Whole Genome Sequencing and Bioinformatics SeqAfrica Training

2-5 June 2025
Lusaka

Marco van Zwetselaar

Niamh Lacy-Roberts

Day 3

The Fleming Fund Regional Grants

UKaid from the British people

DTU

SeqAfrica

The Fleming Fund
Regional Grants

Introduction to Antimicrobial Resistance (AMR)

The Fleming Fund
Regional Grants

Introduction to antimicrobial resistance (AMR)

- AMR is when bacteria, viruses, fungi or parasites no longer respond to antimicrobial medicines.

- As a result of drug resistance, antibiotics and other antimicrobial medicines become ineffective and infections become difficult or impossible to treat.

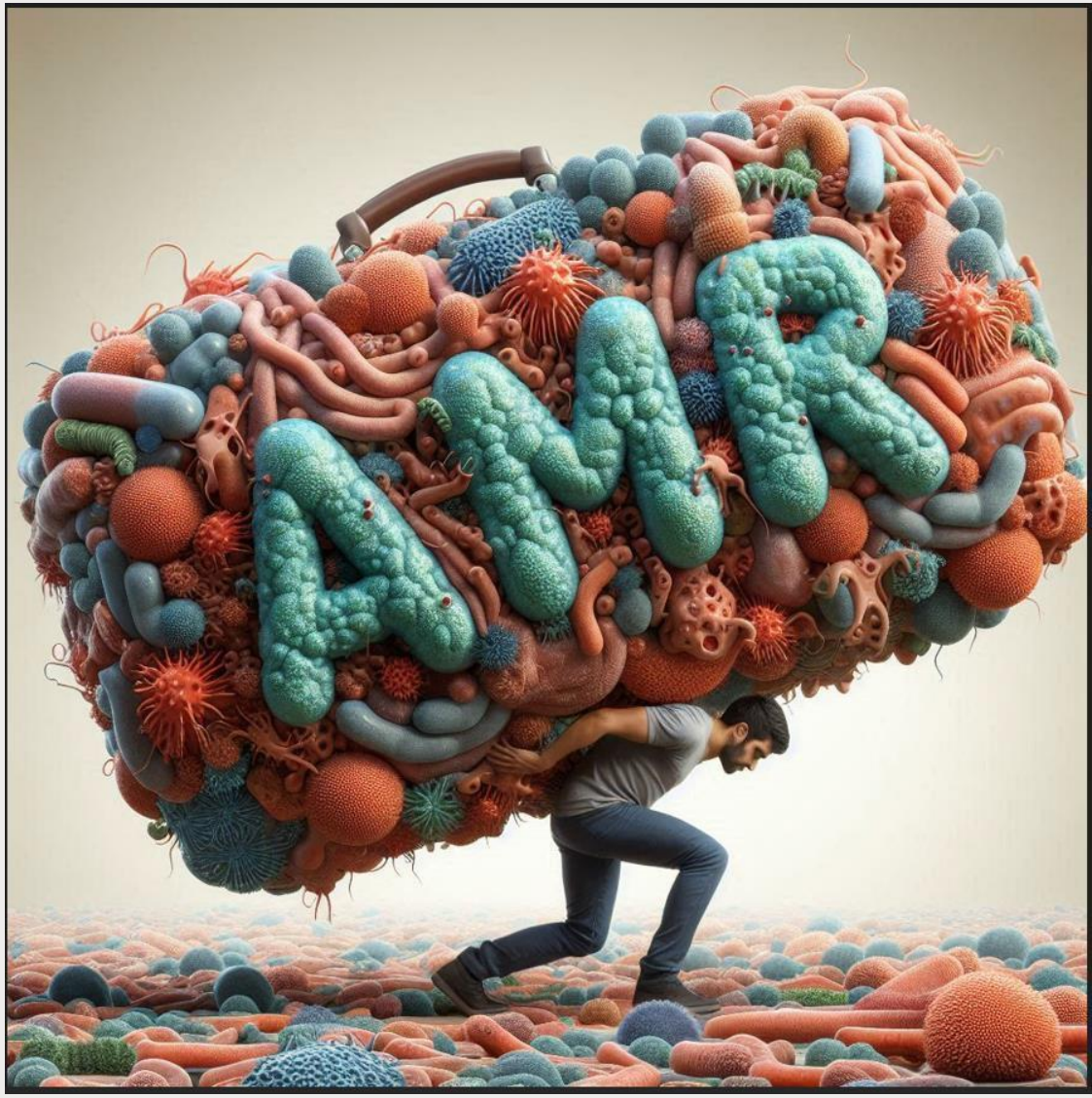
- Increases the risk of disease spread, severe illness, disability and death.

Generated with AI by L.H. Sørensen · Microsoft copilot· September 30,

Burden of AMR

- 1,2 million deaths were directly attributed and 4,9 million associated with AMR in 2021.[1]
- Death attributed to AMR in people above 5 years increase across all regions between 1990-2021, except in Central and Western Europe.[1]
- Main drivers thought to be Antimicrobial usage and lack of antimicrobial stewardship.[2]
- Low and middle income countries are the affected most by AMR, but there are major disparities in AMR reporting.[1]

Generated with AI by L.H. Sørensen · Microsoft copilot· September 30, 2024



1) **Global burden of bacterial antimicrobial resistance 1990–2021: a systematic analysis with forecasts to 2050,** Naghavi, Mohsen et al. The Lancet, Volume 404, Issue 10459, 1199 – 1226

2) **Global antimicrobial-resistance drivers: an ecological country-level study at the human–animal interface,** Allel, Kasim et al., The Lancet Planetary Health,

DTU STATENS SERUM INSTITUT

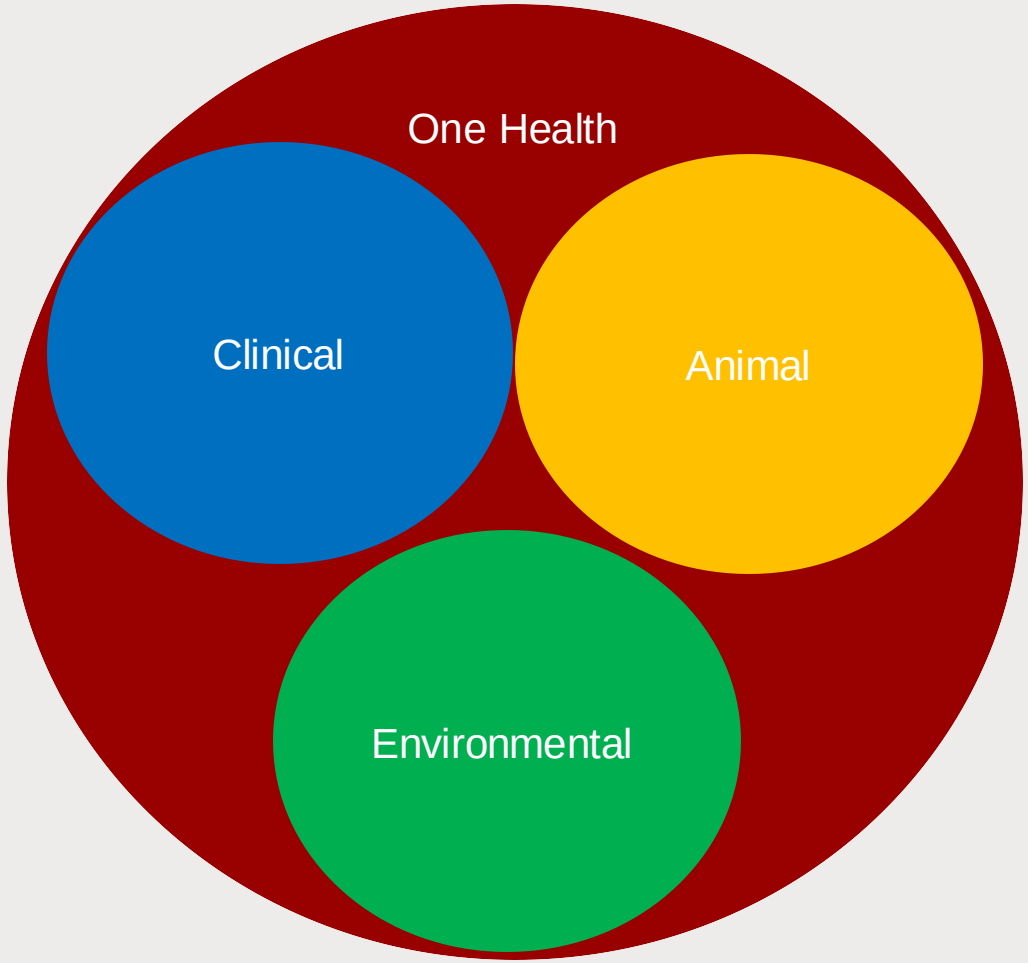
DANMAP
The Danish Integrated Antimicrobial Resistance Monitoring and Research Programme

- https://www.danmap.org/

AMR in Denmark

- Reports on:
  - Antimicrobial consumption in Humans.
  - Antimicrobial consumption in Animals.
  - Resistance in human pathogens.
  - Resistance in indicator bacteria from animal side.
  - Zoonotic pathogens.
  - Animal pathogens.
  - General trends compared to previous years.

- Collaboration with farmers based on voluntary systems.

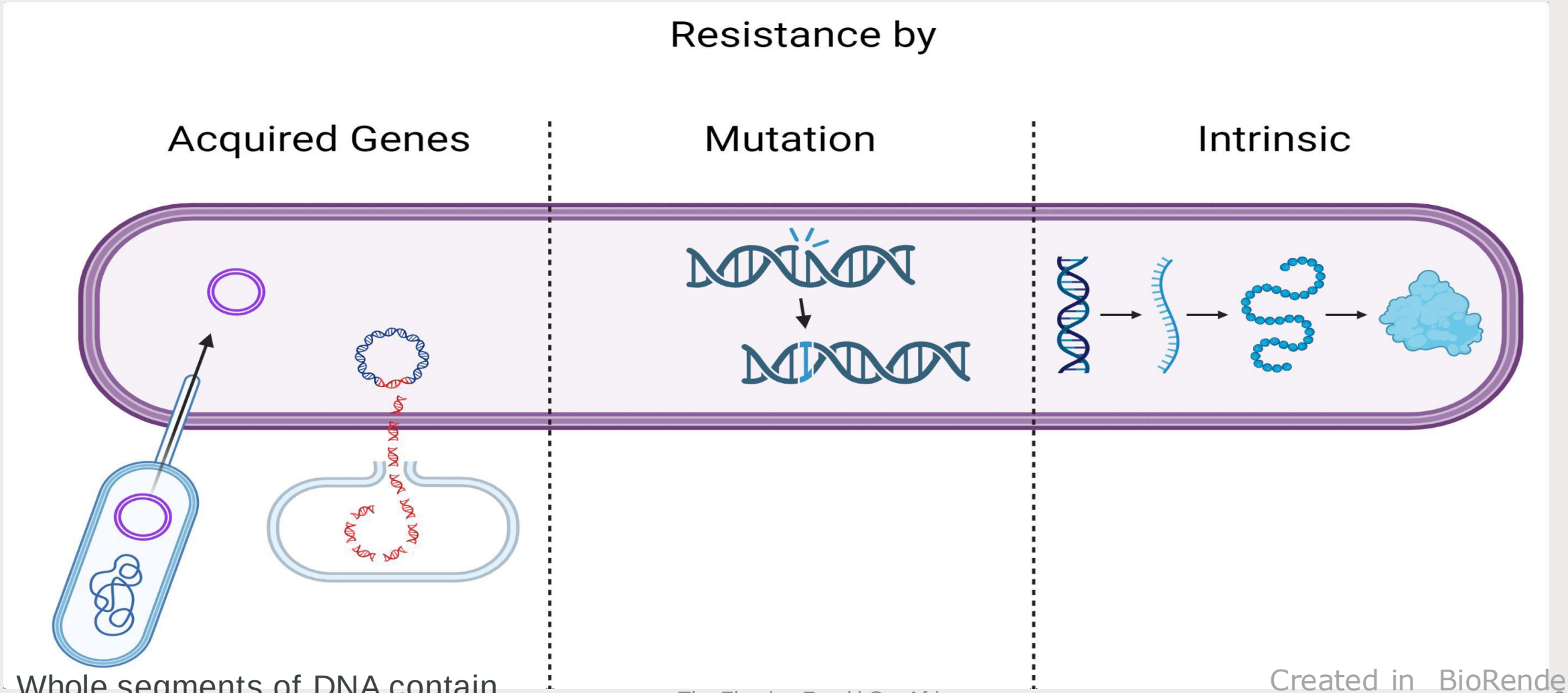


The Fleming Fund
Regional Grants

Importance of AMR surveillance

- AMR surveillance is crucial because it allows for
  - the early detection of resistant bacterial strains,
  - tracking trends in resistance patterns,
  - informing clinical decision-making,
  - guiding policy development,
  - enabling effective interventions to combat the growing threat of AMR
  - mitigating resistance development and spread.

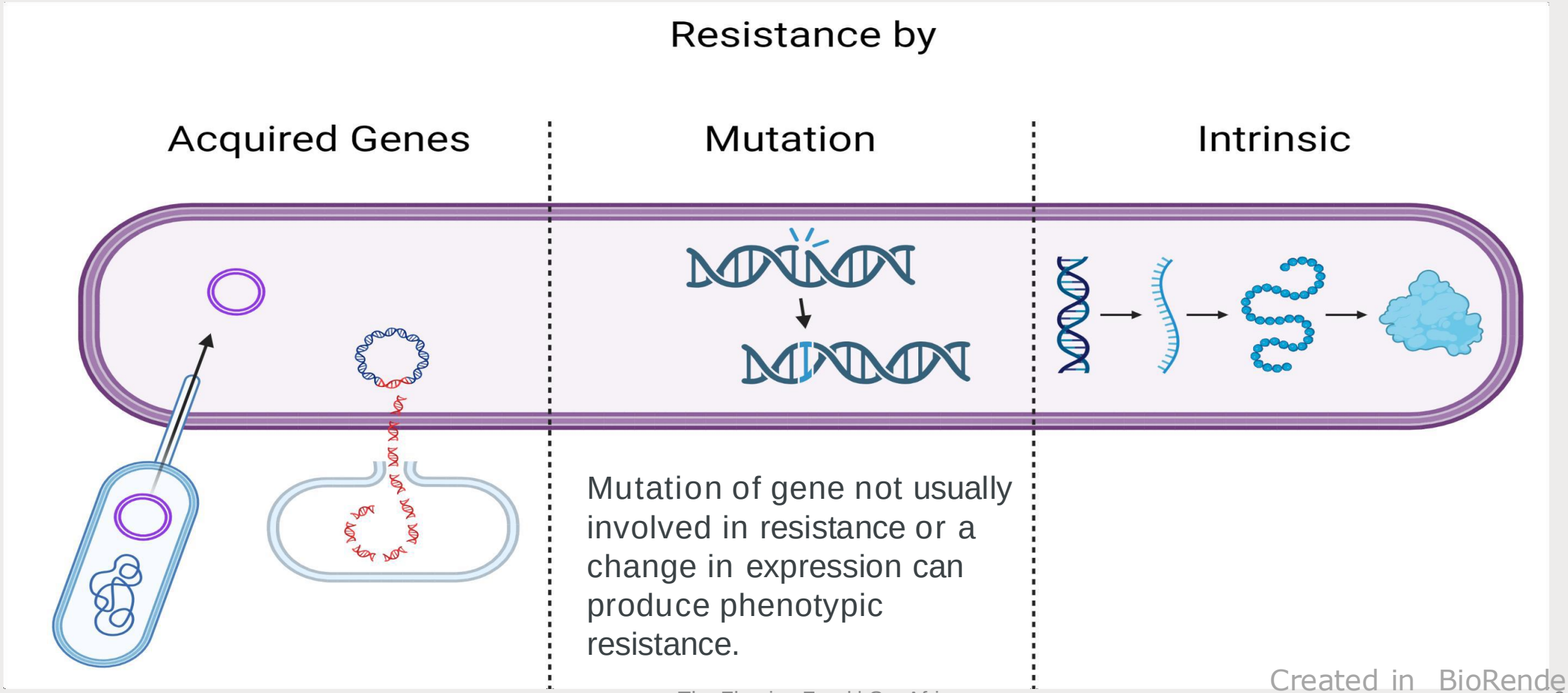
AMR genomic background



Whole segments of DNA contain one or multiple genes are acquired

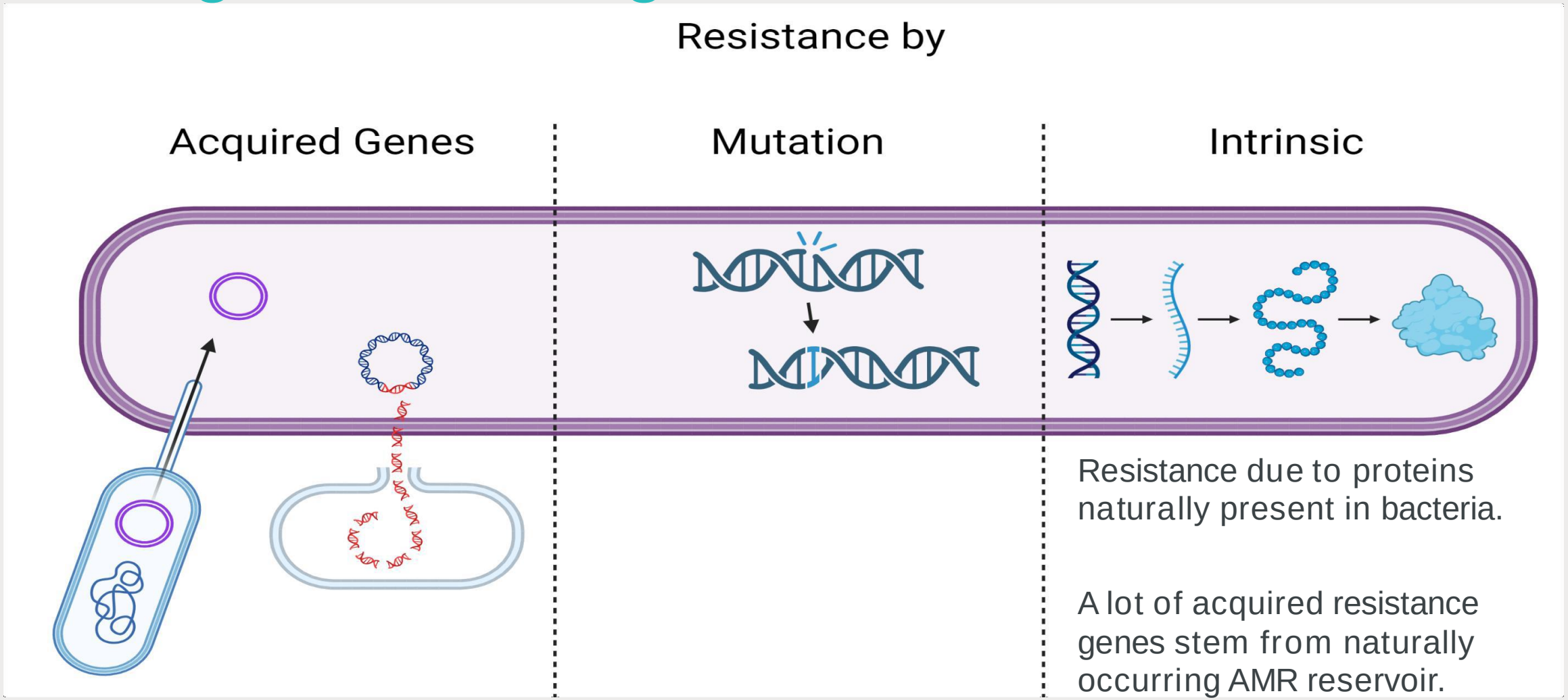
Created in BioRender.com

AMR genomic background



Created in BioRender.com

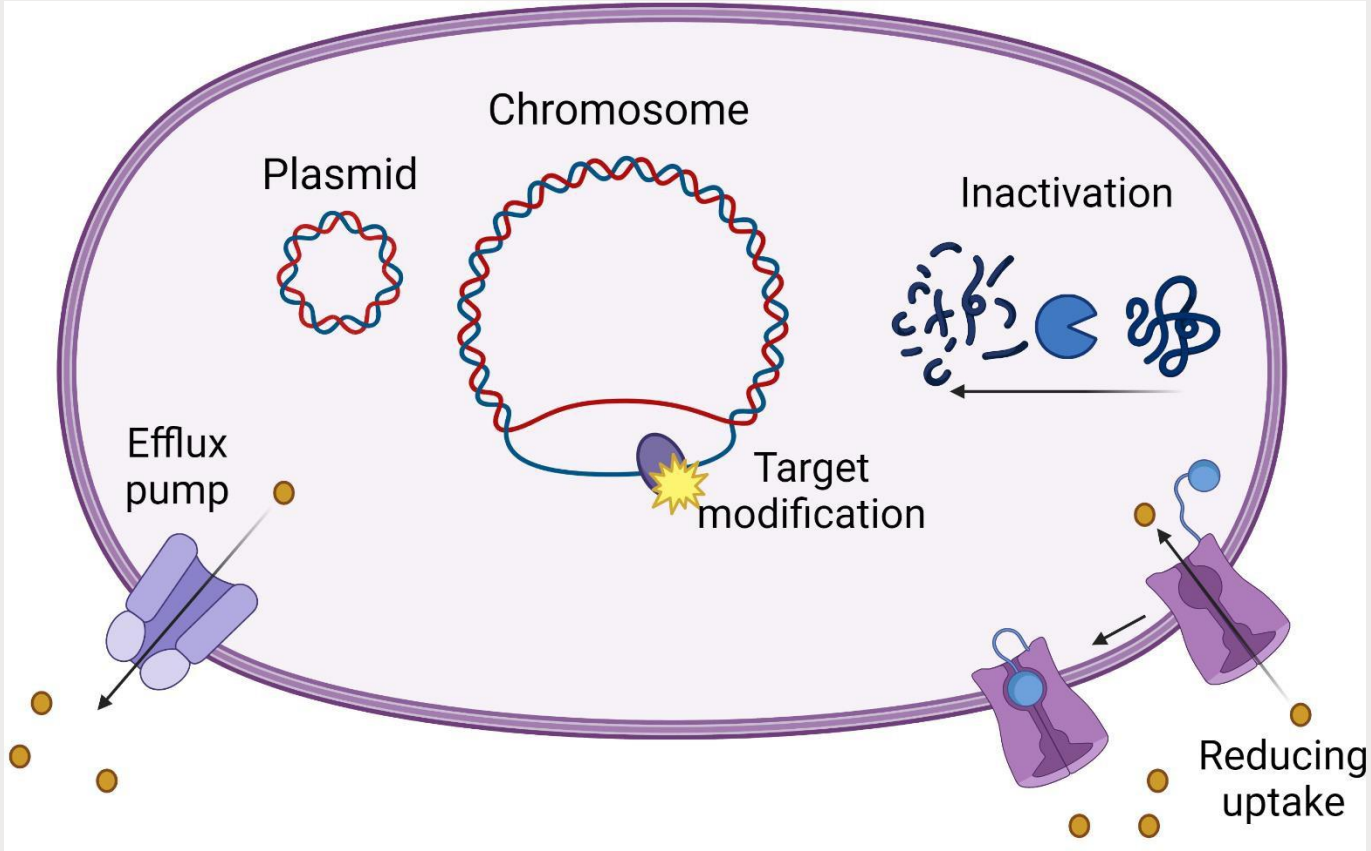
AMR genomic background



AMR in bacteria

- AMR can arise by several mechanisms in the cell:
  - Efflux of antimicrobials
  - Enzyme inactivation
  - Target modification
  - Reducing uptake

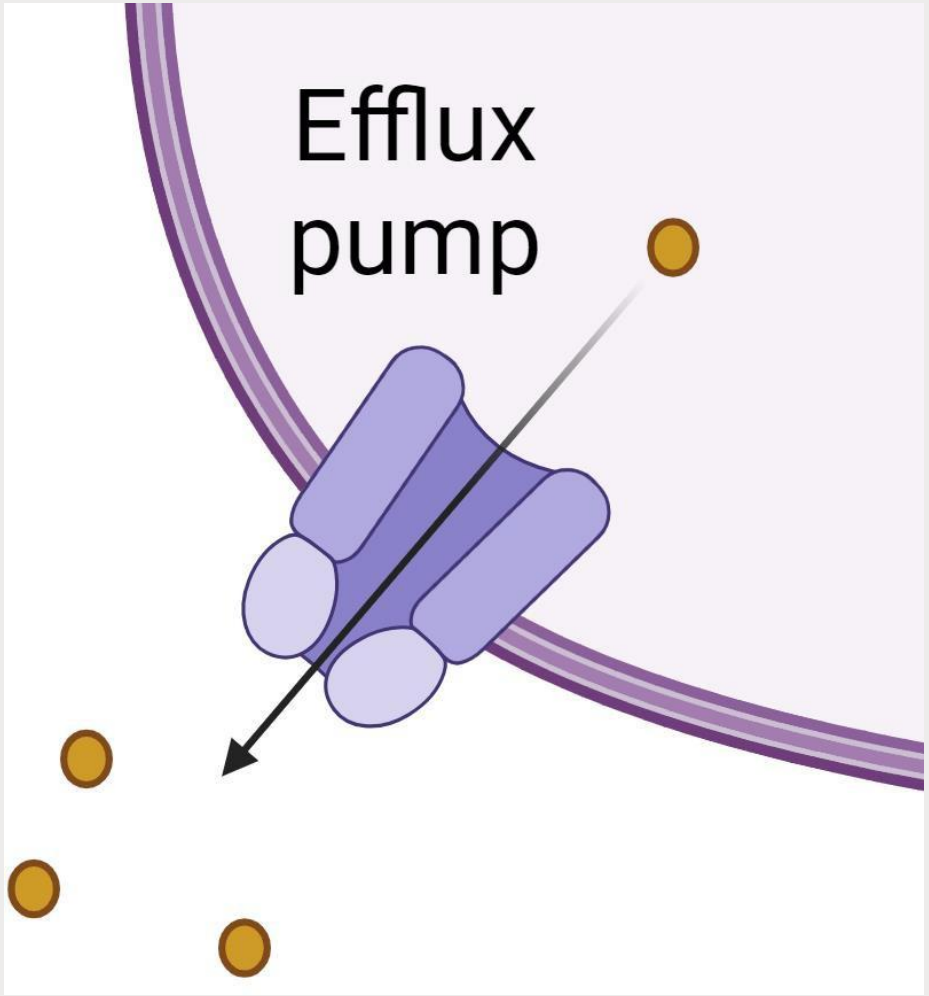
- Phenotypic AMR can be a result of several mechanisms working in tandem.



Created in BioRender.com

Efflux pumps

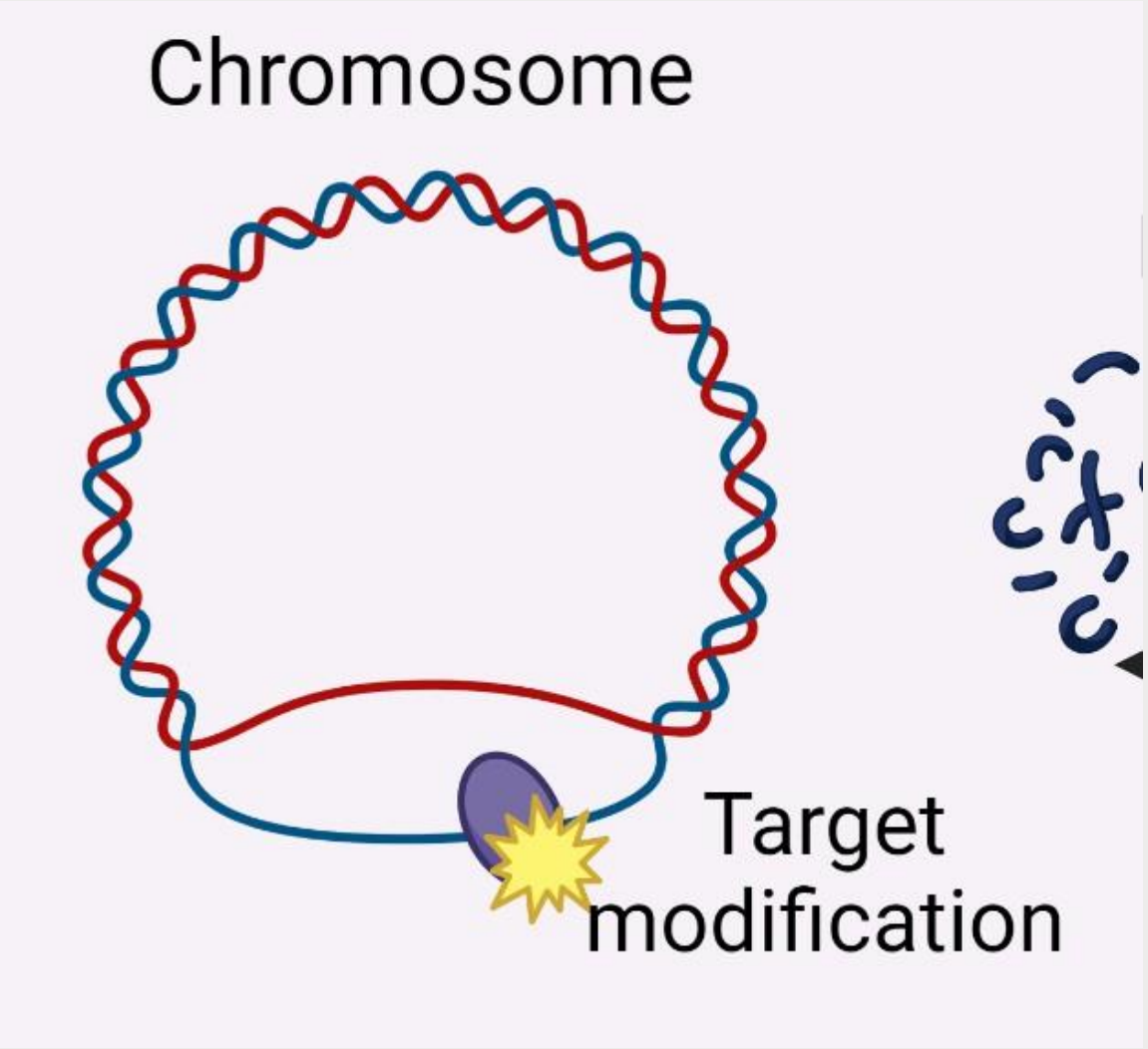
- Active transportation of antibiotics out of the cell.
- Can provide increased tolerance of wide variety of antibiotics, e.g. mdfA in E. coli exports a large number of toxins, including several classes of antimicrobials.
- Increased tolerance can be achieved by a number of mechanisms:
  - High expression
  - High copy number
  - Structural variation
- Synergy with other resistance mechanisms.



Created in BioRender.com

Target modification

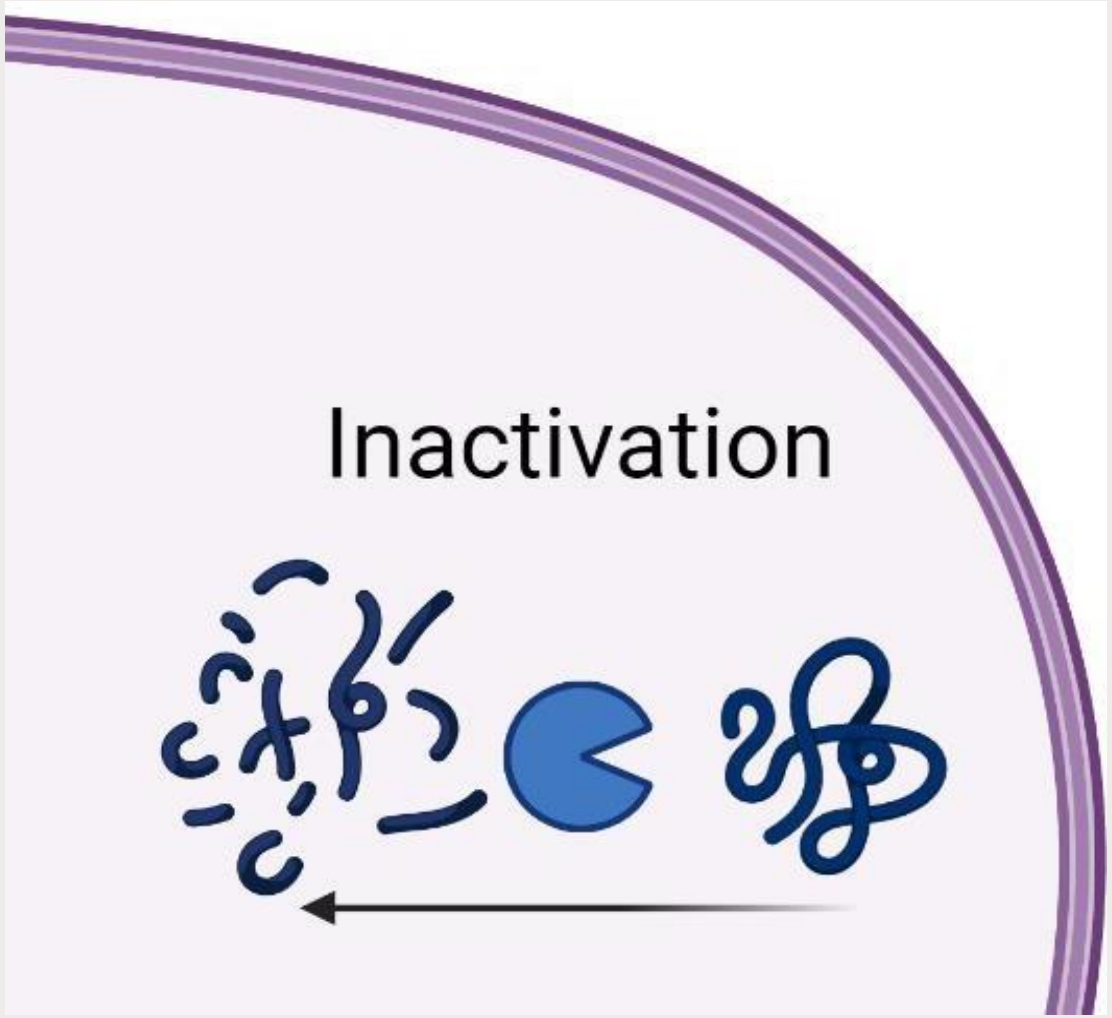
- Modification of target, either by a change in structure or specific motif can prevent binding of antibiotics.
- E.g. gyrase A in a number of pathogens (e.g. E. coli, S. enterica) prevents binding of fluoroquinolones such as ciprofloxacin.
- Several positions in the protein can confer or increase resistance.
- Gyrase A mutation very common in poultry production, and apparently stable (low cost of fitness)



Created in BioRender.com

Inactivation

- Enzymes that break down antimicrobials.
- Includes classes of high priority to public health, such as Extended beta-lactamases (ESBL) genes CTX-M, SHV
- Includes antimicrobial classes of critical importance and last resort drugs in infections such as carbapenemases (e.g. NDM, OXA-48-like, KPC)
- Acquired genes, overexpression of intrinsic genes, mutational gain-of-function in intrinsic genes.

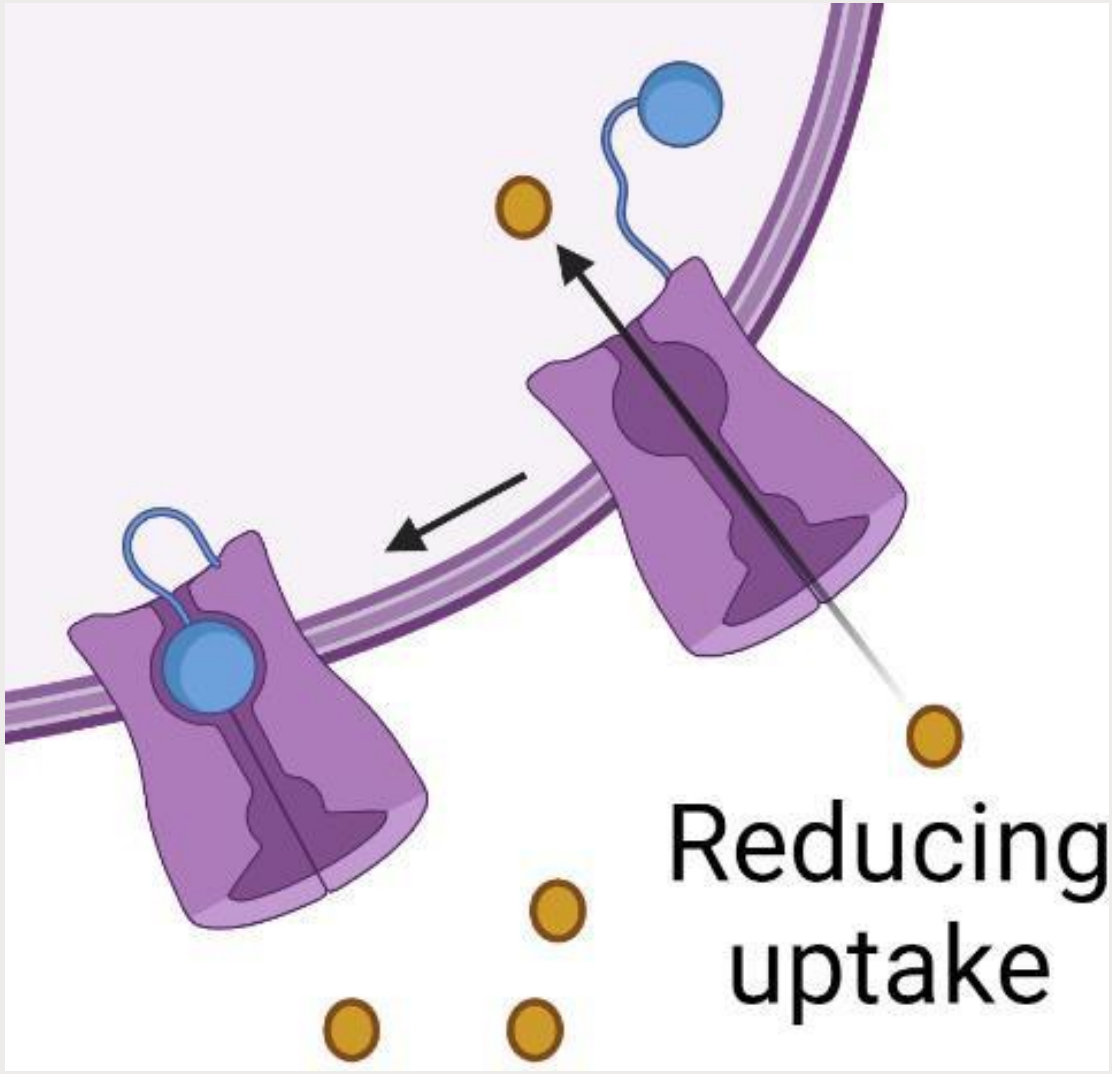


Created in BioRender.com

Reduced uptake

- Reduced uptake can happen by several mechanisms:
  - Mutational loss-of-function.
  - Reduced expression.
  - Natural (intrinsic) regulation.

- Synergy between reduced uptake and natural inactivation of enzymes (e.g. porA mutation in campylobacter increases carbapenem tolerance.



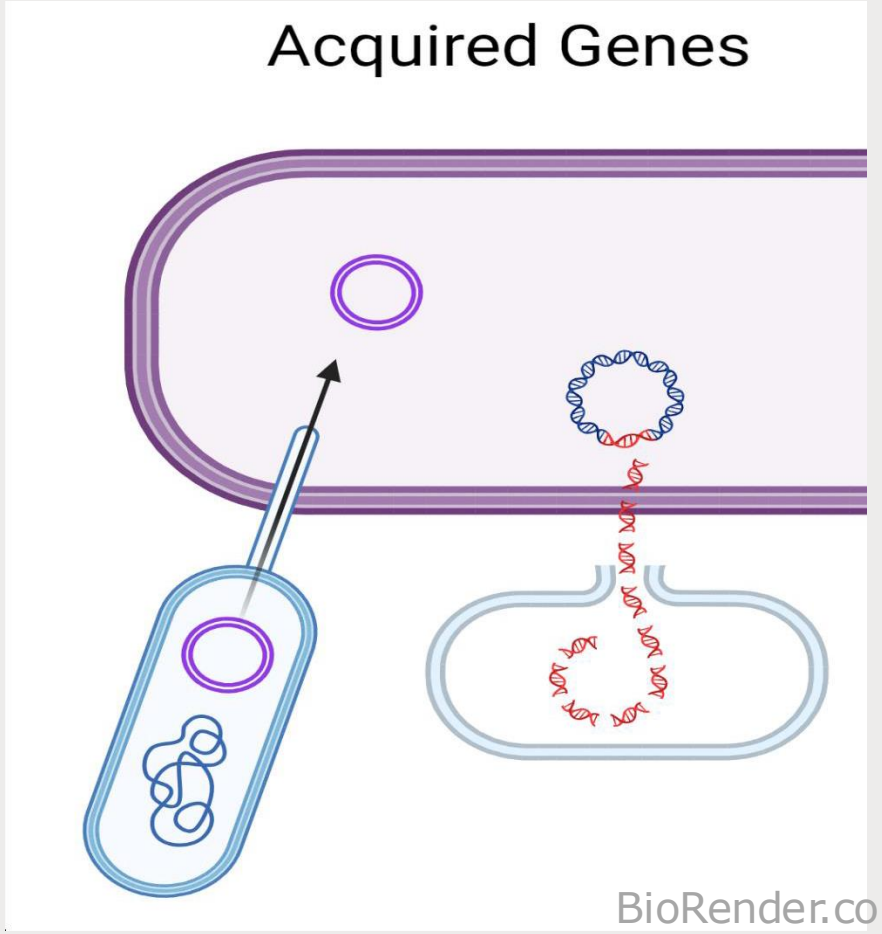
Created in BioRender.com

Horizontal gene transfer (HGT) - plasmids

- Plasmids are a major concern in the dissemination of acquired resistance genes.

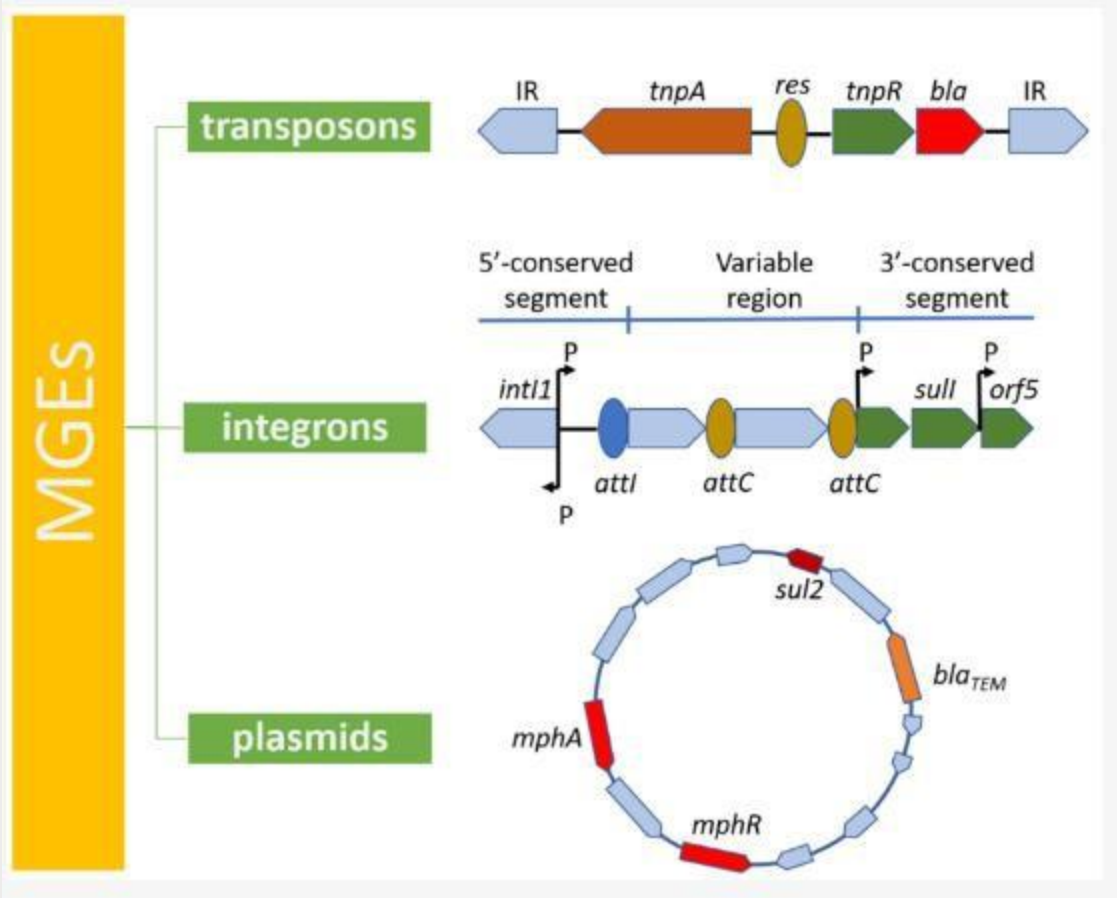
- Plasmids can function as “genomic parasites”, propagating through bacterial populations.

- Self-transmissible plasmids carry conjugative elements and can initiate their own transfer to other cells.
- Mobilizable plasmids carry mobilization genes, but are dependent on other sources of conjugation.
- Some investigation into plasmid transferring, despite lack of mobilization or conjugative elements.



Other mobile genetic elements (MGE)

- Other modes of DNA transfer exists:
  - Transduction by phages
  - Natural transformation by integration of environmental DNA
- Frequency of specific HGTs vary between species.
- Smaller units of mobile genetic elements aid the transfer of AMR genes by integrating into plasmids.
  - Integrons.
  - Transposons.



Rozwadowski, M.; Gawel, D. Molecular Factors and Mechanisms Driving Multidrug Resistance in Uropathogenic *Escherichia coli*—An Update. *Genes* **2022**, *13*, 1397. https://doi.org/10.3390/genes13081397

Full genome annotation

- With Whole genome sequencing (WGS) we capture (almost) everything in the cell
  - Prokka: rapid prokaryotic genome annotation (GitHub - tseemann/prokka: Rapid prokaryotic genome annotation)
  - ANNOVAR: Higher organisms (ANNOVAR Documentation (openbioinformatics.org))
  - NCBI-PGAP: Prokaryotic annotation (NCBI Prokaryotic Genome Annotation Pipeline (nih.gov))
  - Predictive annotation: eggNOG-mapper (eggNOG-mapper (embl.de))

- These pipelines usually generate multiple output files, which can be used for further data handling or visualization

- There are multiple visualization tools, e.g. IGV, which can be installed locally or used online.

E. Coli strain K-12 from NCBI visualized with gff in IGV (IGV)

Annotation: AMR tools

- Resfinder
  – Developed at DTU
- AMRfinderplus
  – Developed at NCBI
- CARD
  – Developed at McMaster University

- Databases and search strategies depends on the tool.

- Curation is a major limiting factor in trustworthy and precise translation of genotype to phenotype.

Generated with AI by L.H. Sørensen · Microsoft copilot· October 1, 2024

ResFinder

Version

4.6.0

ResFinder identifies acquired genes and/or finds chromosomal mutations mediating antimicrobial resistance in total or partial DNA sequence of bacteria.

ResFinder software: (2024-03-22)
ResFinder database: (2024-03-22)
PointFinder database: (2024-03-08)
DisinFinder database: (2023-05-31)

Chromosomal point mutations:

Threshold for %ID

90%

Minimum length

60%

Show unknown mutations
Ignore premature stop codons:
Ignore frameshift indels:

Acquired antimicrobial resistance genes:

Threshold for %ID

90%

Minimum length

60%

Species and input data type:

Select species

Other

Select input type

FASTA (Assembled Genome/Contigs)

Disinfectant:

Run disinfectant

Threshold for %ID

90%

Minimum length

60%

Upload and submit job:

Email (Get email, when finished - Optional):

Enter your email address...

Files (The sum of uploaded file sizes cannot exceed 1 gb):

Choose File No file chosen Choose File No file chosen

Submit Job

EXAMPLE

CARD output:

Data was complete genome of E. Coli strain

44 hits in total!

Let us take a closer look

RGI Criteria	ARO Term	SNP	Detection Criteria	AMR Gene Family	Drug Class	Resistance Mechanism	% Identity of Matching Region	% Length of Reference Sequence
Perfect	acrB		protein homolog model	resistance-nodulation-cell division (RND) antibiotic efflux pump	fluoroquinolone antibiotic, cephalosporin, glycylcycline, penam, tetracycline antibiotic, rifamycin antibiotic, phenicol antibiotic, disinfecting agents and antiseptics	antibiotic efflux	100.0	100.00
Perfect	Escherichia coli acrA		protein homolog model	resistance-nodulation-cell division (RND) antibiotic efflux pump	fluoroquinolone antibiotic, cephalosporin, glycylcycline, penam, tetracycline antibiotic, rifamycin antibiotic, phenicol antibiotic, disinfecting agents and antiseptics	antibiotic efflux	100.0	100.00
Perfect	Escherichia coli emrE		protein homolog model	small multidrug resistance (SMR) antibiotic efflux pump	macrolide antibiotic	antibiotic efflux	100.0	100.00
Perfect	kdpE		protein homolog model	kdpDE	aminoglycoside antibiotic	antibiotic efflux	100.0	100.00
Perfect	mdtA		protein homolog model	ATP-binding cassette (ABC) antibiotic efflux pump	nitroimidazole antibiotic	antibiotic efflux	100.0	100.00
Perfect	mdtG		protein homolog model	major facilitator superfamily (MFS) antibiotic efflux pump	phosphonic acid antibiotic	antibiotic efflux	100.0	100.00
Perfect	mdtH		protein homolog model	major facilitator superfamily (MFS) antibiotic efflux pump	fluoroquinolone antibiotic	antibiotic efflux	100.0	100.00
Perfect	H-NS		protein homolog model	major facilitator superfamily (MFS) antibiotic efflux pump, resistance-nodulation-cell division (RND) antibiotic efflux pump	macrolide antibiotic, fluoroquinolone antibiotic, cephalosporin, cephamycin, penam, tetracycline antibiotic	antibiotic efflux	100.0	100.00
Perfect	marA		protein homolog model	resistance-nodulation-cell division (RND) antibiotic efflux pump, General Bacterial Porin with reduced permeability to beta-lactams	fluoroquinolone antibiotic, monobactam, carbapenem, cephalosporin, glycylcycline, cephamycin, penam, tetracycline antibiotic, rifamycin antibiotic, phenicol antibiotic, penem, disinfecting agents and antiseptics	antibiotic efflux, reduced permeability to antibiotic	100.0	100.00
Perfect	ugd		protein homolog model	pmr phosphoethanolamine transferase	peptide antibiotic	antibiotic target alteration	100.0	100.00
Perfect	mdtA		protein homolog model	resistance-nodulation-cell division (RND) antibiotic efflux pump	aminocoumarin antibiotic	antibiotic efflux	100.0	100.00
Perfect	mdtB		protein homolog model	resistance-nodulation-cell division (RND) antibiotic efflux pump	aminocoumarin antibiotic	antibiotic efflux	100.0	100.00
Perfect	mdtC		protein homolog model	resistance-nodulation-cell division (RND) antibiotic efflux pump	aminocoumarin antibiotic	antibiotic efflux	100.0	100.00
Perfect	baeS		protein homolog model	resistance-nodulation-cell division (RND) antibiotic efflux pump	aminoglycoside antibiotic, aminocoumarin antibiotic	antibiotic efflux	100.0	100.00
Perfect	baeR		protein homolog model	resistance-nodulation-cell division (RND) antibiotic efflux pump	aminoglycoside antibiotic, aminocoumarin antibiotic	antibiotic efflux	100.0	100.00
Perfect	YojI		protein homolog model	ATP-binding cassette (ABC) antibiotic efflux pump	peptide antibiotic	antibiotic efflux	100.0	100.00
Perfect	PmrF		protein homolog model	pmr phosphoethanolamine transferase	peptide antibiotic	antibiotic target alteration	100.0	100.00
Perfect	emrY		protein homolog model	major facilitator superfamily (MFS) antibiotic efflux pump	tetracycline antibiotic	antibiotic efflux	100.0	100.00
Perfect	emrK		protein homolog model	major facilitator superfamily (MFS) antibiotic efflux pump	tetracycline antibiotic	antibiotic efflux	100.0	110.26
Perfect	evgA		protein homolog model	major facilitator superfamily (MFS) antibiotic efflux pump, resistance-nodulation-cell division (RND) antibiotic efflux pump	macrolide antibiotic, fluoroquinolone antibiotic, penam, tetracycline antibiotic	antibiotic efflux	100.0	100.00
Perfect	evgS		protein homolog model	major facilitator superfamily (MFS) antibiotic efflux pump, resistance-nodulation-cell division (RND) antibiotic efflux pump	macrolide antibiotic, fluoroquinolone antibiotic, penam, tetracycline antibiotic	antibiotic efflux	100.0	100.00
Perfect	acrD		protein homolog model	resistance-nodulation-cell division (RND) antibiotic efflux pump	aminoglycoside antibiotic	antibiotic efflux	100.0	100.00
Perfect	emrR		protein homolog model	major facilitator superfamily (MFS) antibiotic efflux pump	fluoroquinolone antibiotic	antibiotic efflux	100.0	100.00
Perfect	emrA		protein homolog model	major facilitator superfamily (MFS) antibiotic efflux pump	fluoroquinolone antibiotic	antibiotic efflux	100.0	100.00
Perfect	emrB		protein homolog model	major facilitator superfamily (MFS) antibiotic efflux pump	fluoroquinolone antibiotic	antibiotic efflux	100.0	100.00

The Fleming Fund
Regional Grants

EXAMPLE CARD output:

- EmrY, emrK and emrB
- Perfect hits!
  - Expect for emrK, ID and COV are 100%
- Should we expect resistance to tetracycline and fluoroquinolones in this isolate?

RGI Criteria	ARO Term	SNP	Detection Criteria	AMR Gene Family
Perfect	emrY		protein homolog model	major facilitator superfamily (MFS) antibiotic efflux pump
Perfect	emrK		protein homolog model	major facilitator superfamily (MFS) antibiotic efflux pump
Perfect	emrB		protein homolog model	major facilitator superfamily (MFS) antibiotic efflux pump

Drug Class	Resistance Mechanism	% Identity of Matching Region	% Length of Reference Sequence
tetracycline antibiotic	antibiotic efflux	100.0	100.00
tetracycline antibiotic	antibiotic efflux	100.0	110.26
fluoroquinolone antibiotic	antibiotic efflux	100.0	100.00

The Fleming Fund
Regional Grants

Lets try a different tool for the strain: ResFinder

- No resistance at all?

ResFinder-4.1 Server - Results

Input Files: *GCF_000005845.2_ASM584v2_genomic.fna*

Warning:
One or more resistance genes does not exist in the phenotype database. The Summary table does not take this into account.

escherichia coli | complete

Antimicrobial	Class	WGS-predicted phenotype	Genetic background
amikacin	aminoglycoside	No resistance	
tigecycline	tetracycline	No resistance	
tobramycin	aminoglycoside	No resistance	
cefepime	beta-lactam	No resistance	
chloramphenicol	amphenicol	No resistance	
piperacillin+tazobactam	beta-lactam	No resistance	
cefoxitin	beta-lactam	No resistance	
ampicillin	beta-lactam	No resistance	
ampicillin+clavulanic acid	beta-lactam	No resistance	
cefotaxime	beta-lactam	No resistance	
ciprofloxacin	quinolone	No resistance	
colistin	polymyxin	No resistance	
sulfamethoxazole	folate pathway antagonist	No resistance	
imipenem	beta-lactam	No resistance	
trimethoprim	folate pathway antagonist	No resistance	
nalidixic acid	quinolone	No resistance	
ertapenem	beta-lactam	No resistance	
tetracycline	tetracycline	No resistance	
fosfomycin	fosfomycin	No resistance	
ceftazidime	beta-lactam	No resistance	
temocillin	beta-lactam	No resistance	
gentamicin	aminoglycoside	No resistance	
meropenem	beta-lactam	No resistance	
azithromycin	macrolide	No resistance	

The Fleming Fund
Regional Grants

Lets try a different tool for the strain: ResFinder

- No resistance at all?
- No resistance to tetracycline or quinolones?

ResFinder-4.1 Server - Results

Input Files: *GCF_000005845.2_ASM584v2_genomic.fna*

Warning:
One or more resistance genes does not exist in the phenotype database. The Summary table does not take this into account.

escherichia coli | complete

Antimicrobial	Class	WGS-predicted phenotype	Genetic background
amikacin	aminoglycoside	No resistance	
tigecycline	tetracycline	No resistance	
tobramycin	aminoglycoside	No resistance	
cefepime	beta-lactam	No resistance	
chloramphenicol	amphenicol	No resistance	
piperacillin+tazobactam	beta-lactam	No resistance	
cefoxitin	beta-lactam	No resistance	
ampicillin	beta-lactam	No resistance	
ampicillin+clavulanic acid	beta-lactam	No resistance	
cefotaxime	beta-lactam	No resistance	
ciprofloxacin	quinolone	No resistance	
colistin	polymyxin	No resistance	
sulfamethoxazole	folate pathway antagonist	No resistance	
imipenem	beta-lactam	No resistance	
trimethoprim	folate pathway antagonist	No resistance	
nalidixic acid	quinolone	No resistance	
ertapenem	beta-lactam	No resistance	
tetracycline	tetracycline	No resistance	
fosfomycin	fosfomycin	No resistance	
ceftazidime	beta-lactam	No resistance	
temocillin	beta-lactam	No resistance	
gentamicin	aminoglycoside	No resistance	
meropenem	beta-lactam	No resistance	
azithromycin	macrolide	No resistance	

The Fleming Fund
Regional Grants

Lets try a different tool for the strain: ResFinder

- No resistance at all?
- No resistance to tetracycline or quinolones?
- One tool gives 44 hits, another gives 0
  what is the truth?

ResFinder-4.1 Server - Results

Input Files: *GCF_000005845.2_ASM584v2_genomic.fna*

Warning:
One or more resistance genes does not exist in the phenotype database. The Summary table does not take this into account.

escherichia coli | complete

Antimicrobial	Class	WGS-predicted phenotype	Genetic background
amikacin	aminoglycoside	No resistance	
tigecycline	tetracycline	No resistance	
tobramycin	aminoglycoside	No resistance	
cefepime	beta-lactam	No resistance	
chloramphenicol	amphenicol	No resistance	
piperacillin+tazobactam	beta-lactam	No resistance	
cefoxitin	beta-lactam	No resistance	
ampicillin	beta-lactam	No resistance	
ampicillin+clavulanic acid	beta-lactam	No resistance	
cefotaxime	beta-lactam	No resistance	
ciprofloxacin	quinolone	No resistance	
colistin	polymyxin	No resistance	
sulfamethoxazole	folate pathway antagonist	No resistance	
imipenem	beta-lactam	No resistance	
trimethoprim	folate pathway antagonist	No resistance	
nalidixic acid	quinolone	No resistance	
ertapenem	beta-lactam	No resistance	
tetracycline	tetracycline	No resistance	
fosfomycin	fosfomycin	No resistance	
ceftazidime	beta-lactam	No resistance	
temocillin	beta-lactam	No resistance	
gentamicin	aminoglycoside	No resistance	
meropenem	beta-lactam	No resistance	
azithromycin	macrolide	No resistance	

Differences in output example

- The strain run in this example is a standard laboratory strain E. coli K-12 substrain MG1655

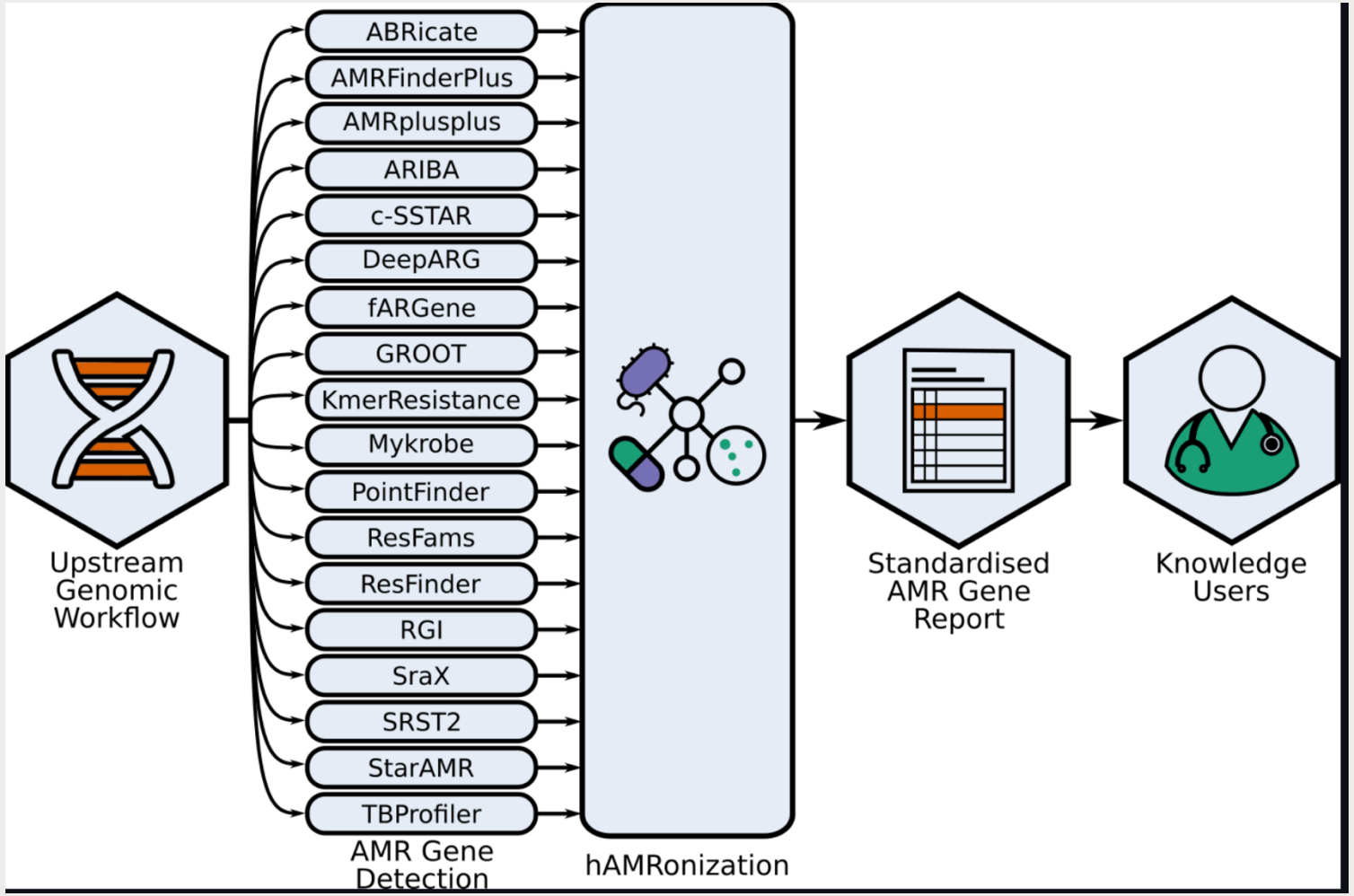
- It is not expected to have any phenotypic resistance to tetracycline (Zhang et al., 2022)
  - Not actually expected to have any particular phenotypic resistance different from wild-type
    - e. coli

- If run on AMRfinderplus, no resistance genes are found either.

- Approach databases with care and select based on your scope
  - How does results translate to the laboratory, genotypic =/= phenotypic
  - How much expertise is demanded to utilize findings
  - What is the aim of your analysis

The Fleming Fund
Regional Grants

hAMRonization



https://github.com/pha4ge/hAMRonization

The Fleming Fund
Regional Grants

Let’s take a break ☺

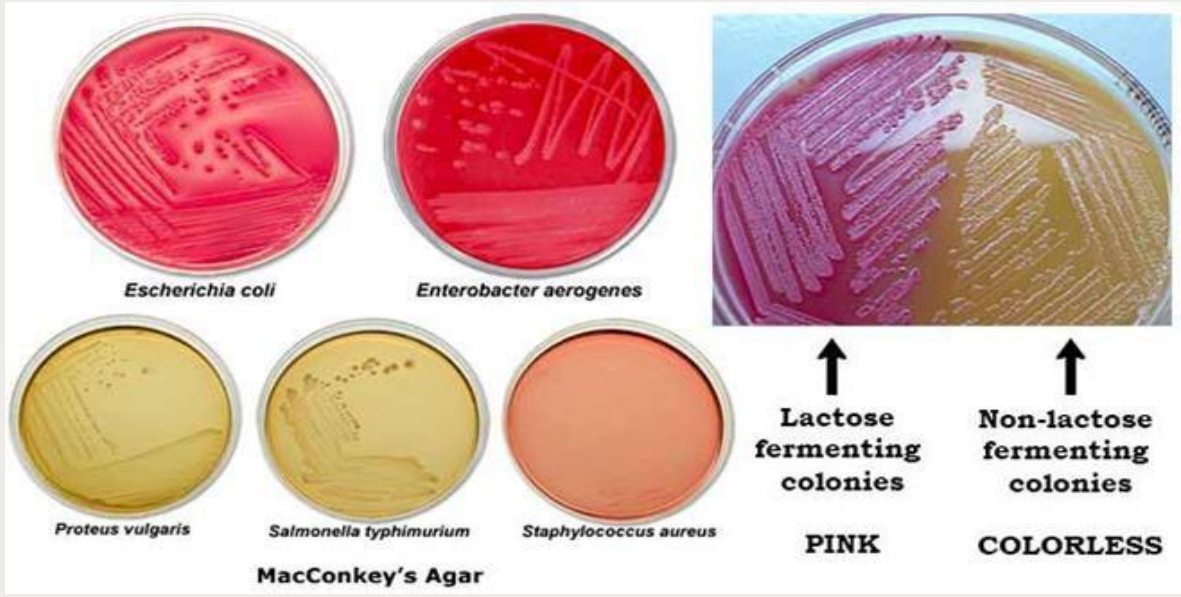
The Fleming Fund
Regional Grants

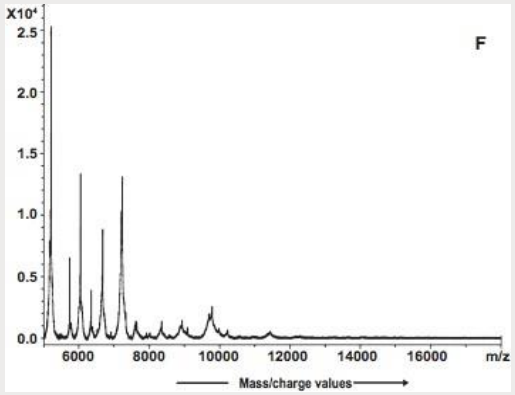
Typing methods

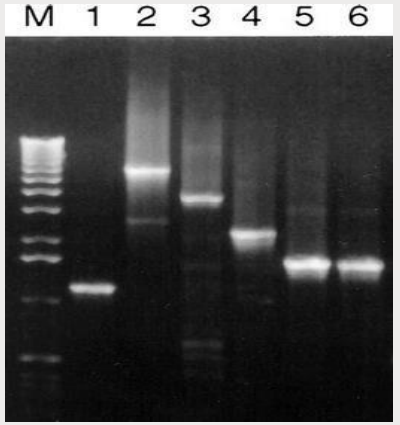
Introduction to typing methods

- Phenotypic identification
  - Biochemical/metabolic analysis
  - Chromogenic media
  - AST
  - CIM test

- Molecular identification
  - PCR (genus/species/AST)
  - MALDI-TOF MS
  - Microarray (AMR)
  - MLST (PCR/Sequencing)







Working with typing and outbreak detection

Purpose of subtyping?

- Genus/Species determination
- Serotyping and MLST
  – Characterization and grouping of isolates
- cgMLST and SNP analysis
  – Comparison
- Resistance patterns
  – pMLST – plasmids
  – Specific genes or combinations

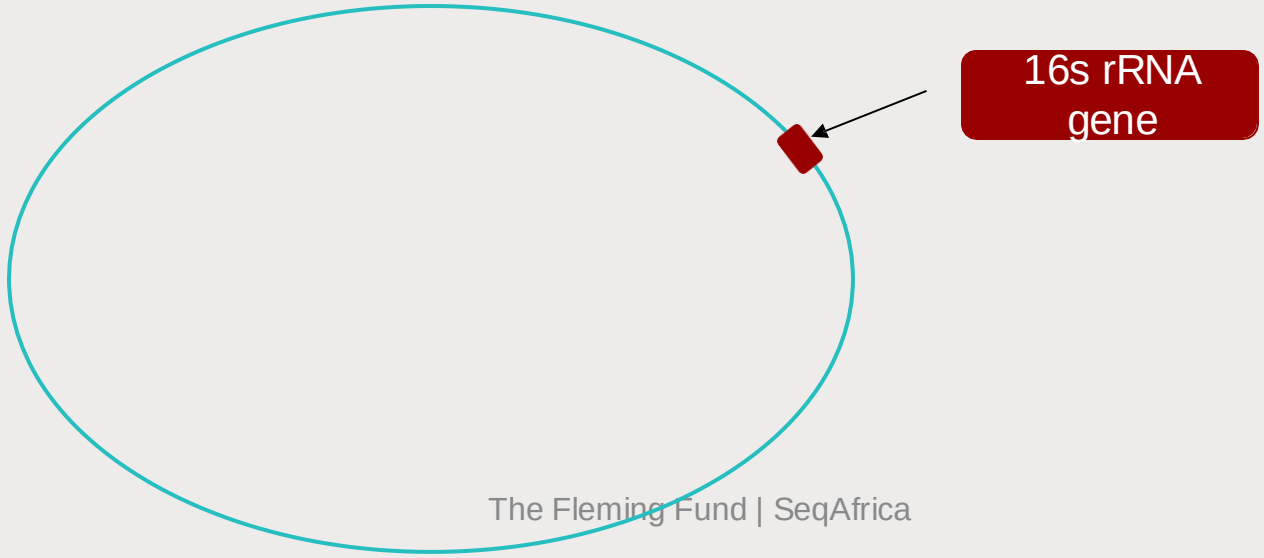
Higher resolution / discriminatory power

Typing tools on CGE server

- Species
  - KmerFinder (full genome) and/or SpeciesFinder (16s rRNA)
- Sub-typing
  - Serotyping (*E. coli, P. aeruginosa, Salmonella*)
- Typing
  - MLST
  - cgMLSTFinder
    - *Campylobacter, Clostridum, E. coli, Listeria, Salmonella, Yersinia*
  - pMLST
  - Plasmidfinder
  - ViruelIncefinder
  - MGE
- Cluster analysis
  - CSIPhylogeny & MinTyper

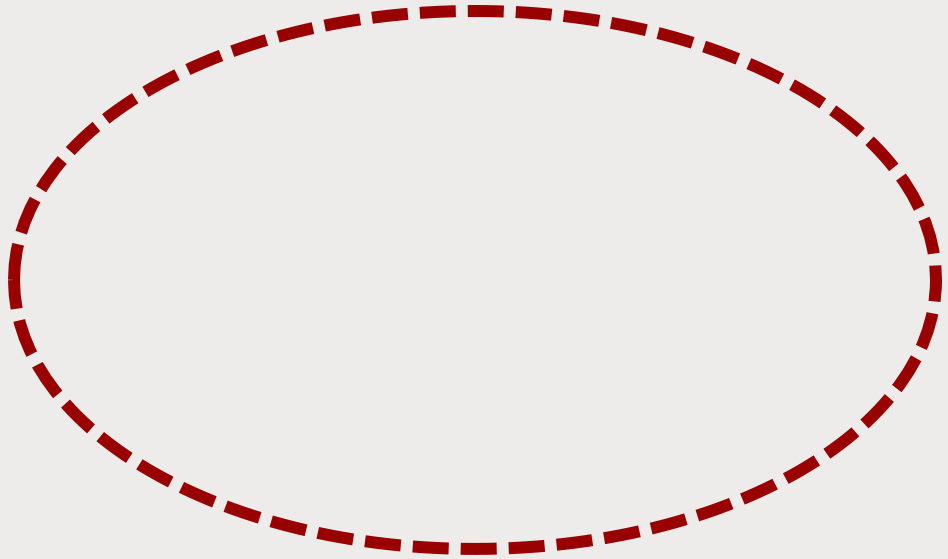
Genotypic species verification

- 16s rRNA gene formed the basis as the first method for sequenced based taxonomy
- Other approaches:
  - gyrB gene, rMLST, species-specific functional domain profiles
    - Only represents a small fraction of the entire genome
  - WGS data can provide higher discriminatory power e.g. between *Shigella* and *Escherichia* spp.



Prediction of species - Kmerfinder

- With WGS we can use all the genetic information to predict the species
- Kmerfinder works by breaking a genome into little pieces (k-mers) and identifying the species from these pieces (k-mers)



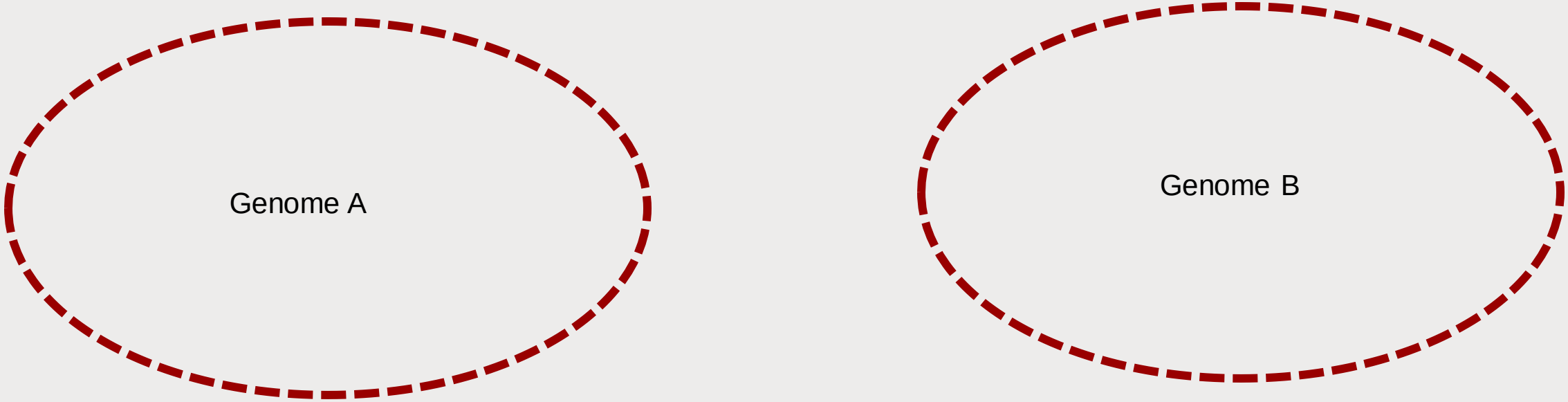
Reminder: k-mers

- A k-mer is a continuous sequence of k bases
  – e.g a certain length of DNA, RNA or protein
- There are 4^k combinations of a k-mer
- Using long k-mers provides a highly unique sequence
- Sequences with high similarity must share k-mers

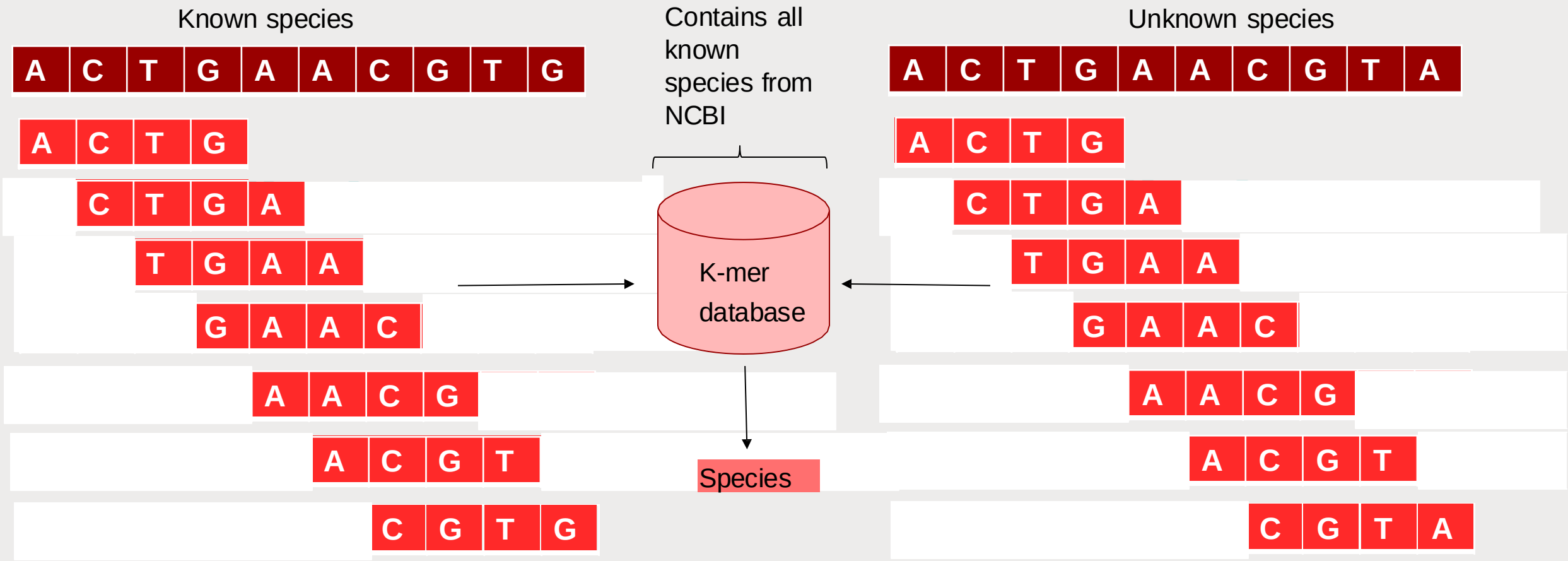
- We can extract all the 4-mers (substrings of length 4) in this DNA sequence

Species prediction with k-mers

- Sequences with high similarity must share k-mers
- We can break genomes up into k-mers and compare them



Species prediction with KmerFinder



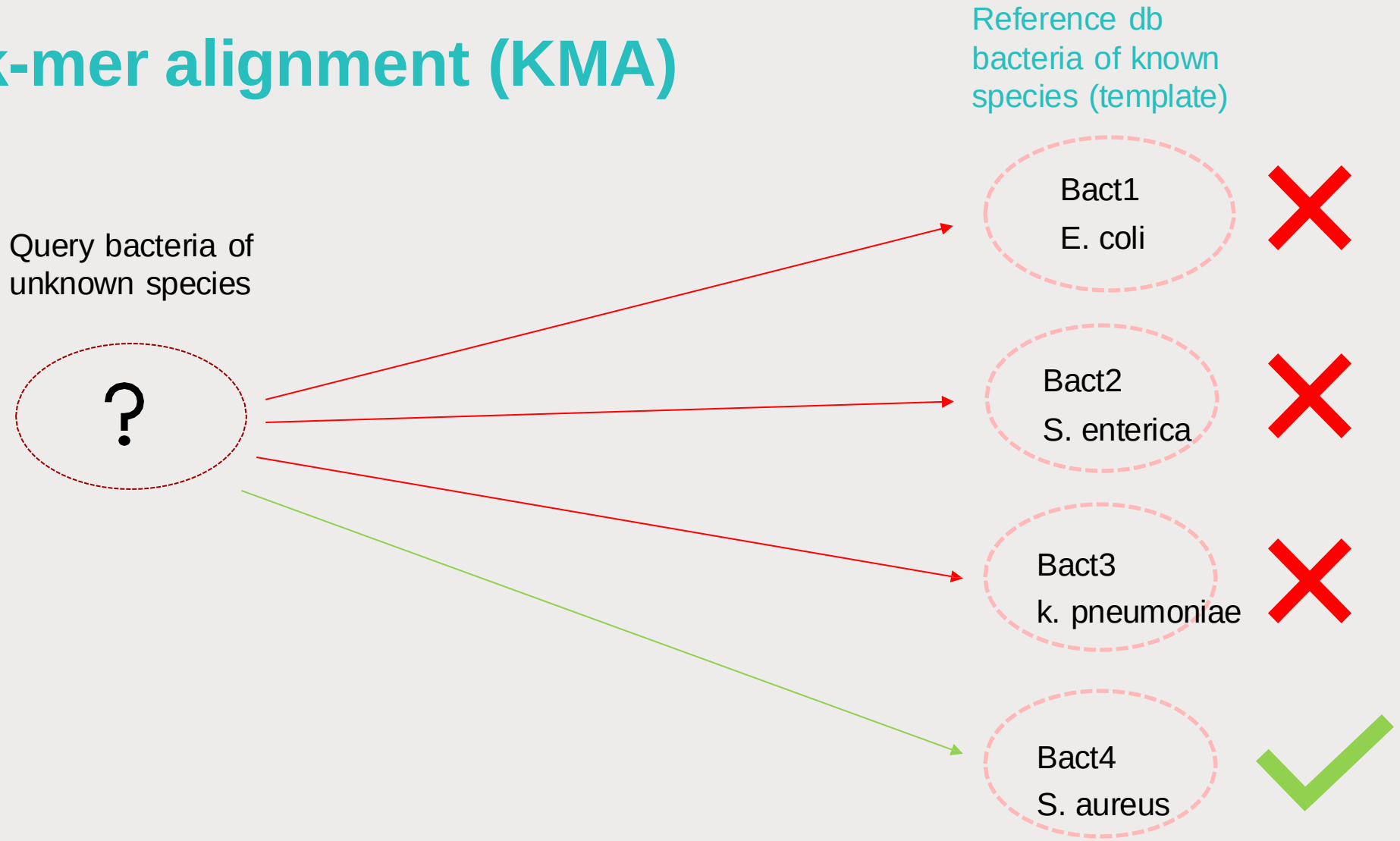
Species prediction with KmerFinder

- Genomes are spilt into 16-mers
  - 4.3 billion combinations
  - ~10.000 recognized spp.
  - ~600.000 bacterial operation taxonomic units.

- Only 16-mers with specific prefixes are kept e.g ATGAG
  - Reducing k-mers reduces size of database.
  - Speed-up computing time.
  - Reduces redundancy.

- But how does the tool the compare k-mers?

k-mer alignment (KMA)



Adapted from: Introduction to CGE tools, presentation by Pimpalas Leekitcharoenphon

The Fleming Fund
Regional Grants

KmerFinder webtool

Select database

Bacteria organisms

Upload file(s)

To input the sequences, upload a single FASTA file, or one/two FASTQ file(s), or one interleaved FASTQ file on your local disk by using the applet below. Both assembled genome (in FASTA format) and raw reads single end or paired end (in FASTQ format) are supported. Gzipped FASTA/FASTQ files are also supported.

If you get an "Access forbidden. Error 403": Make sure the start of the web adress is https and not just http. Fix it by clicking here.

Choose File(s)

Name	Size	Progress	Status

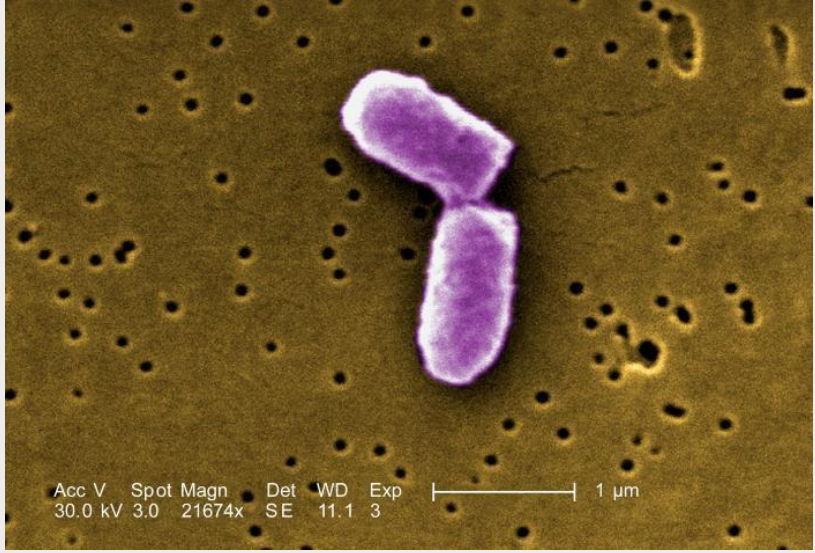
Upload Remove

Subtyping of bacteria

- Information of bacteria below species level
  - Outbreak detection, clusters, common contamination source, transmission routes…

  - E. coli/Salmonella - traditional subtyping:
    - serotyping using antisera against the ca. 186 O-antigens and 53 H-flagellar antigens for *E. coli* or 46 O-antigens and 114 H-antigens for *Salmonella* (ca 2600 serovars).
      - Requires anti-sera and trained personnel.
      - Time consuming and not always accurate or inconclusive.

    - Phagetyping:
      - Golden standard method for surveillance of *Salmonella* Typhimurium and *S.*
        - Enteritidis – also used for *E. coli* and other bacteria.
      - Requires a comprehensive panel of different phages, considerable technical expertise.

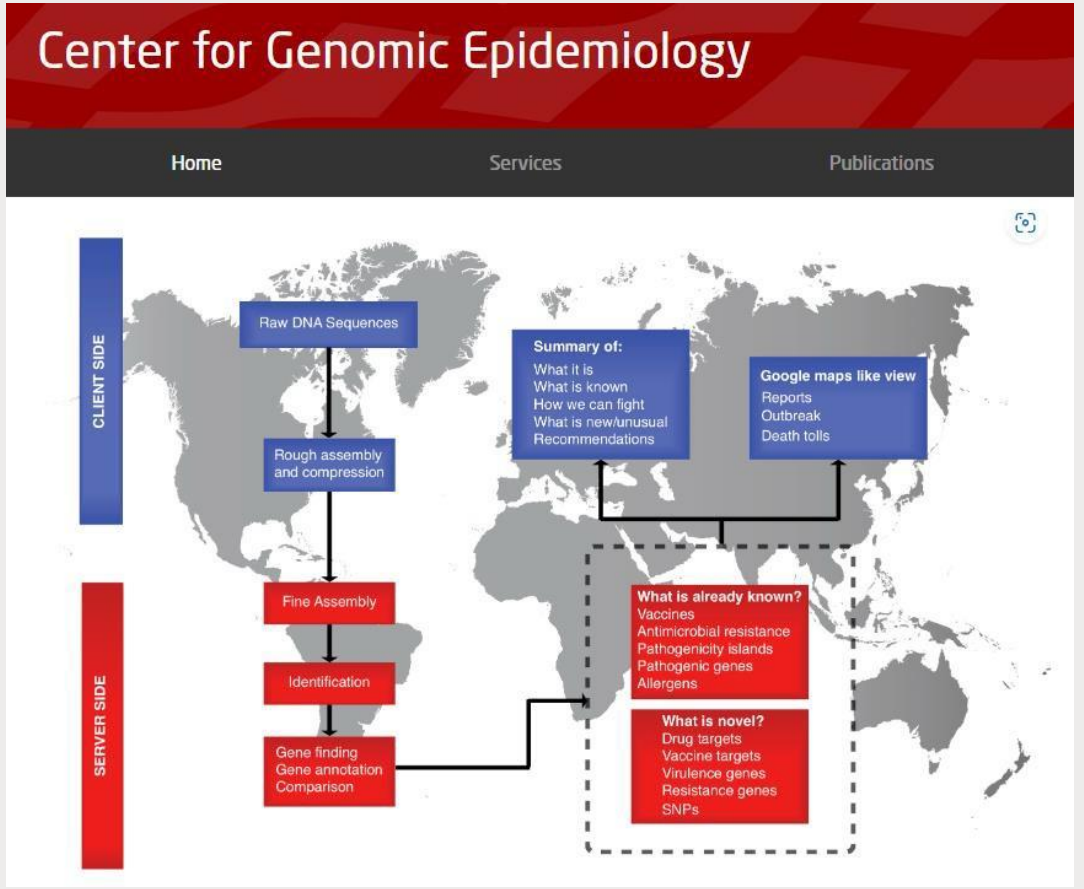


E. coli in scanning electron microscopic image, CDC/ Evangeline Sowers, Janice Haney Carr, 2005, Public domain image, https://phil.cdc.gov/Details.aspx?pid=10042

Genotypic determination of classical methods

- E. coli serotyping:
- DTU developed tool:
- https://cge.food.dtu.dk/services/SerotypeFinder/

- Salmonella subtyping:
- DTU hosted tool:
- https://cge.food.dtu.dk/services/SeqSero/



SerotypeFinder 2.0

SerotypeFinder identifies the serotype in total or partial sequenced isolates of E. coli.
Fasta file with test sequence: Test_sequence

The database is curated by:
Flemming Scheutz, SSI
(click to contact)

Software version: 2.0.1 (2020-07-27)
Database version: 1.0.0 (2022-05-16)

Select organism
Select multiple items, with Ctrl-Click (or Cmd-Click on Mac)
E. coli

Select threshold for %ID
85 %

Select minimum length
The minimum length is the number of nucleotides a sequence must overlap a serotype gene to count as a hit for that gene. Here represented as a percentage of the total serotype gene length.
60 %

Select type of your reads
Only data from one single isolate should be uploaded. If raw sequencing reads are uploaded KMA will be used for mapping. KMA supports the following sequencing platforms: Illumina, Ion Torrent, Roche 454, SOLiD, Oxford Nanopore, and PacBio.
Assembled or Draft Genome/Contigs* (fasta)

Isolate File

Name	Size	Progress	Status

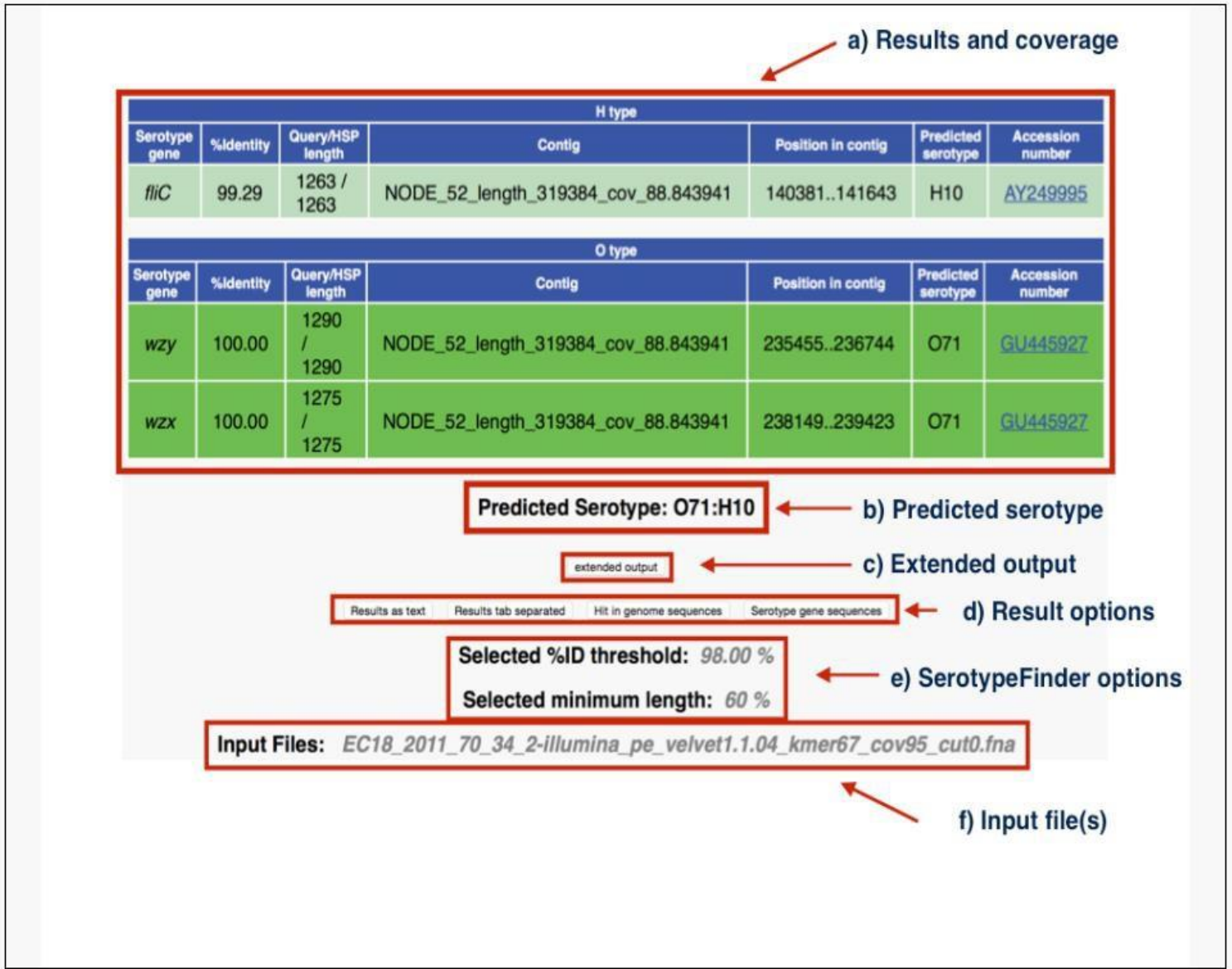
Upload Remove

The Fleming Regional Grants

SerotypeFinder 2.0 Output Guide

Graphical output example and explanation

Once the SerotypeFinder server has finished running the job you submitted, it will display an output similar to the below example:



Analysis of mobile genetic elements

- PlasmidFinder.
  - Tool for identification of replicons
  - Plasmid replicons are divieded into incompability groups.
  - Plasmids which share the same replication mechanisms cannot be maintained in the same cell
  - Use fasta files as input to gain insight into linkage to AMR genes

- MGE (Mobile Genetic Element Finder)
  - Identifies MGEs in the genome
  - Provides information on virulence and AMR genes contained in identified MGEs
  - Takes fasta files as input

PlasmidFinder 2.1

Service | Instructions | Output | Article abstract | Citations

Software version: 2.0.1 (2020-07-01)
Database version: (2023-01-18)
Test sequence

The database is curated by:
Henrik Hasman and Alessandra Carattoli
(click to contact)

Select database

Gram Positive
Enterobacteriales

Select threshold for minimum % identity

95 %

Select minimum % coverage

60 %

Select type of your reads

Only data from one single isolate should be uploaded. If raw sequencing reads are uploaded KMA will be used for mapping. KMA supports the following sequencing platforms: Illumina, Ion Torrent, Roche 454, SOLiD, Oxford Nanopore, and PacBio.

Assembled or Draft Genome/Contigs*

Choose File(s)

Name	Size	Progress	Status

Upload Remove

The Fleming Fund
Regional Grants

If the replicon is found on the same contig as a AMR gene, it indicates the gene is on a plasmid

PlasmidFinder-2.0 Server - Results

Organism(s): *Enterobacteriaceae*

Enterobacteriaceae,Acenitobacter baumannii						
Plasmid	Identity	Query / Template length	Contig	Position in contig	Note	Accession number
IncFIB(AP001918)	96.84	538 / 682	NODE_151_length_1547_cov_574.472534	1..538		AP001918
IncFII(pRSB107)	97.7	261 / 261	NODE_103_length_1790_cov_579.962585	539..799		AJ851089
IncI1-I(Gamma)	97.89	142 / 142	NODE_266_length_500_cov_522.737976	61..202		AP005147

extended output

Input Files: *resfindertest.fa*

Results as text | Results tsv | Hits in genome seqs | Plasmid sequences

The database is curated by:
Markus Johansson
(click to contact)

Software version: v1.0.3 (2020-10-09)
Database version: v1.0.2 (2020-06-09)

MobileElementFinder identifies **mobile genetic elements** and their relation to **antimicrobial resistance genes** and **virulence factors**.

Example sequence

Annotate accessory genes (Optional)
If you want to use databases currently not supported by MobileElementFinder, please download the mobile element sequences and upload them to the service of choice.

- [x] Acquired Antimicrobial Resistance genes (ResFinder)
- [x] Virulence genes (VirulenceFinder)

Isolate File

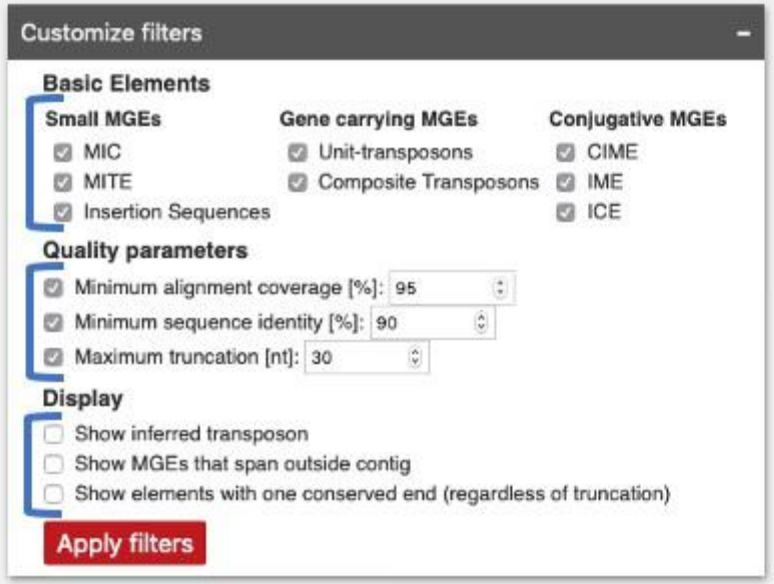
Name	Size	Progress	Status

Upload Remove

1a. Display MGE types

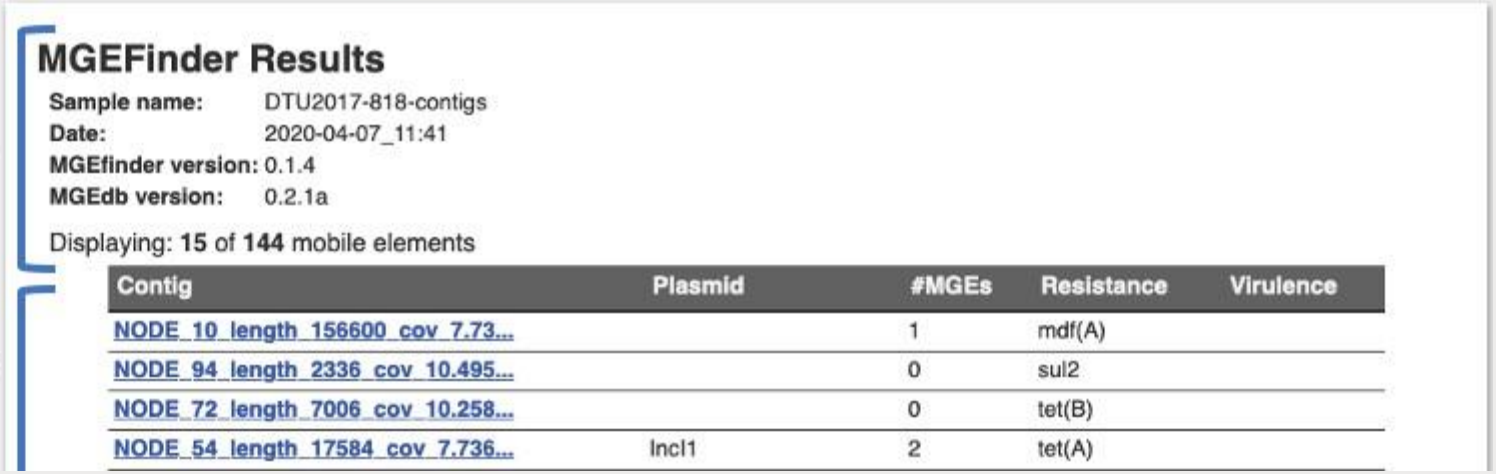
1b. Prediction quality

1c. Display special cases



2. Sample information

3. Result overview



Contig	Plasmid	#MGEs	Resistance	Virulence
NODE_10_length_156600_cov_7.73...		1	mdf(A)	
NODE_94_length_2336_cov_10.495...		0	sul2	
NODE_72_length_7006_cov_10.258...		0	tet(B)	
NODE_54_length_17584_cov_7.736...	IncI1	2	tet(A)	

The Fleming Fund Regiona

Contig result view

1. Genes on contig

2. MGEs on contig

2a. MGE information

2b. Prediction metrics

Contig: NODE_54_length_17584_cov_7.73695_ID_6293

Plasmid results

Plasmid name	Database	Accession	Position in contig	Coverage	Identity
IncI1	Enterobacteriaceae	AP005147	7055-7196	100%	99.3%

Resistance results

Gene name	Phenotype	Accession	Position in contig	Coverage	Identity
tet(A)	Tetracycline resistance	AJ517790	12904-14103	100%	100%

IS26

Synonyms	IS6,IS26L,IS26R,IS46,IS140,IS160
Family	IS6
Type	Insertion sequence
Reference db	isfinder
Accession	X00011
Position in contig	15498-16317
Strand	forward
Read depth	7.74
Alignment coverage	100%; 820 / 820
Sequence identity	100%
Num Substitutions	0
E-value	0

Show MGE alignment

ISSbo1

Family	IS91
Type	Insertion sequence
Reference db	isfinder
Accession	CP001062
Position in contig	9195-10903
Strand	forward
Read depth	7.74
Alignment coverage	100%; 1709 / 1709
Sequence identity	96.02%
Num Substitutions	68
E-value	0

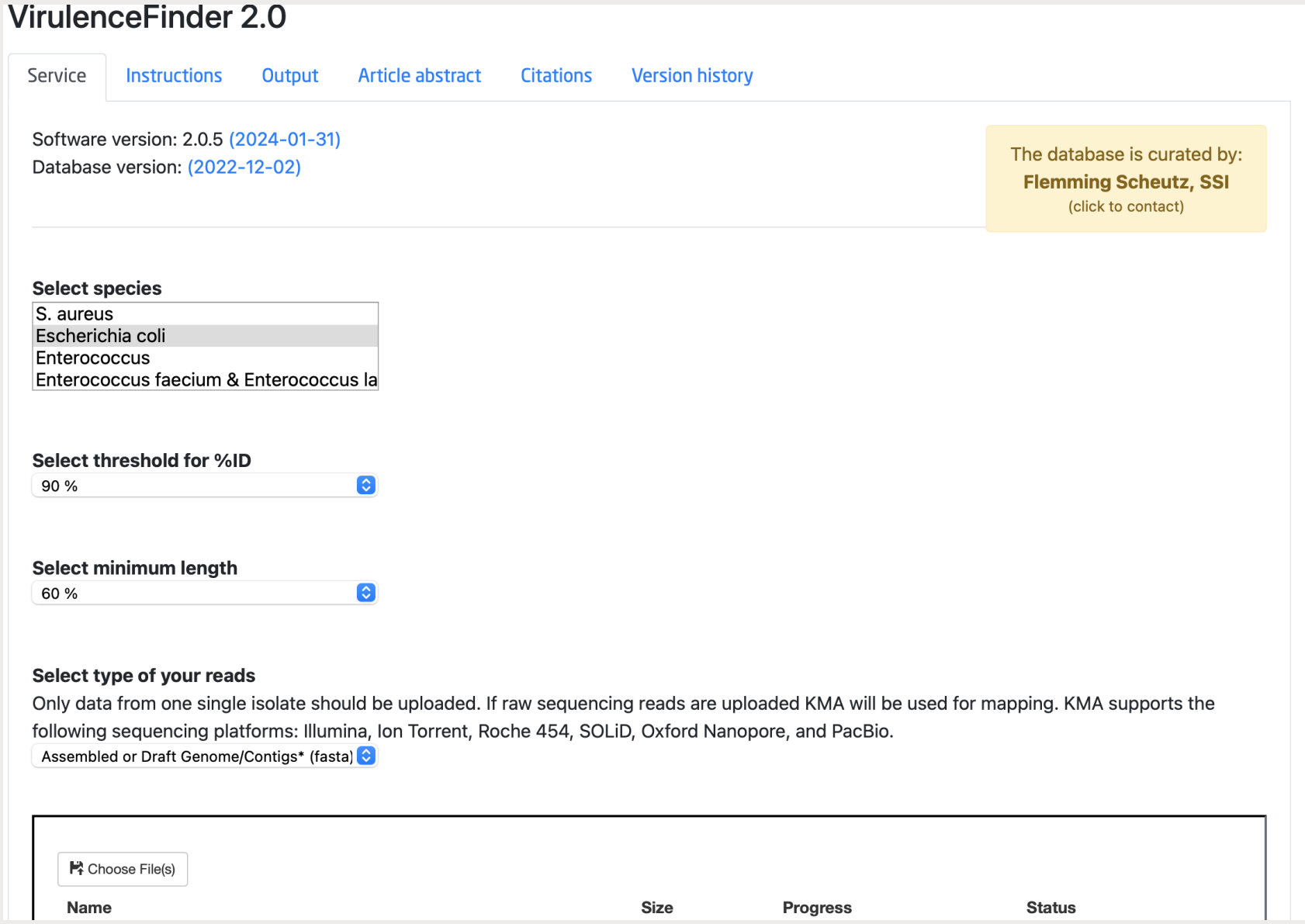
Show MGE alignment

The Fleming Fund
Regional Grants

Virulence Finder

- Detects virulence genes

- Virulence genes are genes that help bacteria establish infections in their hosts.
- These genes encode proteins that help bacteria colonize and survive in the host or damage the host.



The Fleming Fund
Regional Grants

VirulenceFinder-1.2 Server - Results

SETTINGS:
Selected %ID threshold: *98.00*

Virulence - E. coli						
Virulence factor	**%Identity**	**Query/HSP length**	**Contig**	**Position in contig**	**Protein function**	**Accession number**
mcmA	99.64	279 / 279	NODE_17_length_48340_cov_62.616714	40909..41187	Microcin M part of colicin H	AJ515251
lpfA	100.00	573 / 573	NODE_4_length_115337_cov_62.053581	84857..85429	Long polar fimbriae	KC207123
iss	99.71	342 / 342	NODE_195_length_89121_cov_54.610832	87701..88042	Increased serum survival	CU928160
prfB	100.00	882 / 882	NODE_75_length_157387_cov_57.585850	94324..95205	P-related fimbriae regulatory gene	CP002970

extended output

Results as text | Results tab separated | Hit in genome sequences | Virulence gene sequences

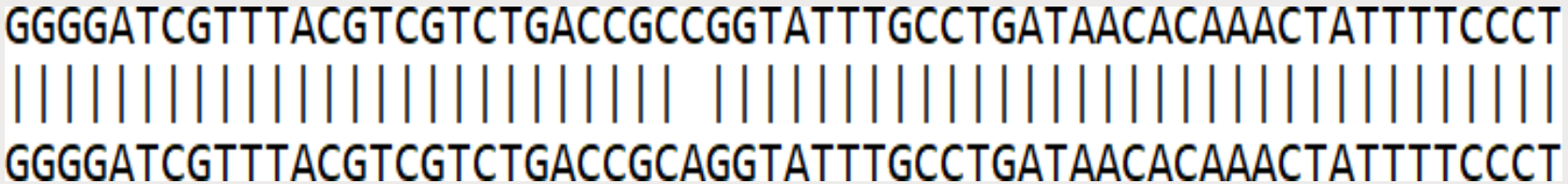
Input Files: *EC19_2011_70_34_3-illumina_pe_velvet1.1.04_kmer63_cov57_cut0.fna*

Sequence identity

- A term we encounter in the cge tools is % identity (ID)

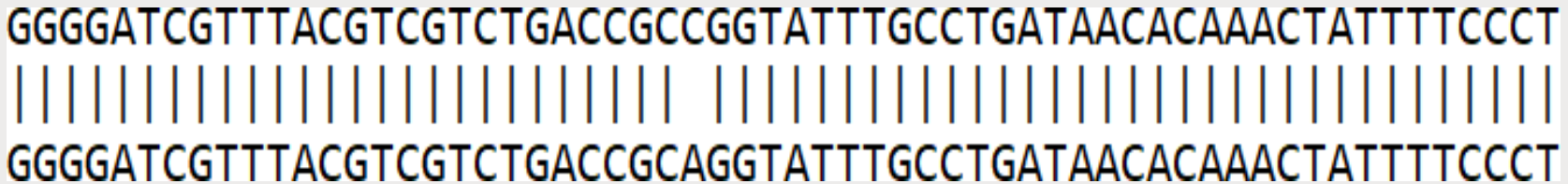
- The identity describes how many bases of the aligned sequences are identical

- Given the alignment:



Sequence identity

- A term we encounter in the cge tools is % identity (ID)
- The identity describes how many bases of the aligned sequences are identical
- Given the alignment:
- Sequence length 60
- Matches 59
- %ID = 59/60*100% = 98.3%



Sequence coverage

- The term sequence % coverage (COV) refers to the proportion of covered gene
- Given the alignment:
- Sequence length 60
- Covered positions are 27
- %COV = 27/60*100% = 45.0%

```
GGGGATCGTTTACGTCGTCTGACCGCCGGTATTTGCCTGATAACACAAACTATTTTCCCT
||||||||||||||||||||||||||
GGGGATCGTTTACGTCGTCTGACCGC
```

The Fleming Fund
Regional Grants

Let’s take a break ☺

The Fleming Fund
Regional Grants

Thank you

This programme is being funded by the UK Department of Health and Social Care.
The views expressed do not necessarily reflect the UK Government's official policies.