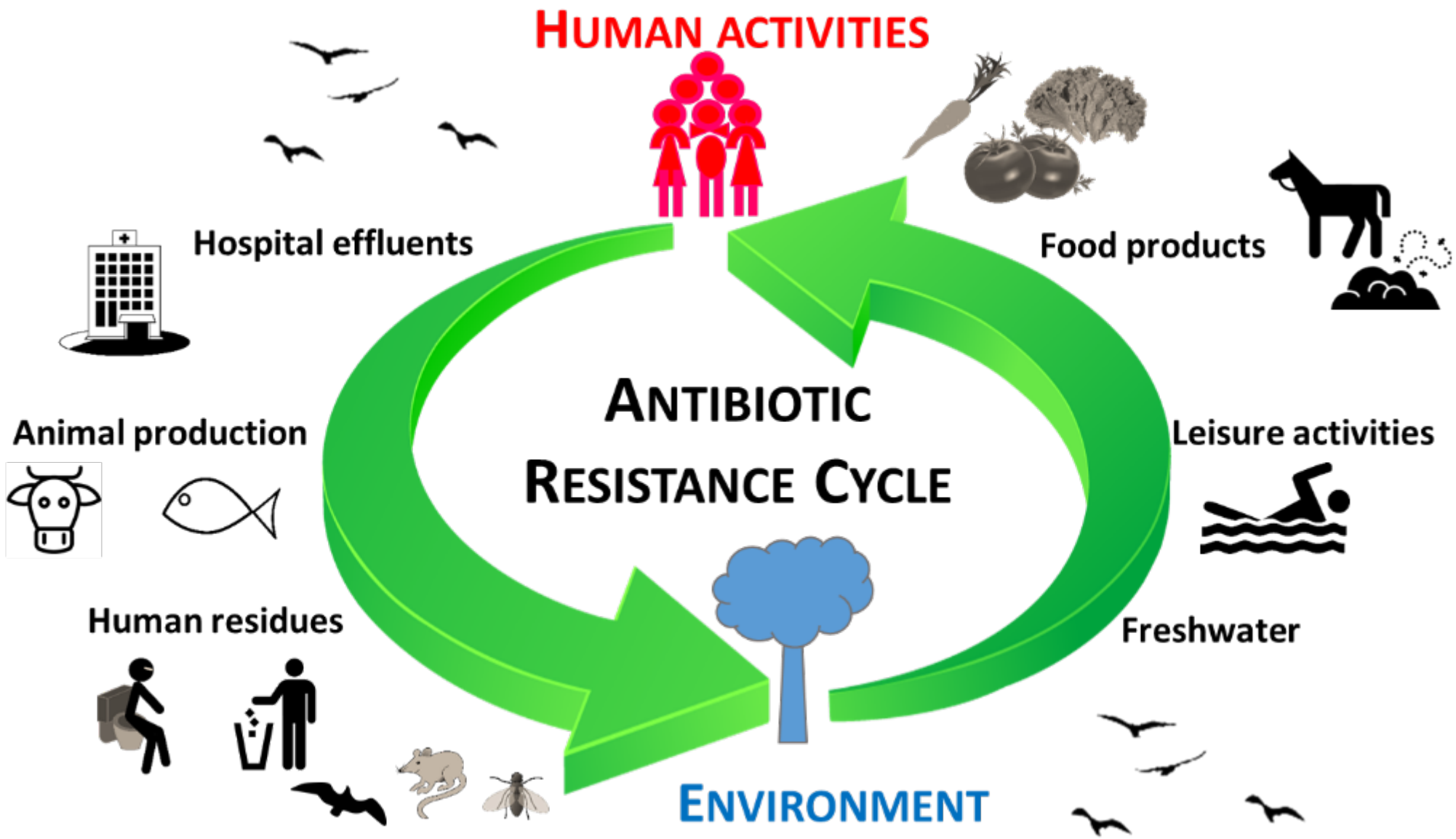
Day 2
AMR Gene Annotation

Arun Gonzales Decano
Senior Bioinformatician
NDM Experimental Medicine
University of Oxford, UK

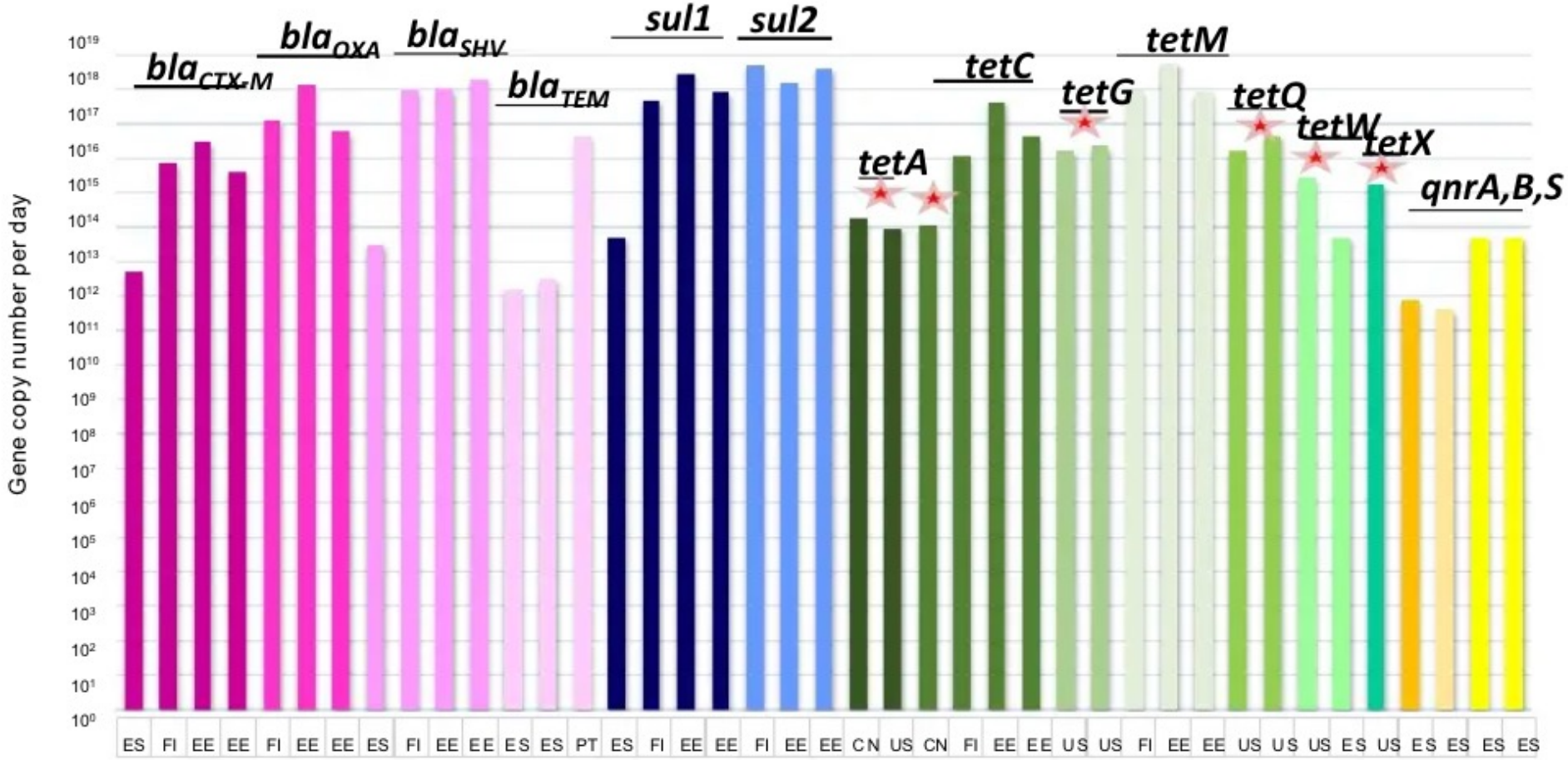
SAGESA

AMR- from nature to environmental contaminant



https://www.envchemgroup.com/celia-manaia.html

It is estimated that more than 10^{10} to 10^{14} copies of genes encoding for tetracycline or beta-lactam resistance are released per minute to the surrounding environment

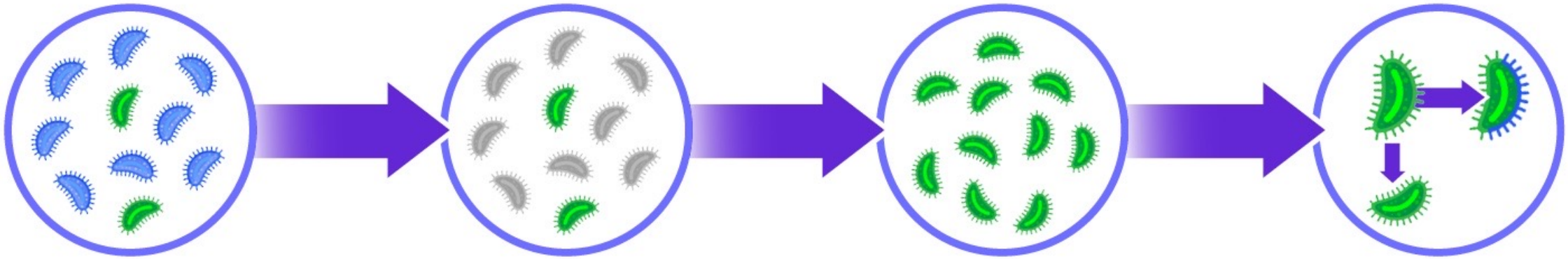


CATÓLICA
ESCOLA SUPERIOR DE BIOTECNOLOGIA
PORTO

Manaia et al., 2016. Applied Microbiology and Biotechnology. 100:1543 –1557

AMERICAN SOCIETY FOR MICROBIOLOGY

Development of Antimicrobial Resistance (AMR)



Antimicrobial products are used to kill or significantly slow the growth of disease-causing microbes.

Under certain conditions, selective pressure drives evolution of mechanisms that allow some microbes to resist antimicrobial activity.

Resistant microbes are able to survive antimicrobial treatment and continue to replicate.

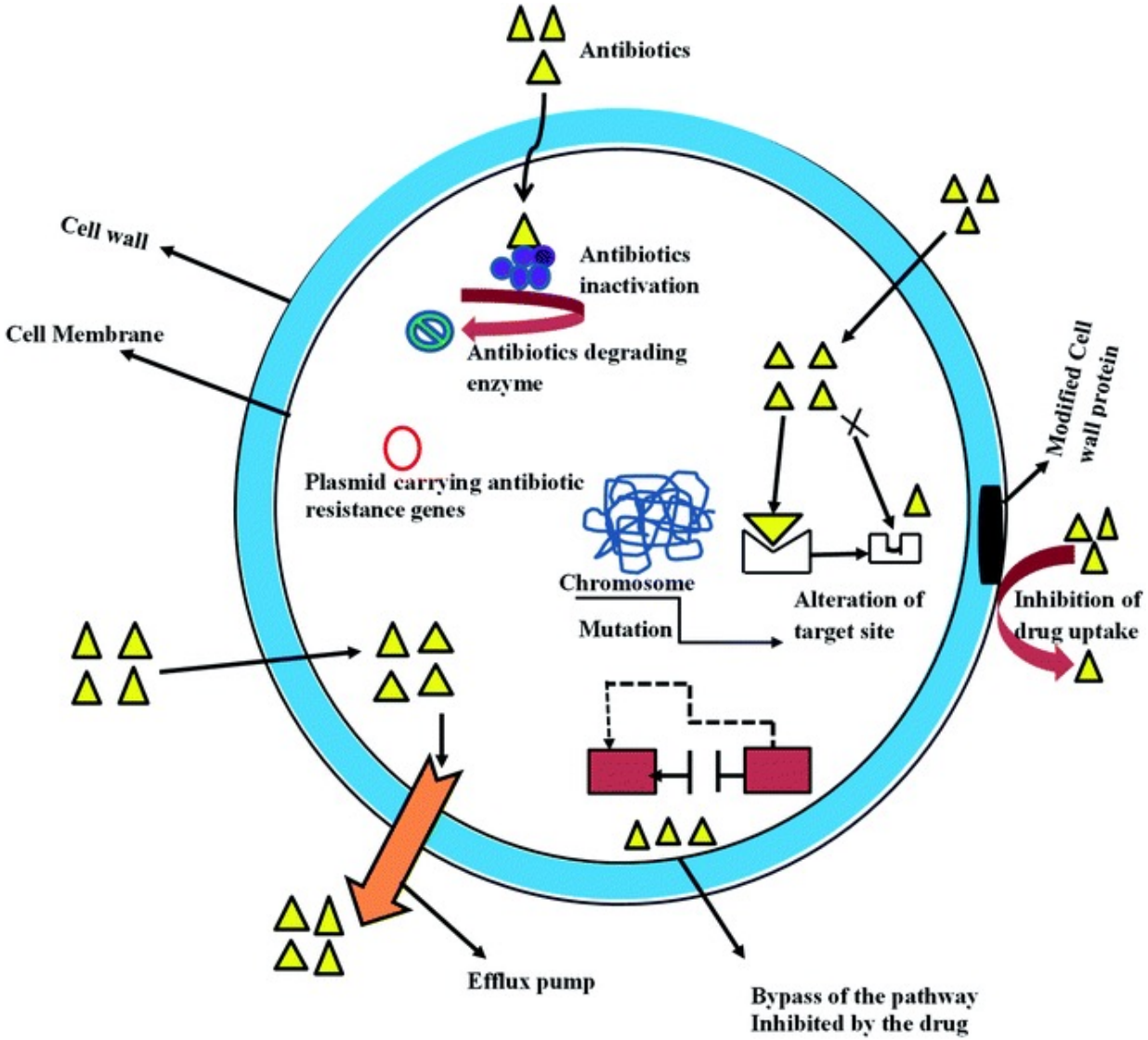
AMR microbes pass resistance genes to other microbes via vertical and/or horizontal transfer, increasing both the quantity and type of resistant pathogens.

Key causes of AMR:

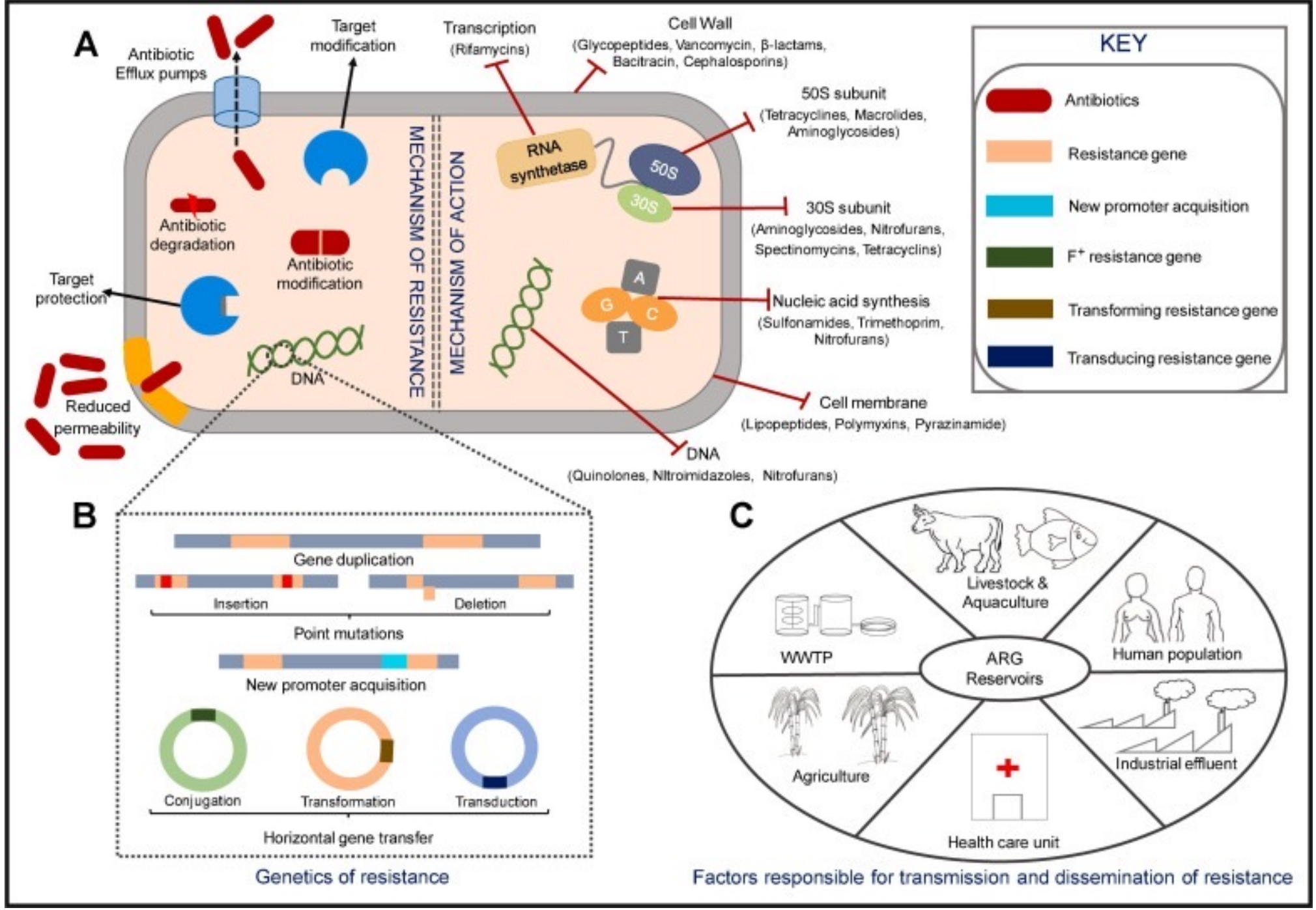
- Over-prescription of antimicrobials.
- Shortened courses or incomplete compliance with antimicrobial treatment.
- Antimicrobial overuse in livestock and fish farming.
- Poor infection control in health care settings.
- Poor hygiene and sanitation.
- Limited discovery of new antimicrobials.

https://asm.org/Resource-Pages/Antimicrobial-Resistance

AMR Mechanisms



https://pubs.rsc.org/en/content/articlehtml/2020/ra/d0ra01484b



https://www.sciencedirect.com/science/article/pii/S0160412019342424

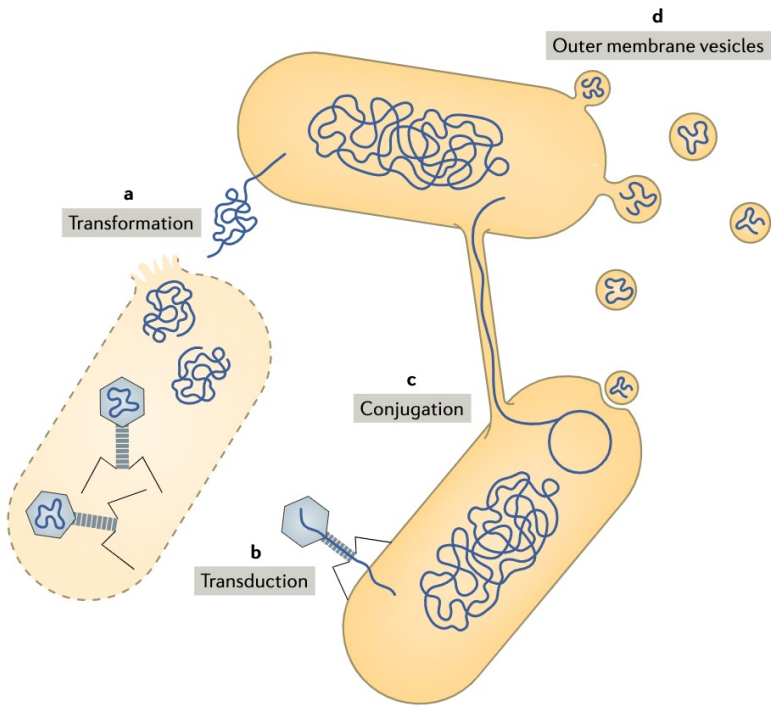
Horizontal gene/DNA transfer

HGT, also known as lateral gene transfer

Non-vertical, ie
not parent -> offspring

1947: *Escherichia coli* by Tatum and Lederberg

Core genome = set of genes encoding fundamental metabolic functions present in all taxon members



Brito, I.L. Examining horizontal gene transfer in microbial communities. *Nat Rev Microbiol* (2021). https://doi.org/10.1038/s41579-021-00534-7

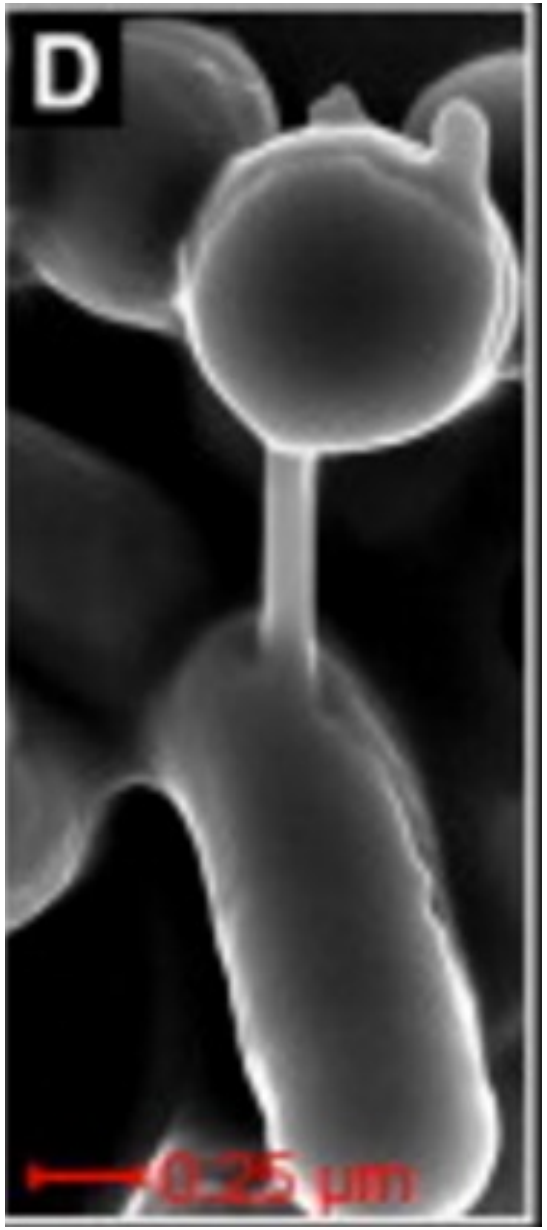
HGT: Nanotube horizontal gene transfer

Nanotubes are attachments
b/w cells to allow HGT

Found in *Bacillus subtilis*, *S aureus*, *E coli* enabling transfer of (GFP) fluorescently-labelled non-conjugative plasmids*

S aureus (circle) and *B subtilis* nanotubing

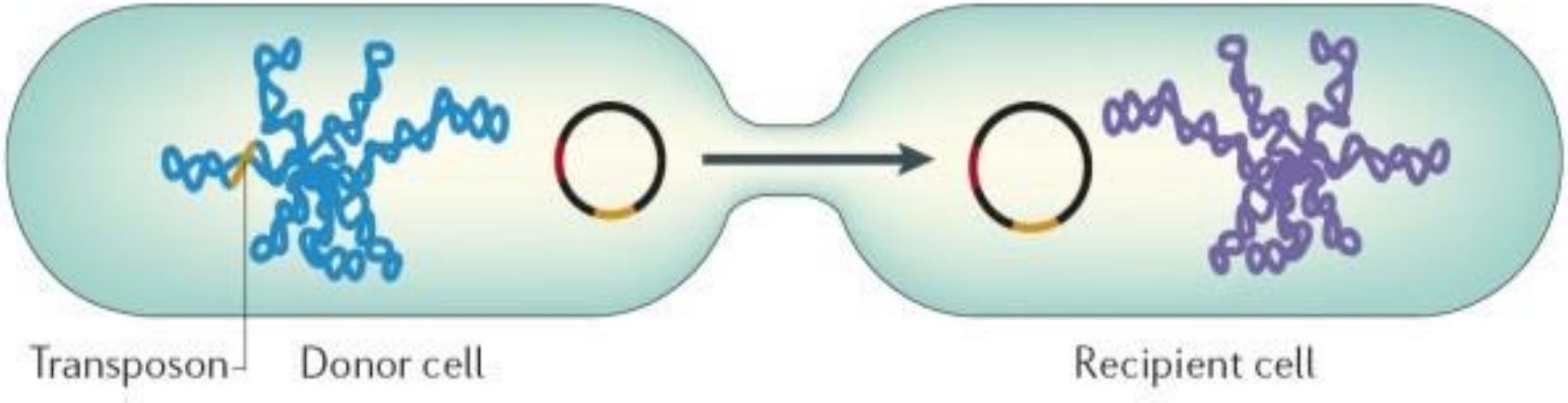
* No *tra* genes for transfer



Dubey and Ben-Yehuda Cell 2011,
video at www.sciencedirect.com/science/article/pii/S009286741100016X

HGT: Conjugation

Conjugation = plasmid transfer via cell-cell contact

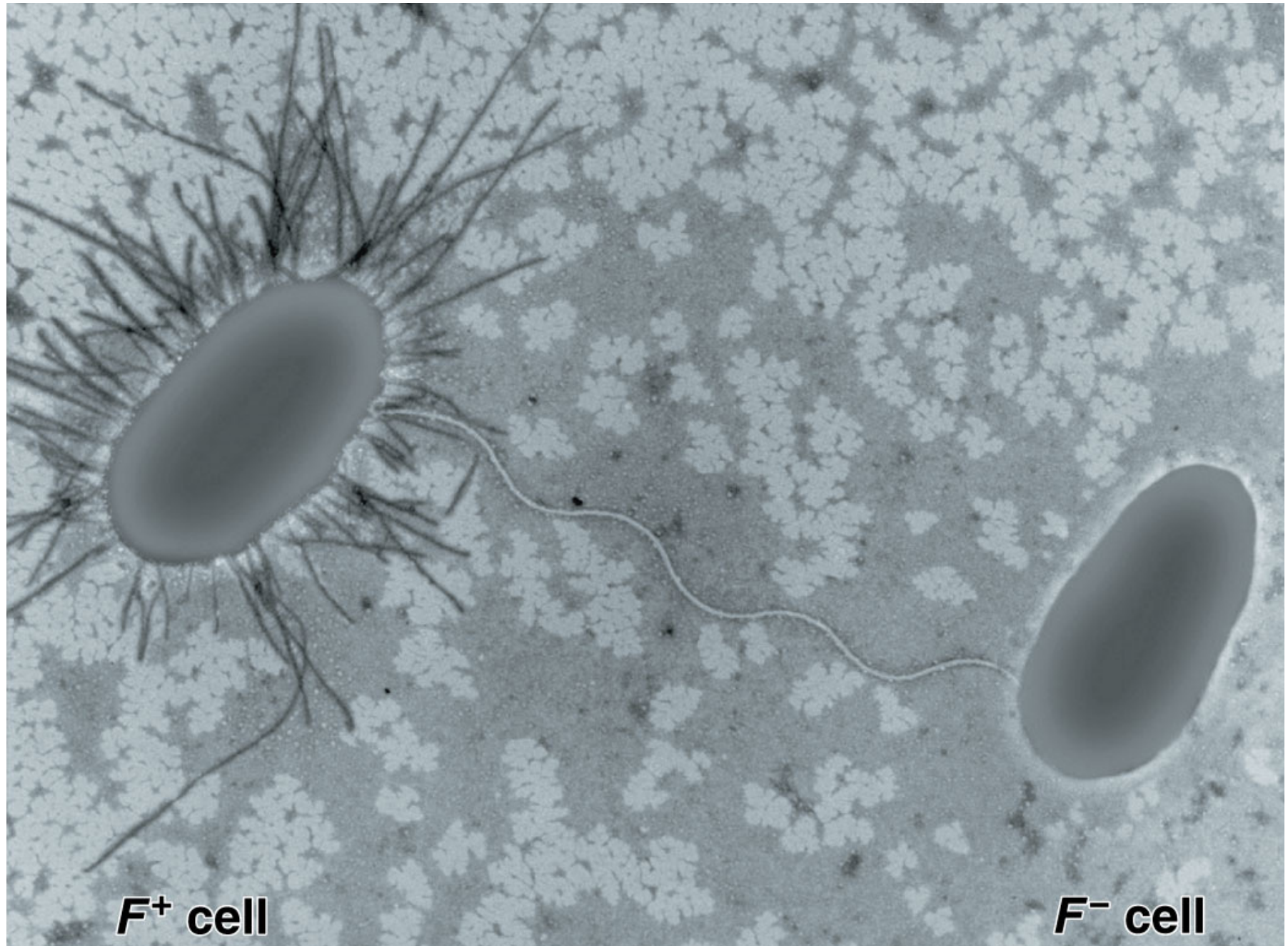


Donor has "F plasmid": F+ (F for fertility)

Recipient was F-, now F+

Conjugative plasmid = self-transmissable

bacterial cells connected by a long, tubular *F*-pilus



Which AMRG Annotation Tool to Use?

Depends on your goal

Goals = Tools

Field Applications

- Surveillance: AMRGs with known effect
- Clinical usage: Outbreak investigation

Fundamental Research

- Gene discovery e.g. novel AMRGs
- Genotype-phenotype prediction

Which AMRG Annotation Tool to Use?

Input data

Assemblies: Quality of draft assemblies

Reads: Lacking positional info (unknown gene position in the genome)

Feldgarden (2021) https://www.climb.ac.uk/wp-content/uploads/Feldgarden_AMR_prediction_tools_compressed.pdf

Which AMRG Annotation Tool to Use?

Output data

- Closest match vs Best estimate ID
- Point mutation
- ‘Broken gene’
- Description of gene => phenotype?
- Online/Web interface/GUI

Feldgarden (2021) https://www.climb.ac.uk/wp-content/uploads/Feldgarden_AMR_prediction_tools_compressed.pdf

Which AMRG Annotation Tool to Use?

Tool Features

- Nucleotide databases vs. Amino acid databases
  - Amino acid describes function
  - Nucleotide-based analyses can be faster, but sometimes inaccurate at fine scale
  - Hybrid (e.g., point mutations of 23S and protein detection)

Feldgarden (2021) https://www.climb.ac.uk/wp-content/uploads/Feldgarden_AMR_prediction_tools_compressed.pdf

How Are Genes Detected: BLAST, kmers, and HMMs

- BLAST (and similar methods)
  - Straightforward to implement
  - Easy to understand how it works
  - Nucleotide-based methods
- K-mers
  - Speed–can search large read sets such as microbiome data
  - Usually mechanism-agnostic (for good and bad)
  - Often tied into phenotype prediction
- Hidden Markov Models (HMMs)
  - Alignments of known proteins are used to build HMMs that identify conserved domains of structure and function
  - Typically use protein sequence for speed/computational reasons
  - Based on biological structure, not arbitrary identity thresholds
- Manually curated cutoffs/rules versus One Rule to Bind Them All

Feldgarden (2021) https://www.climb.ac.uk/wp-content/uploads/Feldgarden_AMR_prediction_tools_compressed.pdf

Choosing a Reliable AMR Database

Things to Look for in a Database

- Is it regularly curated/updated?
- What are the inclusion criteria for genes (and point mutations)?
  - Are only full-length genes included?
    - important for identifying best hit
  - Are start sites are curated?
    - *attC* sites are removed
    - leader peptides verified
- How are gene symbols reported? (hARMonization)
- Are there links to the literature?
- Are possible phenotypes reported?
- **Unfortunately, it's hard to know these things!!**

Feldgarden (2021) https://www.climb.ac.uk/wp-content/uploads/Feldgarden_AMR_prediction_tools_compressed.pdf

Popular Tools and DBs for AMRG Annotation

ResFinder 4 (CGE)

- Can use assemblies or reads
- Nucleotide vs. nucleotide BLAST-based
- A single identity and a single length threshold
- Fast
- Can misassign alleles as closest amino acid hit is not necessarily the closest nucleotide hit
- Online GUI:

RGI (CARD)

- Protein database
- Option for broadening scope to identify novel mechanisms; emphasis on efflux
- Will accept nucleotide sequence or protein sequence
- BLAST-based but manual cutoffs
- Online GUI and ontology

Feldgarden (2021) https://www.climb.ac.uk/wp-content/uploads/Feldgarden_AMR_prediction_tools_compressed.pdf

Popular Tools and DBs for AMRG Annotation

AMRFinderPlus (NCBI)

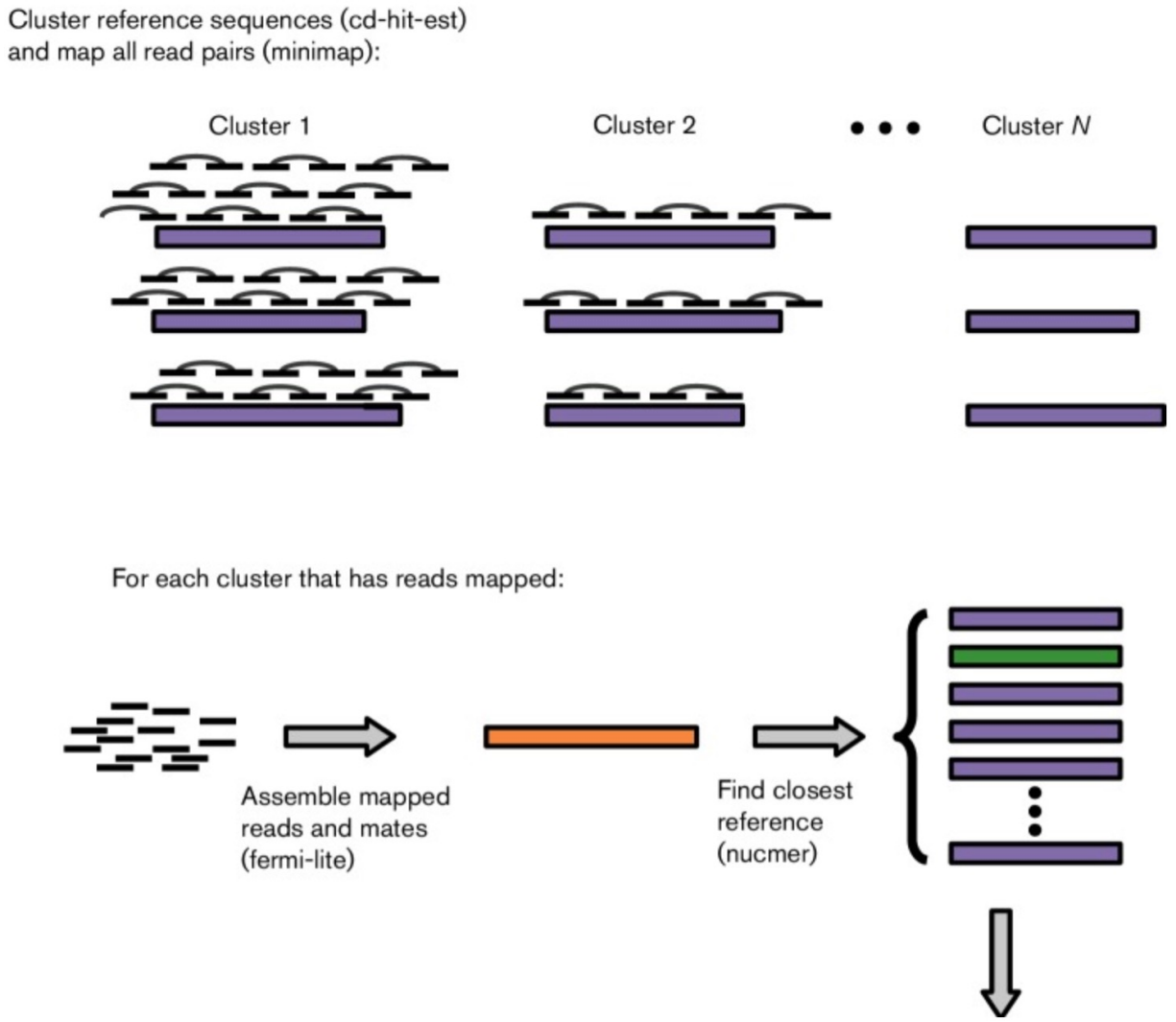
- Protein database
- Will accept nucleotide sequence or protein sequence
- Uses BLAST and HMMs to identify AMR genes
- Manually curated BLAST and HMM cutoffs • Explicit partial and internal stop identification
- No online GUI but data for >780,000 isolates are available in MicroBIGG-E

Abricate

- Only only supports contigs, not FASTQ reads
- Detects acquired resistance genes, NOT point mutations
- Uses a DNA sequence database, not protein

Feldgarden (2021) https://www.climb.ac.uk/wp-content/uploads/Feldgarden_AMR_prediction_tools_compressed.pdf

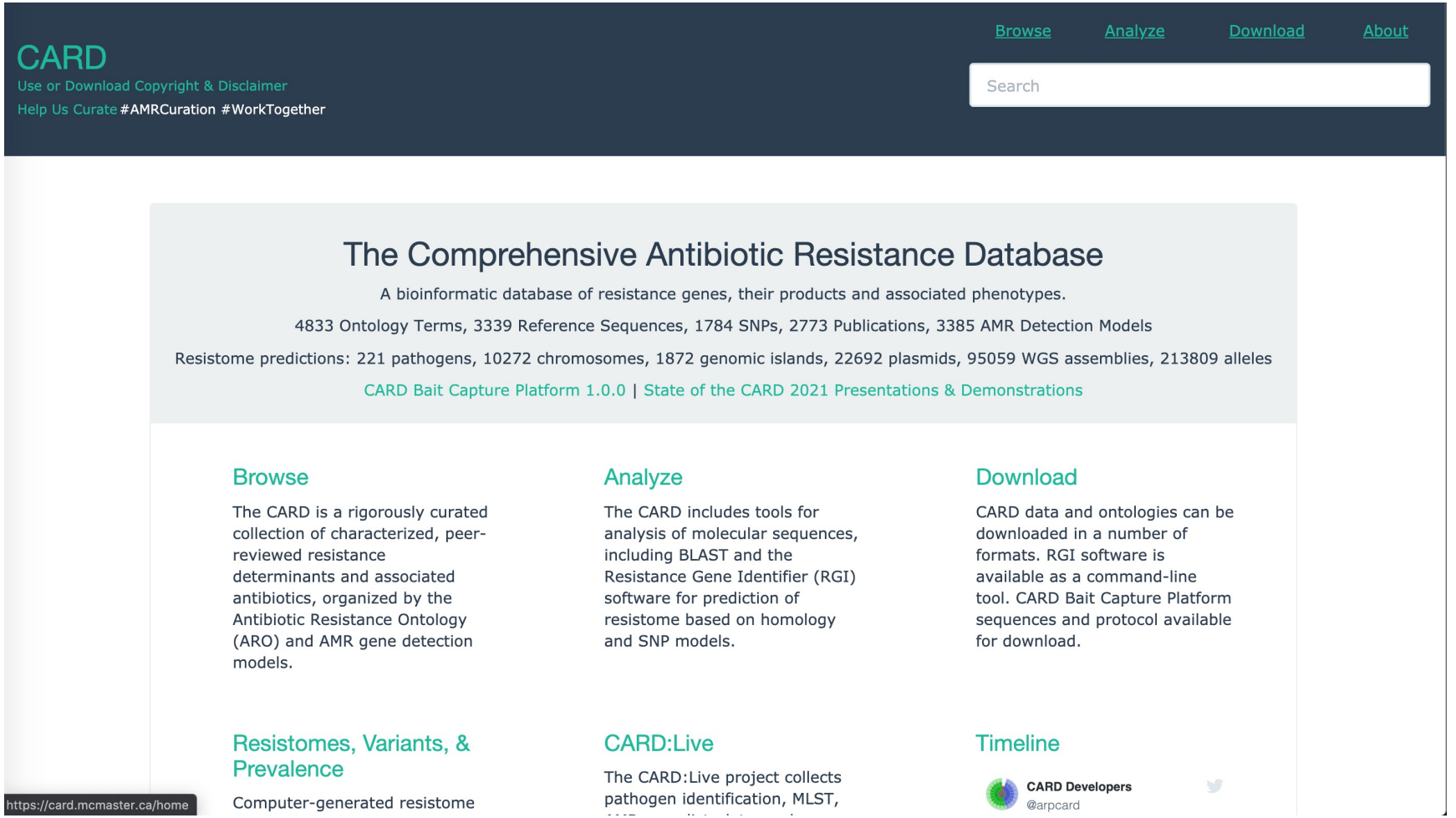
Tool: ARIBA (https://github.com/sanger-pathogens/ariba)



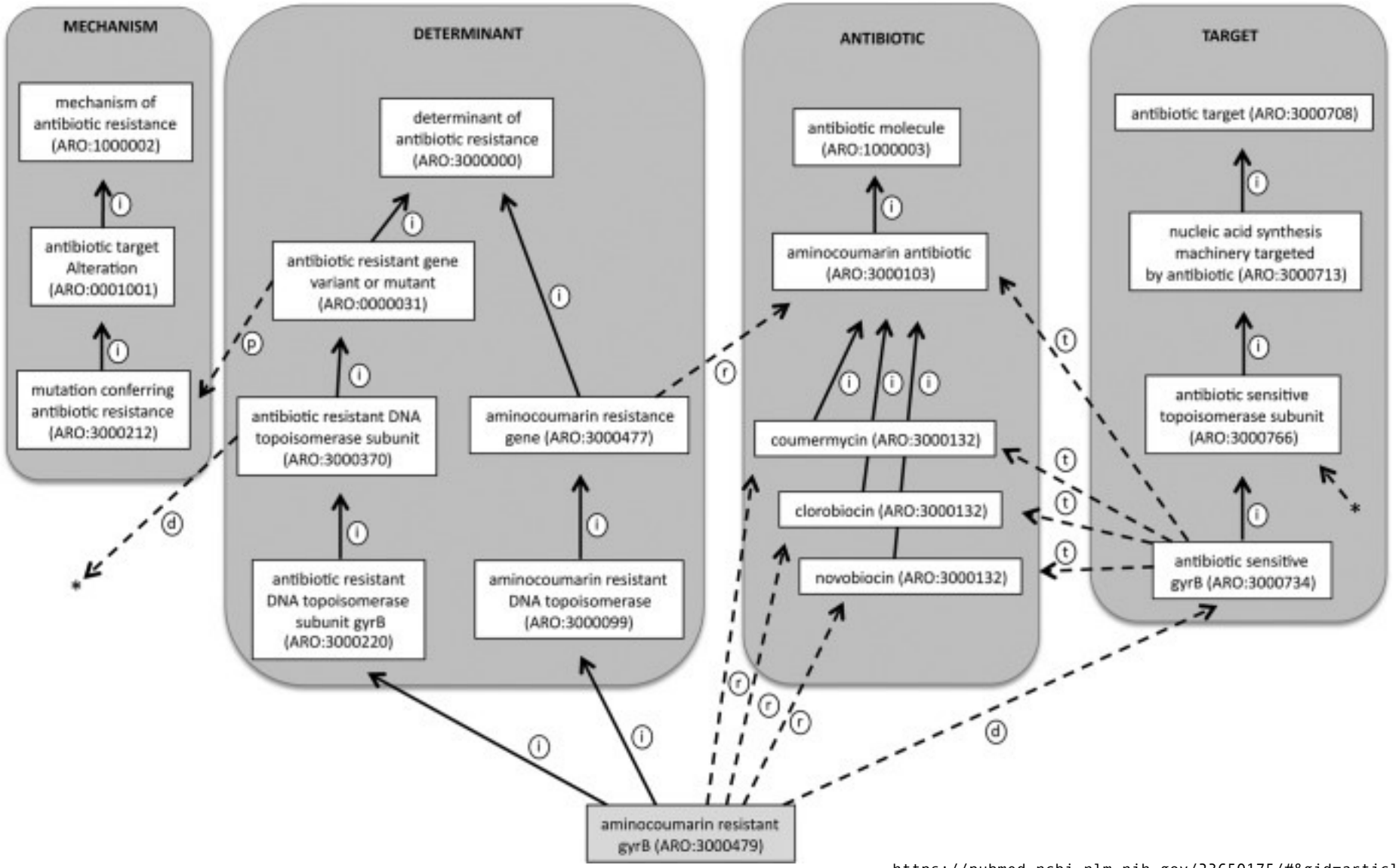
Overview of the ARIBA mapping and targeted assembly pipeline.

AMR genes (% ID)

Comprehensive Antibiotic Resistance Database (CARD)

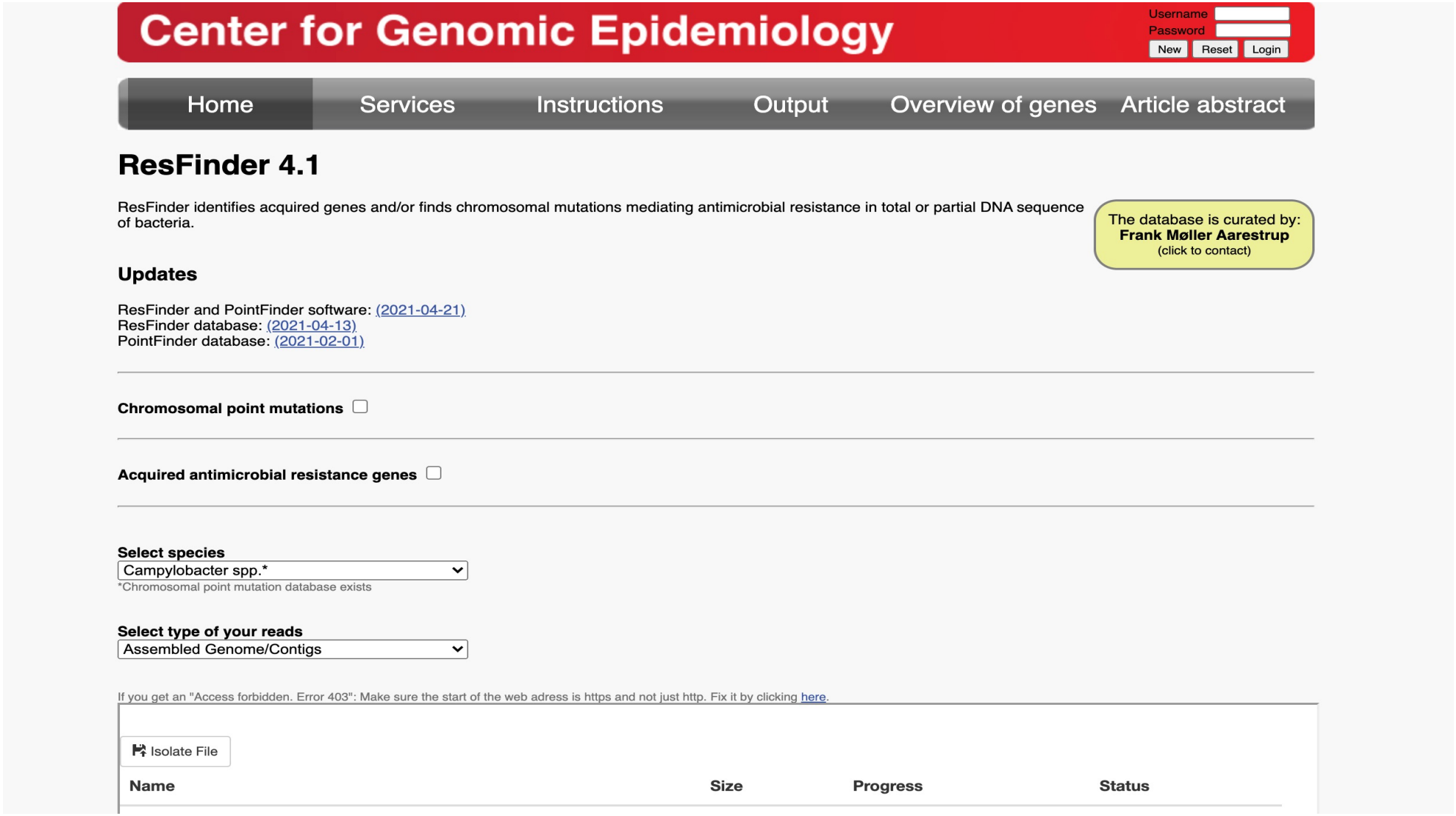


CARD: Use of ontological relationships to describe knowledge about the gene



https://pubmed.ncbi.nlm.nih.gov/23650175/#&gid=article-figures&pid=fig-2-uid-1

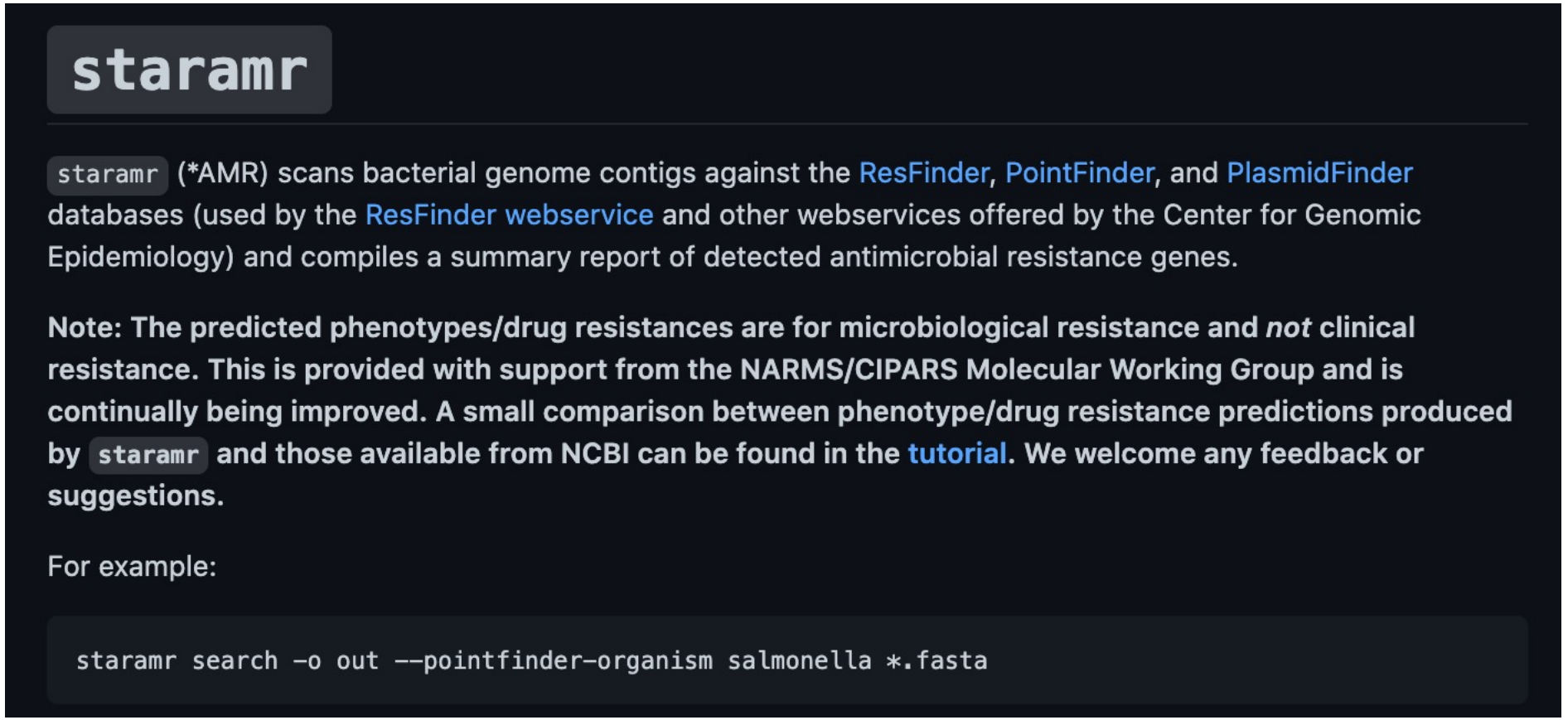
Resfinder


Center for Genomic Epidemiology
Username
Password
New
Reset
Login
Home
Services
Instructions
Output
Overview of genes
Article abstract
ResFinder 4.1
ResFinder identifies acquired genes and/or finds chromosomal mutations mediating antimicrobial resistance in total or partial DNA sequence of bacteria.
The database is curated by:
Frank Møller Aarestrup
(click to contact)
Updates
ResFinder and PointFinder software: (2021-04-21)
ResFinder database: (2021-04-13)
PointFinder database: (2021-02-01)
Chromosomal point mutations
Acquired antimicrobial resistance genes
Select species
Campylobacter spp.*
*Chromosomal point mutation database exists
Select type of your reads
Assembled Genome/Contigs
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.
Isolate File
Name
Size
Progress
Status

Others:

- Interproscan: https://github.com/ebi-pf-team/interproscan
- Groot: https://groot-documentation.readthedocs.io/en/latest/

Staramr pipeline: https://github.com/phac-nml/staramr



Outputs 7 Files including an Excel file containing the summary

AMR Annotation: Practical

ABRicate

Mass screening of contigs for antimicrobial resistance or virulence genes. It comes bundled with multiple databases: NCBI, CARD, ARG-ANNOT, Resfinder, MEGARES, EcOH, PlasmidFinder, Ecoli_VF and VFDB.

Is this the right tool for me?

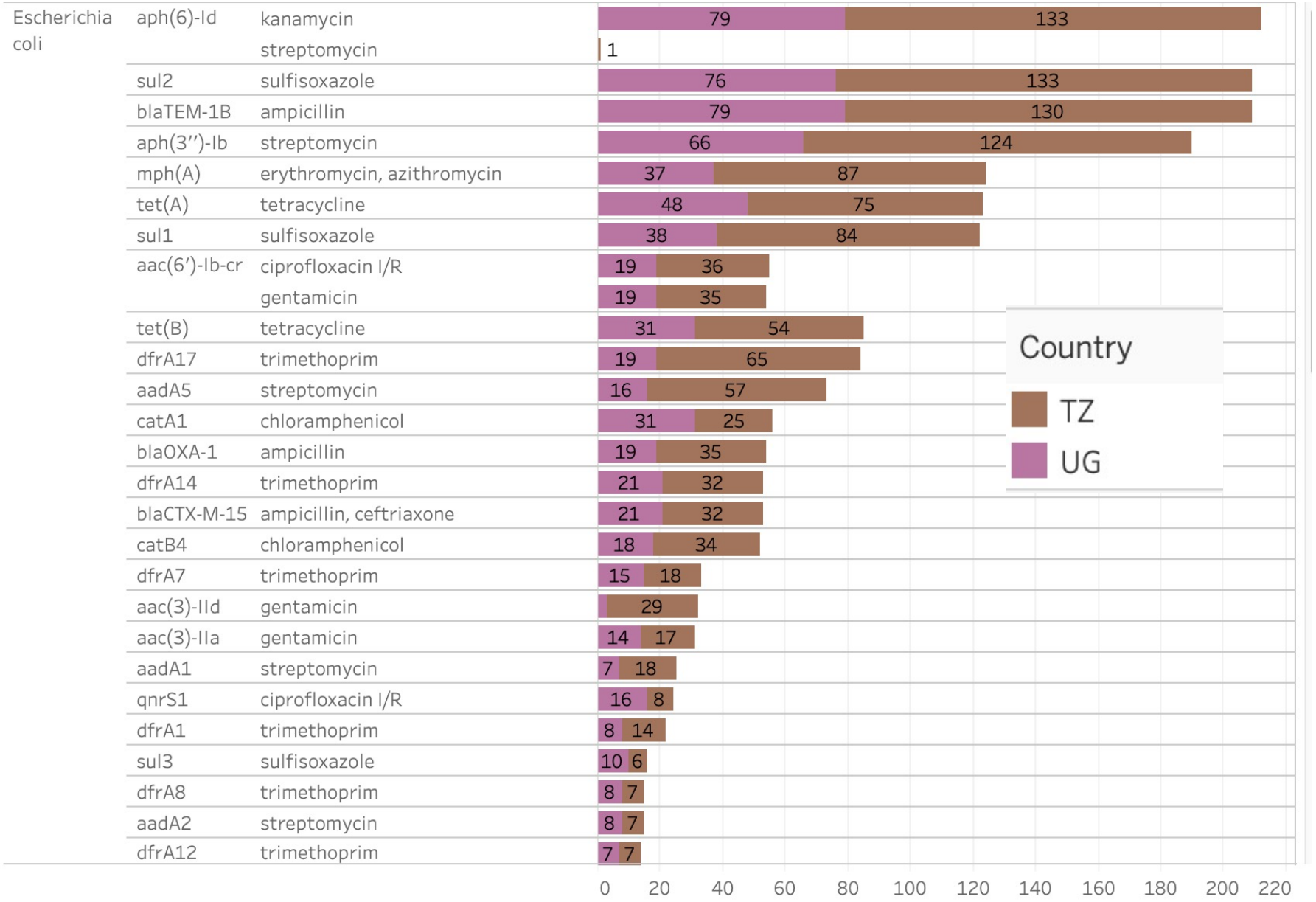
1. It only supports contigs, not FASTQ reads
2. It only detects acquired resistance genes, **NOT** point mutations
3. It uses a DNA sequence database, not protein
4. It needs BLAST+ >= 2.7 and `any2fasta` to be installed
5. It's written in Perl 🐫

If you are happy with the above, then please continue! Otherwise consider using Ariba, Resfinder, RGI, SRST2, AMRFinderPlus, *etc.*

https://github.com/tseemann/abricate

Generic Name	TZ	UG	KY	Class	Oral
Amikacin	AMK		AMK	Aminoglycosides	No
Amoxicillin/clavulanate	AMC	AMC	AMC	Penicillins + β-lactamase inhibitors	Yes
Ampicillin	AMP	AMP	AMP	Penicillins	Yes
Cefepime	FEP		FEP	ESBL Cephalosporins	Yes
Cefotaxime	CTX		CTX	ESBL Cephalosporins	Yes
Cefoxitin		FOX		Anti-staphylococcal β-lactams	No
Ceftazidime	CAZ	CAZ	CAZ	ESBL Cephalosporins	Yes
Ceftriaxone	CRO	CRO	CRO	ESBL Cephalosporins	Yes
Chloramphenicol	C		C/CHL	Phenicols	Yes
Ciprofloxacin	CIP		CIP	Fluoroquinolones	Yes
Erythromycin	E	E	ERY	Macrolides	Yes
Gentamycin	CN			Aminoglycosides	No
Linezolid	LZD	LZD	LZD/LNZ	Oxazolidinones	Yes
Nalidixic Acid		NA	NAL	Quinolone	Yes
Nitrofurantoin	F		NIT/F	Nitrofuran	Yes
Sulfamethoxazole	RL	R	SMX/R	Folate pathway inhibitors	Yes
Tetracycline	TET	TET	TET/TCY	Tetracycline	Yes
Trimethoprim	W		W/TMP	Folate pathway inhibitors	Yes
Vancomycin	VA	VA	VA/VAN	Glycopeptides	Yes
Aztreonam			ATM	Monobactams	No
Clindamycin	CC			Lincosamides	Yes
Fosfomycin			FOS	Phosphonic acids	Yes
Gentamicin			GEN	Aminoglycosides	No
Meropenem	MEM			Carbapenems	No
Piperacillin-tazobactam	TZP			Antipseudomonal penicillins + β-lactamase inhibitors	Yes
Rifampin			RIF	Ansamycins	Yes

Local alignment to AMR databases (FASTA vs FASTA)



AMR Gene Annotation of *de novo* Assembled Genomes using Abricate

(1) Run the Abricate pipeline using CARD and Resfinder as reference databases on the 2 *E. coli* and 2 *S. typhi* genomes you've assembled using Shovill. Save the output to a .csv file as below.

```
abricate  --csv --db resfinder [sample_name.fa] > [sample_name]_resf.csv
abricate  --csv --db card [sample_name.fa] > [sample_name]_card.csv
```

(2) Combine reports aligned with CARD and Resfinder and save each set to a .csv file.

```
abricate  --csv --summary [sample_name-1]_card.csv [sample_name-2]_card.csv > summary_CARD.csv
abricate  --csv --summary [sample_name-1]_resf.csv [sample_name-2]_resf.csv > summary_RESF.csv
```

(3) Plot the number of AMR genes of the genomes you characterised. Only include results with 95-100% identity match with those in the two databases.