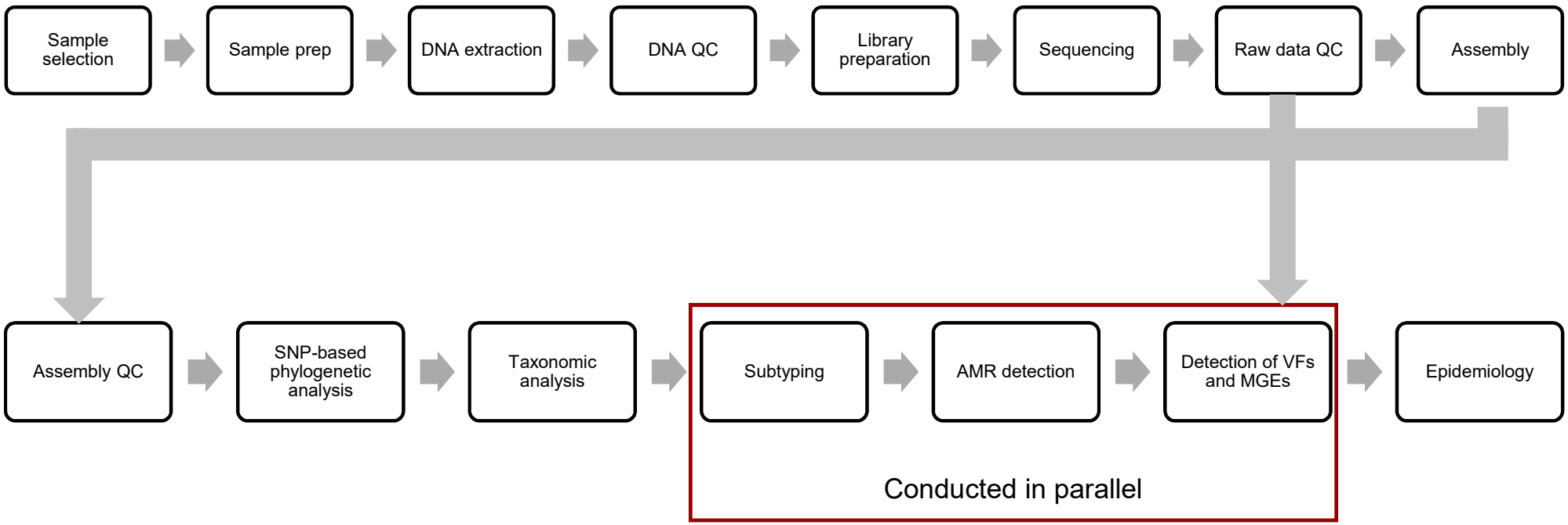
Day 7. Bridging the gaps in bioinformatics

# Annotation and AMR

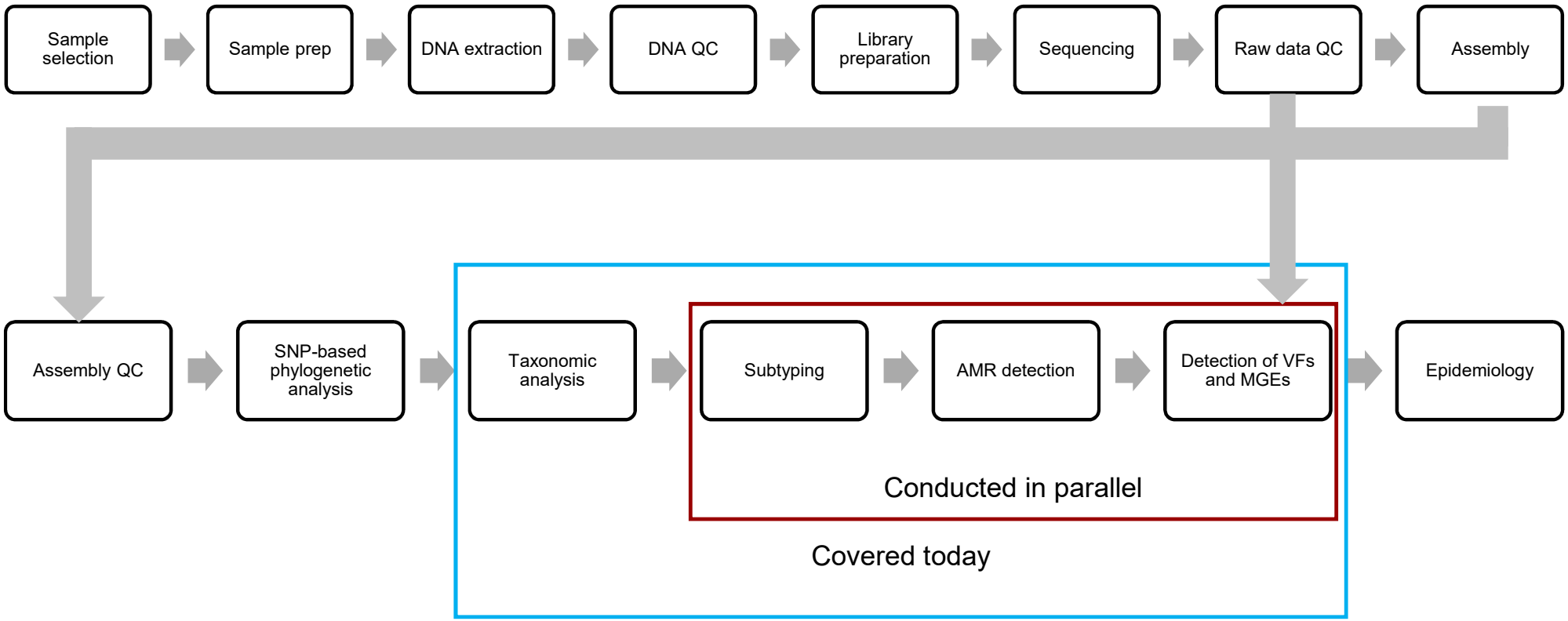
# Subjects this morning

- Annotation
  - Full genome annotation
  - Functional annotation
  - How do we predict?
    - Kmer
    - Seed and extend
  - Identification of genes of interest
    - Database considerations
    - AMR
    - How to make your own database
    - Replicons

# Example of workflow



# Example of workflow



# Annotation

- “The annotation process infers the structure and function of the assembled sequences. Protein-coding genes are often annotated first, but other features, such as non-coding RNAs or presence of regulatory or repetitive sequences, can also be annotated.” (Dominguez Del Angel et al, 2018)

- Once we have the assembled the genome annotation is straight forward:
  - Conformation of species
  - Prediction of genes
  - Prediction of function
    - Conserved modules
  - Typing
  - Other features

- If only certain features are of interest, we can use raw sequence data directly

# 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)

# Data is everything

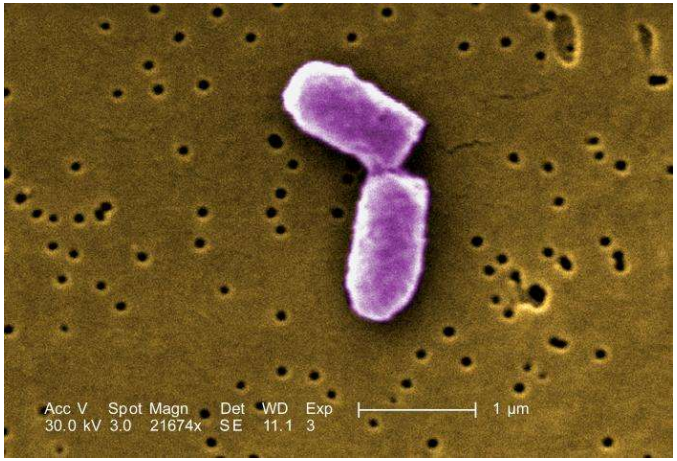
- We can produce a lot of data, but is it useful?
  - Bioinformatics relies heavily on findings produced in the wet lab
    - We cannot prove biological functions purely *in silico*
    - We can make prediction based on previous findings
    - We can track changes down to the nucleotide level
  - We are limited by our databases
    - Some genes have no annotation information
      - Gene with unknown function, but recognized modules or motives
    - Pseudo genes
  - We are only interested in a small subset of information
    - If we are interested in a given subject, we can reduce the amount of data
    - In E. coli K-12 there are 4639 predicted genes of which 3 are related to AMR
    - Specialized databases
    - Simplifies analysis and clarifies message

# Data is everything

- Translating classic biochemical methods to bioinformatics
  - Serotyping of *E. coli* is based on testing the O- and H-antigens with antisera
  - 186 O-antigens and 53 H-flagellar antigens determines the serotype
  - Time consuming and sometimes inconclusive

- O- and H- genes have been extracted and their associated phenotype determined
  - Predicting the Serotype based on genetic profile showed >98% concordance with phenotype
  - Analysis can be conducted in minutes rather than hours
  - It is still only an estimate

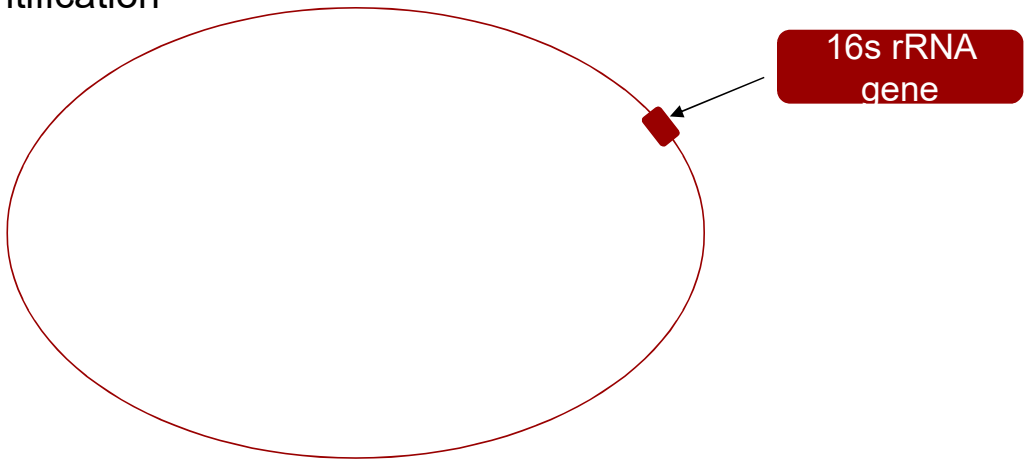
(Joensen et al., 2015, Fratamico et al., 2016)




*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*

# Prediction of species

- 16s rRNA gene formed the basis as the first method for sequenced based taxonomy
- Other approaches:
  - gyrB gene, rMLST, species-specific functional domain profiles
  - Shortcomings:
    - Only represents a small fraction of the entire genome
    - Today we have WGS data, better to ultilize the whole genome instead of just one gene for identification



# 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)

# k-mers

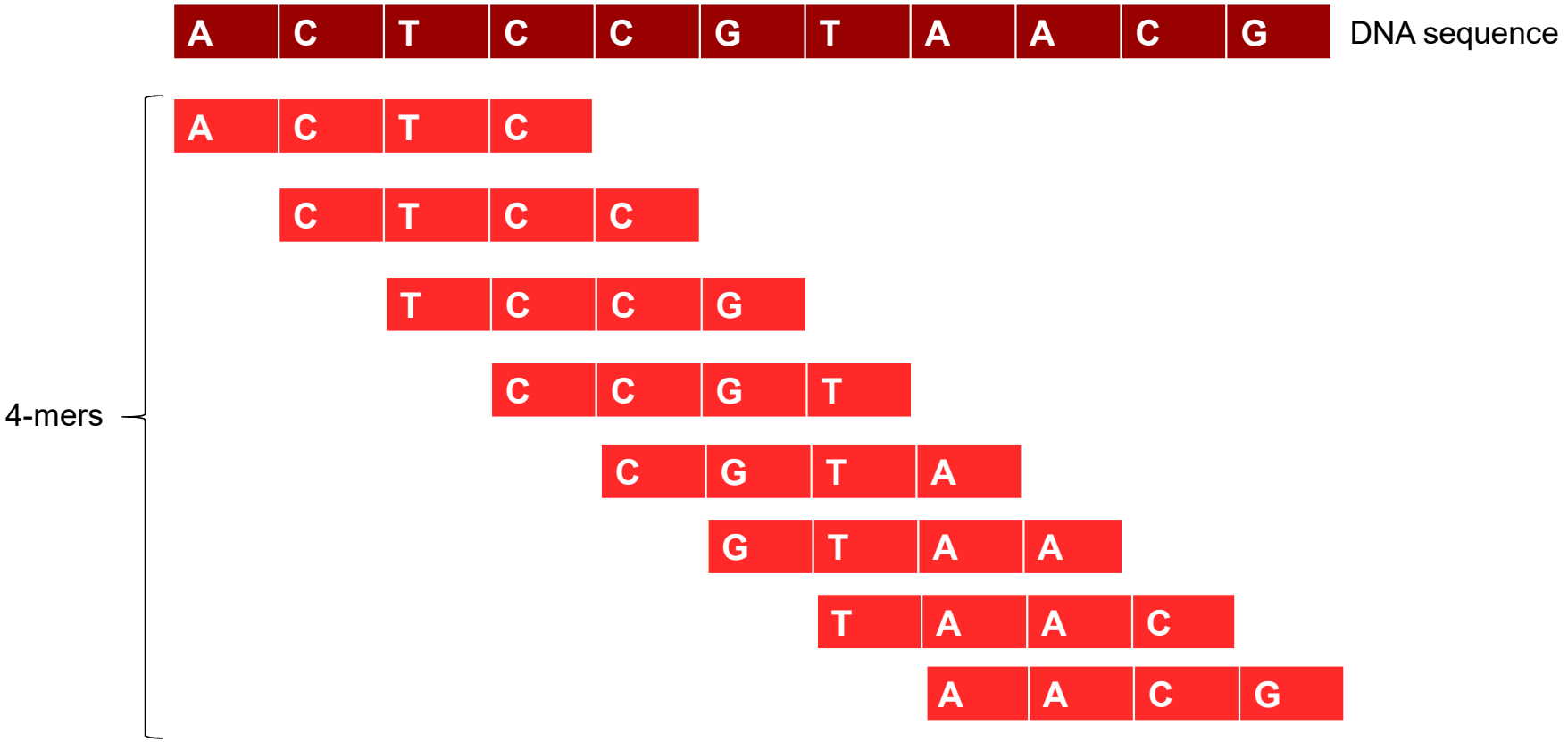
- A k-mer is a continuous sequence of k bases
  - e.g a certain length of DNA, RNA or protein
- k is any positive integer
- Sequences with high similarity must share k-mers

- Consider the nucleotide sequence 'ACTCCGTAACG'.

| A | C | T | C | C | G | T | A | A | C | G |
|---|---|---|---|---|---|---|---|---|---|---|

- We have this long sequence that we can cut into smaller pieces or k-mers
- We can extract all the 4-mers (substrings of length 4) in this DNA sequence

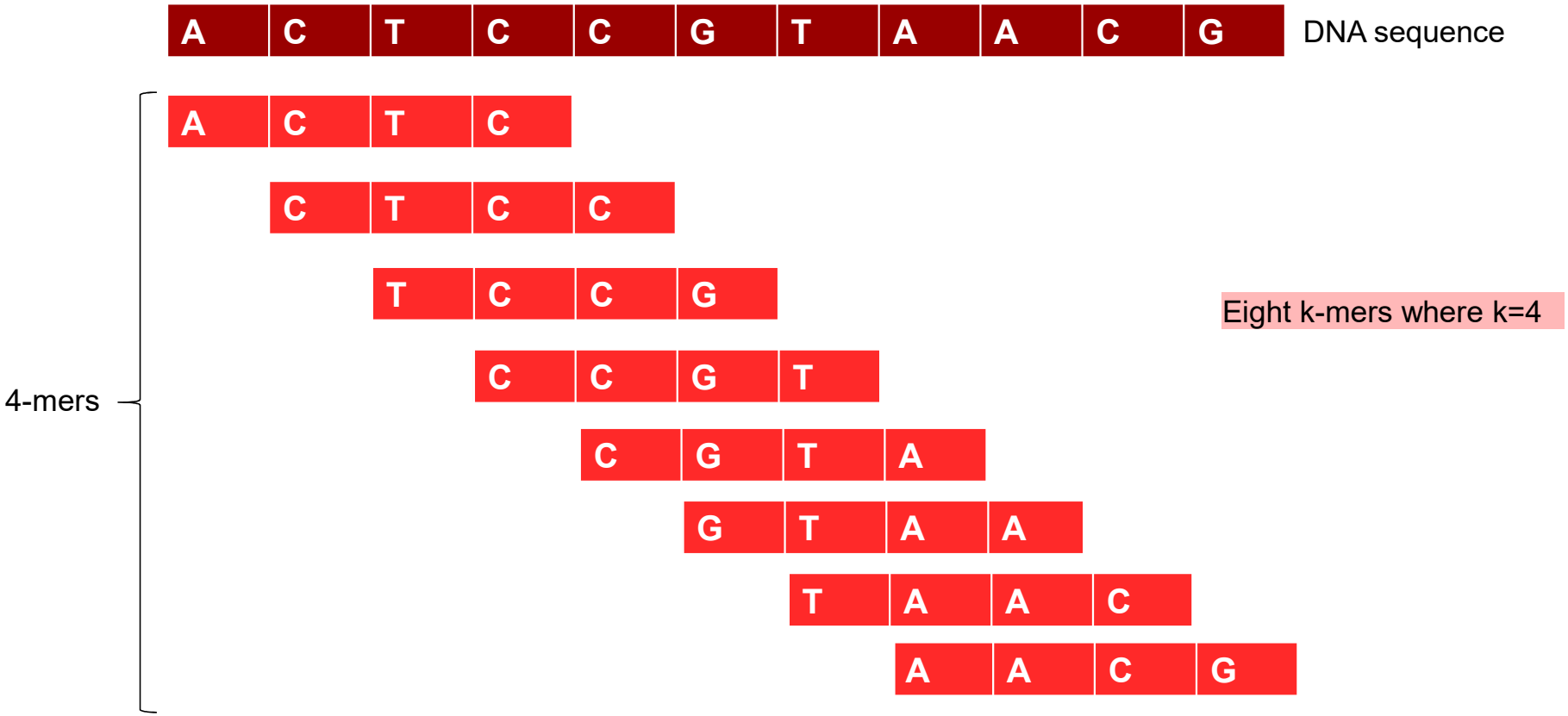
# k-mers


A C T C C G T A A C G
DNA sequence
4-mers
A C T C
C T C C
T C C G
C C G T
C G T A
G T A A
T A A C
A A C G

# k-mers

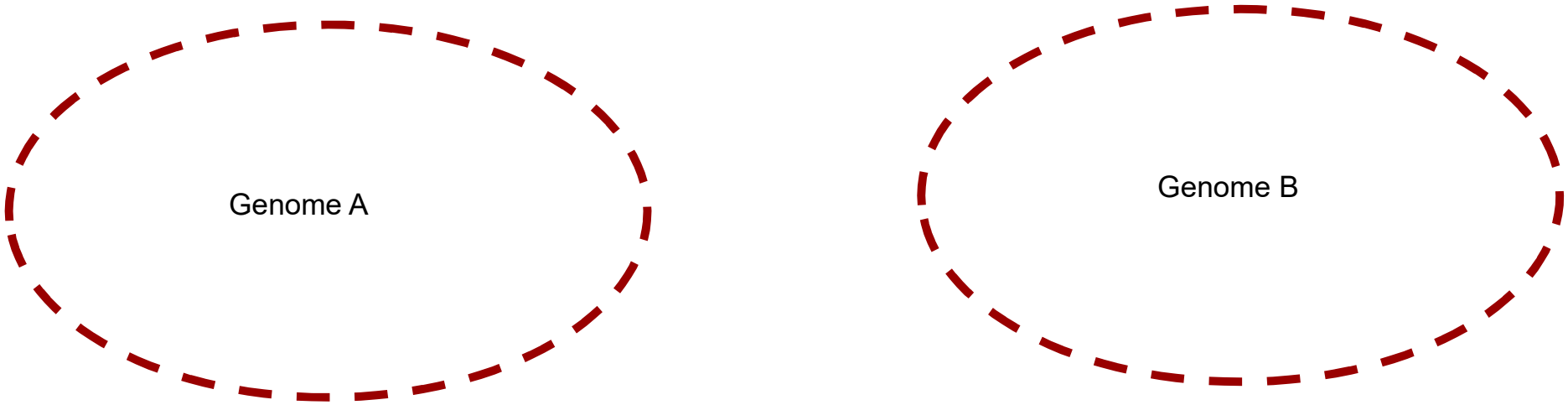
- We can extract all the 4-mers (substrings of length 4) in this DNA sequence.
- We created a window of length 4 and slide it from left to right, shifting one character at a time.
- If the length of a given DNA sequence is N, we end up with N - k+1 k-mers
- Total number of k-mers = N – k + 1
- In the previous slide, the DNA sequence is 11 characters long (N = 11) and k = 4, so we get eight k-mers (11- 4 + 1)

# k-mers

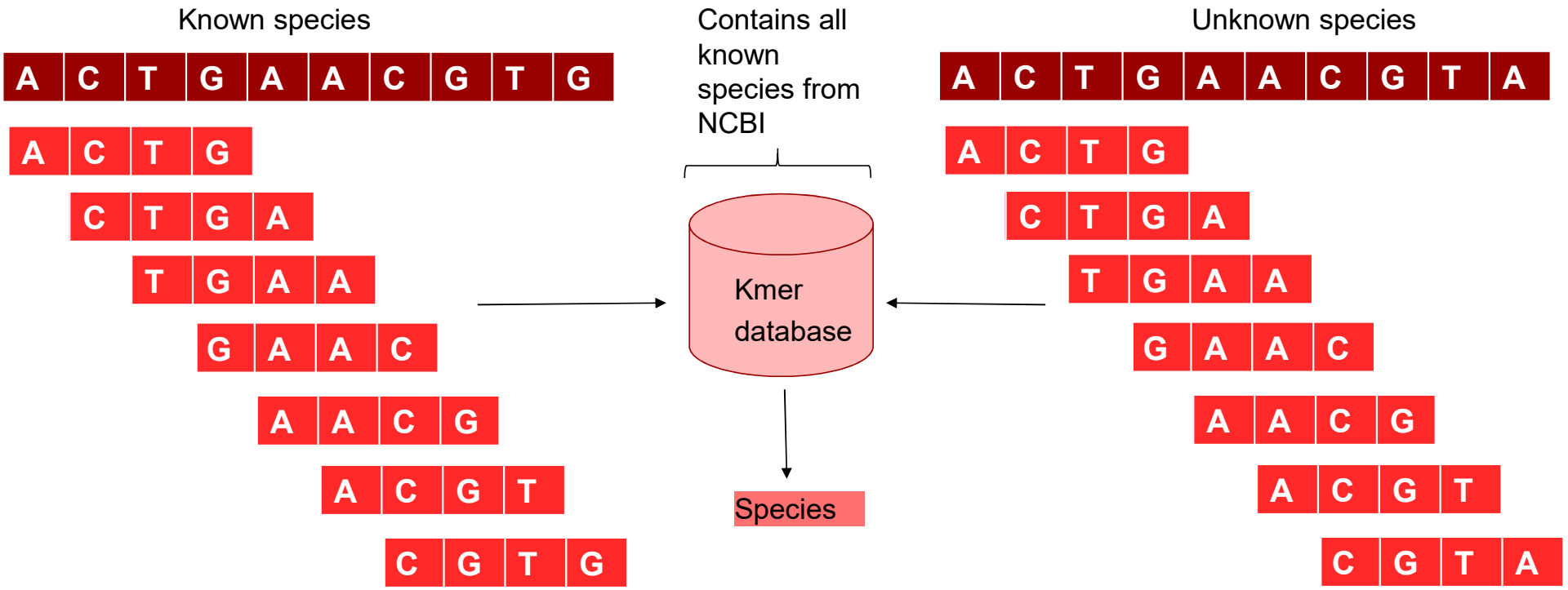

A C T C C G T A A C G
DNA sequence
4-mers
A C T C
C T C C
T C C G
C C G T
C G T A
G T A A
T A A C
A A C G
Eight k-mers where k=4

# Species prediction with k-mers

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



# Species prediction with KmerFinder

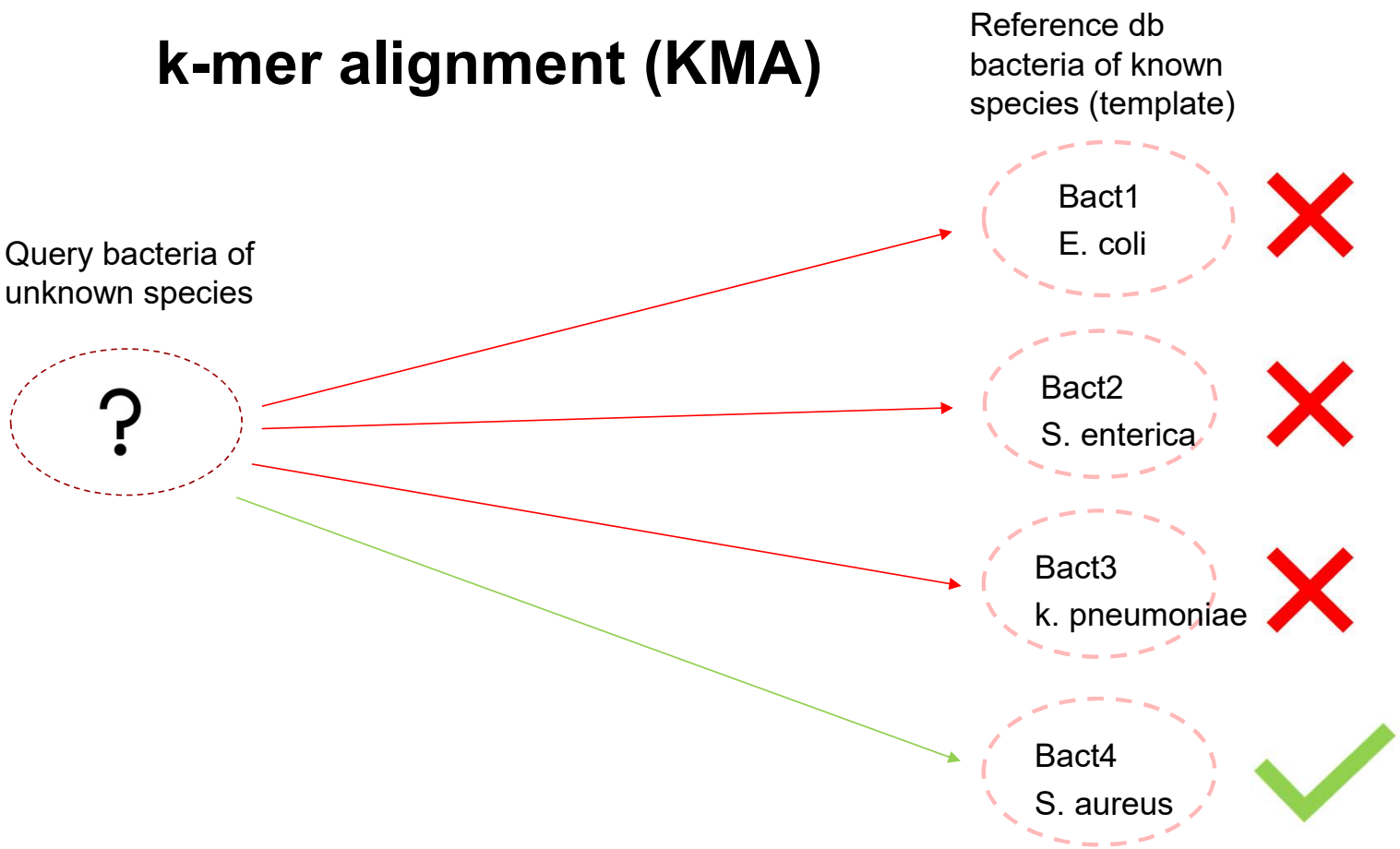


# Species prediction with KmerFinder

- Genomes are spilt into 16-mers
- But all these k-mers won't be in the database
- Tool will take too long to compute
- So redundant data is reduced
- Only 16-mers with specific prefixes are kept e.g ATGAG
- Reduced redundancy speeds up the tool

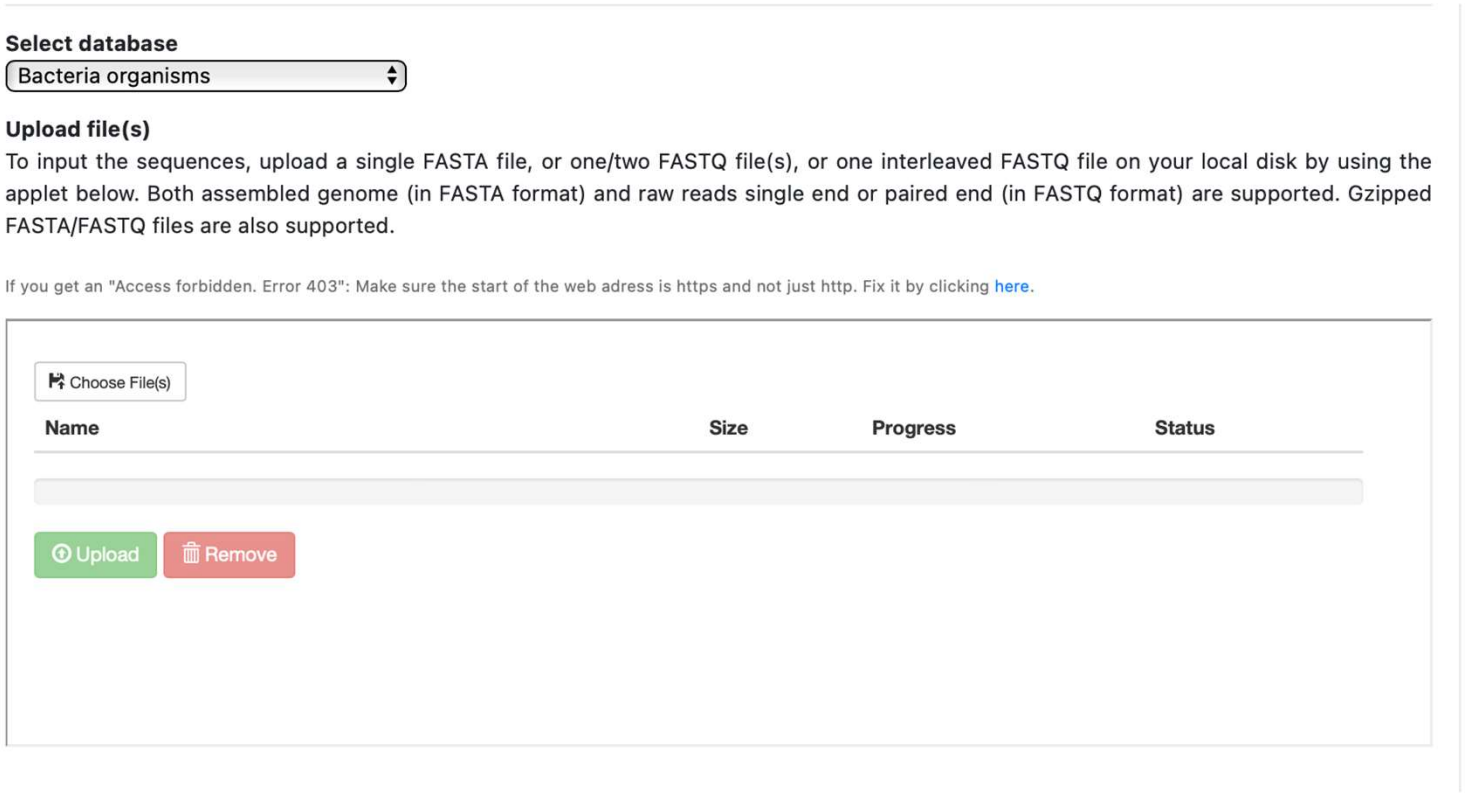
- But how does the tool the compare k-mers?

# k-mer alignment (KMA)




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

# KmerFinder webtool




*https://cge.food.dtu.dk/services/KmerFinder/*

# Kmer webtool output

| Hit | Score | z-score | Query Coverage [%] | Template Coverage [%] | Depth | Total Query Coverage [%] | Total Template Coverage [%] | Total Depth |
|---|---|---|---|---|---|---|---|---|
| Escherichia coli, Escherichia coli O104:H4, Escherichia coli O104:H4 str. 2011C-3493 get sequence | 10879 | 527.6 | 97.08 | 99.79 | 1.00 | 97.08 | 99.79 | 1.00 |
| Escherichia coli, Escherichia coli NA114 get sequence | 16 | 9.3 | 0.14 | 0.15 | 0.00 | 54.87 | 58.92 | 0.60 |
| Salmonella enterica, Salmonella enterica subsp. enterica, Salmonella enterica subsp. enterica serovar Typhimurium, Salmonella enterica subsp. enterica serovar Typhimurium str. T000240 get sequence | 12 | 6.9 | 0.11 | 0.11 | 0.00 | 6.59 | 6.87 | 0.08 |

*Source: Introduction to CGE tools, presentation by Pimpalas Leekitcharoenphon*

# Kmer webtool output

| Hit | Score | | | | Depth | Total Query Coverage [%] | Total Template Coverage [%] | Total Depth |
|---|---|---|---|---|---|---|---|---|
| Escherichia coli, Escherichia coli O104:H4, Escherichia coli O104:H4 str. 2011C-3493 get sequence | 10879 | 527.6 | 97.08 | 99.79 | 1.00 | 97.08 | 99.79 | 1.00 |
| Escherichia coli, Escherichia coli NA114 get sequence | 16 | 9.3 | 0.14 | 0.15 | 0.00 | 54.87 | 58.92 | 0.60 |
| Salmonella enterica, Salmonella enterica subsp. enterica, Salmonella enterica subsp. enterica serovar Typhimurium, Salmonella enterica subsp. enterica serovar Typhimurium str. T000240 get sequence | 12 | 6.9 | 0.11 | 0.11 | 0.00 | 6.59 | 6.87 | 0.08 |

**Score**
number of k-mers

*Source: Introduction to CGE tools, presentation by Pimpalas Leekitcharoenphon*

# Kmer webtool output

| Hit | Score | z-score | | | | Total Query Coverage [%] | Total Template Coverage [%] | Total Depth |
|---|---|---|---|---|---|---|---|---|
| Escherichia coli, Escherichia coli O104:H4, Escherichia coli O104:H4 str. 2011C-3493 get sequence | 10879 | 527.6 | 97.08 | 99.79 | 1.00 | 97.08 | 99.79 | 1.00 |
| Escherichia coli, Escherichia coli NA114 get sequence | 16 | 9.3 | 0.14 | 0.15 | 0.00 | 54.87 | 58.92 | 0.60 |
| Salmonella enterica, Salmonella enterica subsp. enterica, Salmonella enterica subsp. enterica serovar Typhimurium, Salmonella enterica subsp. enterica serovar Typhimurium str. T000240 get sequence | 12 | 6.9 | 0.11 | 0.11 | 0.00 | 6.59 | 6.87 | 0.08 |

z-score
statistical significant

*Source: Introduction to CGE tools, presentation by Pimpalas Leekitcharoenphon*

# Kmer webtool output

| Hit | Score | z-score | Query Coverage [%] | [illegible] | | | | [illegible]pth |
|---|---|---|---|---|---|---|---|---|
| Escherichia coli, Escherichia coli O104:H4, Escherichia coli O104:H4 str. 2011C-3493 get sequence | 10879 | 527.6 | 97.08 | 99.79 | 1.00 | 97.08 | 99.79 | 1.00 |
| Escherichia coli, Escherichia coli NA114 get sequence | 16 | 9.3 | 0.14 | 0.15 | 0.00 | 54.87 | 58.92 | 0.60 |
| Salmonella enterica, Salmonella enterica subsp. enterica, Salmonella enterica subsp. enterica serovar Typhimurium, Salmonella enterica subsp. enterica serovar Typhimurium str. T000240 get sequence | 12 | 6.9 | 0.11 | 0.11 | 0.00 | 6.59 | 6.87 | 0.08 |

**Query Coverage**
percentage of k-mers from unknown sample that map to a template

*Source: Introduction to CGE tools, presentation by Pimpalas Leekitcharoenphon*

# Kmer webtool output

| Hit | Score | z-score | Query Coverage [%] | Template Coverage [%] | | | | |
|---|---|---|---|---|---|---|---|---|
| Escherichia coli, Escherichia coli O104:H4, Escherichia coli O104:H4 str. 2011C-3493 get sequence | 10879 | 527.6 | 97.08 | 99.79 | 1.00 | 97.08 | 99.79 | 1.00 |
| Escherichia coli, Escherichia coli NA114 get sequence | 16 | 9.3 | 0.14 | 0.15 | 0.00 | 54.87 | 58.92 | 0.60 |
| Salmonella enterica, Salmonella enterica subsp. enterica, Salmonella enterica subsp. enterica serovar Typhimurium, Salmonella enterica subsp. enterica serovar Typhimurium str. T000240 get sequence | 12 | 6.9 | 0.11 | 0.11 | 0.00 | 6.59 | 6.87 | 0.08 |

**Template Coverage**
percentage of k-mers from template genome that covered by k-mers from an unknown sample

*Source: Introduction to CGE tools, presentation by Pimpalas Leekitcharoenphon*

# Functional annotation AMR

- Attaching biological, chemical or otherwise functional information to a DNA sequence

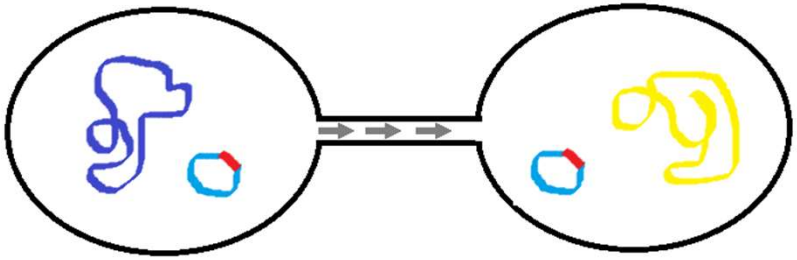
- Often you are only interested in a limited set of genes, we will look further into AMR

- AMR is a large threat to public health
  - Carried on mobile genetic elements (MGE) -> horizontal gene transfer
  - Estimated 1.27 million people died due to AMR in 2019 and estimated up to 10 million deaths by 2050 (Murray et al., 2019)
  - Development of new drugs is slow (Norrby et al., 2005)

# Genetic basis of AMR

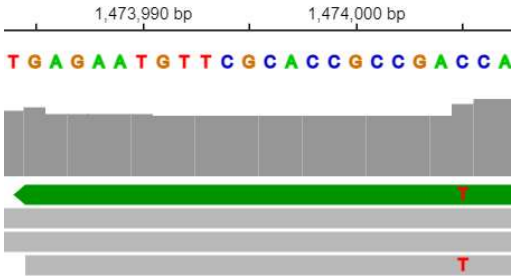
- AMR is conferred by different mechanisms:
  - Acquired resistance genes
  - Mutation
  - (Copy numbers)

- MGEs can transfer resistance genes between isolates closely or distantly related
- Resistance genes tend to aggregate, meaning MGEs often confer resistance to multiple classes
- May integrate into host chromosome



Transfer of plasmid with resistance gene

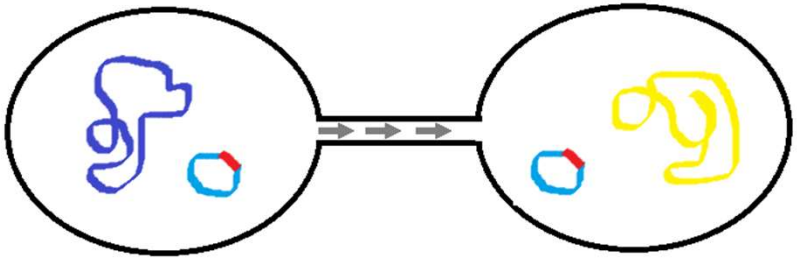
- Point mutations can confer resistance by various mechanisms:
  - Change the target of a drug, making the strain resistant
  - Upregulate the expression of a gene
  - Downregulate the expression of a gene
  - Change target specificity of protein
  - Usually species specific




Possible point mutation

# Genetic basis of AMR

- AMR is conferred by different mechanisms:
  - Acquired resistance genes
  - Mutation
  - (Copy numbers)
- MGEs can transfer resistance genes between isolates closely or distantly related
- Resistance genes tend to aggregate, meaning MGEs often confer resistance to multiple classes
- May integrate into host chromosome

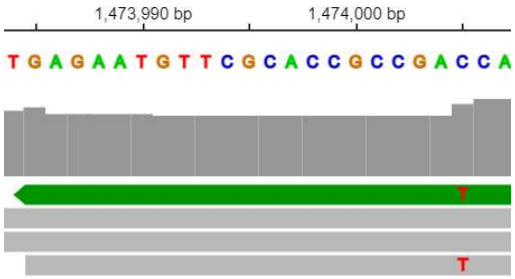


Transfer of plasmid with resistance gene

Note!

We also have intrinsic resistance in certain species, e.g. *Mycobacterium tuberculosis* inherently possess erm(37) protecting against macrolides, lincosamide and streptogramin

- Point mutations can confer resistance by various mechanisms:
  - Change the target of a drug, making the strain resistant
  - Upregulate the expression of a gene
  - Downregulate the expression of a gene
  - Change target specificity of protein
  - Usually species specific




Possible point mutation

# AMR tools and databases

- There are multiple tools which all utilize their own and/or each others databases for predicting antimicrobial resistance
  - Resfinder (ResFinder 4.1 (dtu.dk)), AMRfinderplus (Releases · ncbi/amr (github.com)) , CARD (https://card.mcmaster.ca/home), KmerResistance, ARIBA
  - Differences exists due to
    - How the database is created and curated
    - How the database is maintained
    - How the tool conducts its search
  - The correct tool/database will likely depend on the type of analysis or workflow you are using
  - Approach results from tools with a critical mindset!

**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 | msbA | | 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 |

| Filename |
| --- |
| GCF_000005845.2_ASM584v2_genomic |

**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 |

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 | |

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 | |

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

# Resfinder

- Refinder is a tool and database for detection of genes and point mutations conferring AMR.

- The tool has (3) databases:
- Poinfinder_db
- Resfinder_db
- (DesinFinder_db)

https://cge.food.dtu.dk/services/ResFinder/

Databases with antimicrobial resistance genes and chromosomal point mutations

**ResFinder database** **PointFinder database**

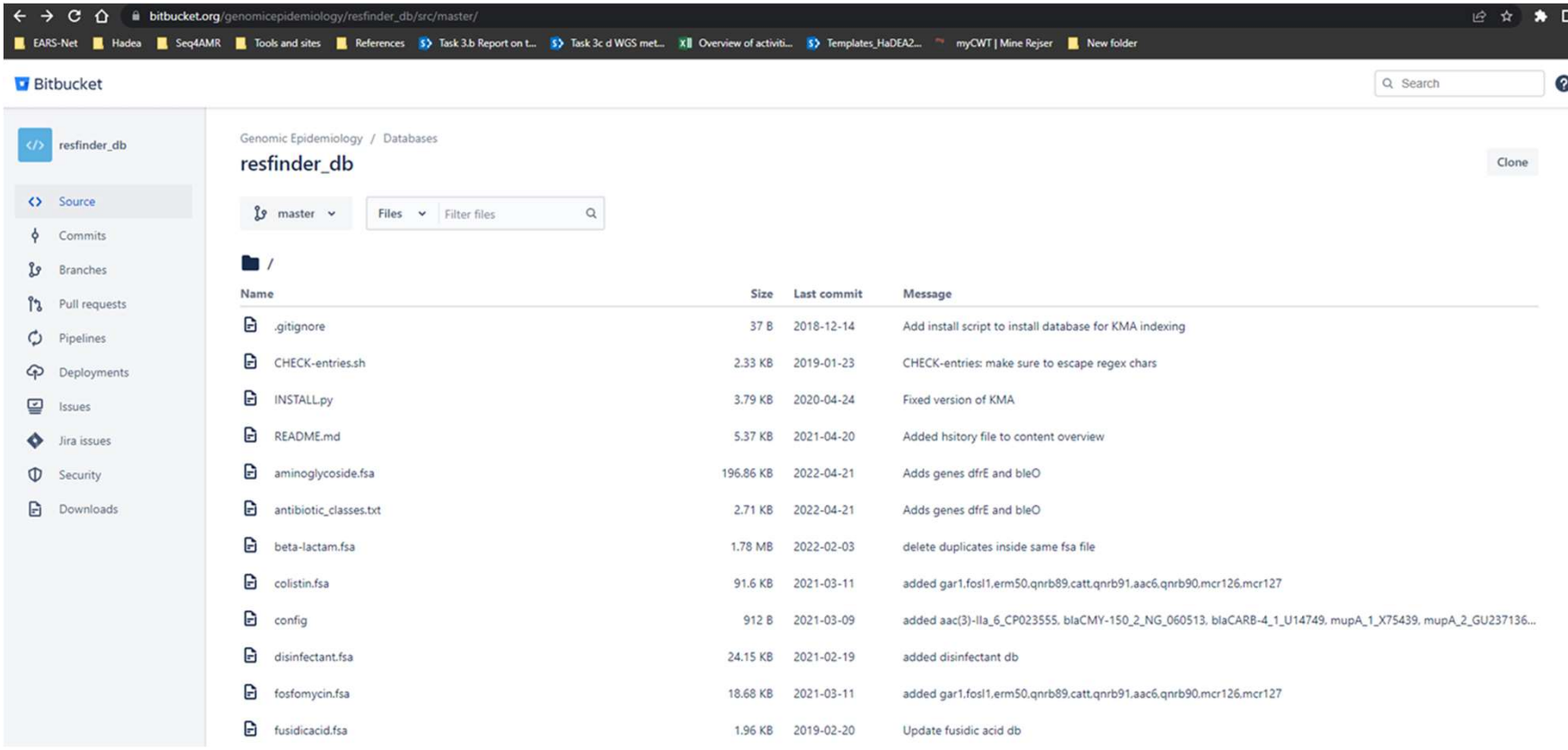
https://bitbucket.org/genomicepidemiology/resfinder_db/src/master/

https://bitbucket.org/genomicepidemiology/pointfinder_db/src/master/

*Source: Genotypic detection of AMR in Salmonella, presentation by Ana Rita Rebelo*

# Resfinder

- The tool and databases are freely available on github
- The main page contains instructions for installation of the tool and how to download and install the associated databases
- We will be using this tool today

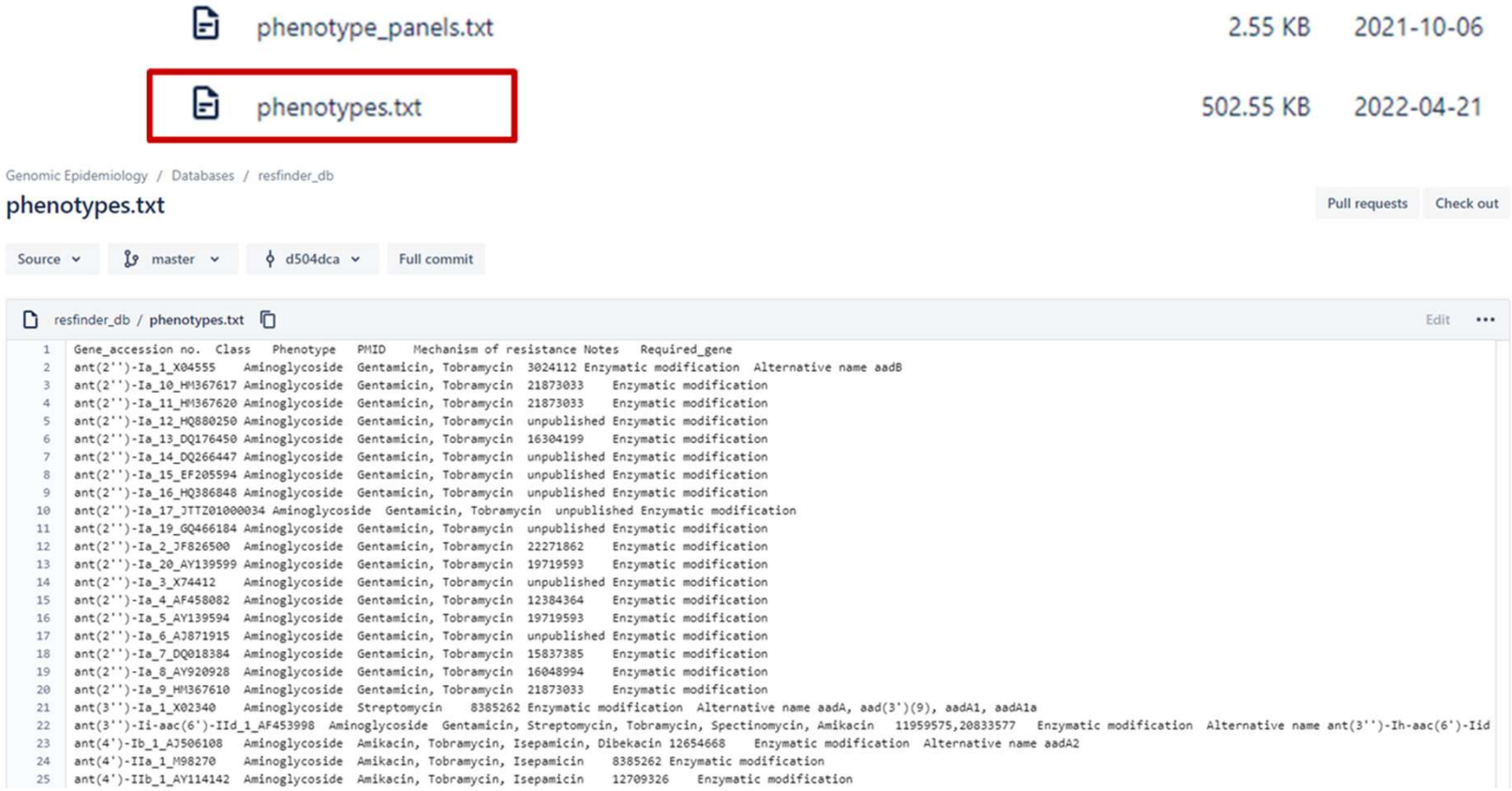



*Source: Genotypic detection of AMR in Salmonella, presentation by Ana Rita Rebelo*

# Resfinder

- The databases contain:
  - Nucleotide references for AMR elements
  - Curation for each genes
  - Phenotypes associated with specific species

- We can get:
  - Class of AMR gene
  - Specific phenotypes
  - Supporting sources
  - Mechanism
  - Other notes

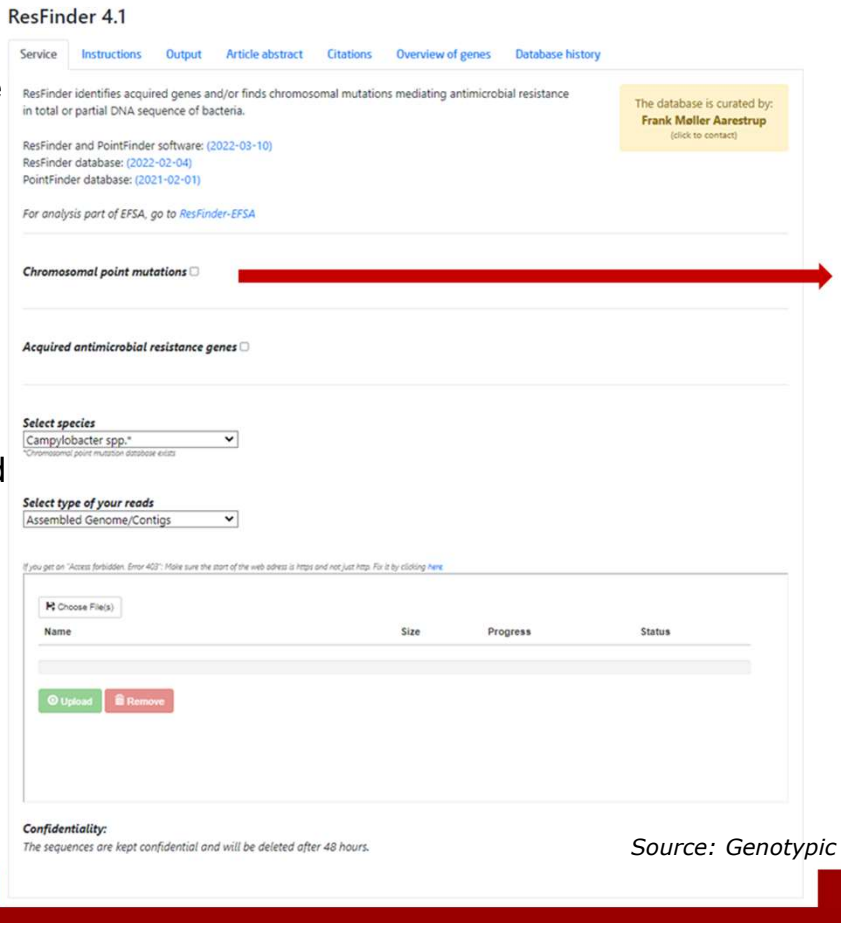


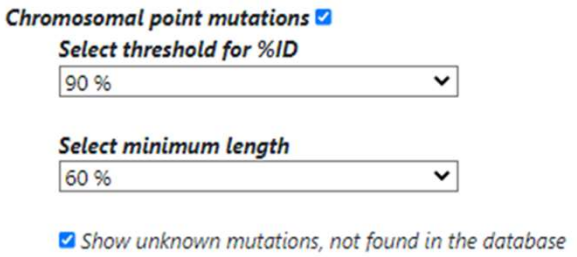

| Gene_accession no. | Class | Phenotype | PMID | Mechanism of resistance | Notes | Required_gene |
|---|---|---|---|---|---|---|
| ant(2'')-Ia_1_X04555 | Aminoglycoside | Gentamicin, Tobramycin | 3024112 | Enzymatic modification | Alternative name aadB | |
| ant(2'')-Ia_10_HM367617 | Aminoglycoside | Gentamicin, Tobramycin | 21873033 | Enzymatic modification | | |
| ant(2'')-Ia_11_HM367620 | Aminoglycoside | Gentamicin, Tobramycin | 21873033 | Enzymatic modification | | |
| ant(2'')-Ia_12_HQ880250 | Aminoglycoside | Gentamicin, Tobramycin | unpublished | Enzymatic modification | | |
| ant(2'')-Ia_13_DQ176450 | Aminoglycoside | Gentamicin, Tobramycin | 16304199 | Enzymatic modification | | |
| ant(2'')-Ia_14_DQ266447 | Aminoglycoside | Gentamicin, Tobramycin | unpublished | Enzymatic modification | | |
| ant(2'')-Ia_15_EF205594 | Aminoglycoside | Gentamicin, Tobramycin | unpublished | Enzymatic modification | | |
| ant(2'')-Ia_16_HQ386848 | Aminoglycoside | Gentamicin, Tobramycin | unpublished | Enzymatic modification | | |
| ant(2'')-Ia_17_JTTZ01000034 | Aminoglycoside | Gentamicin, Tobramycin | unpublished | Enzymatic modification | | |
| ant(2'')-Ia_19_GQ466184 | Aminoglycoside | Gentamicin, Tobramycin | unpublished | Enzymatic modification | | |
| ant(2'')-Ia_2_JF826500 | Aminoglycoside | Gentamicin, Tobramycin | 22271862 | Enzymatic modification | | |
| ant(2'')-Ia_20_AY139599 | Aminoglycoside | Gentamicin, Tobramycin | 19719593 | Enzymatic modification | | |
| ant(2'')-Ia_3_X74412 | Aminoglycoside | Gentamicin, Tobramycin | unpublished | Enzymatic modification | | |
| ant(2'')-Ia_4_AF458082 | Aminoglycoside | Gentamicin, Tobramycin | 12384364 | Enzymatic modification | | |
| ant(2'')-Ia_5_AY139594 | Aminoglycoside | Gentamicin, Tobramycin | 19719593 | Enzymatic modification | | |
| ant(2'')-Ia_6_AJ871915 | Aminoglycoside | Gentamicin, Tobramycin | unpublished | Enzymatic modification | | |
| ant(2'')-Ia_7_DQ018384 | Aminoglycoside | Gentamicin, Tobramycin | 15837385 | Enzymatic modification | | |
| ant(2'')-Ia_8_AY920928 | Aminoglycoside | Gentamicin, Tobramycin | 16048994 | Enzymatic modification | | |
| ant(2'')-Ia_9_HM367610 | Aminoglycoside | Gentamicin, Tobramycin | 21873033 | Enzymatic modification | | |
| ant(3'')-Ia_1_X02340 | Aminoglycoside | Streptomycin | 8385262 | Enzymatic modification | Alternative name aadA, aad(3')(9), aadA1, aadA1a | |
| ant(3'')-Ii-aac(6')-IId_1_AF453998 | Aminoglycoside | Gentamicin, Streptomycin, Tobramycin, Spectinomycin, Amikacin | 11959575,20833577 | Enzymatic modification | Alternative name ant(3'')-Ih-aac(6')-Iid | |
| ant(4')-Ib_1_AJ506108 | Aminoglycoside | Amikacin, Tobramycin, Isepamicin, Dibekacin | 12654668 | Enzymatic modification | Alternative name aadA2 | |
| ant(4')-IIa_1_M98270 | Aminoglycoside | Amikacin, Tobramycin, Isepamicin | 8385262 | Enzymatic modification | | |
| ant(4')-IIb_1_AY114142 | Aminoglycoside | Amikacin, Tobramycin, Isepamicin | 12709326 | Enzymatic modification | | |

*Source: Genotypic detection of AMR in Salmonella, presentation by Ana Rita Rebelo*

# Resfinder

- The tool is also available online, where isolates can be uploaded
- The stand-alone tool utilize the same inputs, as we will find out in the exercises
- Select the databases to use
- Choose thresholds for ID and COV
- Provide species
- Specify input format
- Give input files

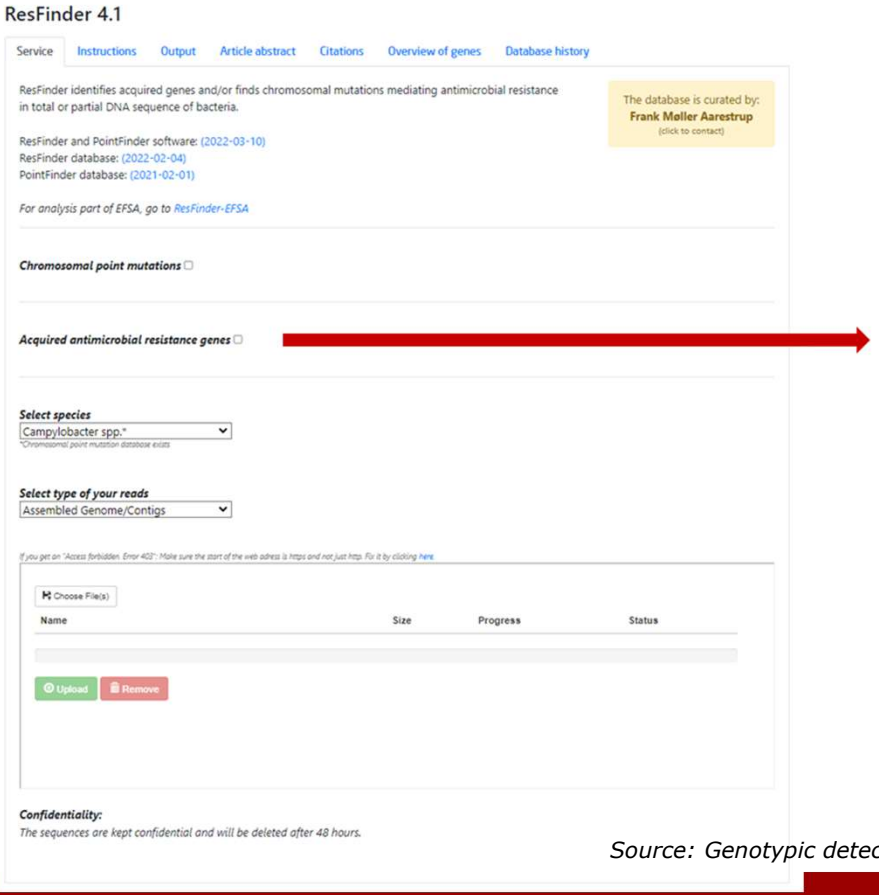


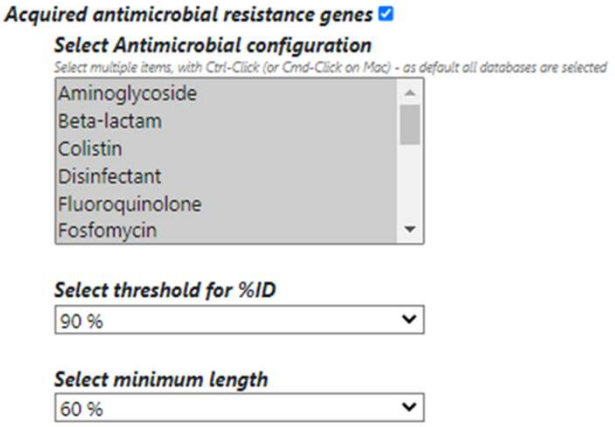




*Source: Genotypic detection of AMR in Salmonella, presentation by Ana Rita Rebelo*

# Resfinder

- The tool is also available online, where isolates can be uploaded
- The stand-alone tool utilize the same inputs, as we will find out in the exercises
- Select the databases to use
- Choose thresholds for ID and COV
- Provide species
- Specify input format
- Give input files
- For AMR genes, specific class can be chosen

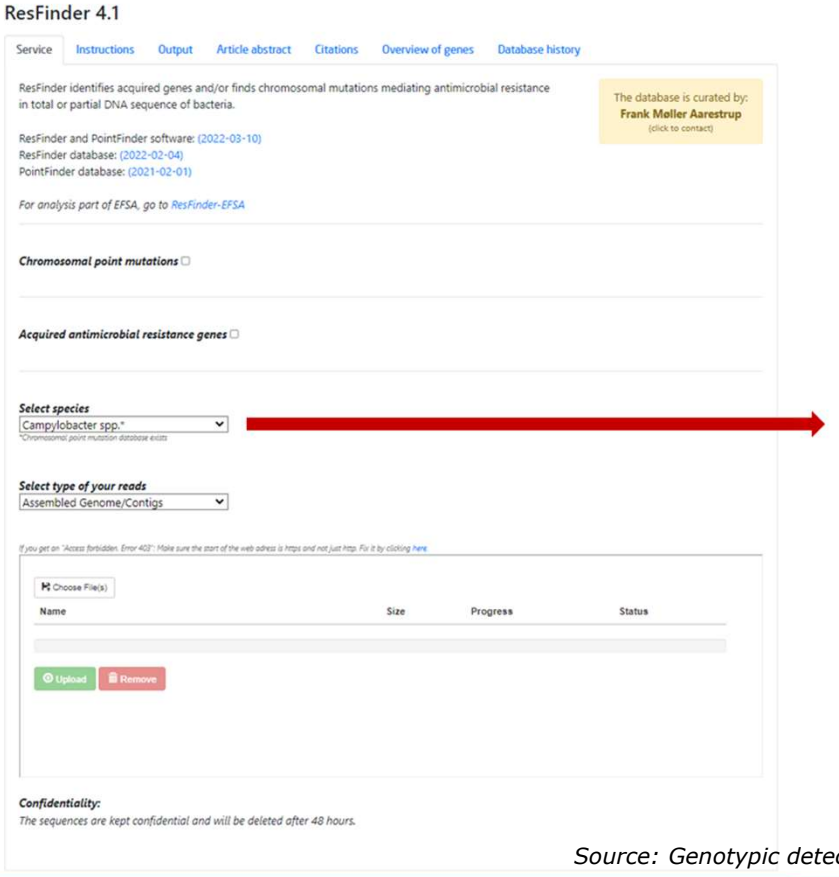


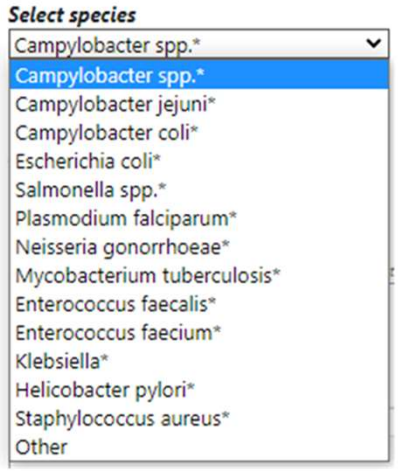




*Source: Genotypic detection of AMR in Salmonella, presentation by Ana Rita Rebelo*

# Resfinder

- Species are neccesary for point mutations
- Not all species are present
- "Other" allows us to search for AMR genes




*Chromosomal point mutation database exists*

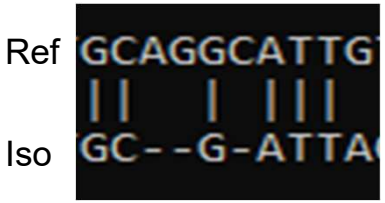
*Source: Genotypic detection of AMR in Salmonella, presentation by Ana Rita Rebelo*

# ID and COV

- What is identity (ID)?
  - Proportion of matching nucleotides:


9/10 bp align = 90% ID

- What is coverage/length (COV)?
  - Proportion of nucleotides covered:




7/10 bp covered = 70% COV, 6/10 bp align = 60% ID

# Note on databases

- Make your own database!
  - MyDBFinder (https://cge.food.dtu.dk/services/MyKMAfinder/)
  - MyKMAFinder (https://cge.food.dtu.dk/services/MyDbFinder/)
  - NCBI-Blast (BLAST+ executables — BLASTHelp documentation (nih.gov))
- Organize genes of interest into a fasta file
  - Upload to online solution
  - Make blast database with makeblastdb (part of standalone blast tools)
- Large range of customization
- Better options likely exist for larger datasets

# Acknowledgements

**The creation of this training material was commissioned by ECDC to Statens Serum Institut (SSI) and produced by The National Food Institute at the Technical University of Denmark (DTU) with the direct involvement of Lauge Holm Sørensen and Niamh Lacy-Roberts**

# References


- Dominguez Del Angel V, Hjerde E, Sterck L *et al.* Ten steps to get started in Genome Assembly and Annotation [version 1; peer review: 2 approved]. *F1000Research* 2018, **7**(ELIXIR):148 (https://doi.org/10.12688/f1000research.13598.1)
- Murray, Christopher J. L., et al. “Global Burden of Bacterial Antimicrobial Resistance in 2019: a Systematic Analysis.” Lancet, vol. 399, no. 10325, Elsevier B.V., 2022, pp. 629–55, doi:10.1016/S0140-6736(21)02724-0.
- Norrby, S. Ragnar, et al. “Lack of Development of New Antimicrobial Drugs: A Potential Serious Threat to Public Health.” Lancet Infectious Diseases, vol. 5, no. 2, Lancet Publishing Group, 2005, pp. 115–19, doi:10.1016/S1473-3099(05)70086-4.
- Zhang, P., Mao, D., Gao, H. et al. Colonization of gut microbiota by plasmid-carrying bacteria is facilitated by evolutionary adaptation to antibiotic treatment. ISME J 16, 1284–1293 (2022). https://doi.org/10.1038/s41396-021-01171-x
- Fratamico PM, DebRoy C, Liu Y, Needleman DS, Baranzoni GM, Feng P. Advances in Molecular Serotyping and Subtyping of Escherichia coli. Front Microbiol. 2016 May 3;7:644. doi: 10.3389/fmicb.2016.00644. PMID: 27199968; PMCID: PMC4853403.
- Joensen, Katrine Grimstrup, et al. “Rapid and Easy In Silico Serotyping of Escherichia Coli Isolates by Use of Whole-Genome Sequencing Data.” Journal of Clinical Microbiology, edited by K. C. Carroll, vol. 53, no. 8, American Society for Microbiology, 2015, pp. 2410–26, doi:10.1128/JCM.00008-15.