SCIENTIFIC WEBINAR #1

GETTING STARTED WITH ONLINE SEQUENCE ANALYSIS BY CGE TOOLS

September 14th 2022
Jette S. Kjeldgaard

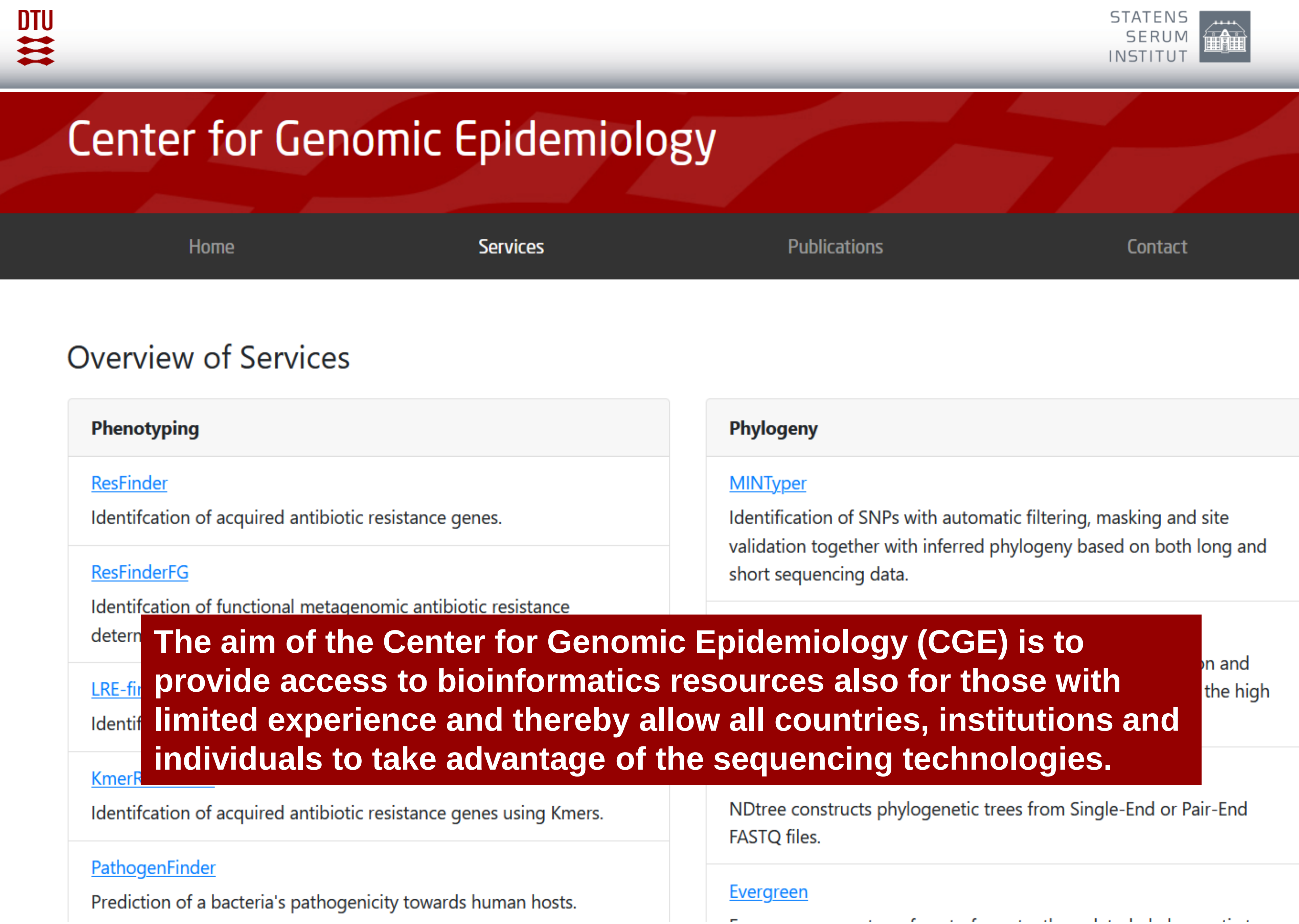
AGENDA

- The concept of CGE tools
- Available tools
- Getting started
- Sequence types

Center for Genomic Epidemiology

THE CONCEPT OF CGE TOOLS?

- Collection of genomic tools with wide range of applications
- Some tools developed by CGE group, others are only linked
  - to have a simple and similar user interface
- Easy to use
  - Submit, get result by email, interpret
- Limited to one sequence/isolate at a time
- WGS – one sequence – numerous applications!
  - Does not (always) require coding skills & super-computers



Center for Genomic Epidemiology

Overview of Services

Phenotyping

ResFinder

Identifcation of acquired antibiotic resistance genes.

ResFinderFG

Identifcation of functional metagenomic antibiotic resistance determ

LRE-fi

Identif

KmerR

Identifcation of acquired antibiotic resistance genes using Kmers.

PathogenFinder

Prediction of a bacteria's pathogenicity towards human hosts.

Phylogeny

MINTyper

Identification of SNPs with automatic filtering, masking and site validation together with inferred phylogeny based on both long and short sequencing data.

n and

the high

NDtree constructs phylogenetic trees from Single-End or Pair-End FASTQ files.

Evergreen

The aim of the Center for Genomic Epidemiology (CGE) is to provide access to bioinformatics resources also for those with limited experience and thereby allow all countries, institutions and individuals to take advantage of the sequencing technologies.

CGE TOOLS COLLECTION - OVERVIEW

- **Phenotyping**
- **Typing**
- **Phylogeny**
- Metagenomics
- PCR tool

Sequence contamination
Species determination
Subspecies typing
Resistance genes and mutations
Outbreak/cluster analysis

- Others:
  - User-defined data bases
  - Identify bacteriophage sequences in metagenomic data
  - Identify the bacterial host of phage genome
  - …
  - …

CGE TOOLS PHENOTYPING – AMR

Phenotyping

- Antimicrobial resistance
  - **ResFinder**
  - ResFinderFG
    - identifies a resistance phenotype based on a functional metagenomic antibiotic resistance determinants database
  - Kmer resistance
  - LRE-Finder
    - Identification of genes and mutations leading to linezolid resistance

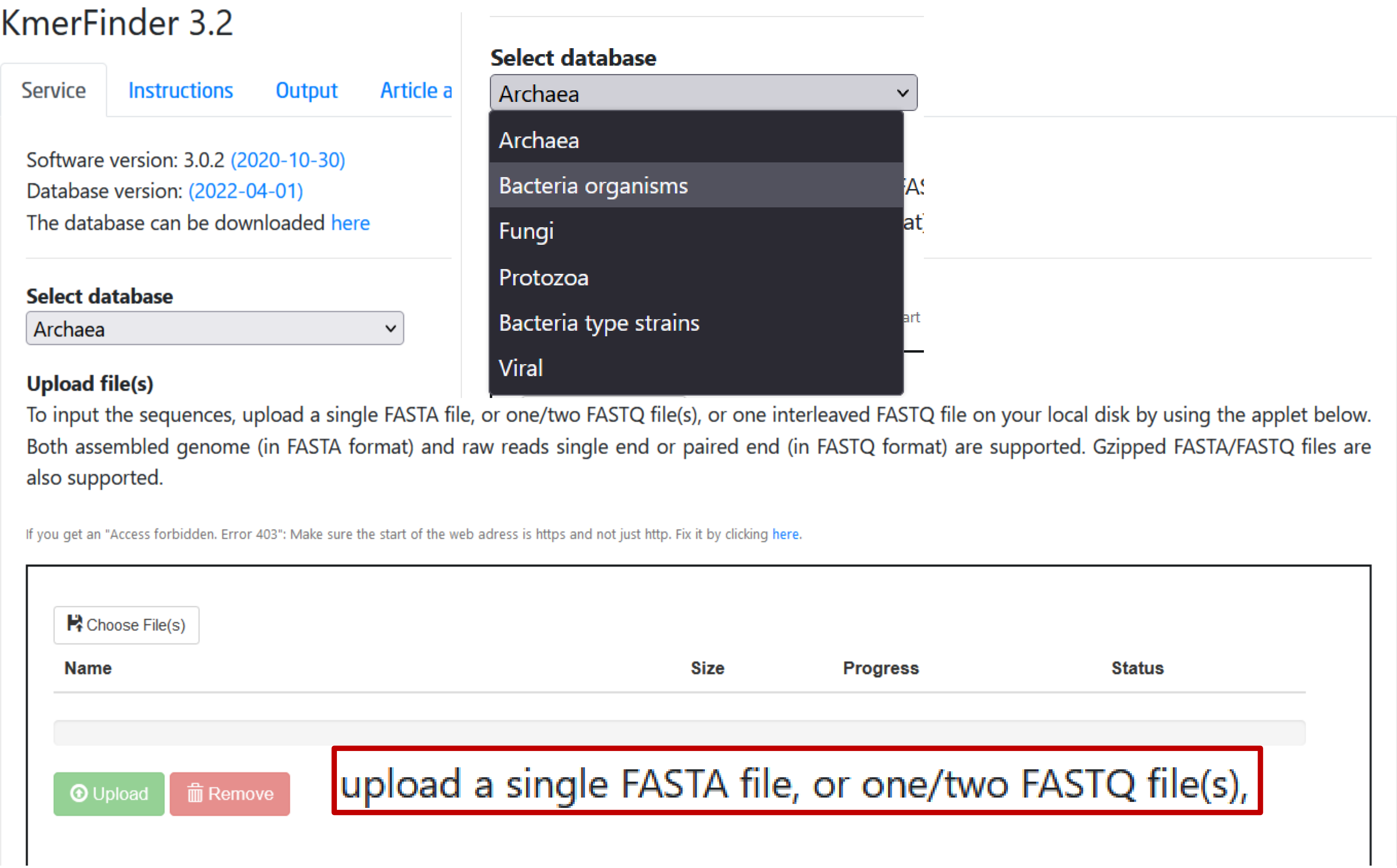
CGE TOOLS TYPING

- **Species identification**
  - **KmerFinder** (kmer-based)
  - SpeciesFinder (16S ribosomal DNA)

- **Typing**
  - MLST
  - cgMLSTFinder
    - *Campylobacter, Clostridum, E. coli, Listeria, Salmonella, Yersinia*
  - pMLST
  - Serotyping (*E. coli, P. aeruginosa, Salmonella*)

  - PlasmidFinder
  - MGE (Mobile genetic elements)

EXAMPLE: KMER-FINDER (SPECIES ID)



KmerFinder 3.2

Software version: 3.0.2 (2020-10-30)
Database version: (2022-04-01)
The database can be downloaded here

Select database

Archaea

Select database

Archaea

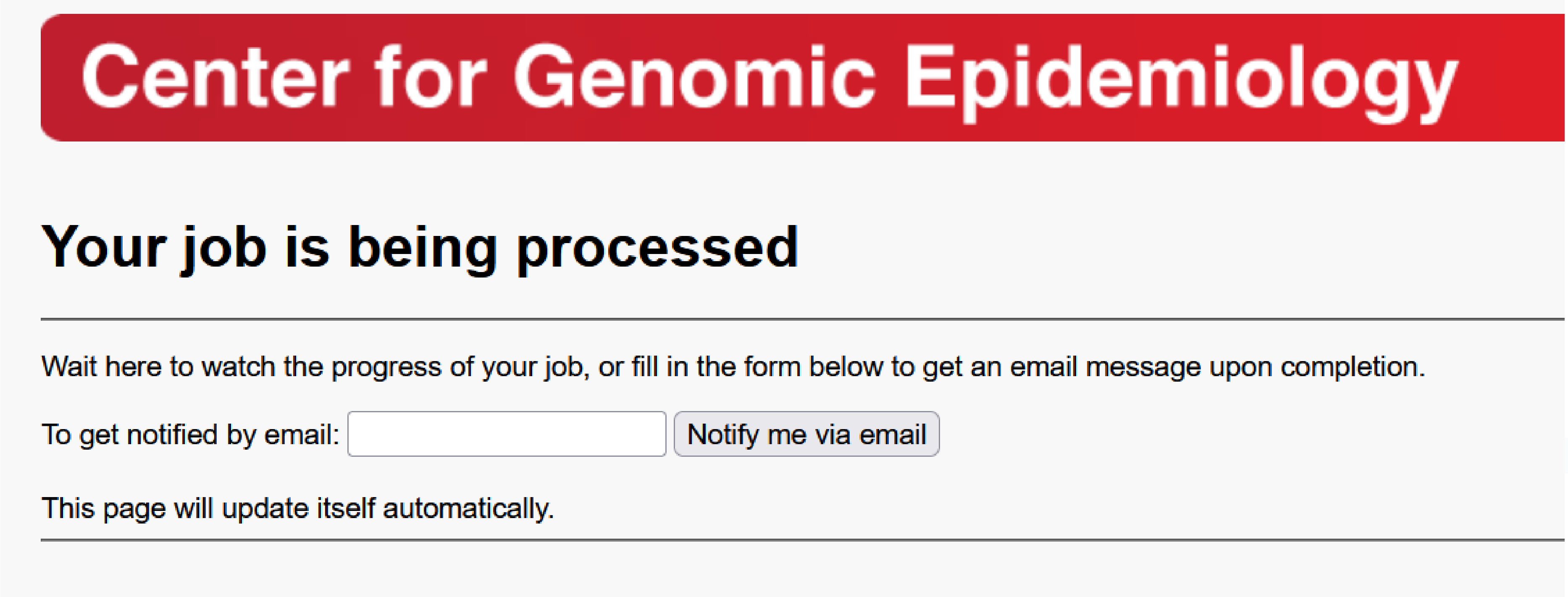
Upload file(s)

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.

upload a single FASTA file, or one/two FASTQ file(s),

EXAMPLE: KMER-FINDER (SPECIES ID)



Center for Genomic Epidemiology

Your job is being processed

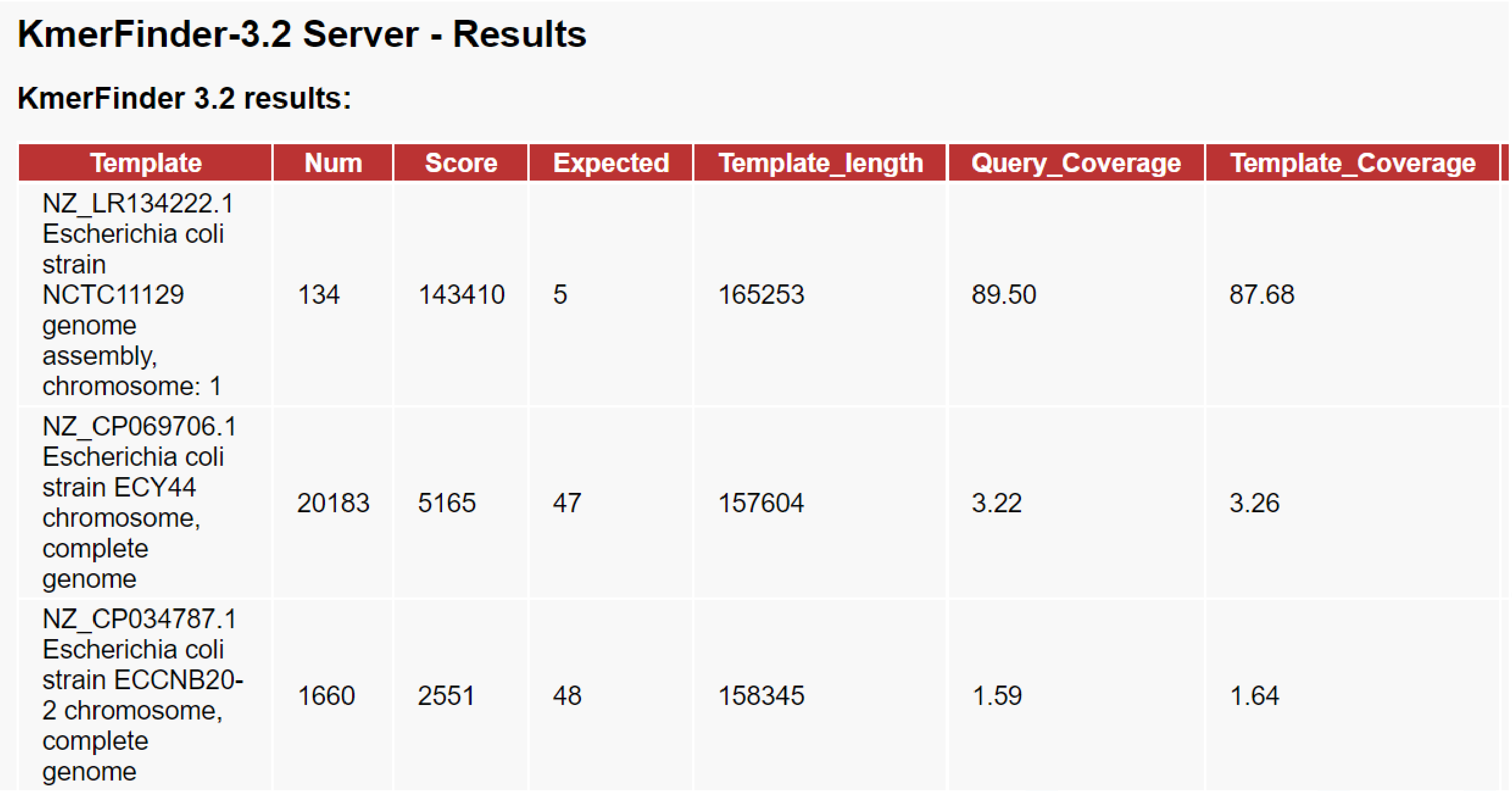
Wait here to watch the progress of your job, or fill in the form below to get an email message upon completion.

To get notified by email:

This page will update itself automatically.

Link to result output can be shared
– valid for a few days or a week

SPECIES ID – MATCH TO DATABASE



KmerFinder-3.2 Server - Results

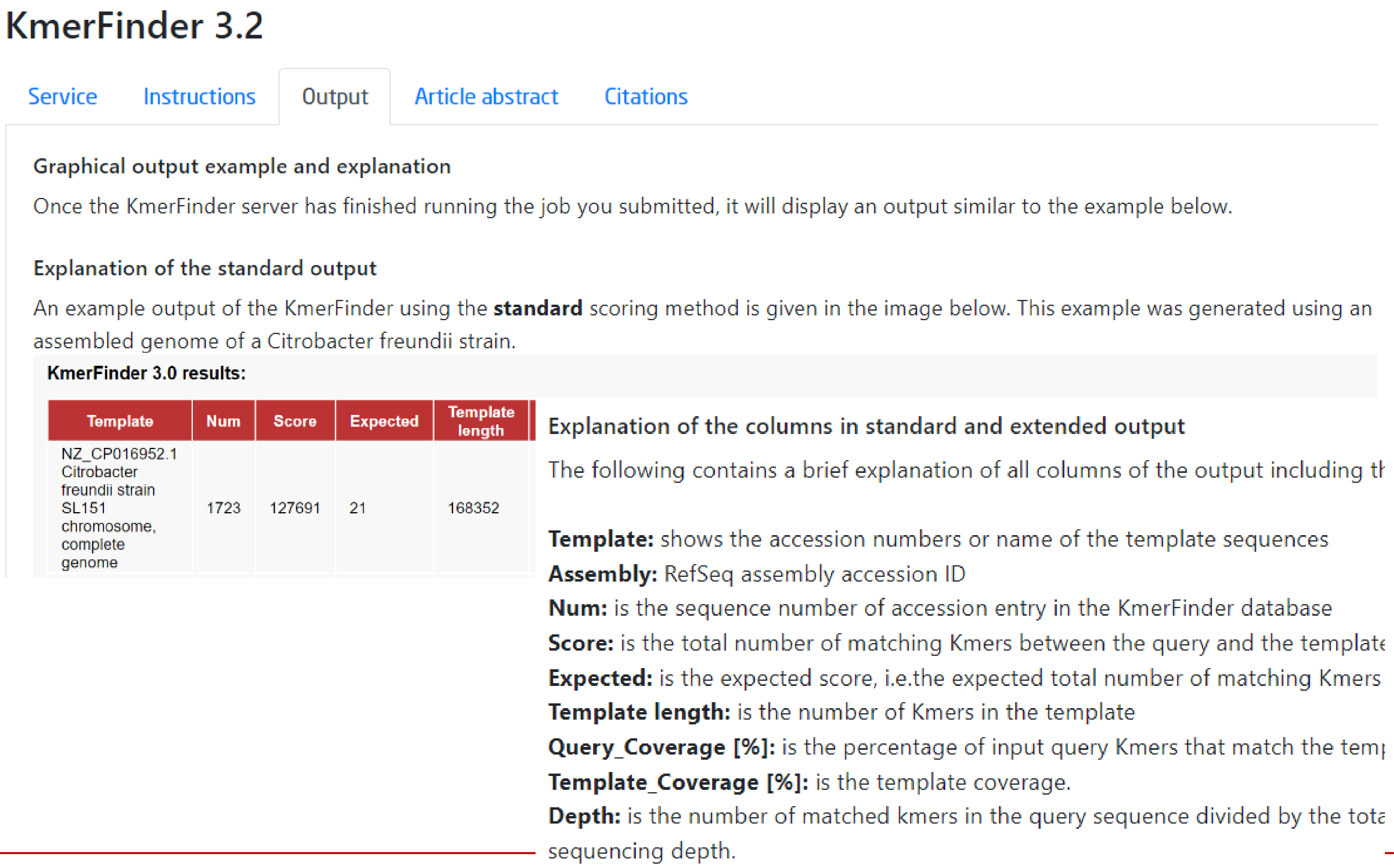
KmerFinder 3.2 results:

Template	Num	Score	Expected	Template_length	Query_Coverage	Template_Coverage
NZ_LR134222.1 Escherichia coli strain NCTC11129 genome assembly, chromosome: 1	134	143410	5	165253	89.50	87.68
NZ_CP069706.1 Escherichia coli strain ECY44 chromosome, complete genome	20183	5165	47	157604	3.22	3.26
NZ_CP034787.1 Escherichia coli strain ECCNB20-2 chromosome, complete genome	1660	2551	48	158345	1.59	1.64

- Good coverage of Query/Template = good match to data base isolate
- A result with several hits is usual
- Check for contamination from other species

INTERPRETATION OF KMER-FINDER RESULT

- Explanations on how to interpret output on website



KmerFinder 3.2

Graphical output example and explanation

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

Explanation of the standard output

An example output of the KmerFinder using the **standard** scoring method is given in the image below. This example was generated using an assembled genome of a Citrobacter freundii strain.

KmerFinder 3.0 results:

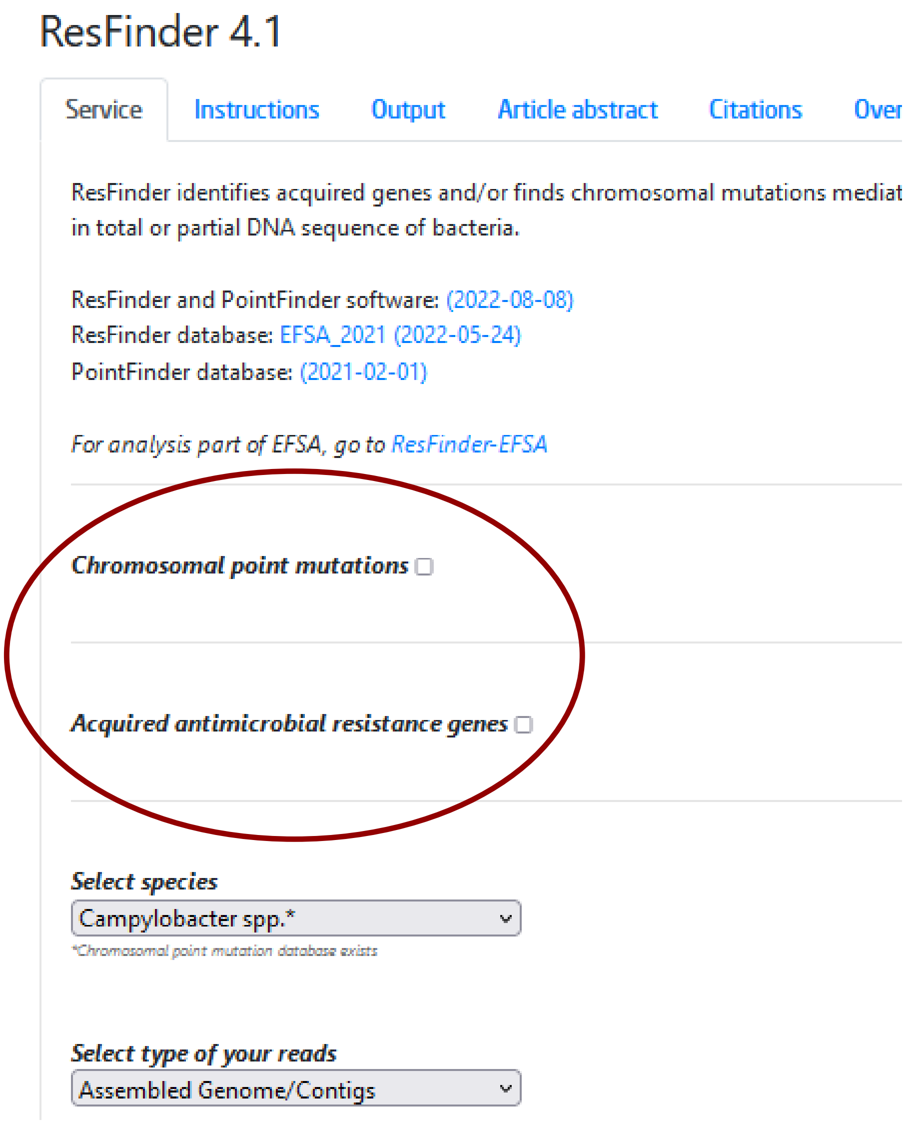
Template	Num	Score	Expected	Template length
NZ_CP016952.1 Citrobacter freundii strain SL151 chromosome, complete genome	1723	127691	21	168352

Explanation of the columns in standard and extended output

The following contains a brief explanation of all columns of the output including th

Template: shows the accession numbers or name of the template sequences
Assembly: RefSeq assembly accession ID
Num: is the sequence number of accession entry in the KmerFinder database
Score: is the total number of matching Kmers between the query and the templat
Expected: is the expected score, i.e.the expected total number of matching Kmers
Template length: is the number of Kmers in the template
Query_Coverage [%]: is the percentage of input query Kmers that match the tem
Template_Coverage [%]: is the template coverage.
Depth: is the number of matched kmers in the query sequence divided by the tota sequencing depth.

EXAMPLE: RESFINDER



ResFinder 4.1

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

ResFinder and PointFinder software: (2022-08-08)
ResFinder database: EFSA_2021 (2022-05-24)
PointFinder database: (2021-02-01)

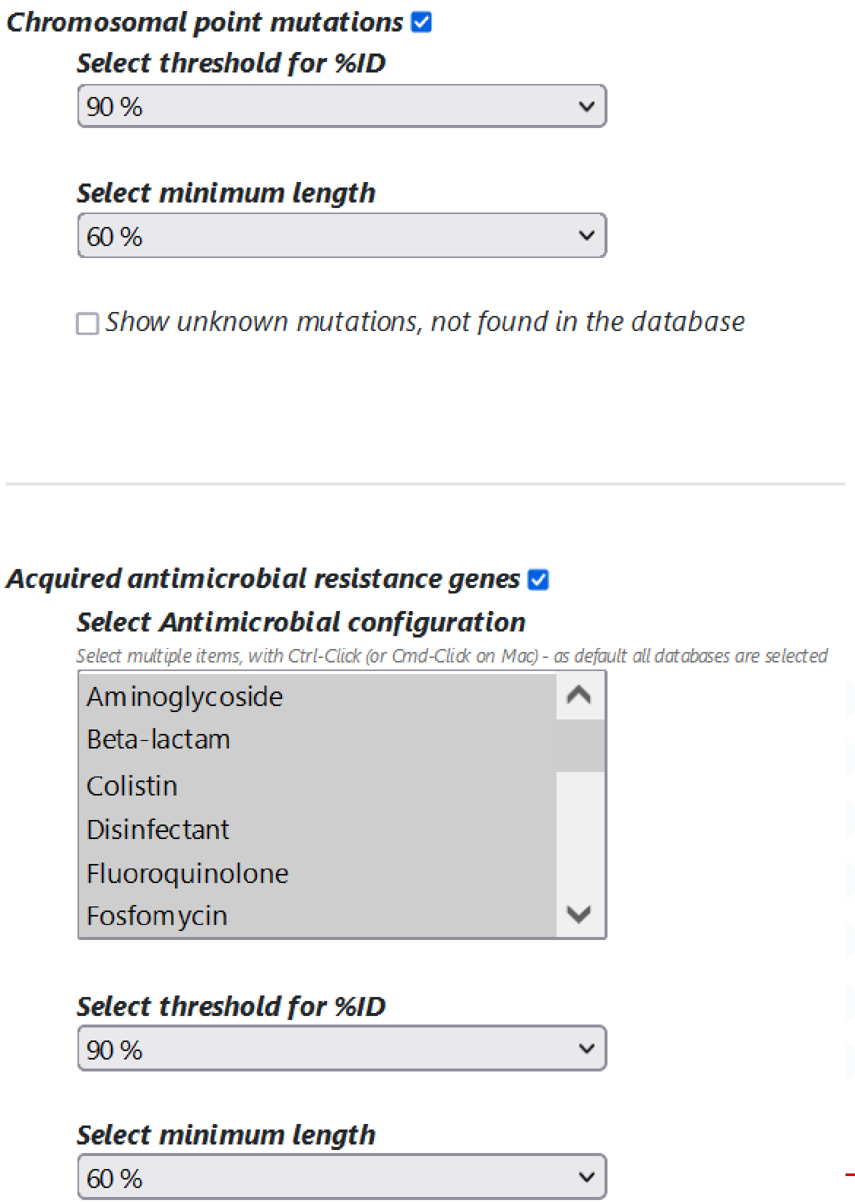
For analysis part of EFSA, go to ResFinder-EFSA

Chromosomal point mutations ☐

Acquired antimicrobial resistance genes ☐

Select species
Campylobacter spp.*
*Chromosomal point mutation database exists

Select type of your reads
Assembled Genome/Contigs



Chromosomal point mutations ☑

Select threshold for %ID
90 %

Select minimum length
60 %

☐ *Show unknown mutations, not found in the database*

Acquired antimicrobial resistance genes ☑

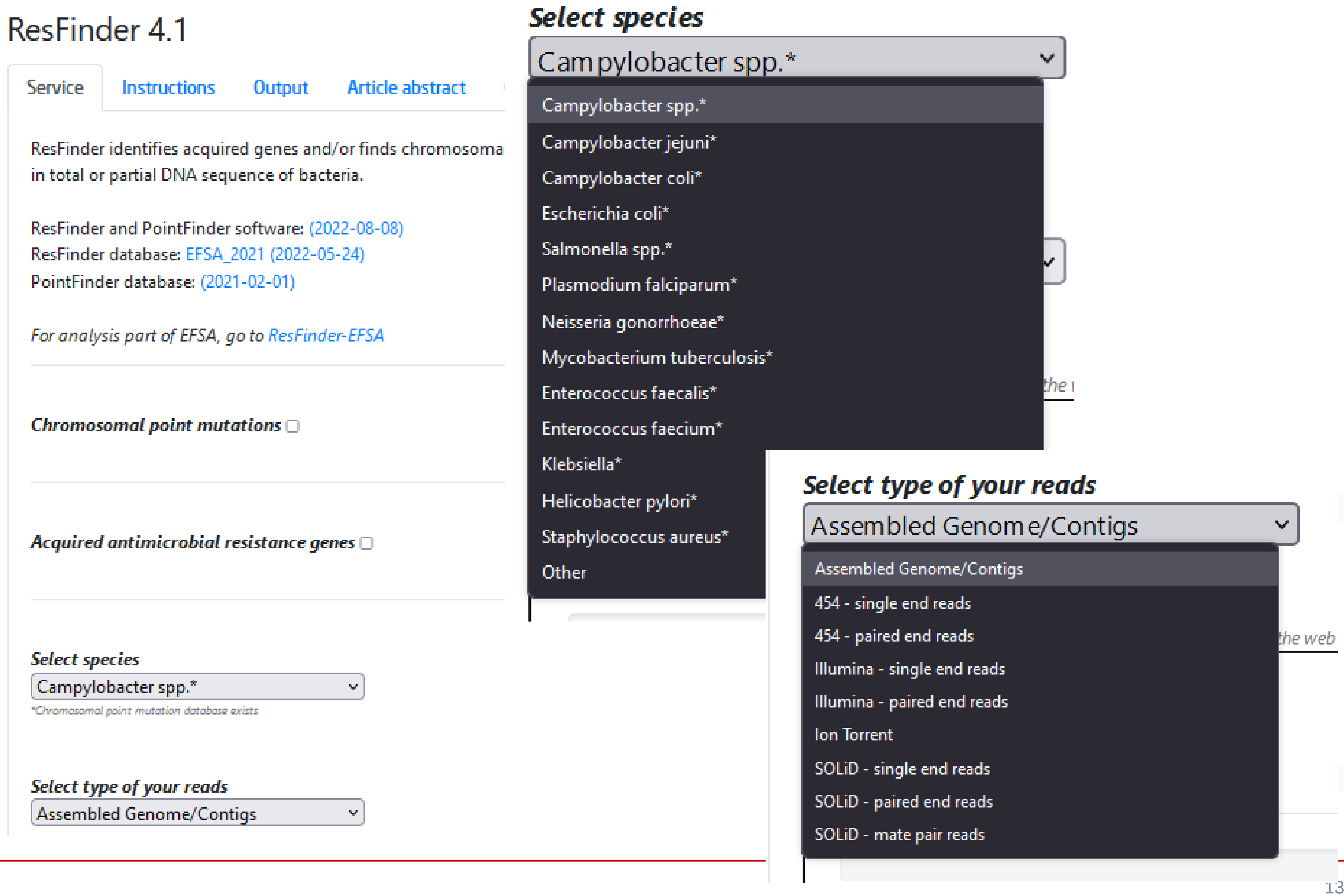
Select Antimicrobial configuration
Select multiple items, with Ctrl-Click (or Cmd-Click on Mac) - as default all databases are selected

- Aminoglycoside
- Beta-lactam
- Colistin
- Disinfectant
- Fluoroquinolone
- Fosfomycin

Select threshold for %ID
90 %

Select minimum length
60 %

EXAMPLE: RESFINDER



ResFinder 4.1

ResFinder identifies acquired genes and/or finds chromosoma
in total or partial DNA sequence of bacteria.

ResFinder and PointFinder software: (2022-08-08)
ResFinder database: EFSA_2021 (2022-05-24)
PointFinder database: (2021-02-01)

For analysis part of EFSA, go to ResFinder-EFSA

Chromosomal point mutations ☐

Acquired antimicrobial resistance genes ☐

Select species
Campylobacter spp.*
**Chromosomal point mutation database exists*

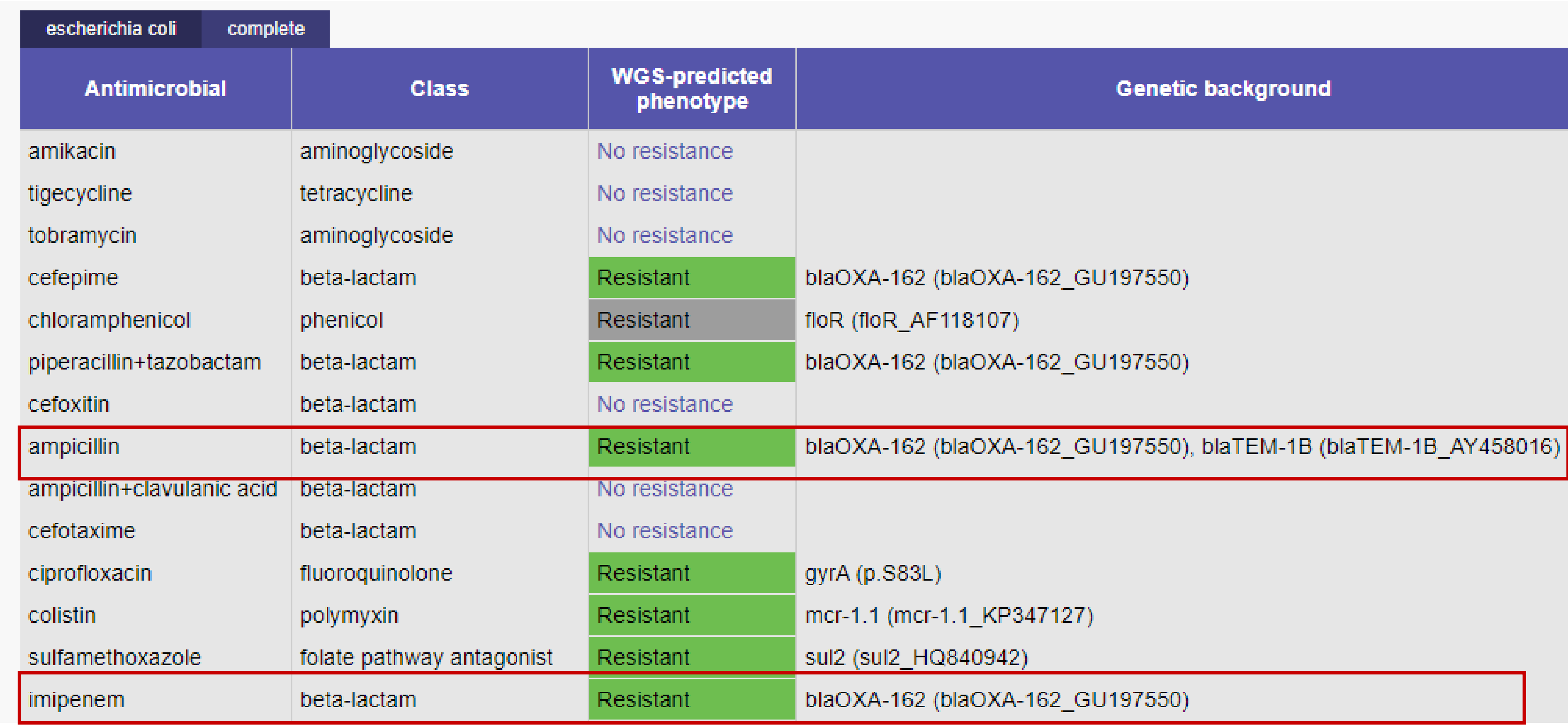
Select type of your reads
Assembled Genome/Contigs

Select species
Campylobacter spp.*

Select type of your reads
Assembled Genome/Contigs

RESFINDER OUTPUT

- Species specific overview of AMR genes present
  - Prediction of class AND phenotype

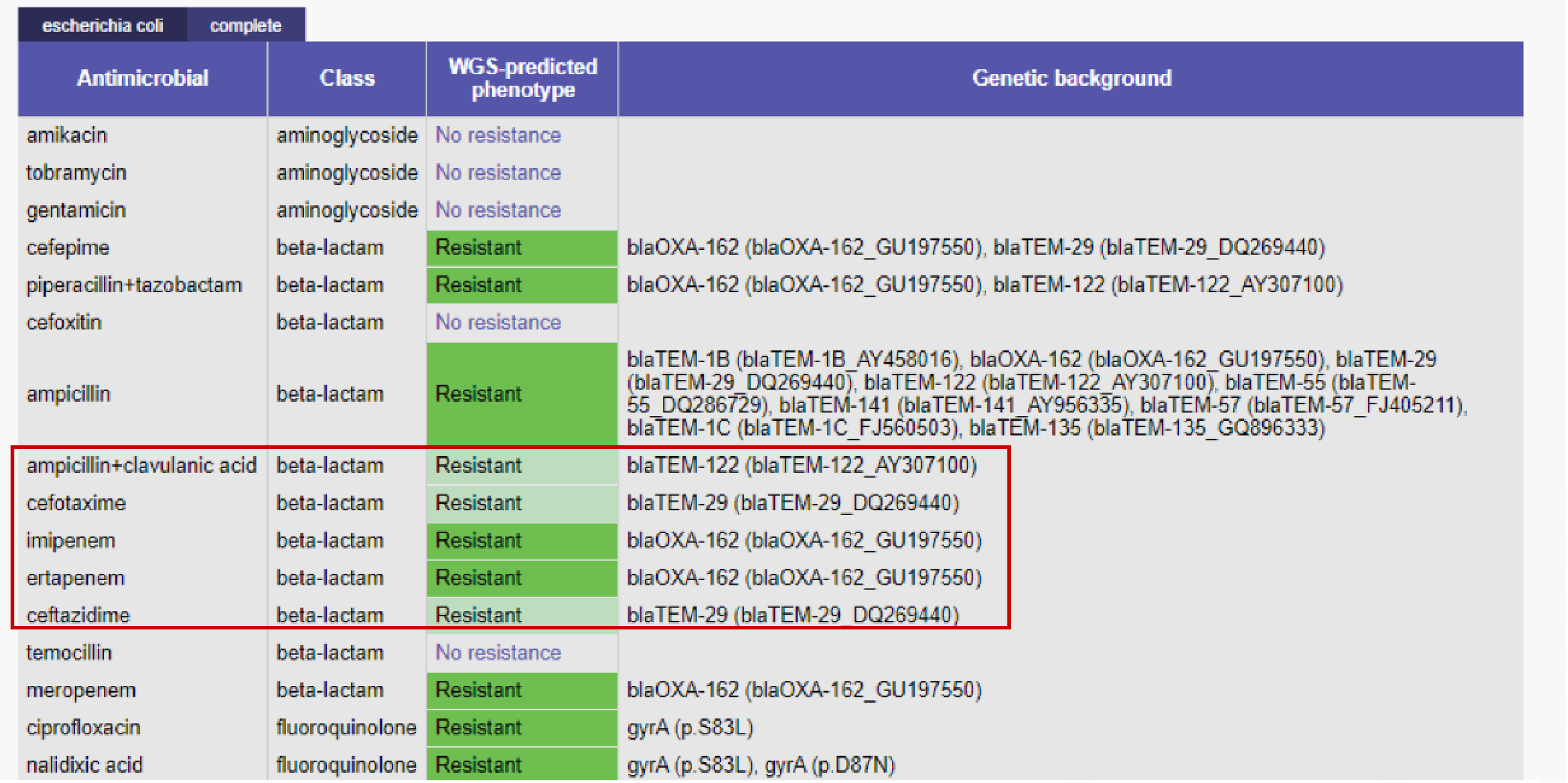


escherichia coli | complete

Antimicrobial	Class	WGS-predicted phenotype	Genetic background
amikacin	aminoglycoside	No resistance	
tigecycline	tetracycline	No resistance	
tobramycin	aminoglycoside	No resistance	
cefepime	beta-lactam	Resistant	blaOXA-162 (blaOXA-162_GU197550)
chloramphenicol	phenicol	Resistant	floR (floR_AF118107)
piperacillin+tazobactam	beta-lactam	Resistant	blaOXA-162 (blaOXA-162_GU197550)
cefoxitin	beta-lactam	No resistance	
ampicillin	beta-lactam	Resistant	blaOXA-162 (blaOXA-162_GU197550), blaTEM-1B (blaTEM-1B_AY458016)
ampicillin+clavulanic acid	beta-lactam	No resistance	
cefotaxime	beta-lactam	No resistance	
ciprofloxacin	fluoroquinolone	Resistant	gyrA (p.S83L)
colistin	polymyxin	Resistant	mcr-1.1 (mcr-1.1_KP347127)
sulfamethoxazole	folate pathway antagonist	Resistant	sul2 (sul2_HQ840942)
imipenem	beta-lactam	Resistant	blaOXA-162 (blaOXA-162_GU197550)

OUTPUT QUALITY DEPENDENT ON INPUT

- Gene variants and non-perfect matches
- Multiple matches in database – multiple matches in output

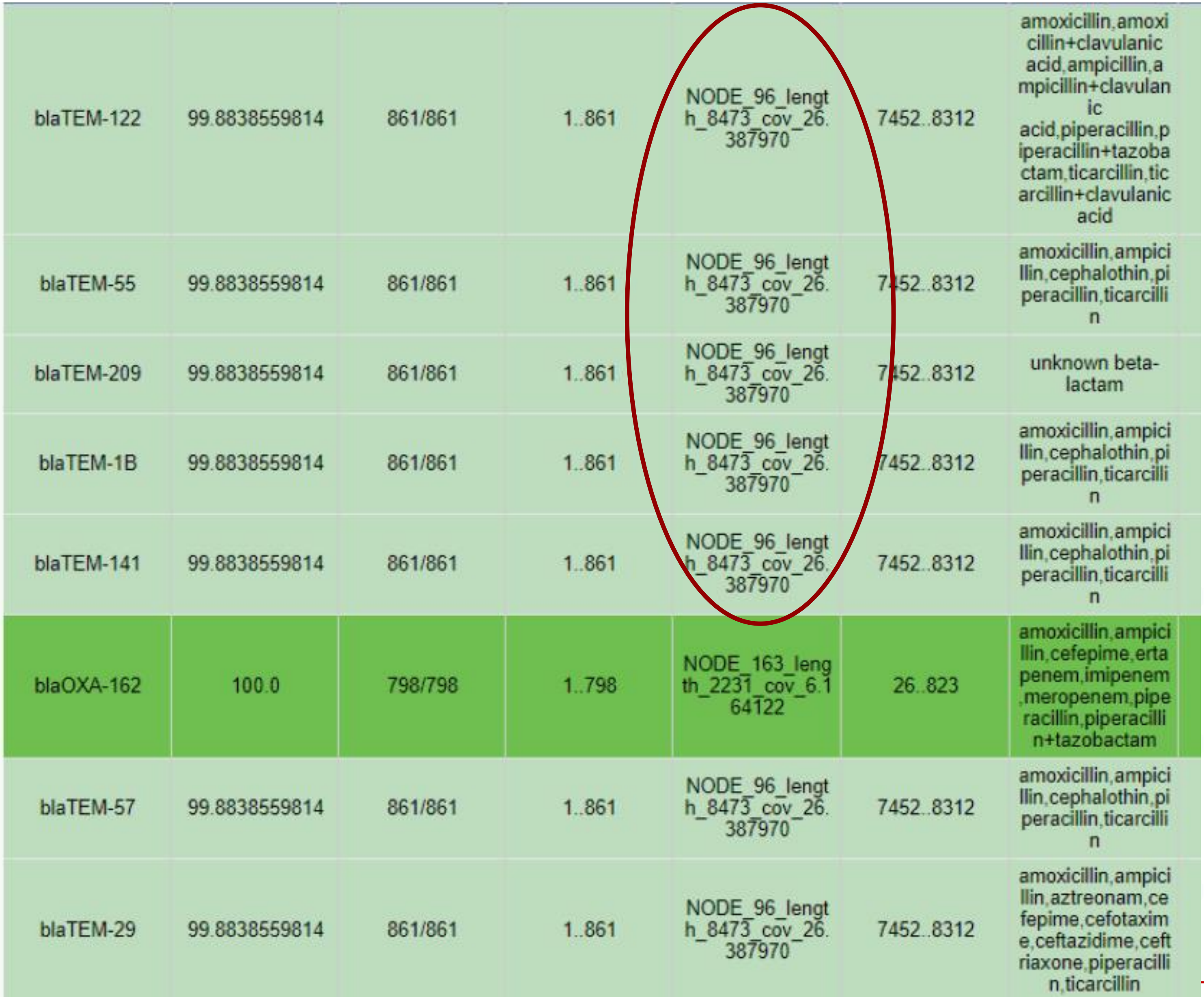


escherichia coli | complete

Antimicrobial	Class	WGS-predicted phenotype	Genetic background
amikacin	aminoglycoside	No resistance	
tobramycin	aminoglycoside	No resistance	
gentamicin	aminoglycoside	No resistance	
cefepime	beta-lactam	Resistant	blaOXA-162 (blaOXA-162_GU197550), blaTEM-29 (blaTEM-29_DQ269440)
piperacillin+tazobactam	beta-lactam	Resistant	blaOXA-162 (blaOXA-162_GU197550), blaTEM-122 (blaTEM-122_AY307100)
cefoxitin	beta-lactam	No resistance	
ampicillin	beta-lactam	Resistant	blaTEM-1B (blaTEM-1B_AY458016), blaOXA-162 (blaOXA-162_GU197550), blaTEM-29 (blaTEM-29_DQ269440), blaTEM-122 (blaTEM-122_AY307100), blaTEM-55 (blaTEM-55_DQ286729), blaTEM-141 (blaTEM-141_AY956335), blaTEM-57 (blaTEM-57_FJ405211), blaTEM-1C (blaTEM-1C_FJ560503), blaTEM-135 (blaTEM-135_GQ896333)
ampicillin+clavulanic acid	beta-lactam	Resistant	blaTEM-122 (blaTEM-122_AY307100)
cefotaxime	beta-lactam	Resistant	blaTEM-29 (blaTEM-29_DQ269440)
imipenem	beta-lactam	Resistant	blaOXA-162 (blaOXA-162_GU197550)
ertapenem	beta-lactam	Resistant	blaOXA-162 (blaOXA-162_GU197550)
ceftazidime	beta-lactam	Resistant	blaTEM-29 (blaTEM-29_DQ269440)
temocillin	beta-lactam	No resistance	
meropenem	beta-lactam	Resistant	blaOXA-162 (blaOXA-162_GU197550)
ciprofloxacin	fluoroquinolone	Resistant	gyrA (p.S83L)
nalidixic acid	fluoroquinolone	Resistant	gyrA (p.S83L), gyrA (p.D87N)

- Beta-lactam resistance - ampicillin
  - TEM-1B, TEM-1C, 1TEM-29, TEM-122, TEM-55, TEM-57, TEM-135, TEM-141

EXTENDED OUTPUT



blaTEM-122	99.8838559814	861/861	1..861	NODE_96_length_8473_cov_26.387970	7452..8312	amoxicillin,amoxicillin+clavulanic acid,ampicillin,ampicillin+clavulanic acid,piperacillin,piperacillin+tazobactam,ticarcillin,ticarcillin+clavulanic acid
blaTEM-55	99.8838559814	861/861	1..861	NODE_96_length_8473_cov_26.387970	7452..8312	amoxicillin,ampicillin,cephalothin,piperacillin,ticarcillin
blaTEM-209	99.8838559814	861/861	1..861	NODE_96_length_8473_cov_26.387970	7452..8312	unknown beta-lactam
blaTEM-1B	99.8838559814	861/861	1..861	NODE_96_length_8473_cov_26.387970	7452..8312	amoxicillin,ampicillin,cephalothin,piperacillin,ticarcillin
blaTEM-141	99.8838559814	861/861	1..861	NODE_96_length_8473_cov_26.387970	7452..8312	amoxicillin,ampicillin,cephalothin,piperacillin,ticarcillin
blaOXA-162	100.0	798/798	1..798	NODE_163_length_2231_cov_6.164122	26..823	amoxicillin,ampicillin,cefepime,ertapenem,imipenem,meropenem,piperacillin,piperacillin+tazobactam
blaTEM-57	99.8838559814	861/861	1..861	NODE_96_length_8473_cov_26.387970	7452..8312	amoxicillin,ampicillin,cephalothin,piperacillin,ticarcillin
blaTEM-29	99.8838559814	861/861	1..861	NODE_96_length_8473_cov_26.387970	7452..8312	amoxicillin,ampicillin,aztreonam,cefepime,cefotaxime,ceftazidime,ceftriaxone,piperacillin,ticarcillin

All the TEM- genes are in fact the same gene

– placed on NODE_96 in position 7452-8312.

None of them have a perfect match, so it is likely a new variant, bad sequence or at least something that wasn't recognized by the database.

Try running on raw reads (fastq) instead of contigs (fasta)

ILLUMINA FASTQ/FASTA

- Almost all the CGE tools are available for Illumina reads/fasta files

Tool	Result
KmerFinder	Species identification/confirmation and check for contamination
ResFinder (Incl. PointFinder)	Identification of acquired resistance genes and point mutations, phenotype prediction
PlasmidFinder/MGEFinder	Predicting presence of plasmids/relation to resistance genes
MLST	Sub-species level typing and check for contamination
CSIPhylogeny	Cluster detection and SNP calling
…etc….	

ONT MINION

- Some tools can also analyse ONT MinION data, including:

Tool	Result
KmerFinder	Species identification/confirmation and check for contamination
KmerResistance	Identification of acquired resistance genes
MLST	Sub-species level typing and check for contamination
MinTyper	Cluster detection and SNP calling

- Subtyping and cluster analysis will be presented on the virtual multidisciplinary exercise on Sept 26th and Oct 10th
  - More information about this next week

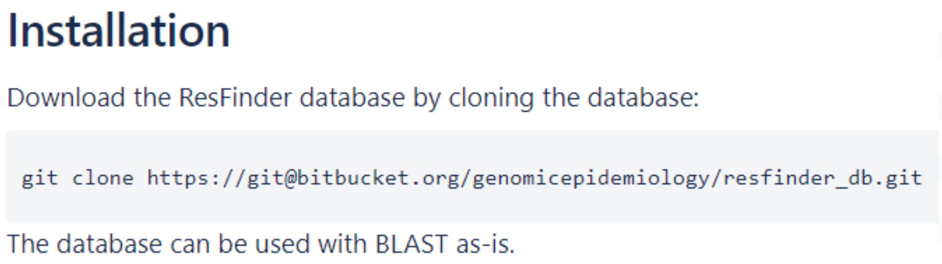
AVAILABILITY OF TOOLS

- For local installation
  - Download tools and databases at

https://bitbucket.org/genomicepidemiology/

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

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



Installation

Download the ResFinder database by cloning the database:

```
git clone https://git@bitbucket.org/genomicepidemiology/resfinder_db.git
```

The database can be used with BLAST as-is.

LINKS

Main site:

https://www.genomicepidemiology.org/

http://www.genomicepidemiology.org/services/

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

Please note that the links have changed recently

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

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

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

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

Thank you for the attention!