



Theme 4 - Resistance elements - 5 June 2024




Plasmid detection & typing

Tools, advantages and implications



Gabriele Arcari



Senior research fellow, Department of Medicine and Technological Innovation, University of Insubria, Italy



Intended Learning Objectives (ILOs)

-  **Bioinformatician:** the most used tools for plasmid detection, typing, and classification
-  **Microbiologist:** advantages (and pitfalls) of bioinformatic plasmid analyses
-  **Epidemiologist:** how to link plasmid types to AMR and virulence genes





Intended Learning Objectives (ILOs)

 **Bioinformatician:** the most used tools for plasmid detection, typing, and classification 

 **Microbiologist:** advantages (and pitfalls) of bioinformatic plasmid analyses 

 **Epidemiologist:** how to link plasmid types to AMR and virulence genes 

Outline

-  Main plasmid reconstruction strategies
 - Short-reads only, long-reads only and hybrid assemblies
-  Plasmid detection and typing methods
 - Conserved-gene based and network analyses
-  Gene-plasmid linking strategies
 - Visual association and plasmid analysis
-  Practical applications
 - Main plasmid typing tools and applications

Plasmid reconstruction

Short-reads sequencing - I - Methods



assembly-based

(e.g., plasmidSPAdes, PLACNET, Recycler and others)



genomic features-based

(e.g., Platon)



k-mer frequencies + learning-based

(e.g., PlasFlow, PlasClass, cBar and others)







hybrid

(e.g. Deeplasmid, plasmidVerify, PLASMe and others)




Plasmid reconstruction

Short-reads sequencing - II - advantages and pitfalls

Advantages

-  Accuracy and coverage
-  Short-read sequencing platforms diffusion
-  Cost-effectiveness
-  Scalability

Pitfalls

-  Reconstruction of either large-low copy or small multi-copy plasmids
-  Issues in identifying chromosomal integrations/transpositions
-  Limited epidemiological value

Plasmid reconstruction

Long-reads sequencing - Most widespread methods



Oxford Nanopore Technologies

(Assemble reads using flye, raven, tricycler and other tools)







PacBio HiFi sequencing assembly

(Assemble reads using FALCON or others)



Plasmid reconstruction

Oxford Nanopore Technologies sequencing - advantages and pitfalls

Advantages

-  *A priori* approach
-  Short turn-around-time
-  Complete assembly and resolution of complex repeats
-  Structural insights (e.g., chromosomal integration)





Pitfalls

-  Low-throughput and lack of scalability
-  Low accuracy and need for corrections




Plasmid reconstruction

PacBio HiFi sequencing - advantages and pitfalls

Advantages

-  *A priori* approach
-  Complete assembly and resolution of complex repeats
-  Structural insights (e.g., chromosomal integration)
-  Accuracy and coverage

Pitfalls

-  Low-throughput and lack of scalability
-  Cost
-  Low diffusion of PacBio sequencing platforms

Plasmid reconstruction

Short- and long-reads sequencing hybrid methods/long-reads only methods

Short-reads first ↓



Unicycler

Long-reads first ↑



Hybracter

Long-reads only ≈

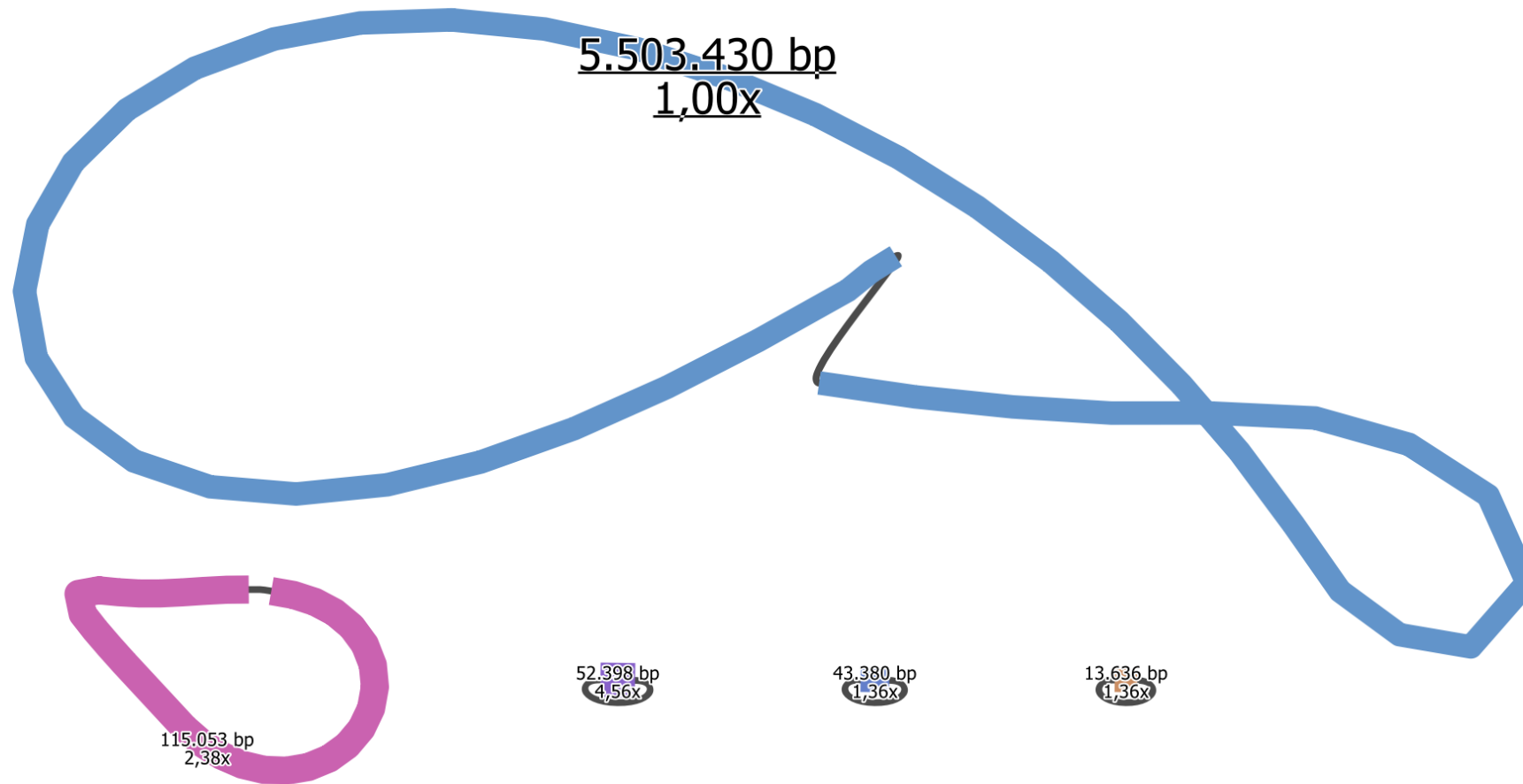


Trycycler



Plasmid reconstruction



Optimal results



Plasmid detection and typing

Basics and approaches - I

Plasmid classification and typing present various issues, such as:



-  Absence of universal marker sequences across diverse plasmid types
-  High frequency of recombination events

Hence, accurate plasmid classification and tracing require a joint effort of all professional figures

Plasmid detection and typing

Basics and approaches - II

Fundamental mechanisms

-  Plasmids have a replicon to autonomously reproduce their genetic material within a host cell, ensuring their independent maintenance and propagation
-  To undergo horizontal transfer, plasmids need a relaxase (mobilizable) and a mating pair formation complex (self-conjugative). Some plasmids lack both systems (non-mobilizable)

Network analyses




Plasmid modularity is a fundamental trait of mobile genetic elements and has a major role in the spread of AMR.

In presence of complete plasmid assemblies, it is possible to compare their whole sequences, revealing association patterns among different plasmids



Plasmid detection and typing

Relaxases, replicons and type IV secretion systems



Replicon

-  Regions activating and controlling replication
-  Correlates with the ≈ 25 major *Enterobacterales* incompatibility (**Inc**) groups
-  ≥ 150 *in silico* probes

Relaxase

-  Catalyzes *oriT* cleavage and ligates transported DNA in the recipient cell
-  Six mobility (**MOB**) families




T4SSs

-  Membrane-associated mating pair formation complexes
-  Four groups (MPF_T, MPF_F, MPF_I and MPF_G)






Plasmid detection and typing

Replicon typing (PlasmidFinder)

Advantages

-  Widespread and portable nomenclature
-  Moderate discriminative power
 -  Allows to trace replicons in plasmid collections

Pitfalls

-  Dependence on known sequences
 -  $\geq 80\%$ identity in small plasmids (≤ 10 kb)
 -  $\geq 95\%$ identity in large plasmids (> 10 kb)
-  Mostly limited to *Enterobacteriaceae* family
-  Plasmids may carry multiple replicons

Plasmid detection and typing

Replicon typing and associated AMR genes (StarAMR)

Command line approach

Install staramr via pip, conda/mamba or source

```
staramr search -o output_name *.fasta
```

GUI approach (Galaxy servers)

Open your preferred Galaxy platform (e.g., <https://galaxy.pasteur.fr/>)

Upload all plasmid .fasta files to be analyzed on Galaxy

Look for starAMR among search tools

Set parameters (e.g., lower the “Percent identity threshold for BLAST” to 95%)

Plasmid detection and typing

Replicon sub-typing (*pMLST*) <https://pubmlst.org/organisms/plasmid-mlst>

Advantages

- 🦠 High discriminative power
 - 🦠 Allows to distinguish allelic variants of “core” plasmid genes




Pitfalls

- 🦠 Limited to only few plasmid/replicon types
 - 🦠 IncI1
 - 🦠 IncHI1 and IncHI2
 - 🦠 IncF (Replicon Sequence Typing → FAB formula)
 - 🦠 IncA and IncC
 - 🦠 IncN
 - 🦠 *Shigella flexneri* virulence plasmid




Plasmid detection and typing

Relaxase typing (MOB-typer)

Advantages

-  Typically only one relaxase/plasmid
-  Wider evolutionary significance
-  Plasmid identification across diverse taxa




Pitfalls

-  Low discriminative power
-  Less diffused and portable nomenclature
-  Unapplicable to non-mobilizable plasmids



Plasmid detection and typing

Fundamental mechanisms-based typing - Recap

Advantages

-  Widespread reference-based nomenclature
-  Allows to form epidemiological links and to back-trace with historical plasmids
-  No need to perform clustering analyses with other plasmids


Pitfalls

-  Identifying plasmids with uncharacterized replicons and/or relaxases is difficult
-  Moderate/low discriminative power


Plasmid detection and typing

Network analyses-based typing - I

COPLA

-  Assigns plasmids to Plasmid Taxonomic Units (PTUs) based on their average nucleotide identity (ANI)





mge-cluster

-  Breaks down plasmid sequences into unitigs (DNA "stretches" that don't branch off into other sequences), 2D embed them into a presence/absence matrix (using openTSNE) and cluster eventual patterns (using HDBSCAN)




Plasmid detection and typing

Network analyses-based typing - II

Advantages

-  High discriminative power
-  Does not depend on known sequences
-  Scalable and applicable to all plasmid types
-  Produce shareable and reusable schemes






Pitfalls

-  Works better using complete plasmid sequences
-  Long computational time (COPLA)
-  Not yet widely disseminated

Plasmids and virulence

General traits


Virulence factors of several bacteria (e.g., *Shigella flexneri*, *Salmonella enterica* serovar subsp. *Enterica* serovar Typhimurium, EH/EP *Escherichia coli*, *Yersinia enterocolitica* and *pseudotuberculosis*) are carried on plasmids sharing common features:

-  Multiple replicons (plausibly to circumvent incompatibility/to alter copy number)
-  Multiple toxin–antitoxin systems (to grant persistence and intracellular survival)
-  Multiple partitioning systems (to grant segregation during replication and incompatibility)
-  Multiple virulence genes
-  Multiple insertion sequences (resulting in the mosaic structures)


Phages, phage-plasmids and ICEs

The complex landscape of mobile genetic elements




Phages

-  Viral particles infecting and replicating within prokaryotic cells

Phage-plasmids

-  Genetic elements characterized by phage-like horizontal transfer (transduction) between cells and plasmid like vertical inheritance (extra-chromosomal phase)

ICEs

-  Integrative Conjugative Elements (conjugative transposons) are modular structures integrated in the host chromosome which can:
 -  be passively propagated during cell division
 -  excise, produce a T4SS machinery and transfer DNA to recipient hosts

Phages, phage-plasmids and ICEs




Clinical and epidemiological relevance

Phages

Largely anecdotal impact, but evidence is emerging about their role in AMR diffusion (e.g., *qnrA*)

ICEs










Albeit a low transfer efficiency ($\approx 1 \times 10^{-8}$) they may supply significant phenotypes to the host cell:

-  virulence (e.g., ICE*Pm1* in members of the *Proteeae* tribe, ICE*HAI*₂ in *Pectobacterium atrosepticum* or ICE*Kp* in *Klebsiella pneumoniae*)
-  AMR (e.g., ICE_{Tn4371}6061 in *Pseudomonas aeruginosa*)
-  heavy metals (e.g., *SGI-4* in *Salmonella typhimurium* ST34)

Phages, phage-plasmids and ICEs

Detection tools

Phages

-  Sequence-based (AA and/or HMM)
 -  PHASTER (web-based GUI)
 -  Phigaro (stand-alone)
 -  ProphET (stand-alone + web-based GUI)
-  Deep learning-based prediction
 -  DeepVirFinder
-  Hybrid (database search-dependent and -independent features)
 -  VIRALVERIFY
 -  VIBRANT

ICEs

-  ICEfinder



Which repositories offer data for...

Antimicrobial resistance genes

CARD <https://card.mcmaster.ca/>

ResFinder <http://genepi.food.dtu.dk/resfinder>

AMRFinderPlus <https://github.com/ncbi/amr>

Which repositories offer data for...

Curated plasmid sequences

PLSDB <https://ccb-microbe.cs.uni-saarland.de/plsdb>

pATLAS <http://www.patlas.site/>

IMG/PR <https://img.jgi.doe.gov/pr>

pMLST <https://pubmlst.org/organisms/plasmid-mlst>

Which repositories offer data for...

Virulence-associated genes

VFDB <http://www.mgc.ac.cn/VFs/main.htm>

EPEC VirulenceFinder <https://cge.food.dtu.dk/services/VirulenceFinder/>

Genome collection screening

Hands-on tools to analyze your data



ARIBA (from paired .FASTQ reads) <https://github.com/sanger-pathogens/ariba>




ABRicate (from .FASTA assemblies) <https://github.com/tseemann/abricate>

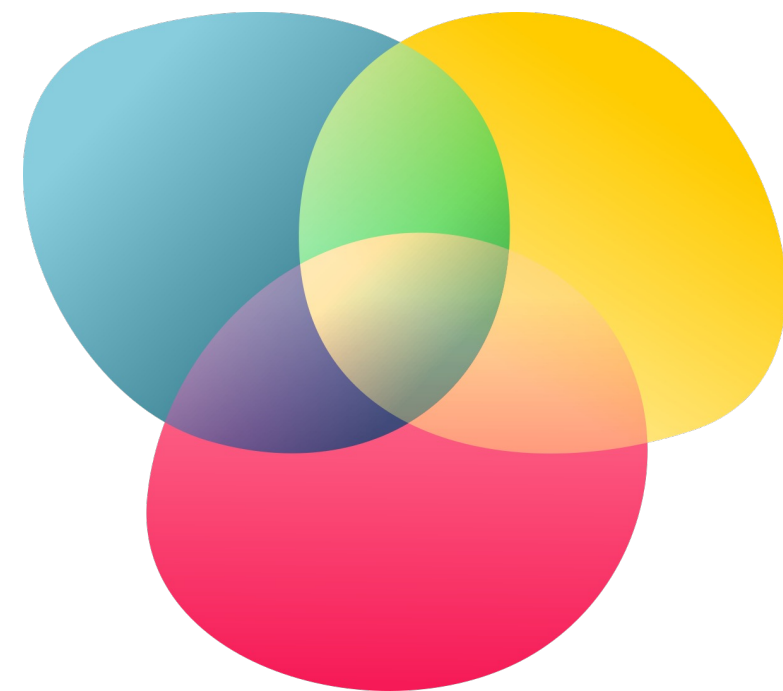
starAMR (from .FASTA assemblies) <https://github.com/phac-nml/staramr>



Summary

An ILOs reminder

-  **Bioinformatician:** the most used tools for plasmid detection, typing, and classification
-  **Microbiologist:** advantages (and pitfalls) of bioinformatic plasmid analyses
-  **Epidemiologist:** how to link plasmid types to AMR and virulence genes













Summary

Take home message

Sharing knowledge and understanding edges and limits of each different professional figure is the best way to perform public health actions.

*Jack of all trades, master of none...
but oftentimes better than a master of one*

Further reading and references

-  **A review on short reads only plasmid prediction/reconstruction tools** <https://doi.org/10.3390/microorganisms9081613>
-  **A practical guide to bacterial genome assembly** <https://github.com/rrwick/Tracycler/wiki/Guide-to-bacterial-genome-assembly>
-  **Bandage: interactive visualization of de novo genome assemblies** <https://doi.org/10.1093/bioinformatics/btv383>
-  **Replicon detection: PlasmidFinder** <https://doi.org/10.1128/2FAAC.02412-14>
-  **Replicon subtyping: the pMLST page** https://pubmlst.org/bigsdb?db=pubmlst_plasmid_seqdef
-  **Relaxase detection and typing: the MOB-suite** <https://doi.org/10.1099/mgen.0.000206>
-  **COPLA: a taxonomic classifier of plasmids** <https://doi.org/10.1186/s12859-021-04299-x>
-  **Mge-cluster: a reference-free approach for typing bacterial plasmids** <https://doi.org/10.1093/nargab/lqad066>
-  **Virulence plasmids in enteric pathogens** <https://doi.org/10.1038/s41579-018-0031-2>
-  **Phage identification and typing: currently available computational tools** <https://doi.org/10.1128/mmbr.00004-21>

Thank you for your attention!

Questions?

Acknowledgements

The creation of this training material was commissioned by ECDC to Institut Pasteur with the direct involvement of Gabriele Arcari