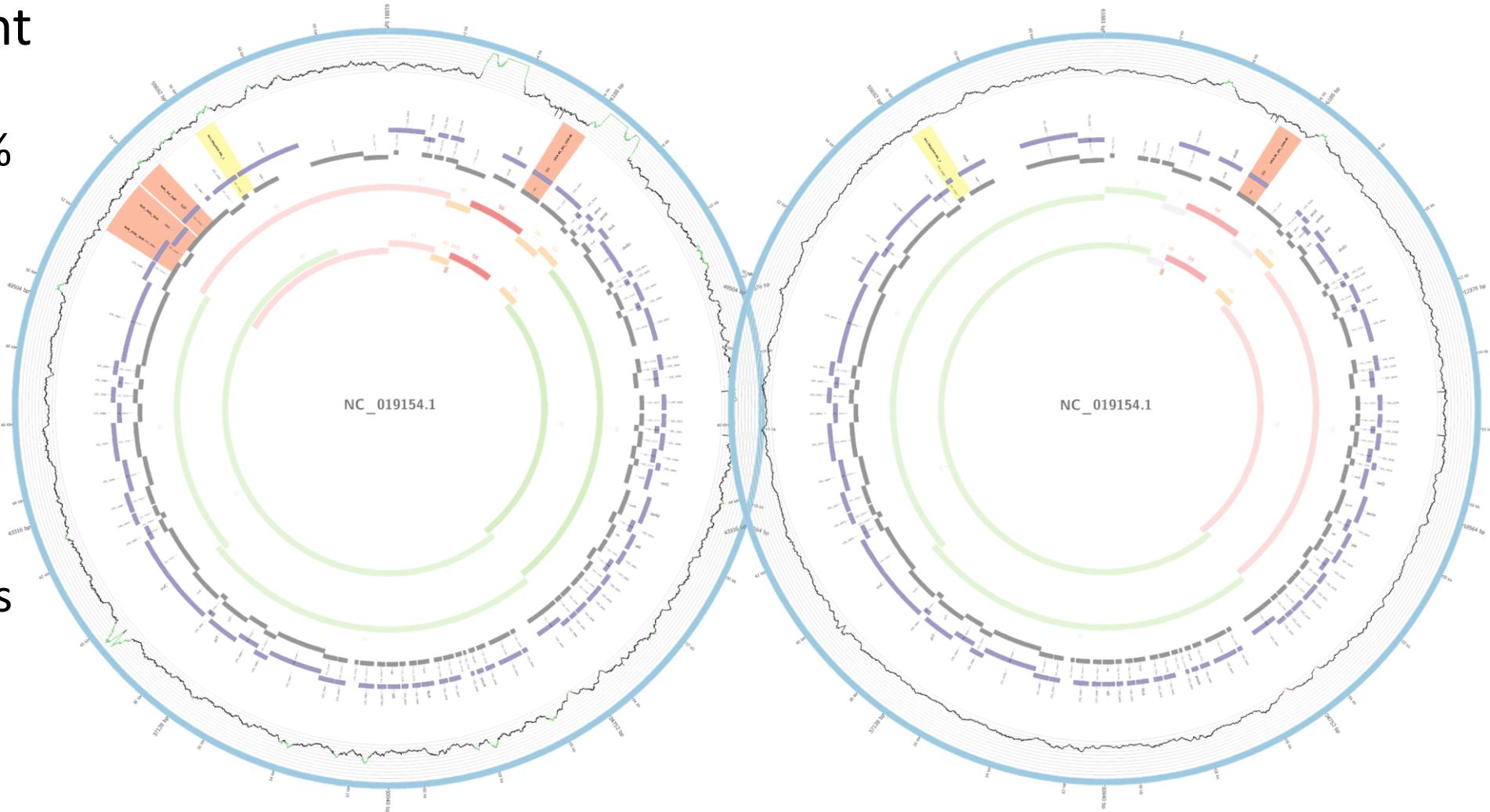
- Length dependent
 - Contigs with alignment > 50%
- No duplicated
- Adapted coords.
- Functions:
 - Additional info
 - Gap filler
 - Test duplications
 - Plasmid proofreading
 - Curation



PlasmidID result

Complete contig track

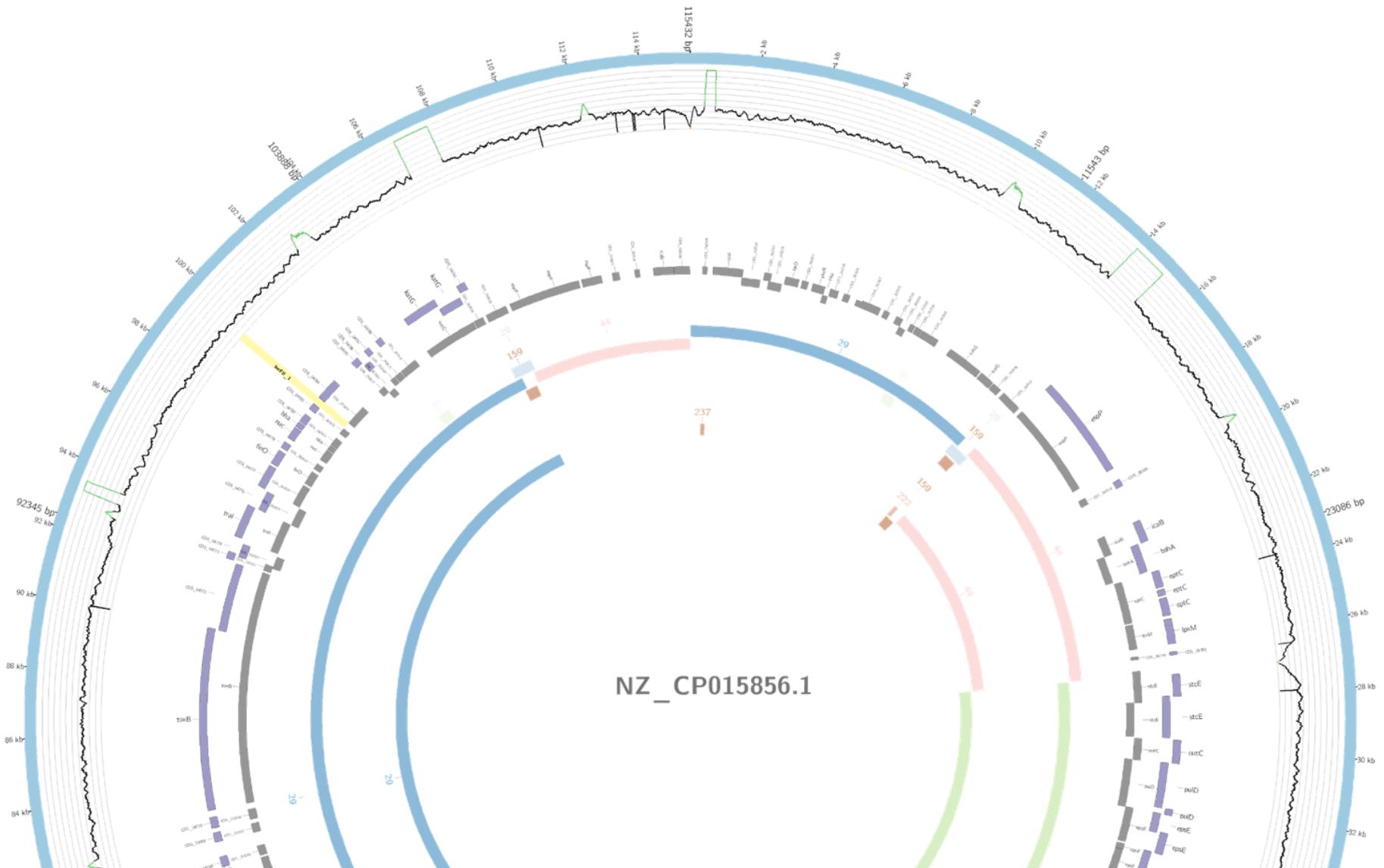
- Length dependent
 - Contigs with alignment > 50%
- No duplicated
- Adapted coords.
- Functions:
 - Additional info
 - Gap filler
 - Test duplications
 - Plasmid proofreading
 - Curation



PlasmidID result

Complete contig track

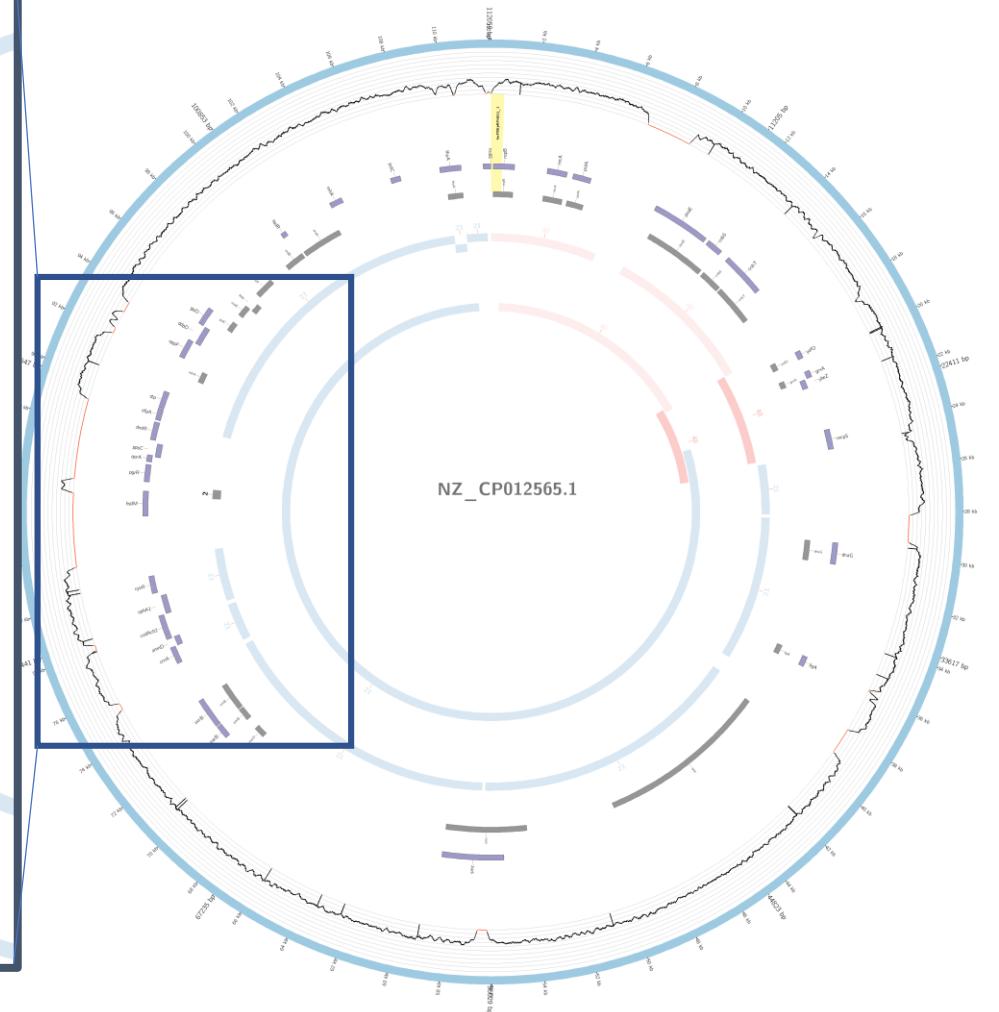
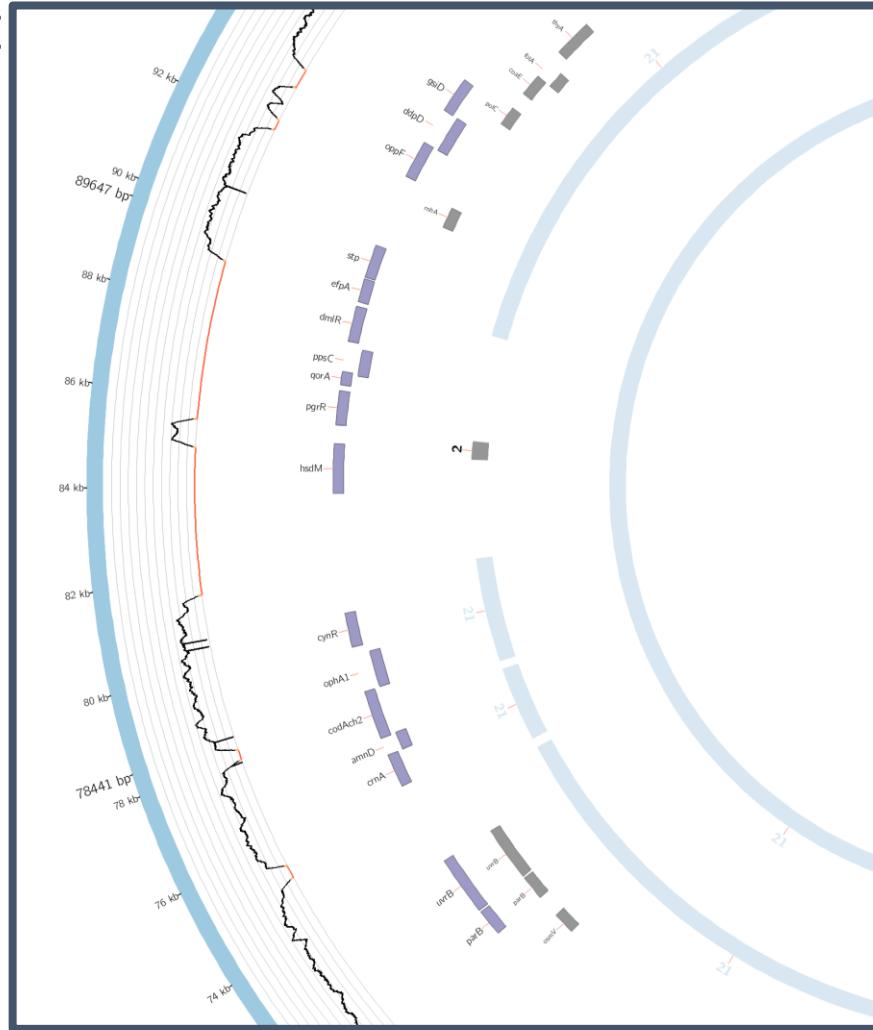
- Length dependent
 - Contigs with alignment > 50%
 - No duplicated
 - Adapted coords.
 - Functions:
 - Additional info
 - Gap filler
 - Test duplications
 - Plasmid proofreading
 - Curation



PlasmidID result

Complete contig track

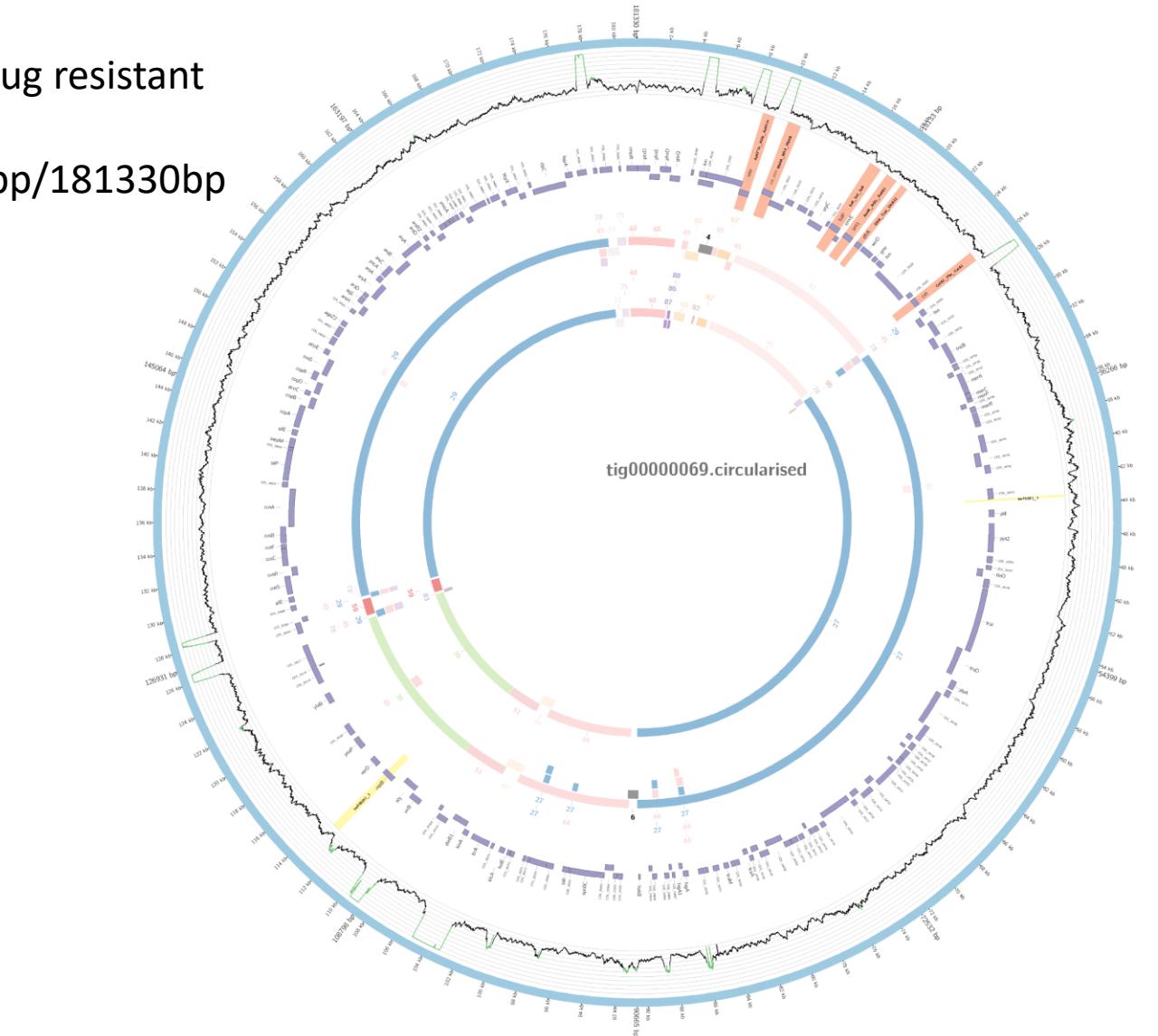
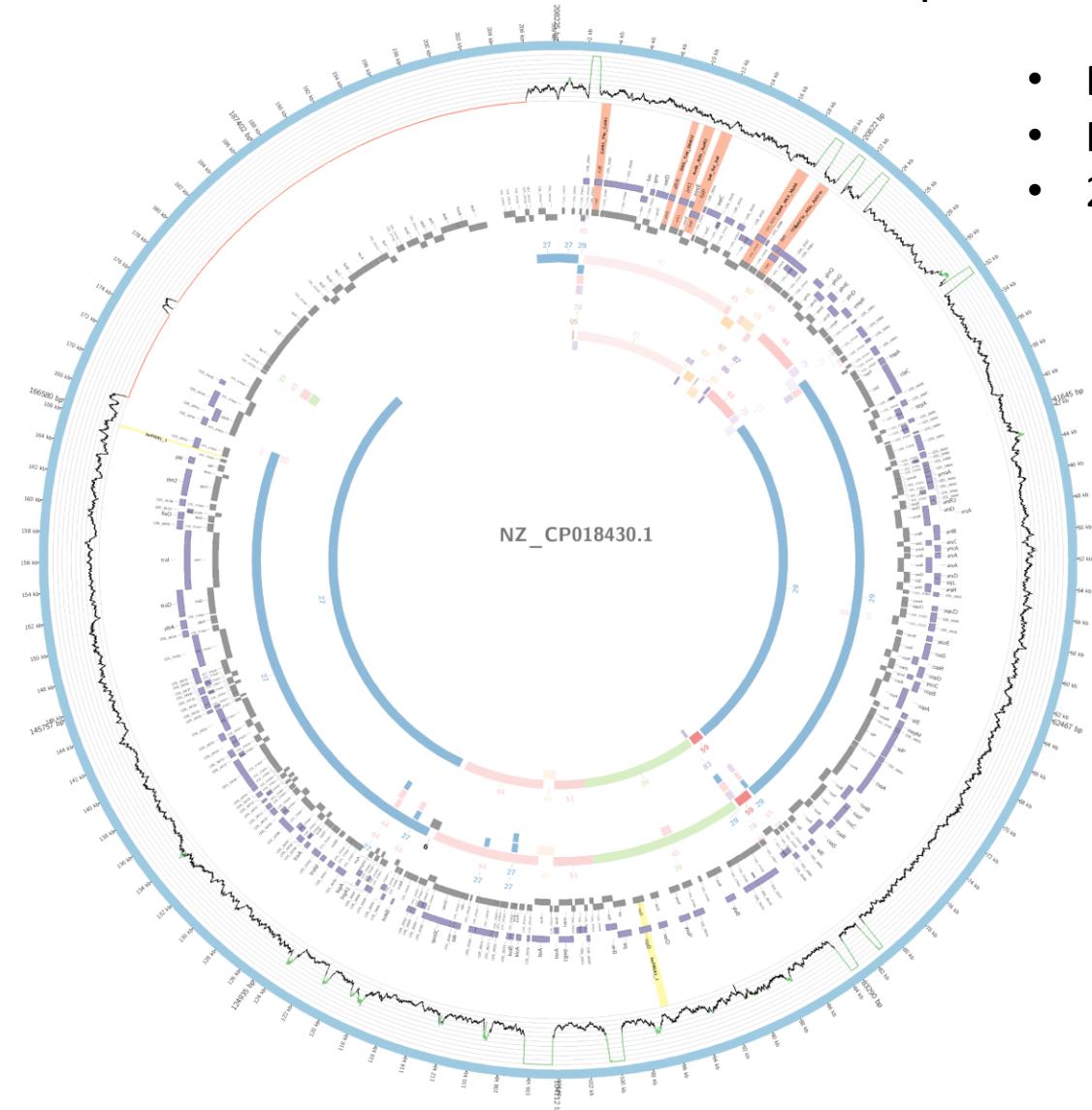
- Length dependent
 - Contigs with alignment > 50%
- No duplicated
- Adapted coords.
- Functions:
 - Additional info
 - **Gap filler**
 - Test duplications
 - Plasmid proofreading
 - Curation



PlasmidID validation

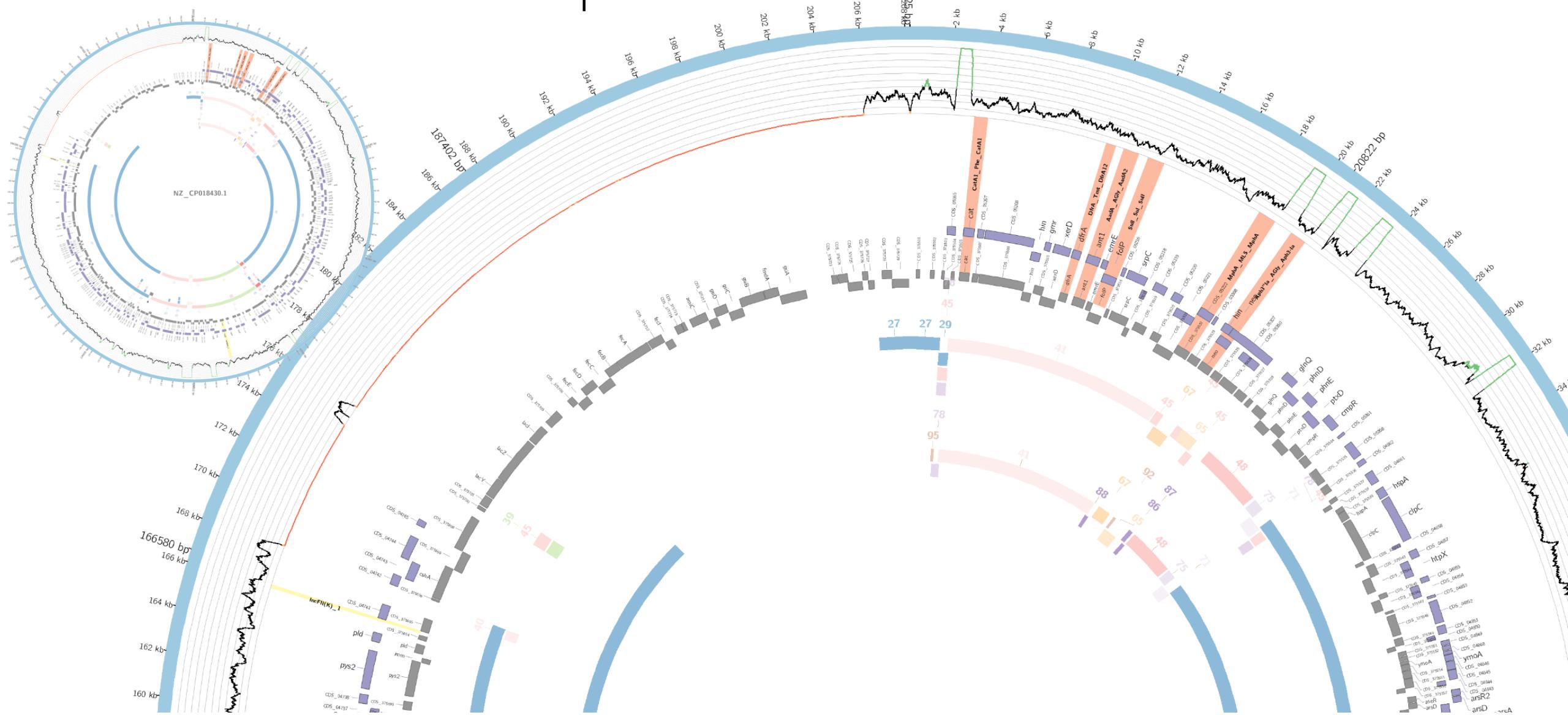
Compare with PacBio

- Multi-drug resistant
 - IncFIB
 - 208221bp/181330bp



PlasmidID validation

Compare with PacBio



PlasmidID web

Comming soon

- Submit R1 and R2
- Optional assembled contigs
- Own database to annotate
- github.com/BU-ISCIII/plasmidID

HOME PIDWEB FAQS CONTACT

The screenshot shows the PlasmidID web beta interface. On the left, there's a sidebar with the PlasmidID logo and a 'Run pidWeb' section containing buttons for 'Add files R1...', 'Add files R2...', and 'Start upload'. Below this is a 'Notes' section with instructions about file uploads. The main content area features a large circular genome map with various colored segments and labels like 'NC_000011'. To the right of the map is a 'PlasmidID' heading and a descriptive text block. Further down are sections for 'Easy installation', 'Plasmid identification', 'Annotation', and 'Reporting', each with a brief description and icons. At the bottom are links for 'Read the docs', 'Download v1.0', and 'PlasmidID on GitHub'.

HOME PIDWEB FAQS CONTACT

PlasmidID

PlasmidID is designed to be a mapping-based contig assisted tool for plasmid identification, annotation, classification and representation.

This tool offers an easy-to-interpret image that answers most of questions a researcher can retrieve from WGS-HTS about plasmids in their sample.

Easy installation Implemented in a version-independent language programming and dependency installation is assisted

Plasmid identification Only tool that is mapping-based for plasmid identification.

Annotation Resistance genes, virulence factors and functional annotation is performed.

Reporting Circos framework is used for creating easy-to-interpret images for researchers.

Read the docs

Download v1.0

PlasmidID on GitHub

Logos

Connect

bioinformatica@isciii.es

Links

- Contact
- Getting started
- FAQs



Conclusions

- PlasmidID **identifies the most likely plasmids** to be present in a sample
- Using their sequence is able to **reconstruct plasmids** using **sample reads**
 - Not necessarily identical
- PlasmidID generates a **summary table** that outlines isolates relationship
- PlasmidID works for **every species and plasmid size**
- Reconstruction method **adapts to plasmid modular structure**
- Plasmid databases needs **curation** process and guidelines for submission



Future tasks

- Extract **full sequence** for every putative sequence
 - Check with scaffolders
- Improve **plasmid database**
- Adapt parameters for metaplasmidic analyses
- Adapt pipeline for SMRT sequences
- Compare all samples with the same plasmid
- Release web service

Thanks



> BU-ISCIII

NCBI

MÉDiterranée
INFECTION



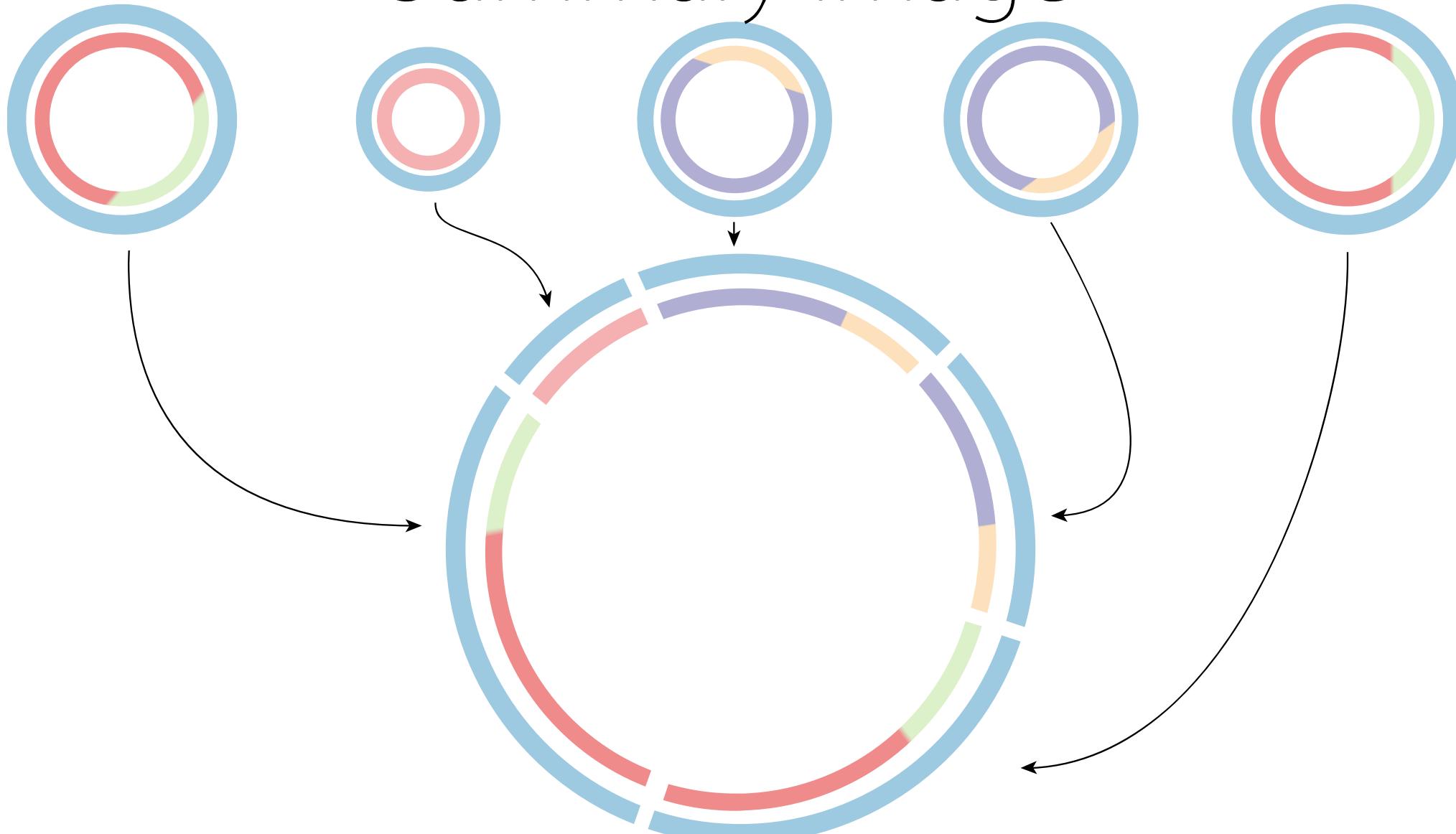
Center for Genomic
Epidemiology

stack overflow

IWBBIO 2018
INTERNATIONAL WORK-CONFERENCE ON
BIOINFORMATICS AND BIOMEDICAL ENGINEERING

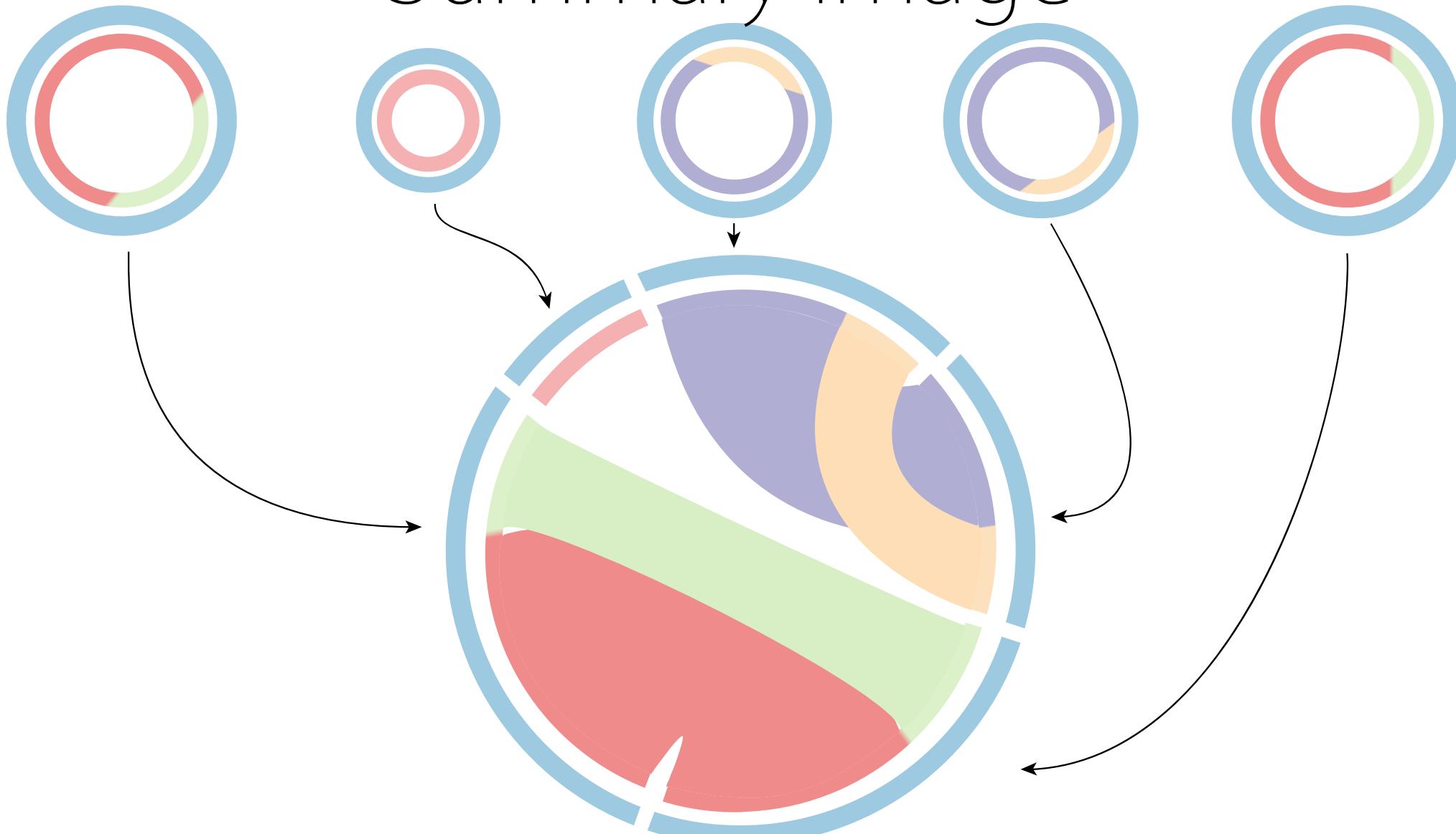
PlasmidID results

Summary image



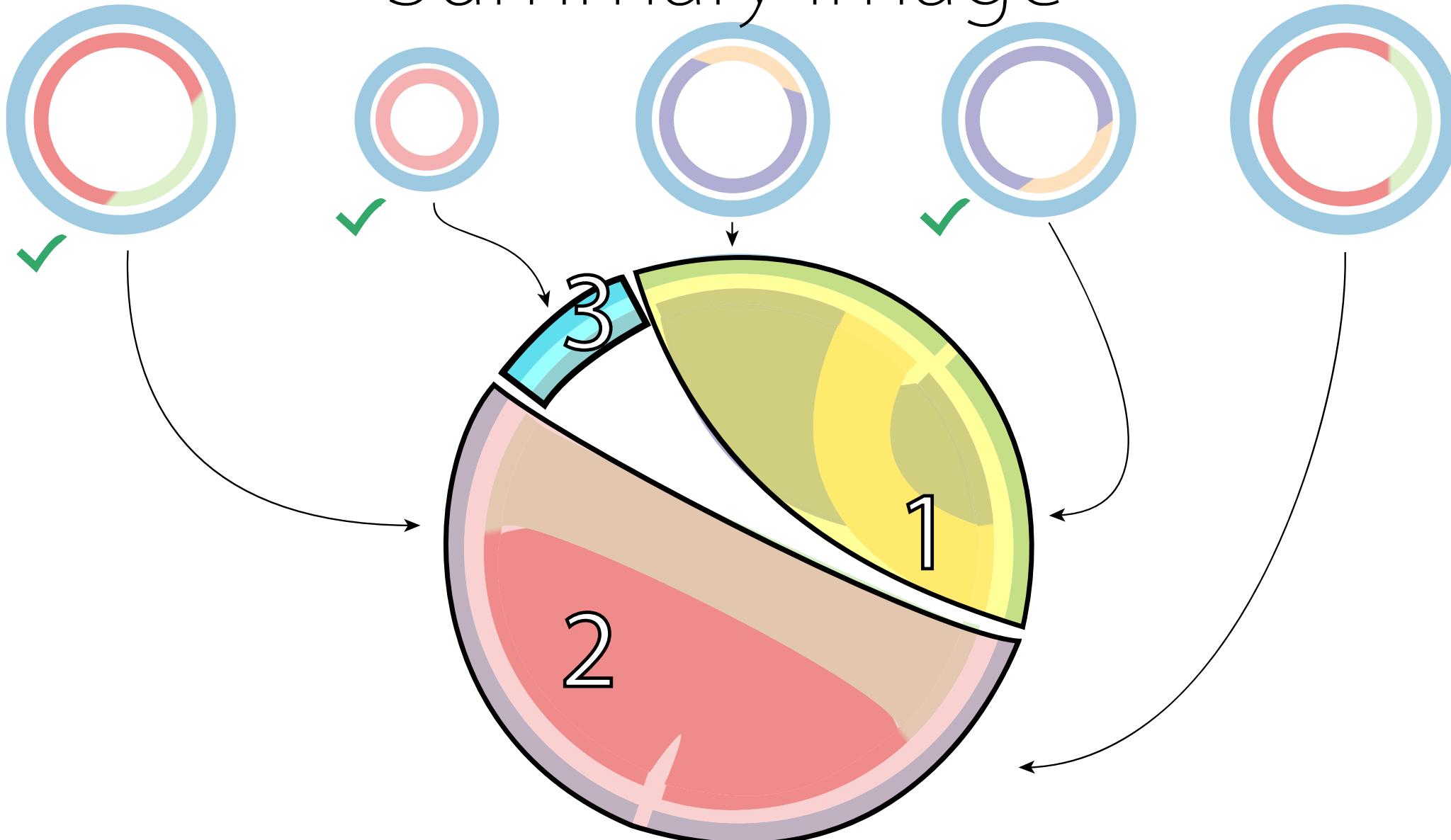
PlasmidID results

Summary image



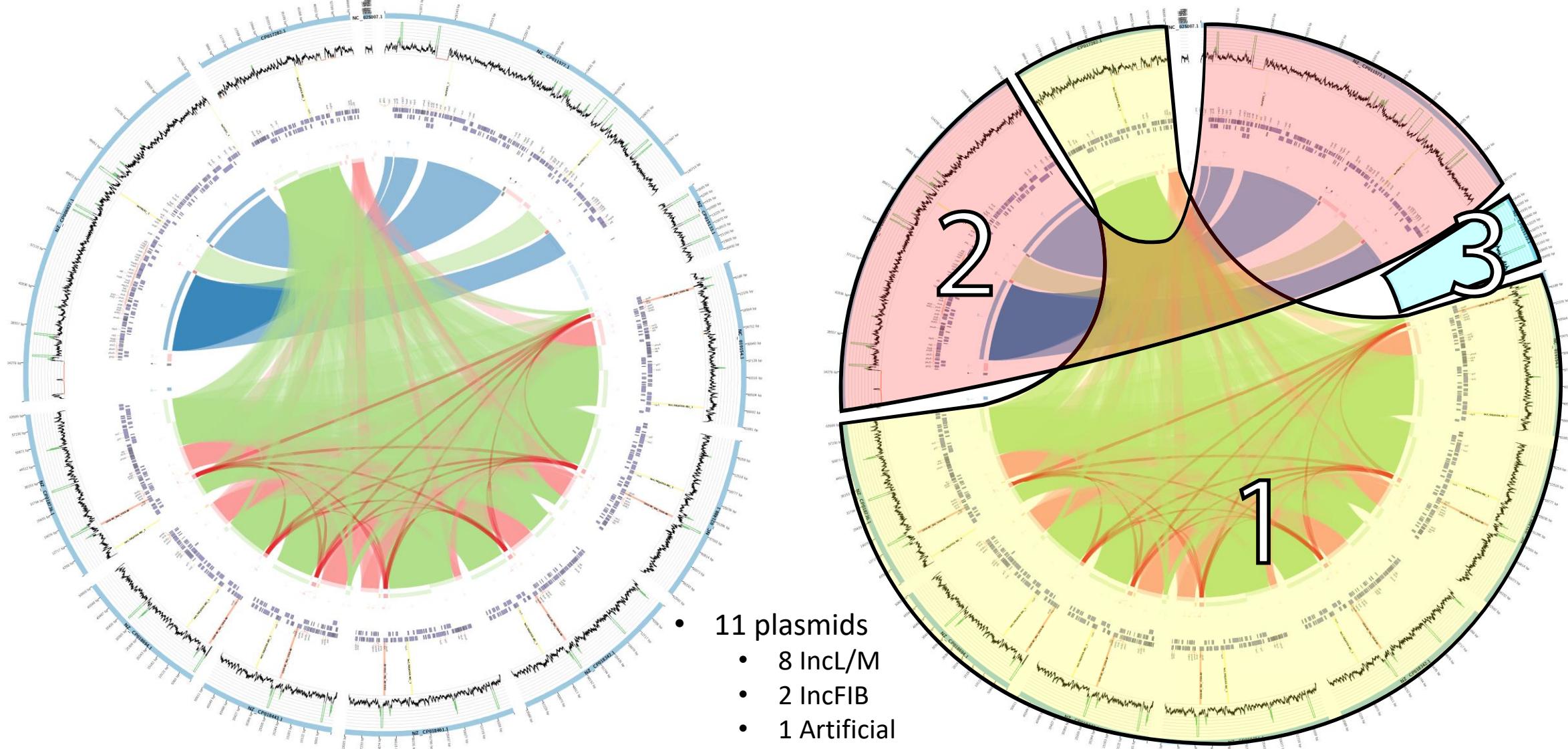
PlasmidID results

Summary image



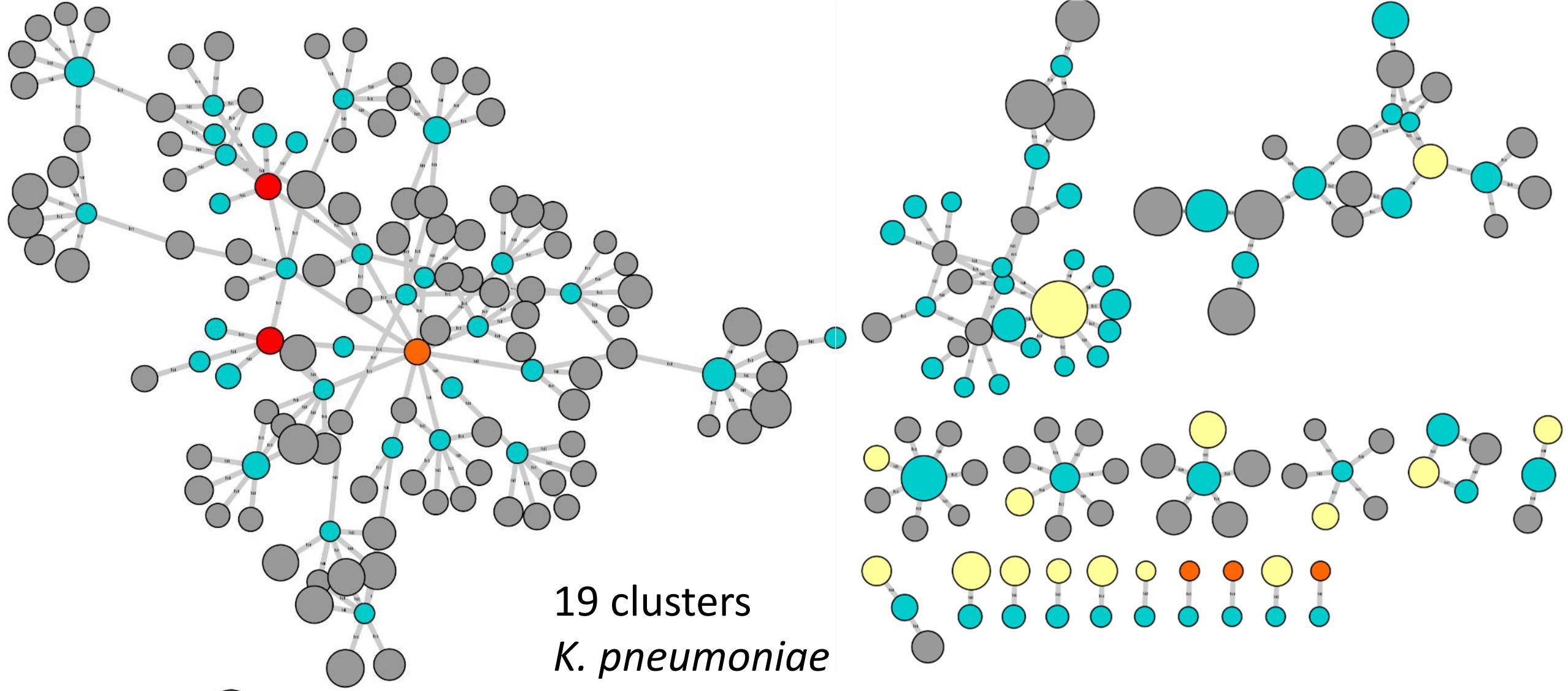
PlasmidID results

Summary image



PLACNET results

On *in silico* sample



PlasmidID results

Summary table

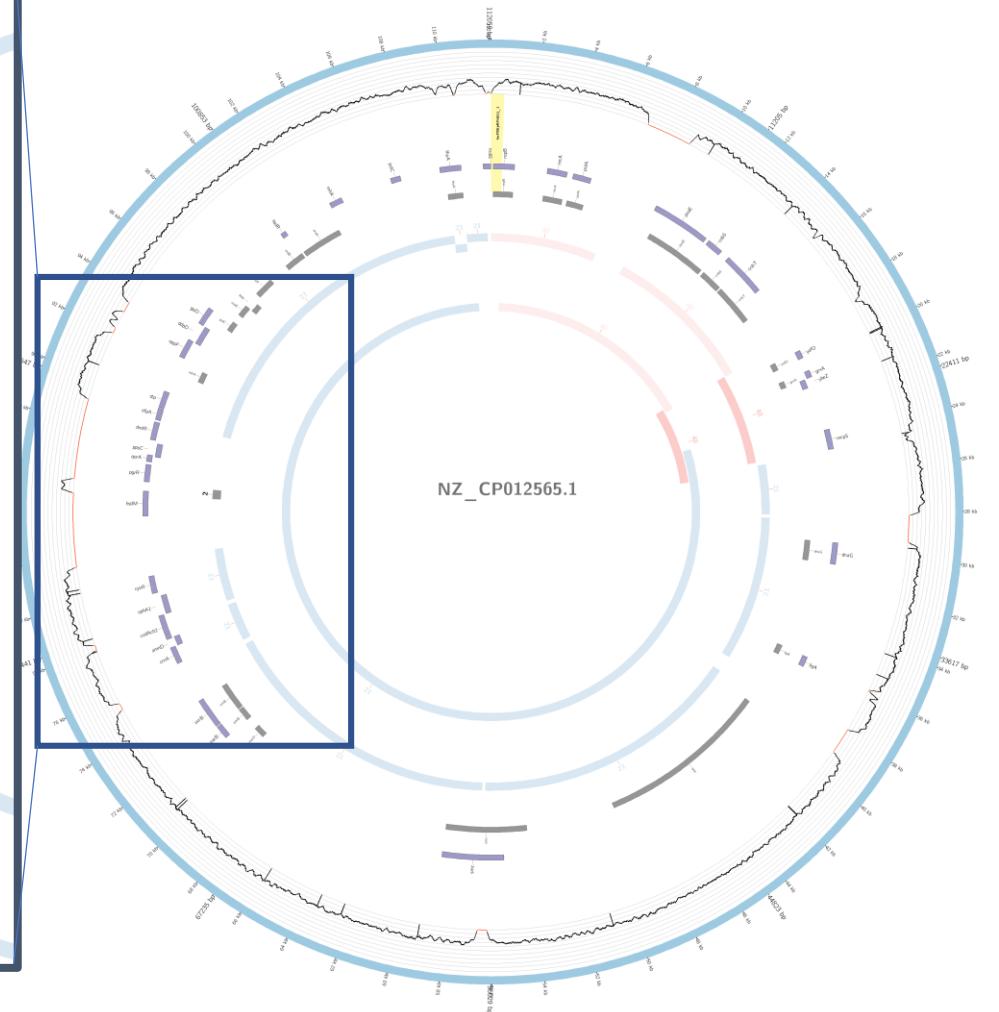
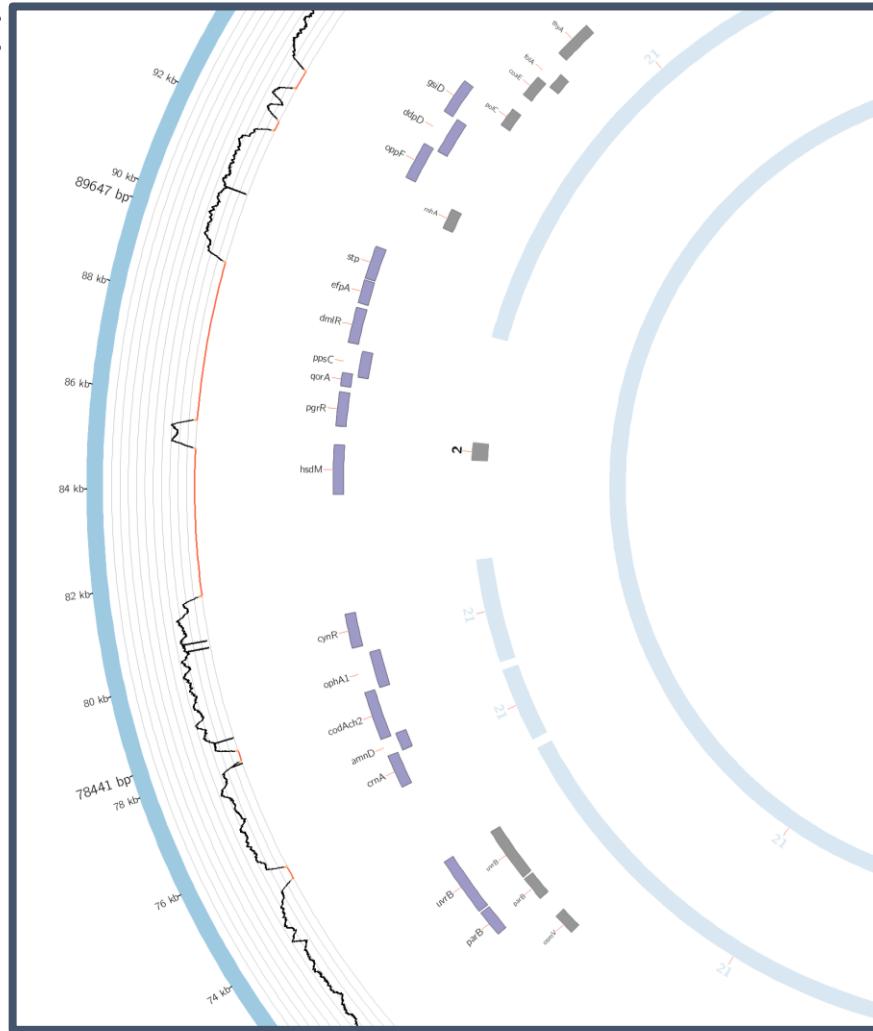


- NDM producing
- 28 *k. pneumoniae* isolates

PlasmidID result

Complete contig track

- Length dependent
 - Contigs with alignment > 50%
- No duplicated
- Adapted coords.
- Functions:
 - Additional info
 - **Gap filler**
 - Test duplications
 - Plasmid proofreading
 - Curation



PlasmidID validation

Compare with PacBio

