BIOMEDICAL CENTER

FACULTY OF MEDICINE IN PILSEN
CHARLES UNIVERSITY IN PRAGUE

NGS and Plasmids 101

Asst. Prof. Ibrahim Bitar, Ph.D.

Department of Microbiology, Faculty of Medicine and University Hospital in Plzen,

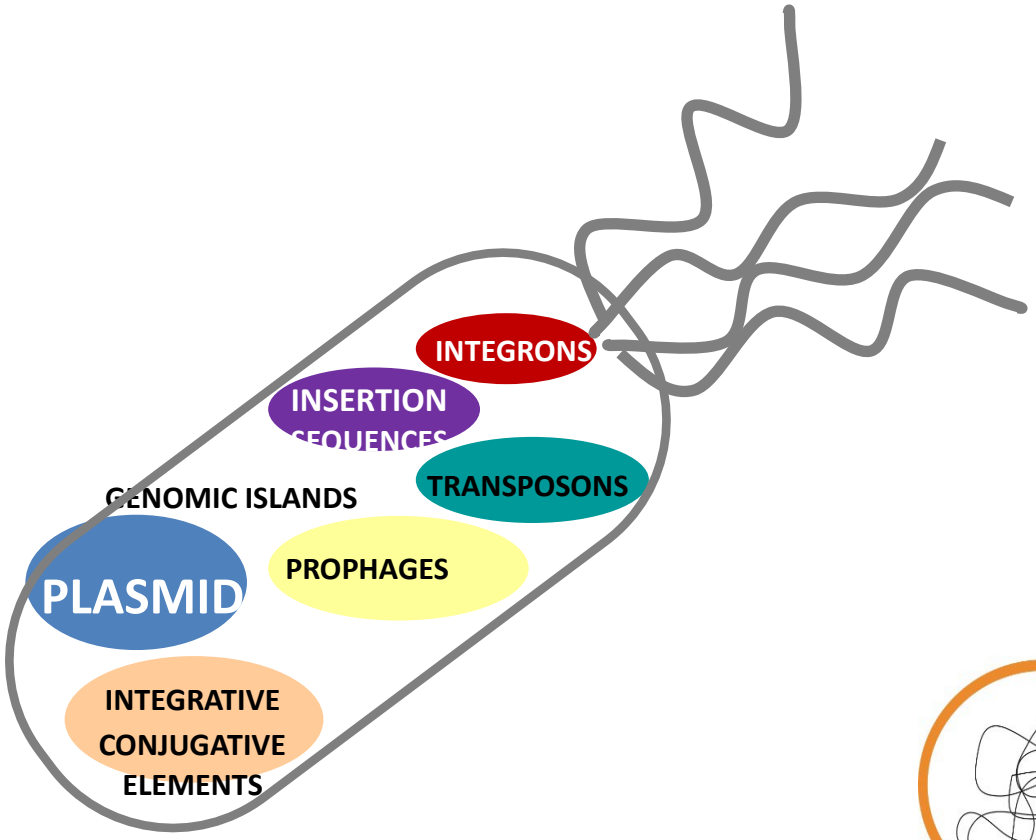
Charles University in Prague,

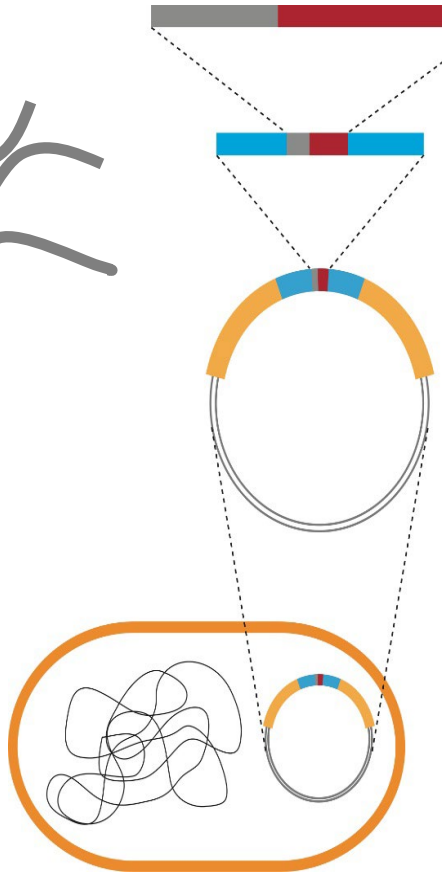
Plzen, Czech Republic

Outline

- Introduction:
  - Plasmids' structure
  - Short reads
- Mainstream Method:
  - In-silico
- Pilsen's Method:
  - Transformation
  - Scaffolding
  - S1 PFGE
  - PCR-Gap closing Technique
- Long reads
- Conclusion

Mobile genetic elements in transmission of MDR



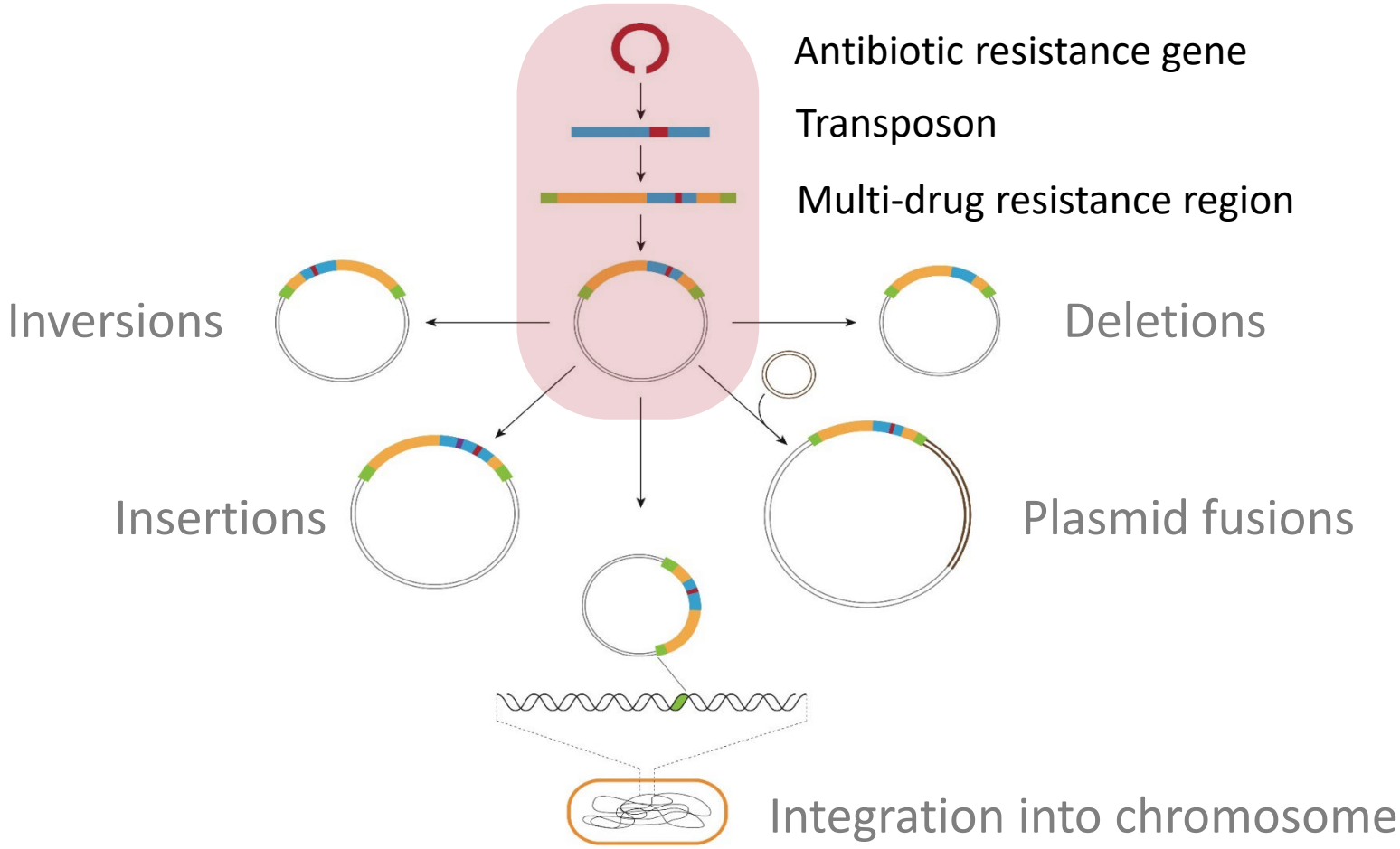


ARG+insertion sequence

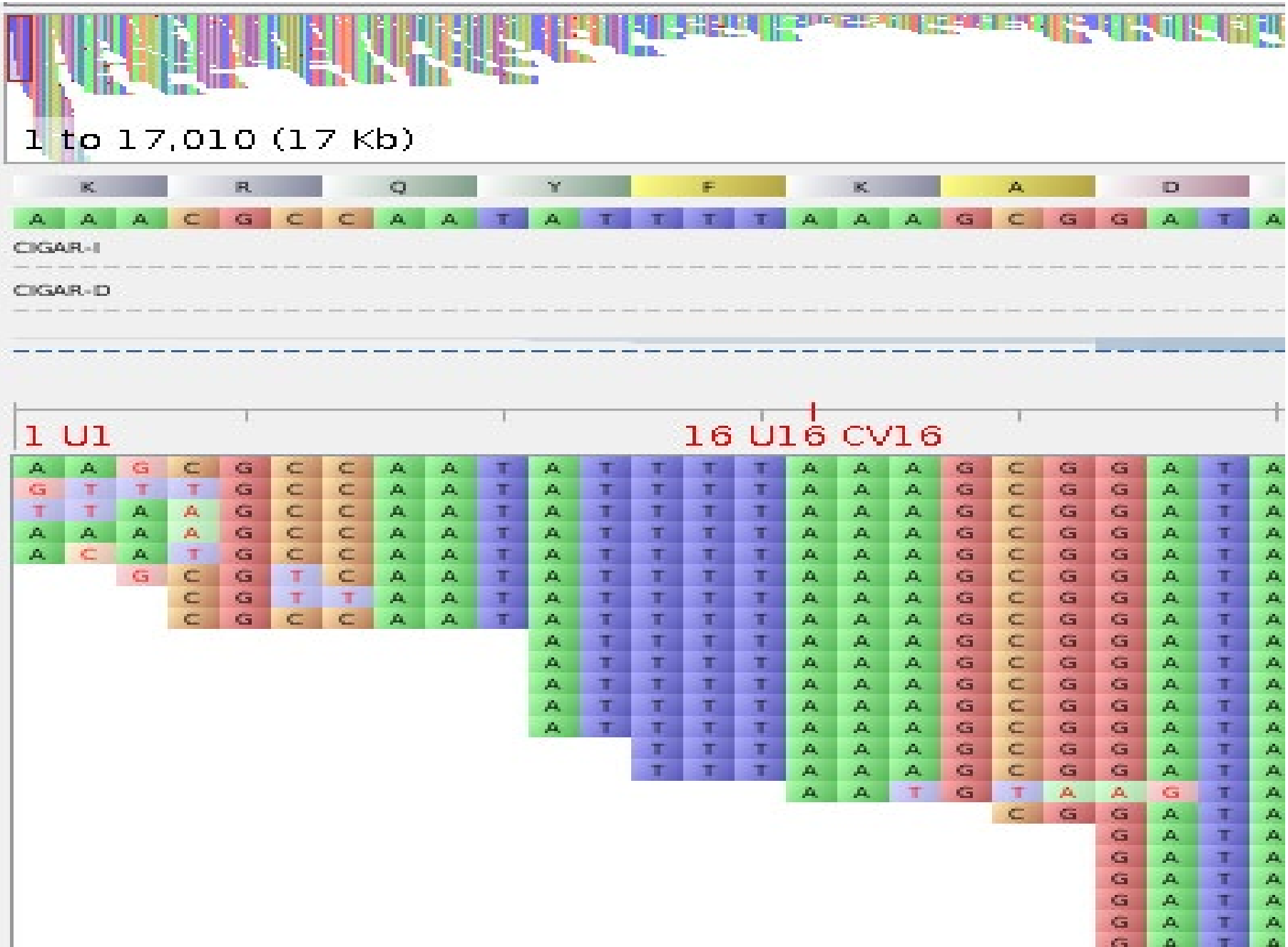
ARG/IS inside transposon

ARG element integration into MDR region

Mosaic structure and evolution of MDR plasmids

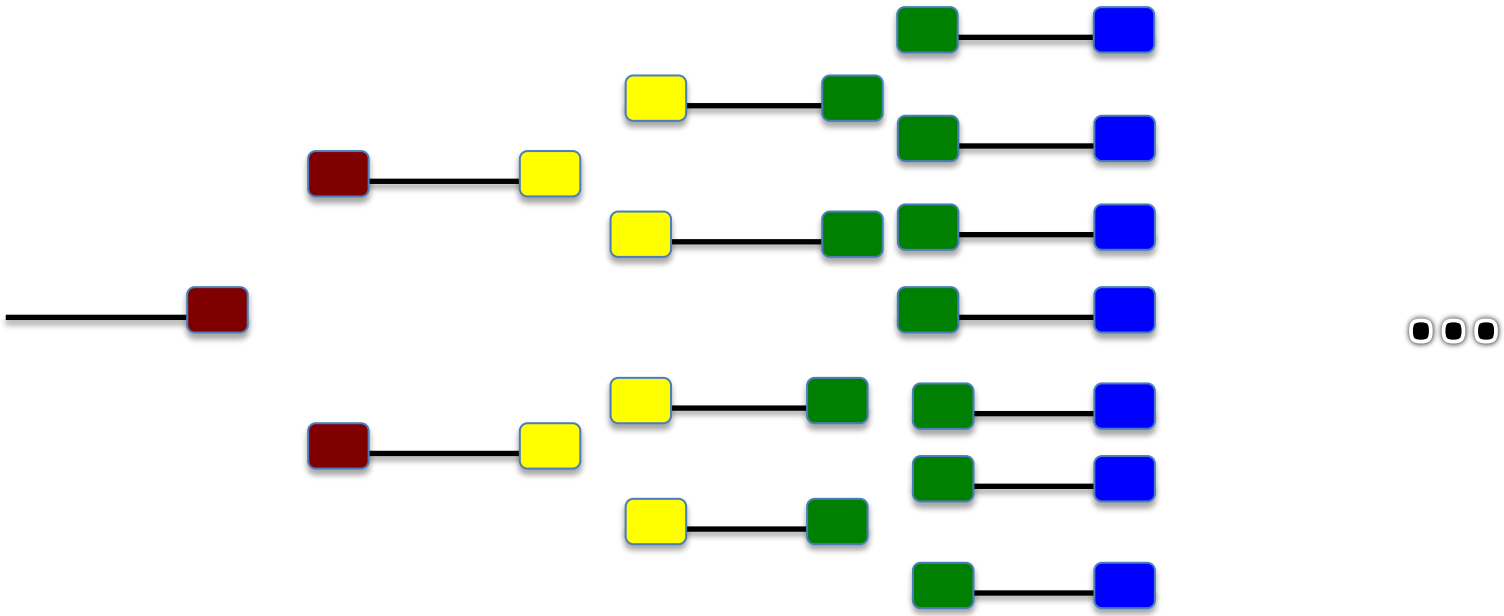


Short Reads

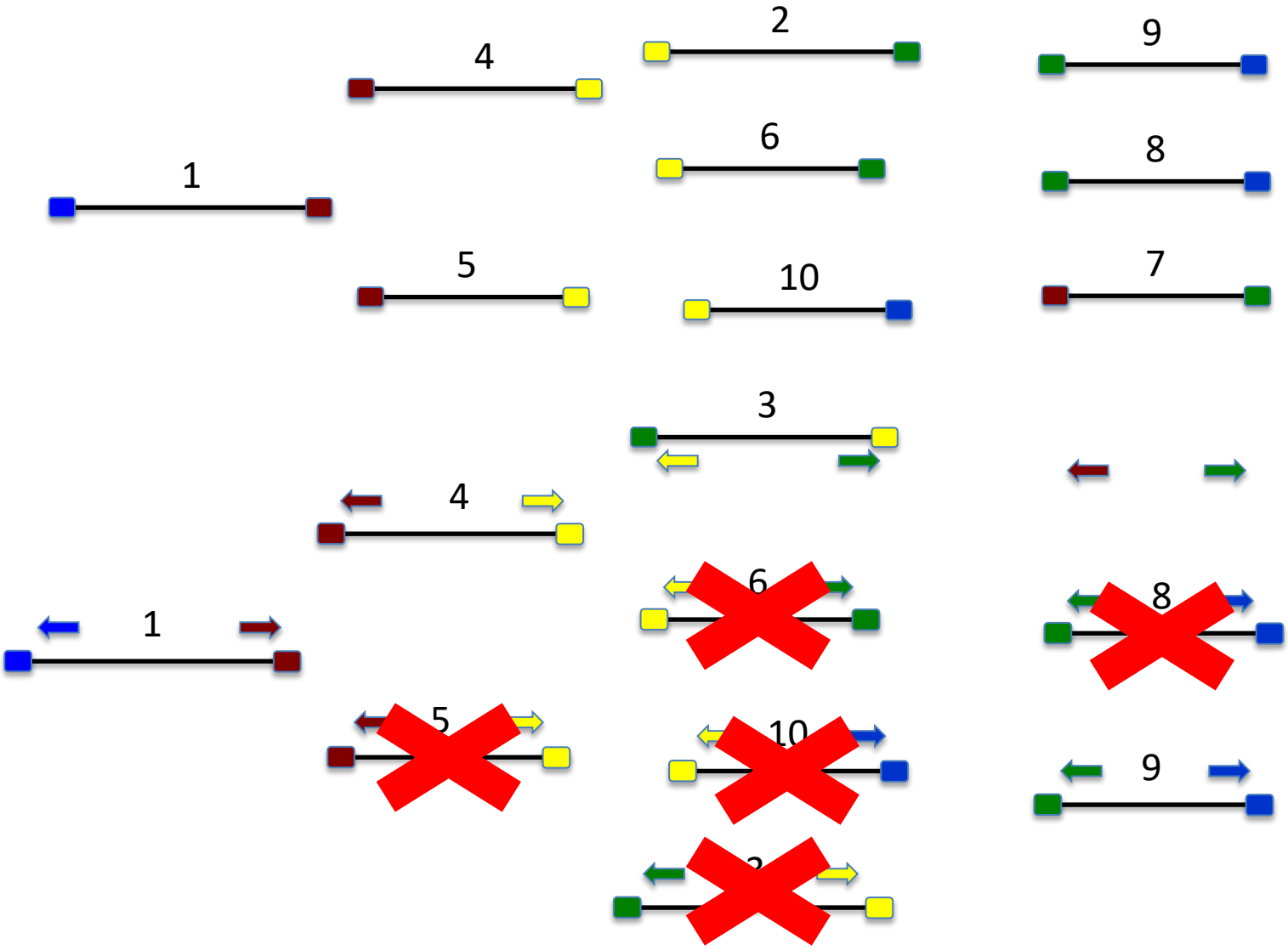


The dilemma !!

- Illumina's short reads:
  1. Repeated elements in the genome and in the plasmid will lead the assembler to break the contig.
  2. Many contigs generated for one isolate.
  3. Linking contigs to close the plasmid will be difficult.



Short Reads



Short Reads

Mainstream Approach

- Center for Genomic Epidemiology

Home | Services | Instructions | Output | Overview of genes | Article abstract

- **ResFinder-4.1 Server - Results**

Polymyxin									
Resistance gene	Identity	Alignment Length/Gene Length	Position in reference	Contig or Depth	Position in contig	Phenotype	PMID	Accession no.	Notes
mcr-1.1	100.0	1626/1626	1..1626	NODE_49_length_30715_cov_40.324670	17564..19189	colistin	26603172	KP347127	

Tetracycline									
Resistance	Identity	Alignment Length/Gene	Position in	Contig or	Position in	Phenotype	PMID	Accession no.	Notes

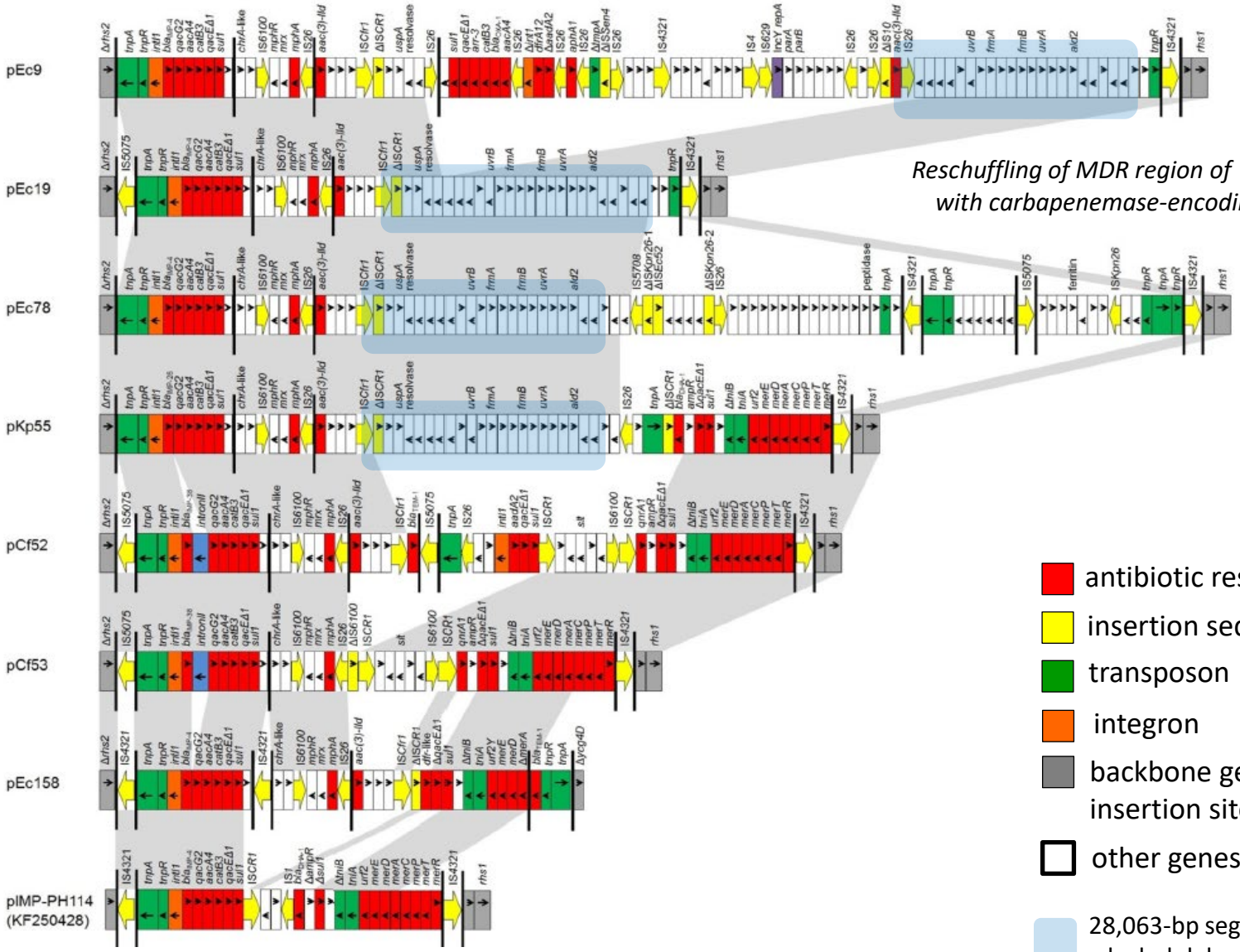
PlasmidFinder-2.0 Server - Results

Organism(s): *Enterobacteriaceae*

Enterobacteriaceae,Acinetobacter baumannii						
Plasmid	Identity	Query / Template length	Contig	Position in contig	Note	Accession number
IncFIB(AP001918)	98.24	683 / 682	NODE_58_length_29447_cov_4.681685	26045..26727		AP001918
IncFIC(FII)	96.62	355 / 499	NODE_161_length_4354_cov_3.615567	1..355		AP001918
IncX4	99.47	375 / 374	NODE_175_length_2632_cov_1.526547	2034..2407		CP002895

Disadvantages of Mainstream Approach

Plasticity of plasmid MDR regions



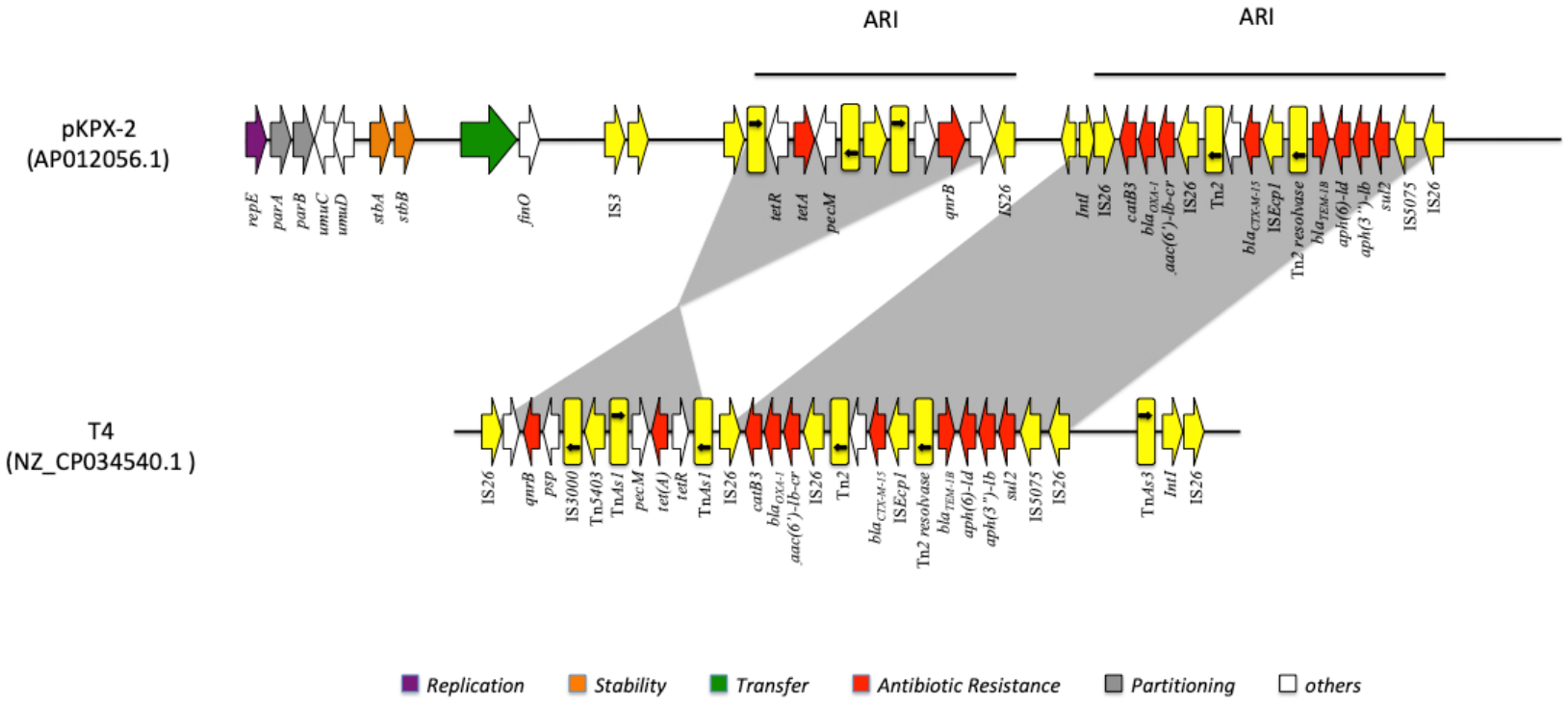
Reschuffling of MDR region of IncA/C_2 plasmids with carbapenemase-encoding gene bla_{IMP-4}

- antibiotic resistance
- insertion sequence
- transposon
- integron
- backbone genes/ insertion site of MDR region
- other genes

28,063-bp segment (sulfate transporter, alcohol dehydrogenase, ABC transporters)

Papagiannitsis et al. 2017 Antimicrob Agents Chemother.

Disadvantages of Mainstream Approach



Bitar et al. 2019 JGAR.

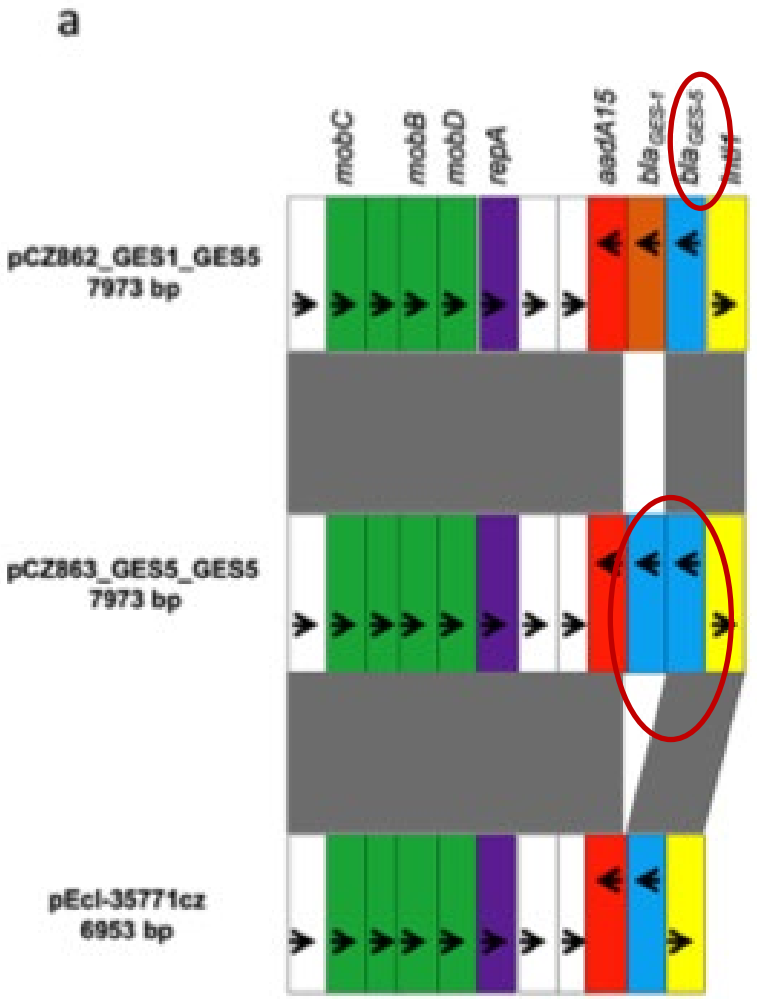
Disadvantages of Mainstream Approach

TABLE 2 | WGS data of *Enterobacterales*, co-producing NDM- and OXA-48-like carbapenemases, isolates recovered from Czech hospitals.

Isolate	ST	Replicons	Size	Plasmid name	Inc group	Carbapenemase encoding-genes	Other resistance genes	GenBank accession no.
Kpn47733	147	Chromosome	5401559 bp	–	–	–	*bla*$_{CTX-M-15}$, *bla*$_{SHV-11}$, *oqxA*, *oqxB*, *fosA*	CP050360
		Plasmid	103085 bp	p47733_NDM-5	IncFII	*bla*$_{NDM-5}$	*aadA2*, *rmtB*, *bla*$_{TEM-1B}$, *erm*(B), *mph*(A), *sul1*, *dfrA12*	CP050367
		Plasmid	6812 bp	p47733_OXA-181	ColKP3	*bla*$_{OXA-181}$	–	CP050368
		Plasmid	119981 bp	p47733_CTX-M-15	R	–	*rmtF*, *bla*$_{CTX-M-15}$, *mph*(A), *catA2*, *aac*(6')-*Ib-cr*, *qnrB1*, *drfA12*, *dfrA14*	CP050364
		Plasmid	107451 bp	p47733_ARR-2	IncFII	–	*rmtF*, *aac*(6')-*Ib-cr*, *arr-2*	CP050361
		Plasmid	115360 bp	p47733_IncFIB	IncFIB	–	–	CP050365
		Plasmid	2101 bp	p47733_Col_BS512	Col	–	–	CP050362
		Plasmid	1546 bp	p47733_Col_MG828	Col	–	–	CP050363
		Plasmid	2056 bp	p47733_Col_PVC	Col	–	–	CP050370
		Plasmid	4715 bp	p47733_S	NT	–	–	CP050369
		Plasmid	55119 bp	p47733_L	NT	–	–	CP050366
Kpn50595	11	Chromosome	5339674 bp	–	–	–	*bla*$_{SHV-11}$, *fosA*, *oqxA*, *oqxB*	CP050371
		Plasmid	193462 bp	p50595_NDM-1	IncFIB-IncFII	*bla*$_{NDM-1}$	*aac*(3)-*IIa*, *aac*(6')-*Ib-cr*, *rmtF*, *bla*$_{CTX-M-15}$, *catB*, *arr-2*	CP050374
		Plasmid	51140 bp	p50595_OXA-181	IncX3-ColKP3	*bla*$_{OXA-181}$	*qnrS1*	CP050375
		Plasmid	76387 bp	p50595_ERM	IncFII	–	*erm*(B), *mph*(A)	CP050372
		Plasmid	127925 bp	p50595	IncFII-IncFIB	–	–	CP050373

Chudejova et al. 2020 Frontiers in Microbiology.

Disadvantages of Mainstream Approach



Finianos et al. 2022 JGAR.

Disadvantages of Mainstream Approach

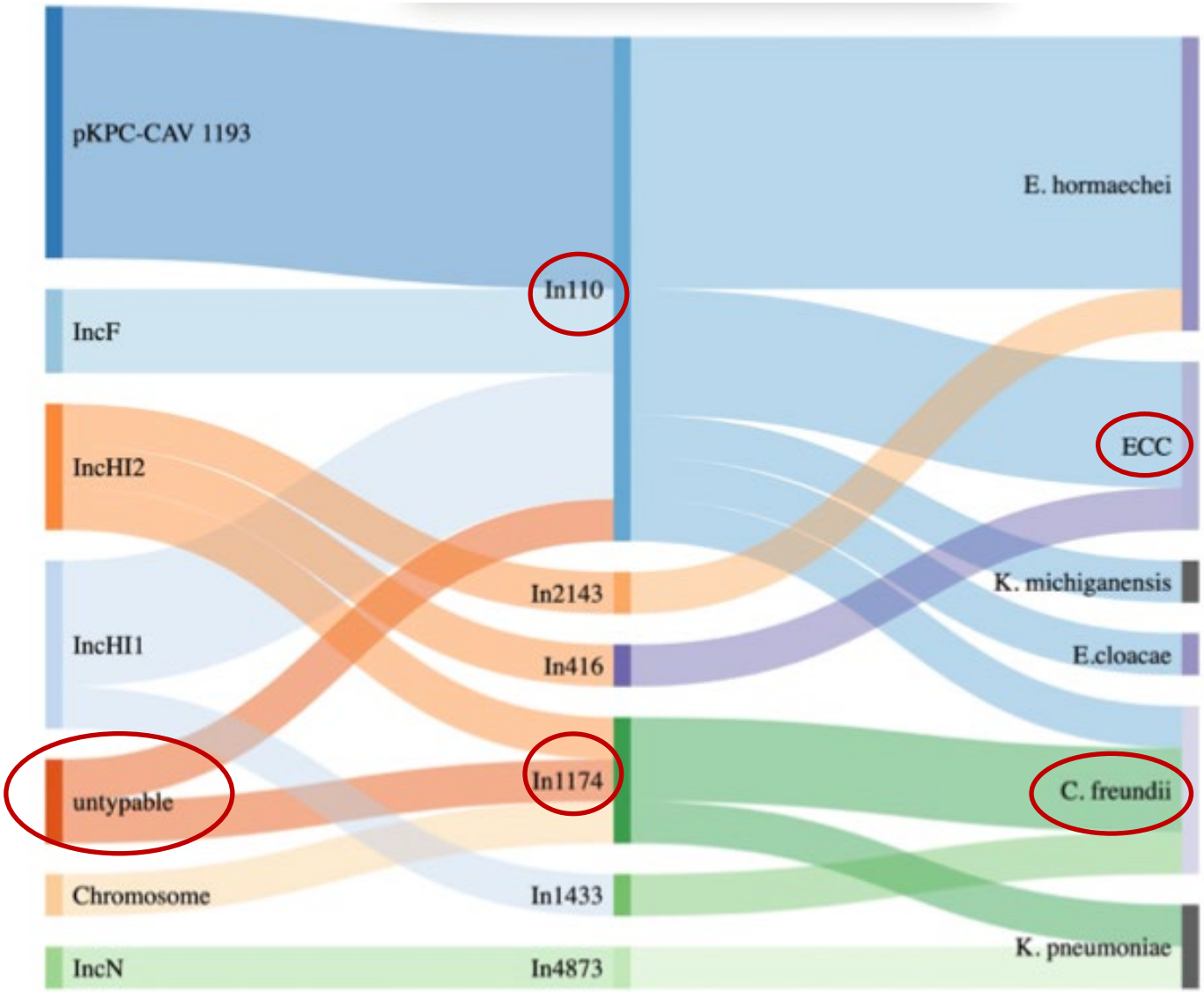
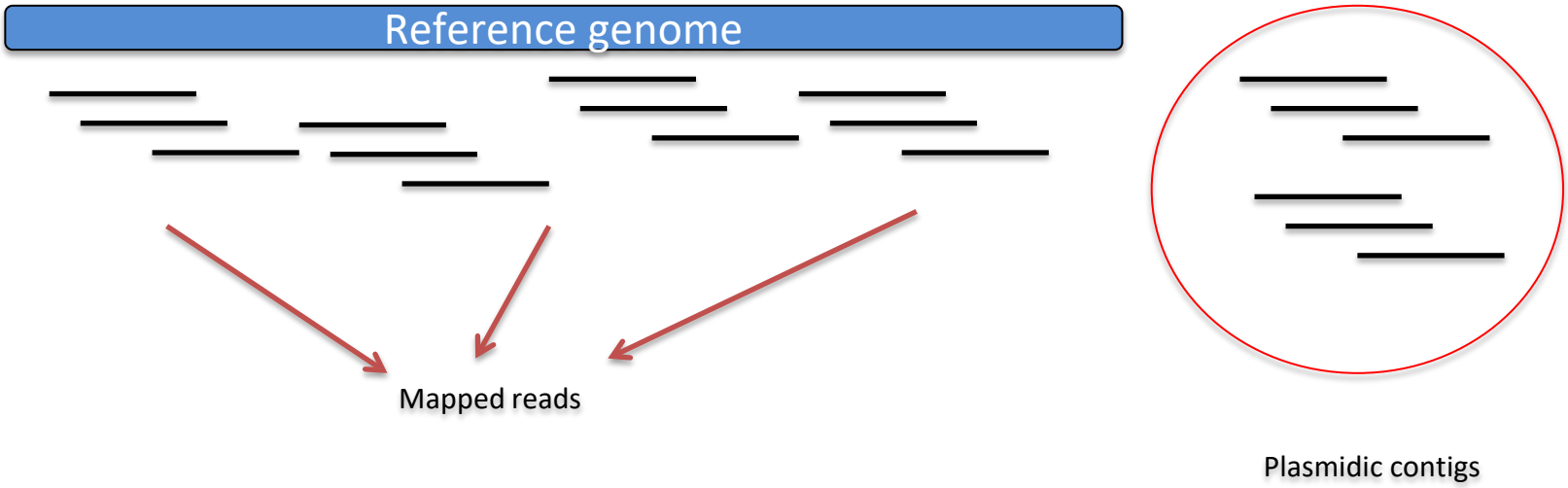


FIGURE 3
Sankey diagram showing the distribution of different integrons linking to the plasmid types (left) and to the bacterial species detected (right).

Bitar et al. 2023 Frontiers in Microbiology.

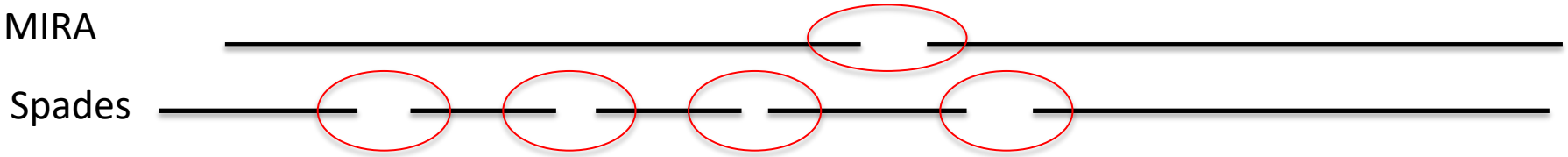
Pilsen Approach

- In Plzen:
  1. Transformation.
  2. S1 PFGE
  3. Sequencing of recipient and transformant.
  4. Assemble Recipient
  5. Map reads from transformant against reference.
  6. Extract unmapped.
  7. Assemble unmapped.
  8. S1 PFGE



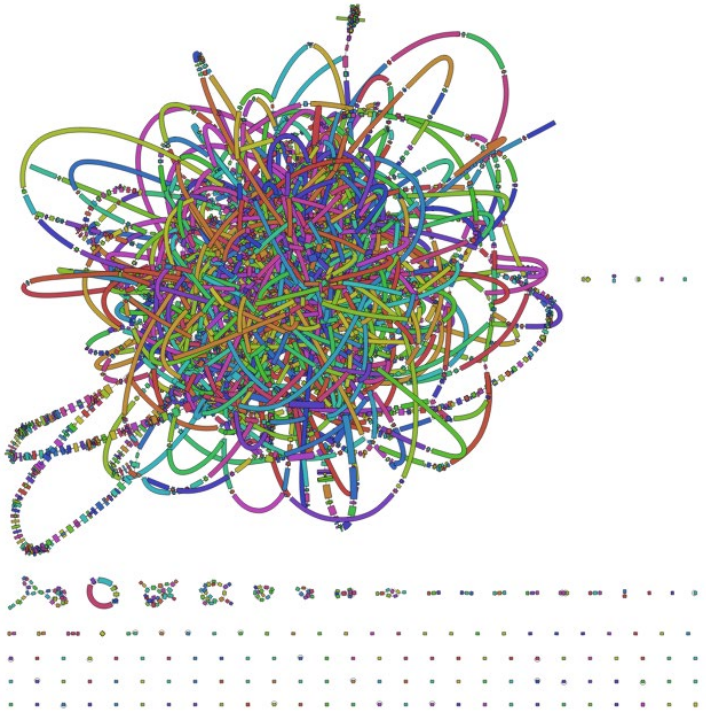
Cutting the Gordian Knot

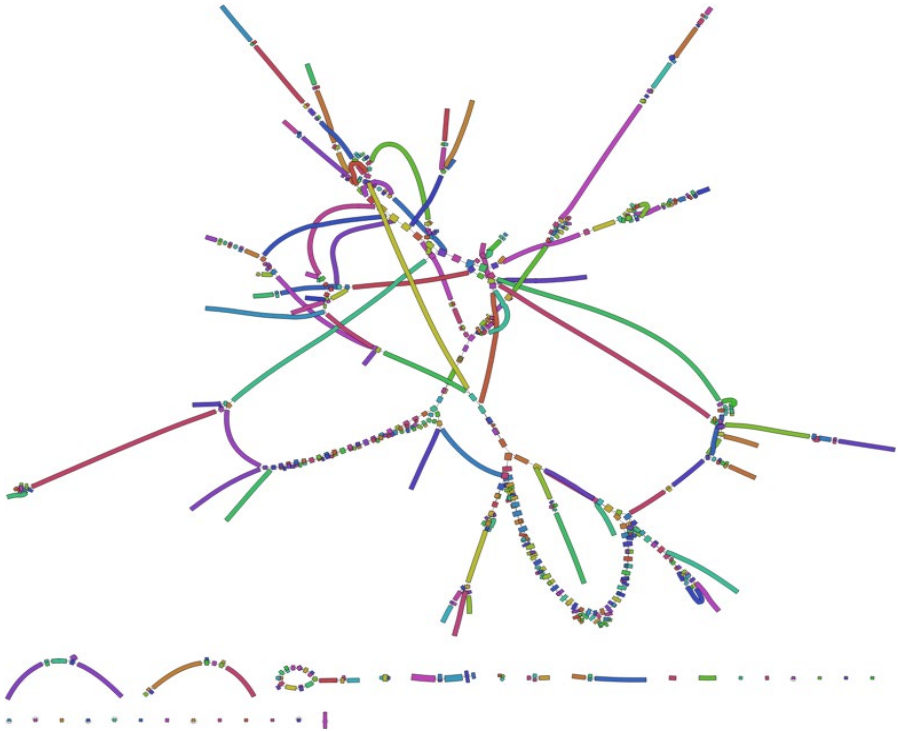
- Scaffolding:
  1. Assembling the genome using 2 assemblers that have different pipelines.
  2. Using software that will visually help solve some knots.
  3. Gap-closing using PCR.



Cutting the Gordian Knot

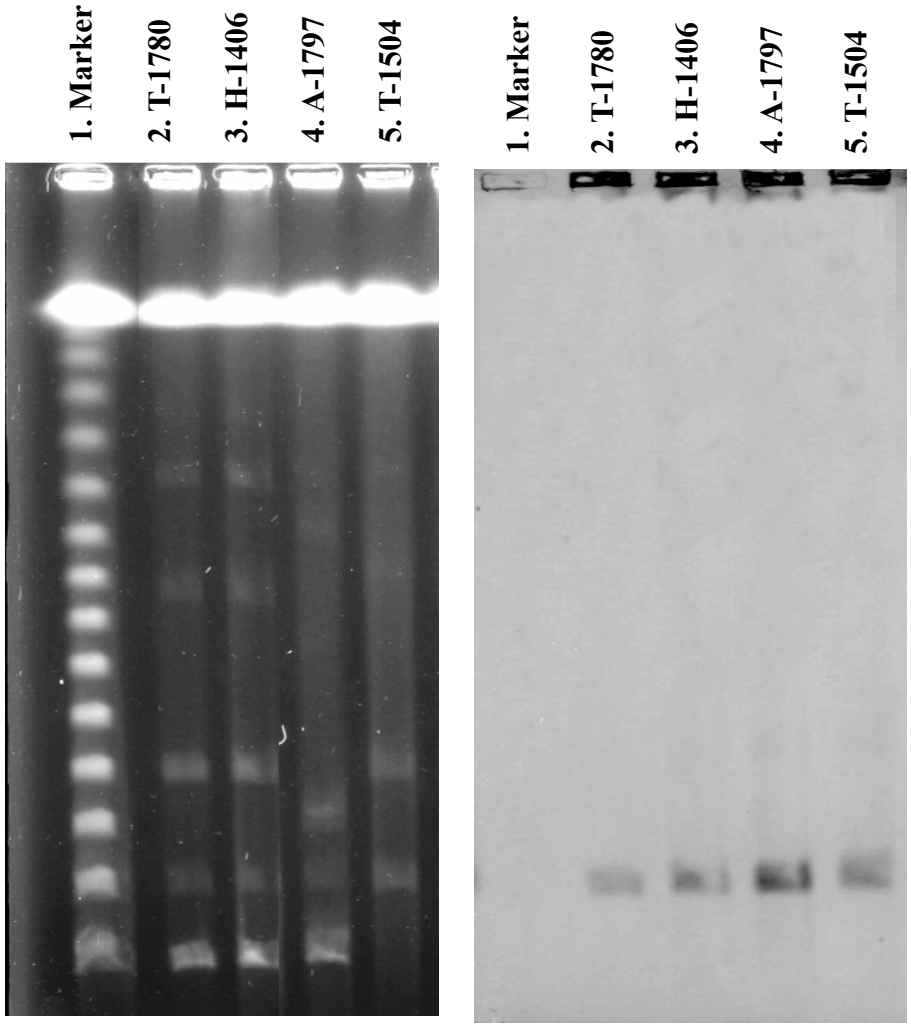
- Scaffolding:
  1. Assembling the genome using 2 assemblers that have different pipelines.
  2. Using software that will visually help solve some knots.
  3. Gap-closing using PCR.





Cutting the Gordian Knot

- S1 analysis of *bla*$_{KPC-2}$-positive *K. pneumoniae* isolates

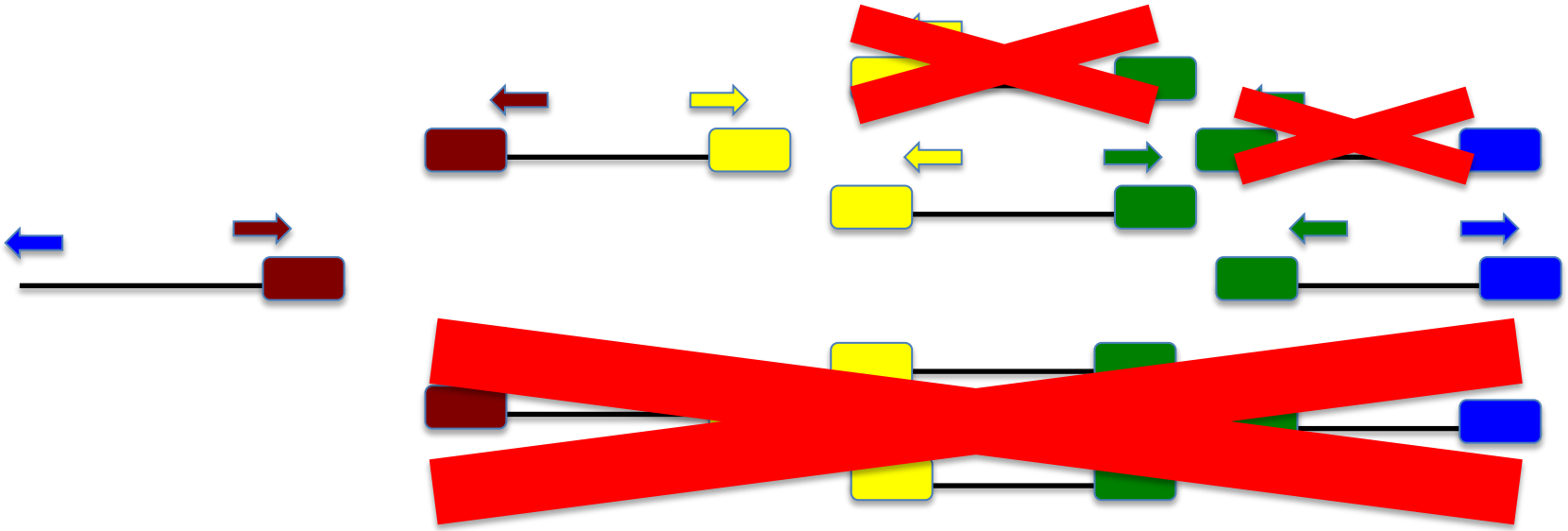


- After the PFGE separation, the DNA is transferred to nitrocyttarine membranes and hybridized with double-labeled bla-specific probes.

→Therefore, the blaKPC gene is located in a ~100-kb patch.

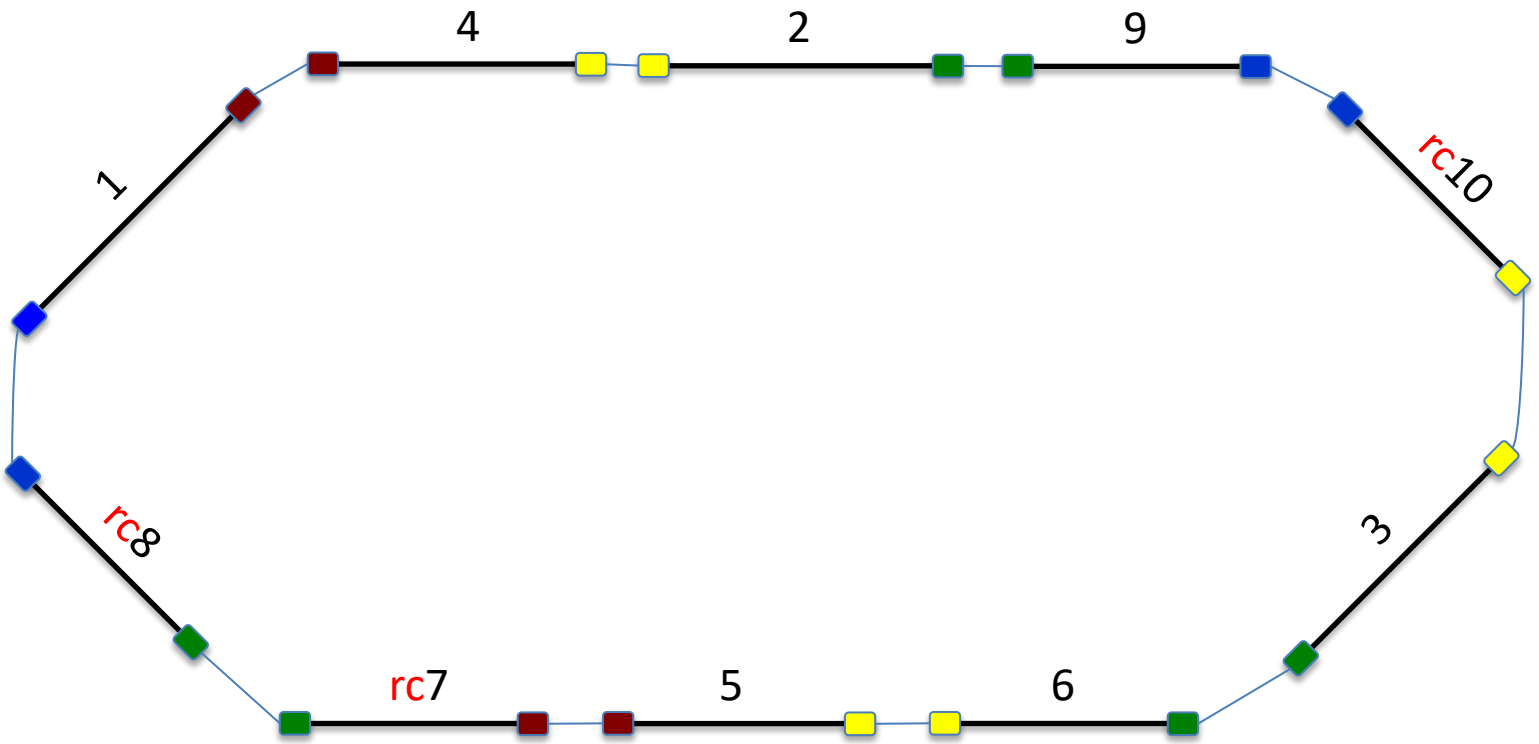
Cutting the Gordian Knot

- Scaffolding:
  1. Assembling the genome using 2 assemblers that have different pipelines.
  2. Using software that will visually help solve some knots.
  3. Gap-closing using PCR.



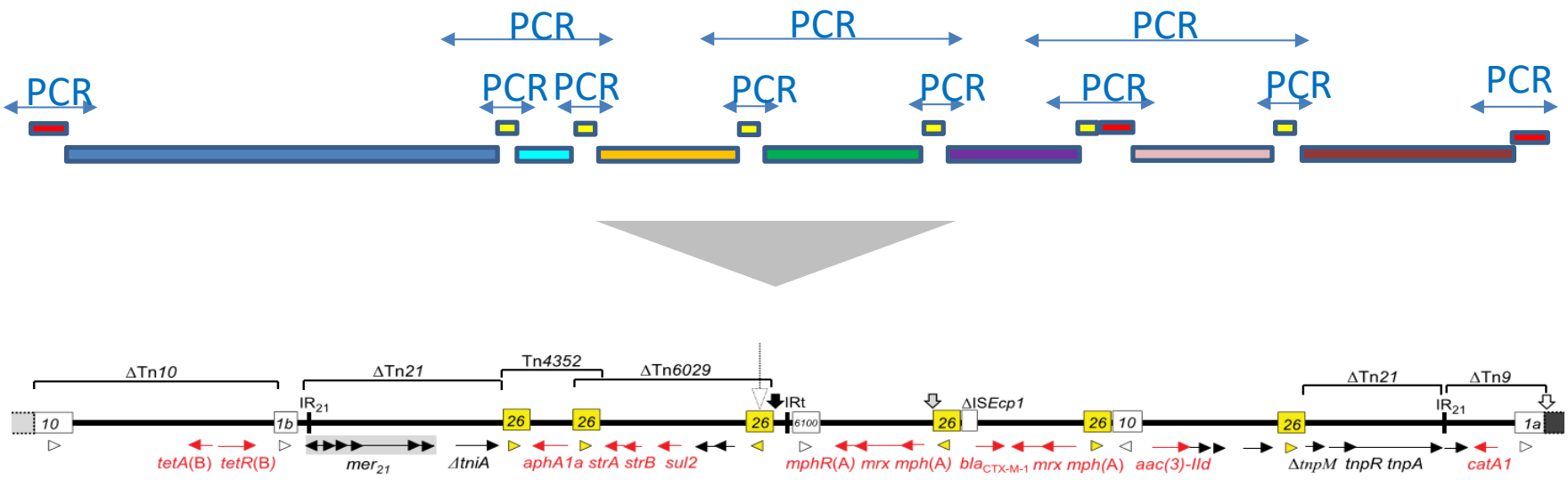
Cutting the Gordian Knot

- Scaffolding:
  1. Assembling the genome using 2 assemblers that have different pipelines.
  2. Using software that will visually help solve some knots.
  3. Gap-closing using PCR.



Cutting the Gordian Knot

43 kb MDR region of 240 kb IncHI1 plasmid from Escherichia coli



Antibiotic resistance genes are shown in red

Insertion sequence IS*26* as yellow box

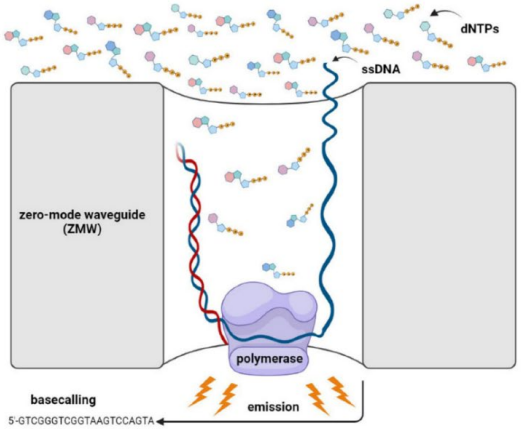
Dolejska et al. J Antimicrob Chemother 2014, 69: 2388-2393.

Long Reads

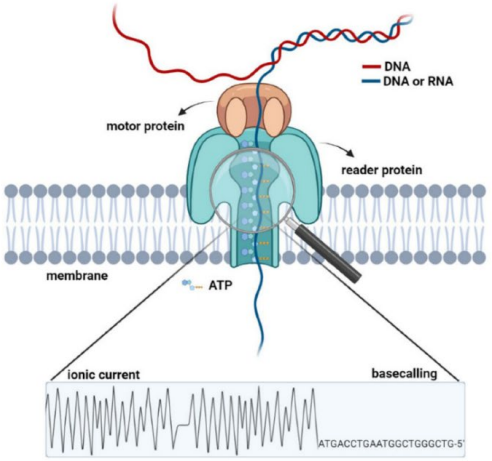
- **Short-read** sequencing produces draft genomes with sequence gaps and incomplete assemblies

- **Long-read** technology produces reads 10,000-100,000 bp in length and allow for the generation of complete genome assemblies

Pacific Biosciences

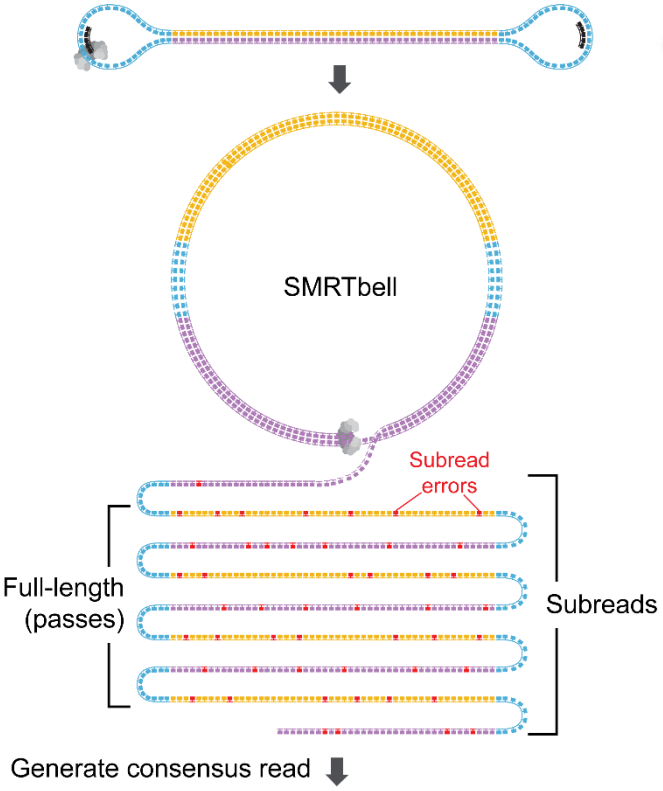


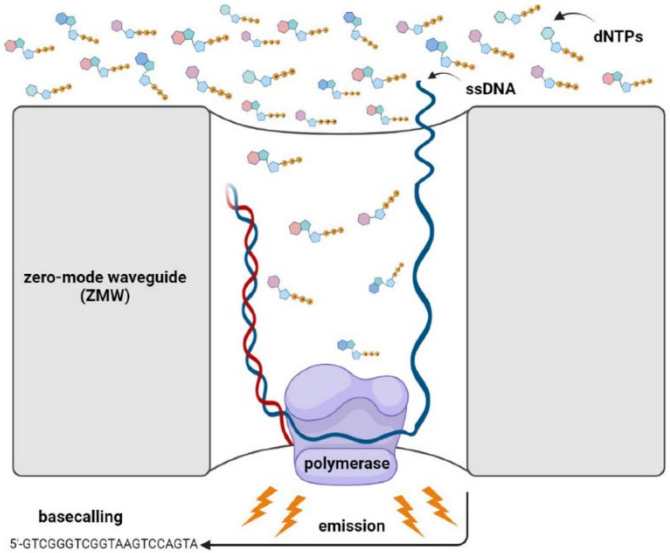
Oxford Nanopore



https://www.pacb.com/blog/the-evolution-of-dna-sequencing-tools/. Athanasopoulou et al. Life 2022, 12(1).

Long Reads



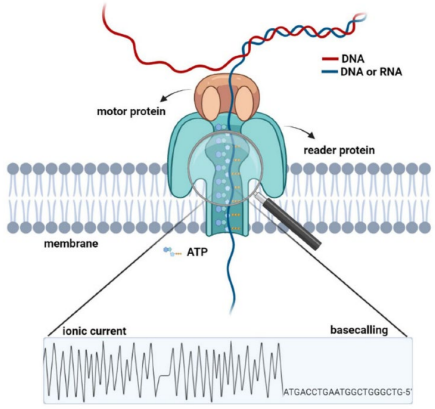


- Rapid, flexible, mobile
- Small and cheap unit
- Inexpensive at low throughput
- Simple user interface and analysis platforms
- Higher base-call and overall error rate

PacBio

- High throughput
  - High accuracy
- Systems are large and more costly
  - Reagents are more costly

ONT

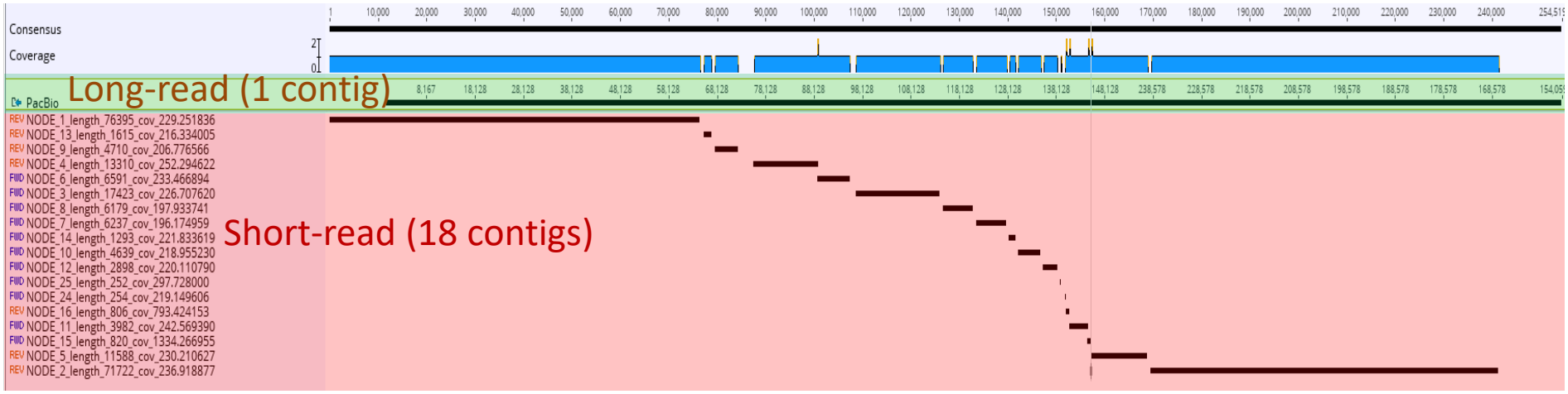


Athanasopoulou et al. Life 2022, 12(1).

Long Reads

Geneious 8.0 software

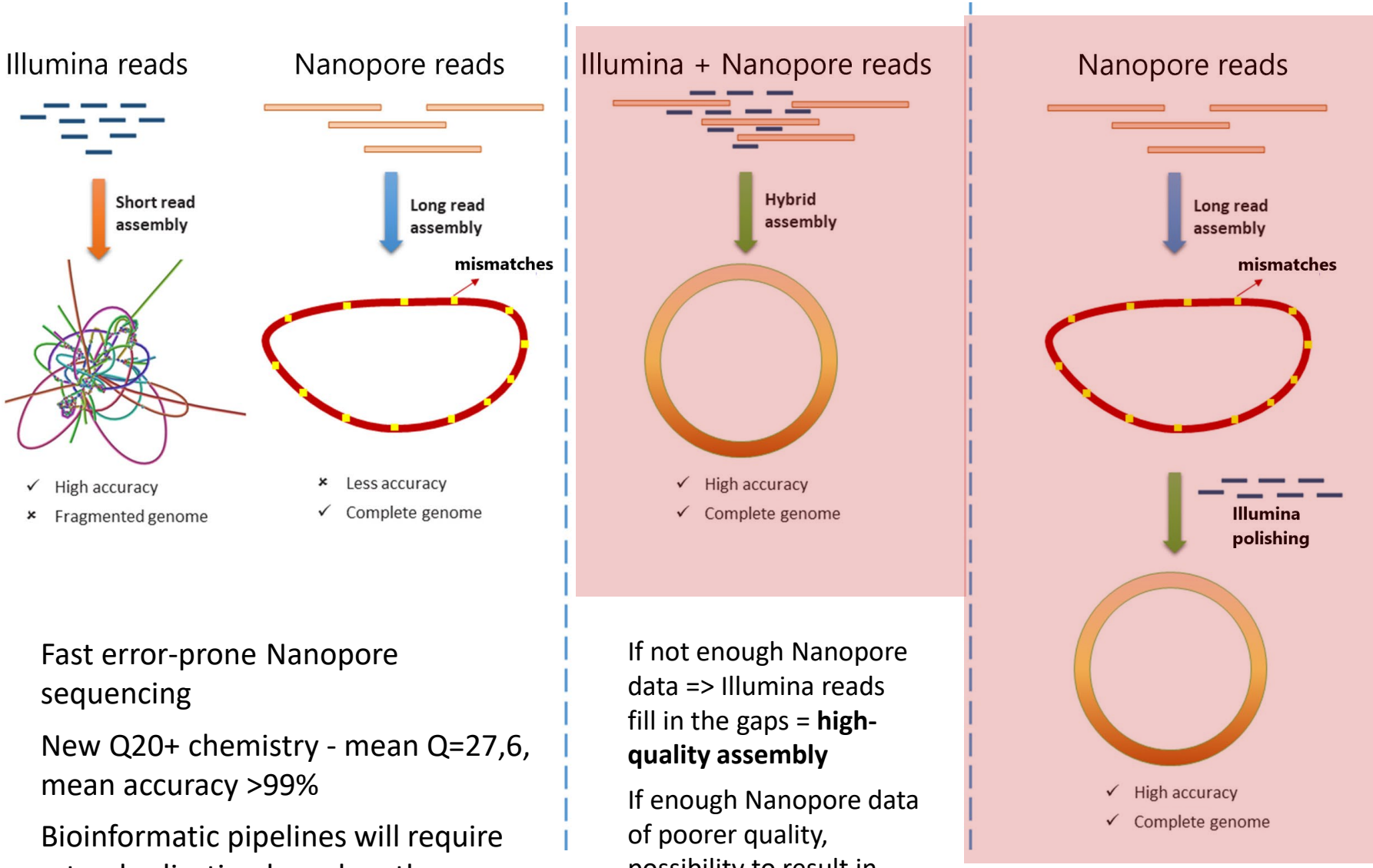
MDR IncHI1 plasmid (240 kb)



Chromosome (4.487)

Valcek et al. Antimicrob Agents Chemother 2023. 65(5):e02556-2

Long Reads



- Fast error-prone Nanopore sequencing
- New Q20+ chemistry - mean Q=27,6, mean accuracy >99%
- Bioinformatic pipelines will require sstandardisation based on the purpose of sequencing

If not enough Nanopore data => Illumina reads fill in the gaps = **high-quality assembly**

If enough Nanopore data of poorer quality, possibility to result in duplicated sequences

If not enough Nanopore data => fragmented to 2-3 pieces

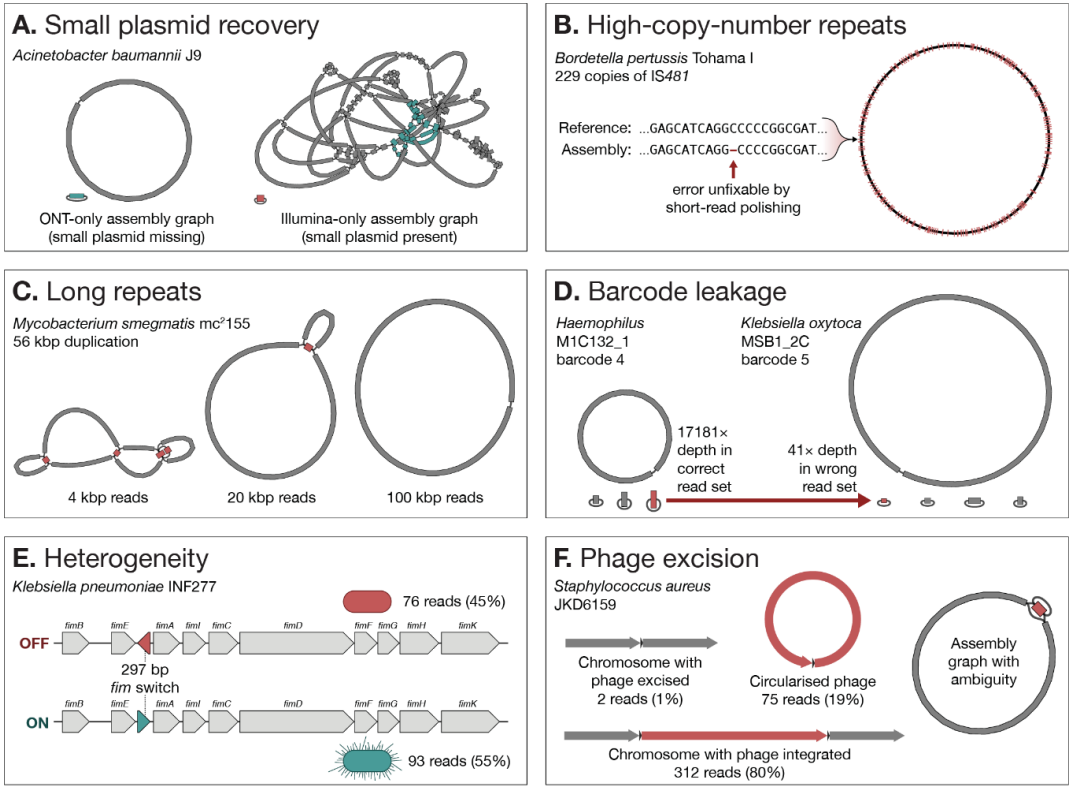
If enough Nanopore data = **high-quality assembly**

Long Reads

PLOS COMPUTATIONAL BIOLOGY 2023. 19(3): e1010905.

Assembling the perfect bacterial genome using Oxford Nanopore and Illumina sequencing

Ryan R. Wick[1]*, Louise M. Judd[2], Kathryn E. Holt[1,3]



Free online tutorial at:

github.com/rrwick/perfect-bacterial-genome-tutorial

Step 1: DNA extraction

isolate DNA extract

- Minimise fragmentation for longer ONT reads
- One DNA extract for both ONT and Illumina
- Save extra DNA in case more sequencing is needed

Step 2: hybrid sequencing

fastq ONT reads

fastq Illumina reads

- Deeper is better: ideally 200× ONT and 200× Illumina
- Best possible ONT reads: R10.4.1 with highest accuracy basecalling

Step 3: long-read assembly

fasta ONT-assembled genome

- Trycycler: combine multiple alternative assemblies into a single consensus
- Goal: genome assembly with zero structural errors (i.e. only small-scale errors)

Step 4: long-read polishing

fasta ONT-polished genome

- Medaka: match model to ONT chemistry and basecaller
- Goal: best possible genome assembly using only ONT reads

Step 5: short-read polishing

fasta hybrid-polished genome

- Polypolish first: low risk of introduced errors
- Then other tools (e.g. POLCA, FMLRC2): sometimes catch errors Polypolish missed

Step 6: manual curation

fasta final curated genome

- Assess changes by visualising read alignments before/after polishing
- Search for errors/misassemblies with variant callers (e.g. freebayes, Clair3, Sniffles2)

- In-silico plasmid closing using Illumina reads is not enough.
- Downstream analysis is needed. These include but not limited to:
    - Transformation (minimizing analyzed plasmids)
    - S1 PFGE (detecting plasmid approximate size)
    - Assembling using different assemblers (elongating contigs).
    - Scaffolding (using BANDAGE, or manually using BLAST)
- Sequencing using Long reads can eliminate the the downstream analysis.
- Correcting sequencing errors and correct assemblies should be taken into consideration

Thank You