Instructor note: In the exercise nine contig files of *Staphylococcus aureus* were used

# Exercise: Bacterial typing for use in public health

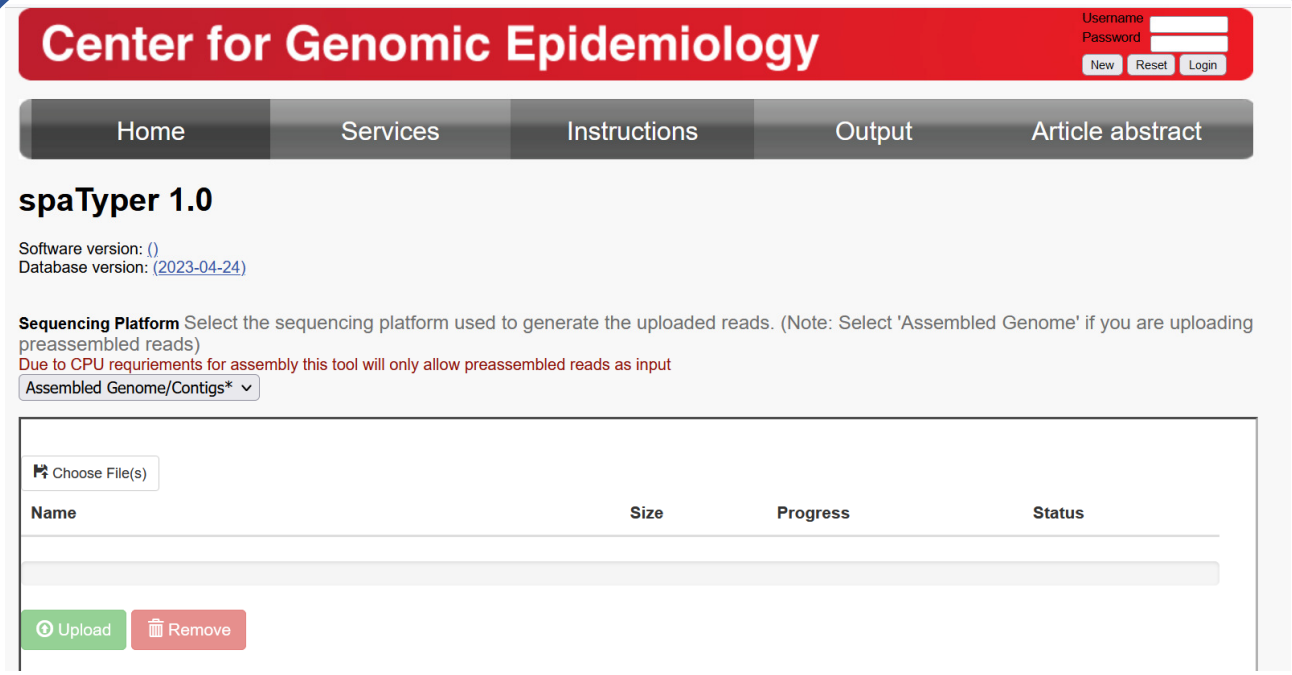
In this exercise you will be using different Finders based at the Center for Genomic Epidemiology. They can be used for many different bacterial species, but in this exercise, we will focus on nine MRSA strains. We will characterize them by three different typing methods that each supplement each other, and we will explore their content of virulence and resistance genes. The finders can be used without any bioinformatic knowledge; however, the interpretation of the results may require microbiological and/or epidemiological knowledge. In this document the internet based version is shown. For three of the finders, the corresponding unix command is also given.

## Exercise 1, *spa* typing

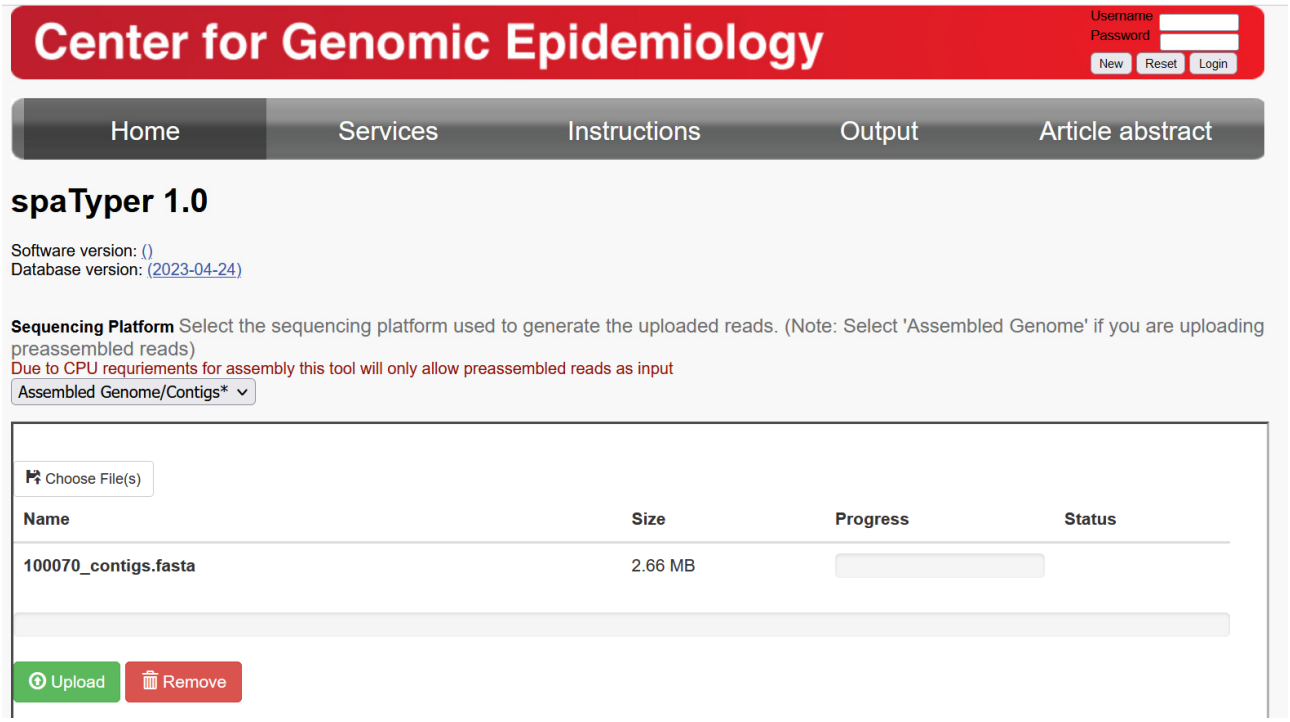
*spa* typing is the gold standard for typing of *Staphylococcus aureus*, including MRSA. It is based on sequence variation in a single gene, *spa*A, staphylococcus protein A. Due to the high variability of a repetitive domain of the gene, a high resolution can be obtained. However, the repetitive nature of the gene sometimes make an assembly from short read WGS difficult.

In the exercise, upload the assembled gene for each strain and register the result. In the result table, a *spa* type is given, based on traditional Sanger sequencing. If you note any discrepancies, try to figure out why.

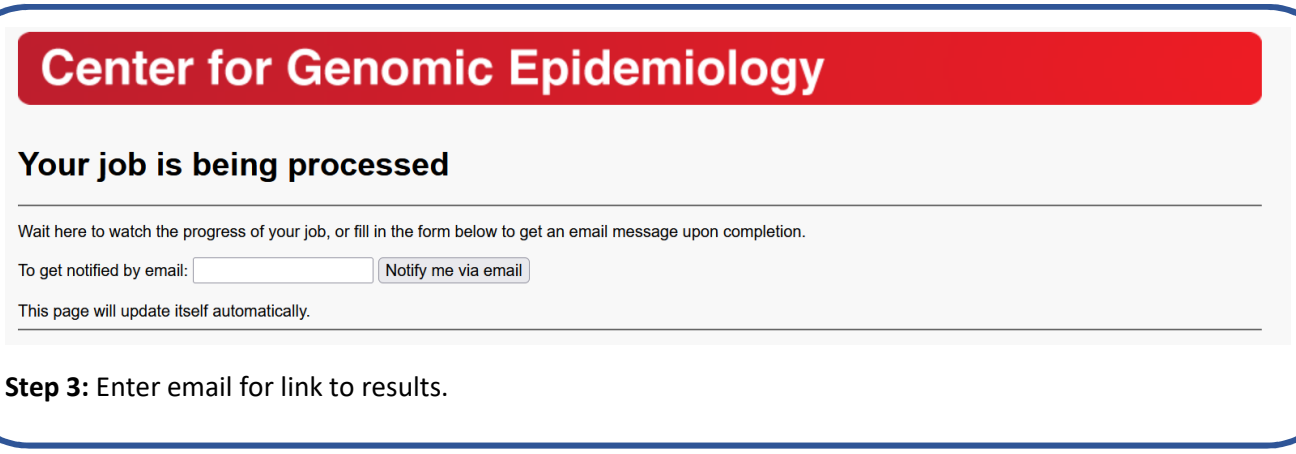
The service can be found at: https://cge.food.dtu.dk/services/spaTyper/



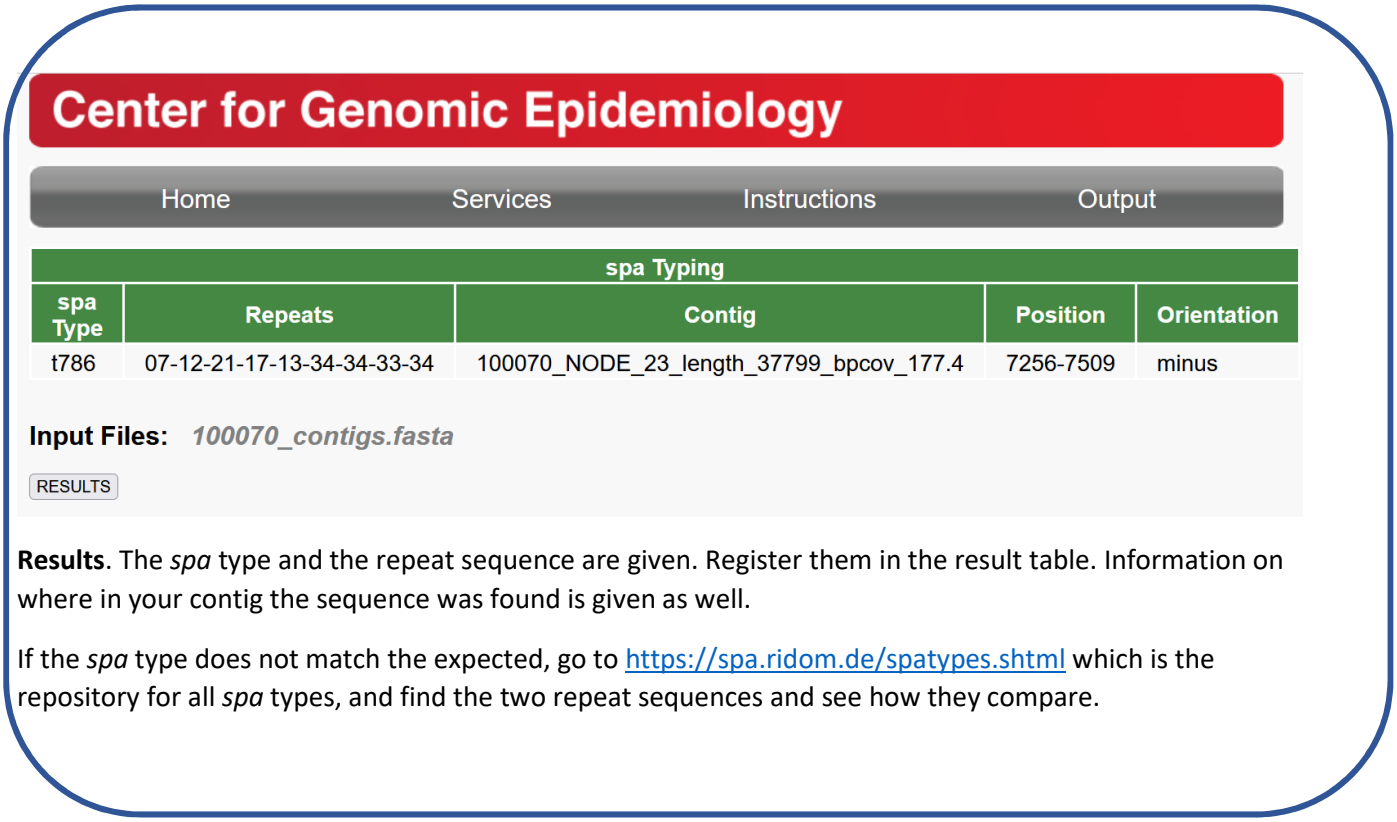

**Step 1:** Choose a file, only assembled contigs are allowed as input for this tool.




**Step 2:** Press upload.




**Step 3:** Enter email for link to results.



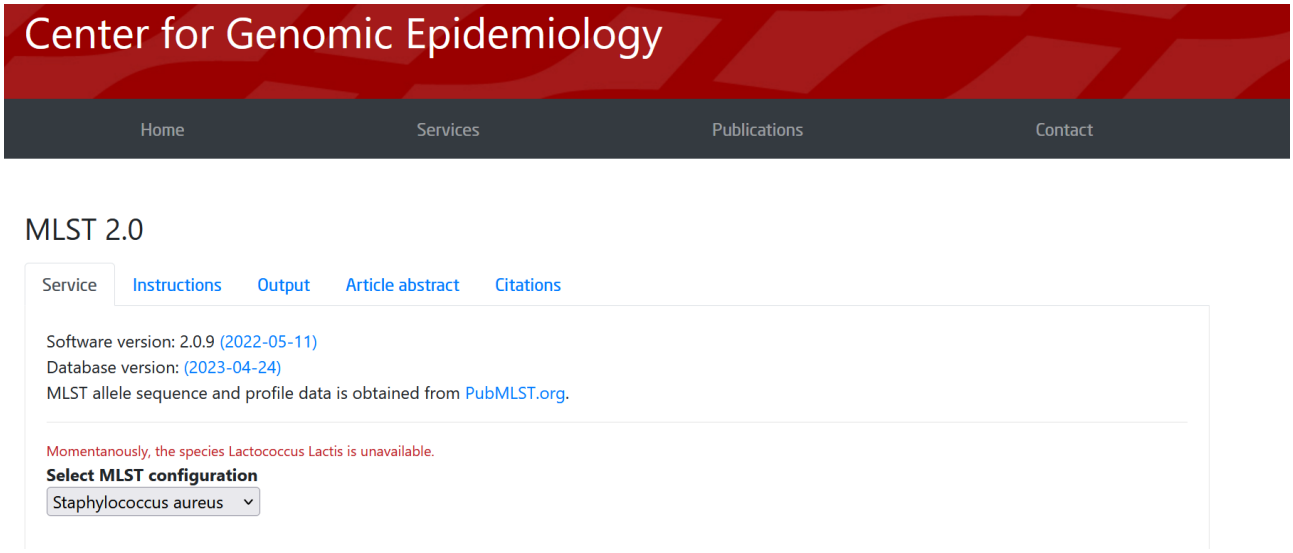

**Results**. The *spa* type and the repeat sequence are given. Register them in the result table. Information on where in your contig the sequence was found is given as well.

If the *spa* type does not match the expected, go to https://spa.ridom.de/spatypes.shtml which is the repository for all *spa* types, and find the two repeat sequences and see how they compare.

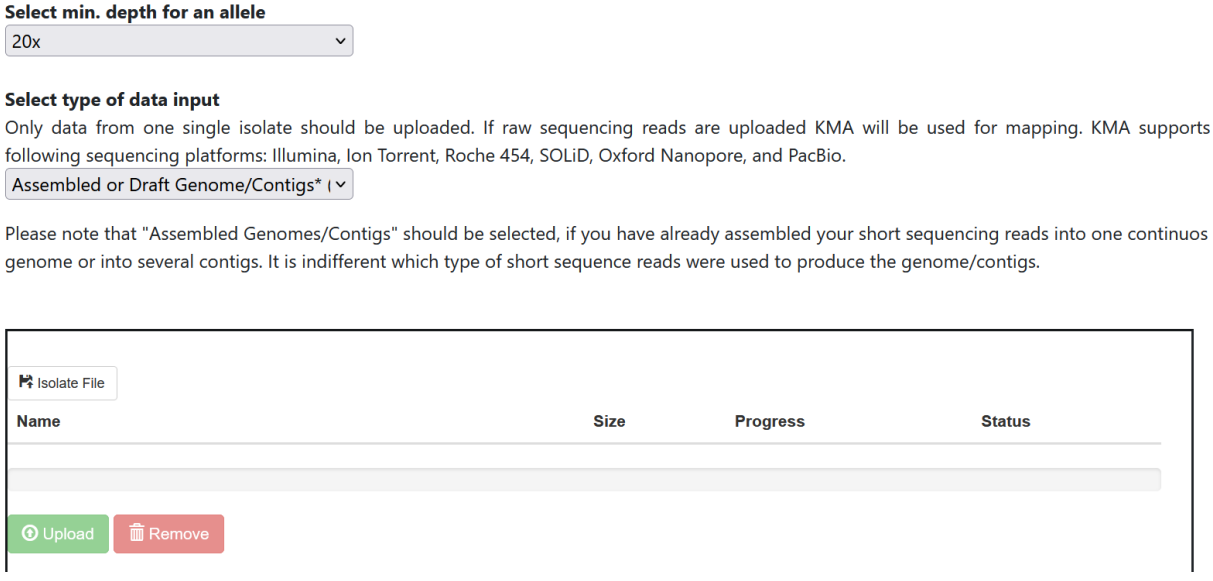
## Exercise 2: MLST typing

Multi locus sequence typing, MLST, is based on the sequence of seven house keeping genes. The sequences can be extracted from the assembled genomes. The typing assists in grouping related *spa* and ST types in larger clonal complexes.

The service can be found at: https://cge.food.dtu.dk/services/MLST/



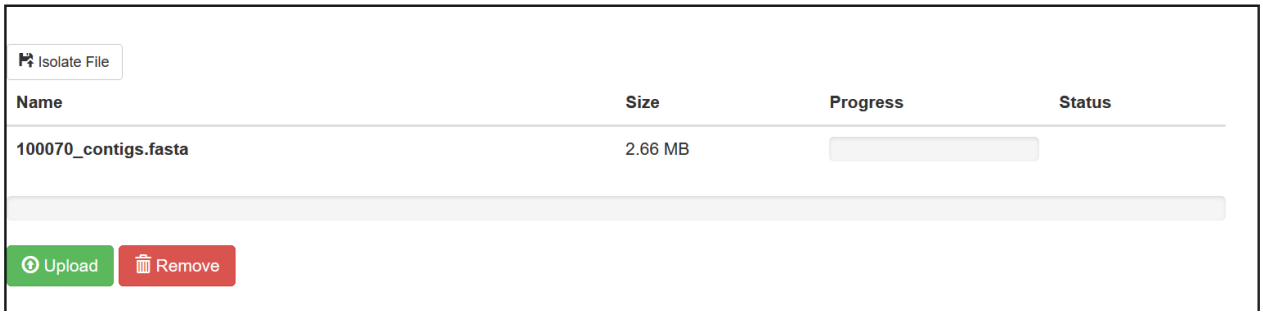

**Step 1:** Choose MLST configuration/species (*Staphylococcus aureus*) from the dropdown menu



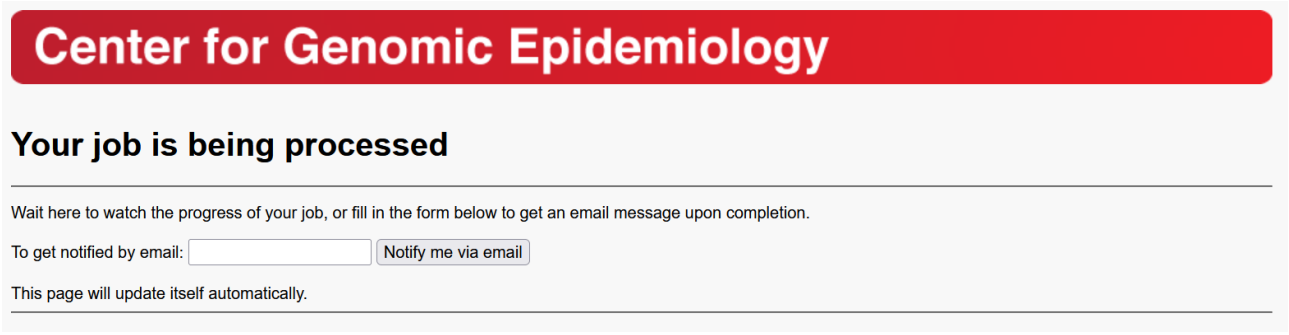

**Step 2a:** Select min. depth for an allele from the dropdown menu. Use 20x in this exercise.

**Step 2b:** Select type of data input. Use assembled or draft genome/contigs in this exercise.

**Step 2c:** Choose a file.




**Step 3:** Press upload.




**Step 4:** Enter email for link to results.

Unix command:

Use following command to explore all options:
python mlst.py -h

Example on running MLST on an *S. aureus* isolate (fill the red text with our own paths):

python [/path/to/mlst.py] -i [path/to/input_file] -o [outdir] -s saureus -mp [blastn/kma]

OSB! Choose between BLASTn or *kma* based on input file format. If readfiles then use BLASTn, if assemblies/draft genomes then use *kma*

Center for Genomic Epidemiology

Home | Services | Instructions | Output

## MLST-2.0 Server - Results

**mlst Profile:** *saureus*

**Organism:** *Staphylococcus aureus*

**Sequence Type:** *88*

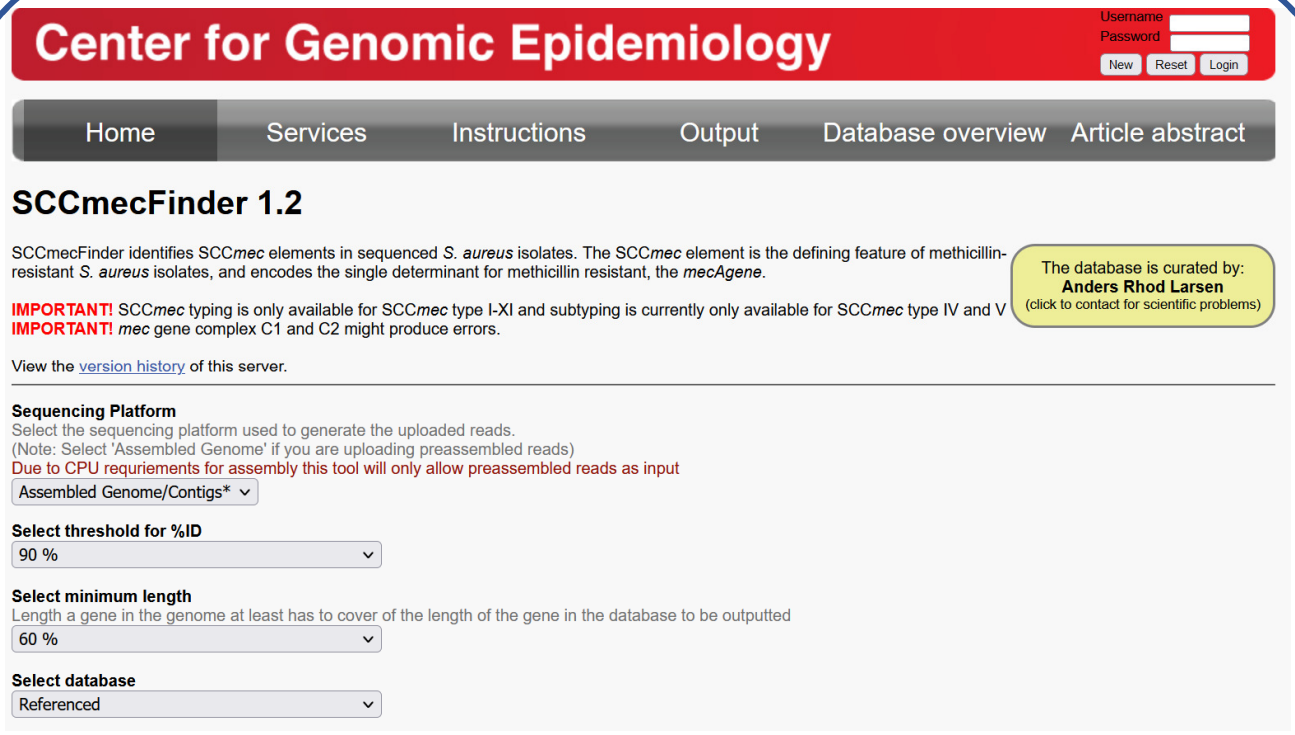
| Locus | Identity | Coverage | Alignment Length | Allele Length | Gaps | Allele |
|---|---|---|---|---|---|---|
| arcC | 100 | 100 | 456 | 456 | 0 | arcC_22 |
| aroE | 100 | 100 | 456 | 456 | 0 | aroE_1 |
| glpF | 100 | 100 | 465 | 465 | 0 | glpF_14 |
| gmk | 100 | 100 | 417 | 417 | 0 | gmk_23 |
| pta | 100 | 100 | 474 | 474 | 0 | pta_12 |
| tpi | 100 | 100 | 402 | 402 | 0 | tpi_4 |
| yqiL | 100 | 100 | 516 | 516 | 0 | yqiL_31 |

**Results**. The sequence type is given, as well as each of the seven alleles numbers. Register them in the result table.

## Exercise 3: SCC*mec* typing

The mobile element SCCmec (Staphylococcal Cassette Chromosome mec) is a genomic island that encodes methicillin resistance. The element has varied considerably during the specific waves of MRSA. For epidemiological and research purposes it can be of interest to determine which type is present.

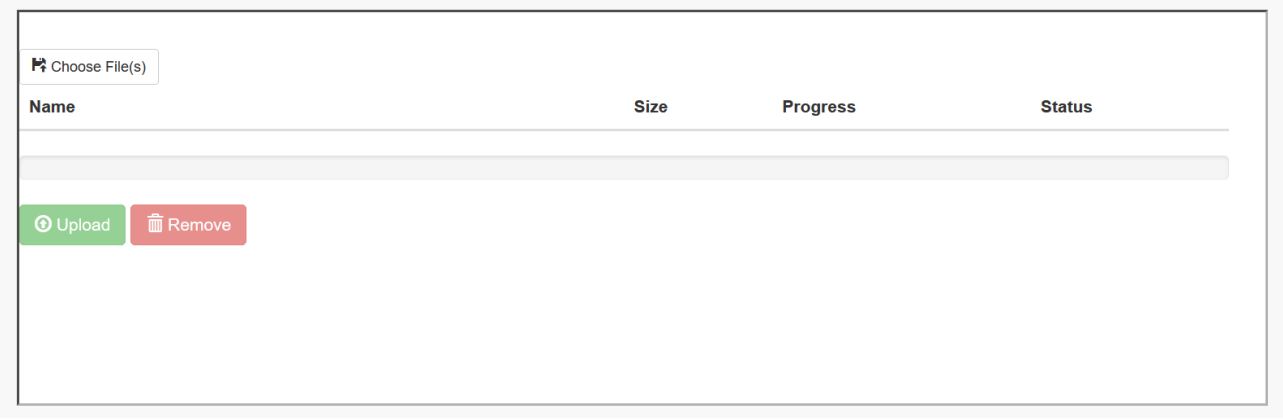
The service can be found at: https://cge.food.dtu.dk/services/SCCmecFinder/



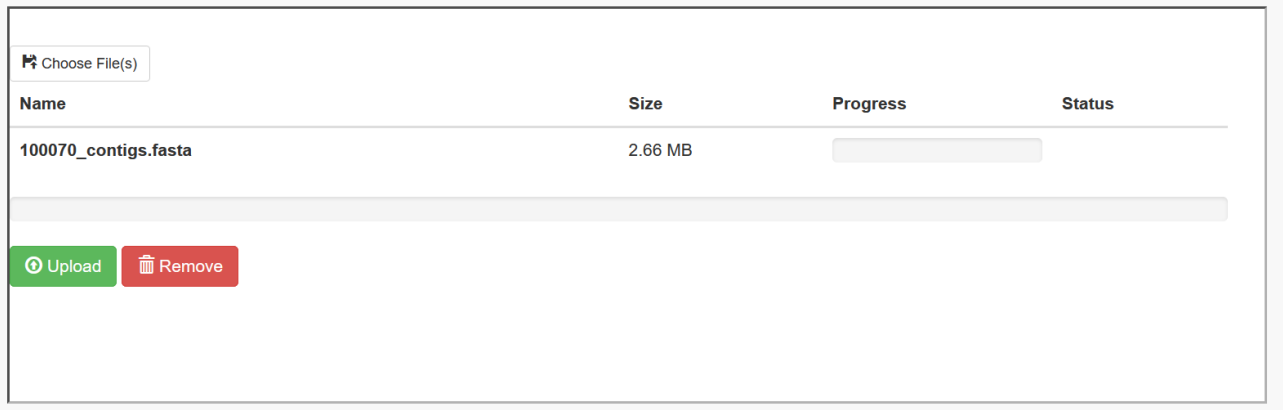

**Step 1a:** Select threshold for %ID. Use default of 90 % in this exercise.

**Step 1b:** Select minimum length. Use default of 60 % in this exercise.

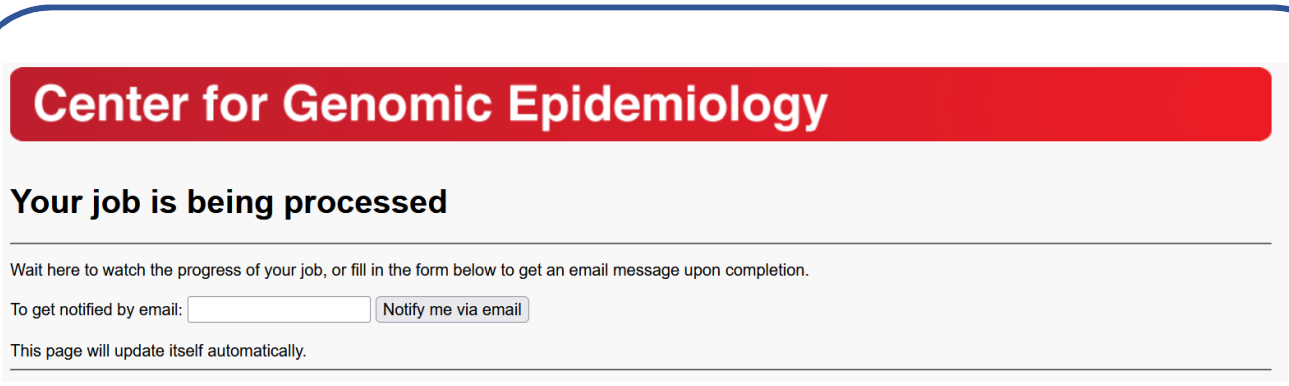
**Step 1c:** Select database. Use default of Referenced in this exercise.




**Step 2:** Choose a file, only assembled contigs are allowed as input for this tool.




**Step 3:** Press upload.




**Step 4:** Enter email for link to results.

# Center for Genomic Epidemiology

Home | Services | Instructions | Output | Database overview

**The input organism was prediced as a MRSA isolate**

**The mecA gene was detected**

One SCCmec element detected.

**Prediction based on genes:**

Predicted SCCmec element: SCCmec_type_IVa(2B)

**Prediction based on homology to whole cassette:**

Predicted whole cassette and %template coverage: SCCmec_type_IVa(2B) 82.80%

**Predicted genes:**

| Fasta header | % Identity | Query/HSP Length | Contig | Position in contig |
|---|---|---|---|---|
| mecA:5:CP000046 | 99.95 | 2007/2007 | 100070_NODE_5_length_180116_bpcov_164.5 | 139006..141012 |
| dmecR1:1:AB033763 | 100.00 | 987/987 | 100070_NODE_5_length_180116_bpcov_164.5 | 141112..142098 |
| IS1272:3:AM292304 | 100.00 | 1843/1843 | 100070_NODE_5_length_180116_bpcov_164.5 | 142087..143929 |
| ccrB2:9:JCSC4469:AB097677 | 99.94 | 1650/1650 | 100070_NODE_5_length_180116_bpcov_164.5 | 145771..147420 |
| ccrA2:7:81108:AB096217 | 99.93 | 1350/1350 | 100070_NODE_5_length_180116_bpcov_164.5 | 147421..148770 |
| subtype-IVa(2B):1:CA05:AB063172 | 100.00 | 1491/1491 | 100070_NODE_5_length_180116_bpcov_164.5 | 152557..154047 |

**Predicted whole SCCmec elements:**

| SCCmec elements | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Template | Score | Expected | z | p_value | query coverage [%] | template coverage [%] | depth | Kmers in Template | Description |
| SCCmec_type_IV(2B)\|SCCmec_type_IVa(2B)\|gb\|AB063172.2 | 42231 | 15706 | 249.30 | 3.6e-25 | 0.77 | 82.80 | 0.87 | 51003 | |
| SCCmec_type_IV(2B)\|SCCmec_type_IVa(2B)\|gb\|BA000033.2 | 41435 | 14660 | 260.70 | 3.6e-25 | 0.76 | 87.04 | 0.91 | 47607 | |
| SCCmec_type_IV(2B)\|SCCmec_type_IVc(2B)\|gb\|AB096217.1 | 31549 | 18315 | 115.50 | 3.6e-25 | 0.58 | 53.05 | 0.55 | 59474 | |
| SCCmec_type_IV(2B)\|SCCmec_type_IVi(2B)\|gb\|AB425823.1 | 29845 | 14091 | 157.10 | 3.6e-25 | 0.54 | 65.22 | 0.70 | 45760 | |
| SCCmec_type_IV(2B)\|SCCmec_type_IVc(2B)\|gb\|AY271717.1 | 28987 | 14762 | 138.60 | 3.6e-25 | 0.53 | 60.47 | 0.66 | 47939 | |

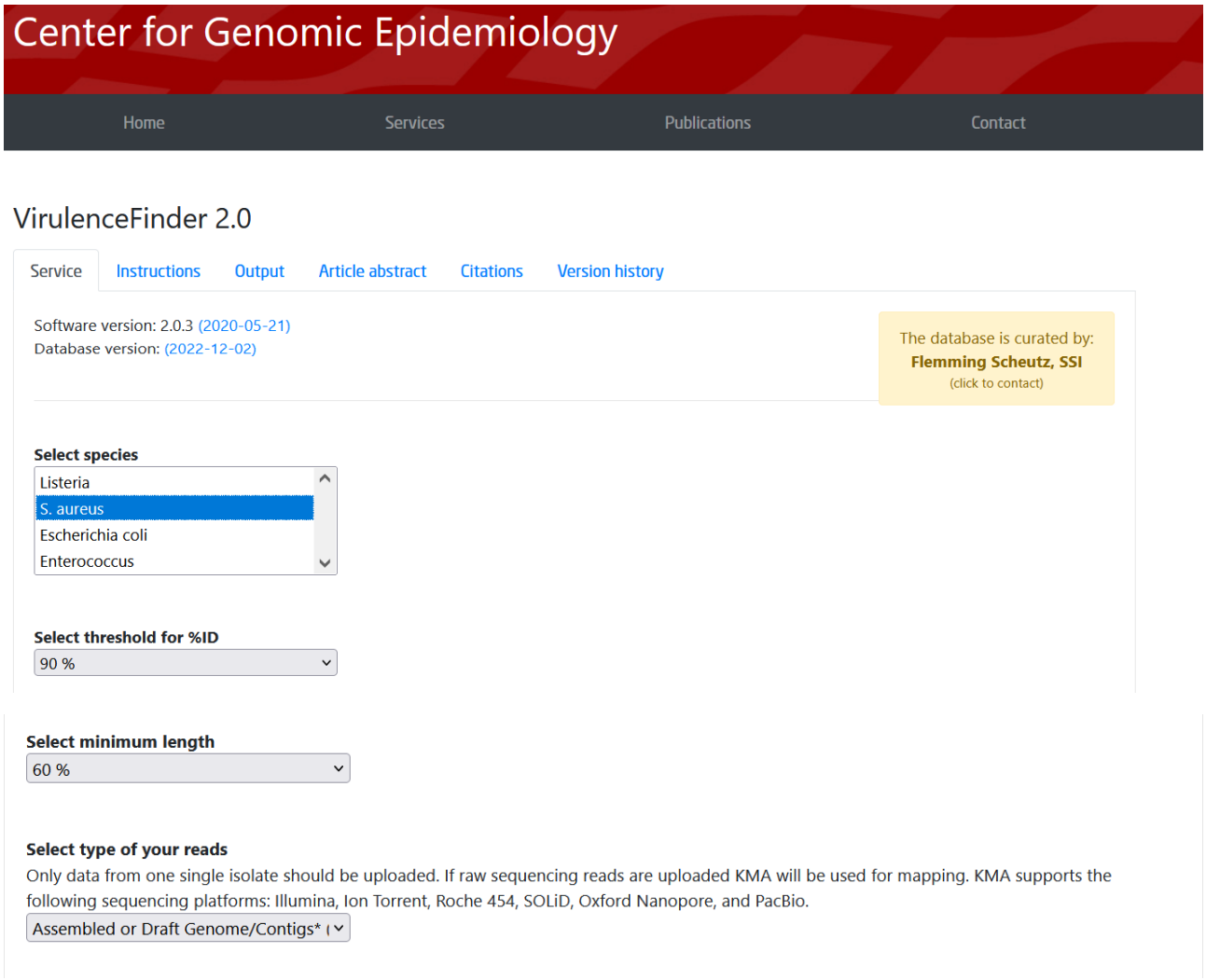
**Results.** Two predictions of the SCC*mec* element is given; one is based on the gene content (BLAST-based approach), the other on homology to the whole cassette (*k*-mer-based approach). The type of SCC*mec* element has to be read from both approaches, as the approaches can give contradicting results or one approach might give an inconclusively typing.

Register the predictions in the result table.

## Exercise 4: VirulenceFinder

Many virulence genes have been described in *Staphylococcus aureus*. The clinical relevance is not always straight-forward. In this exercise we will explore the VirulenceFinder.

The service can be found at: https://cge.food.dtu.dk/services/VirulenceFinder/

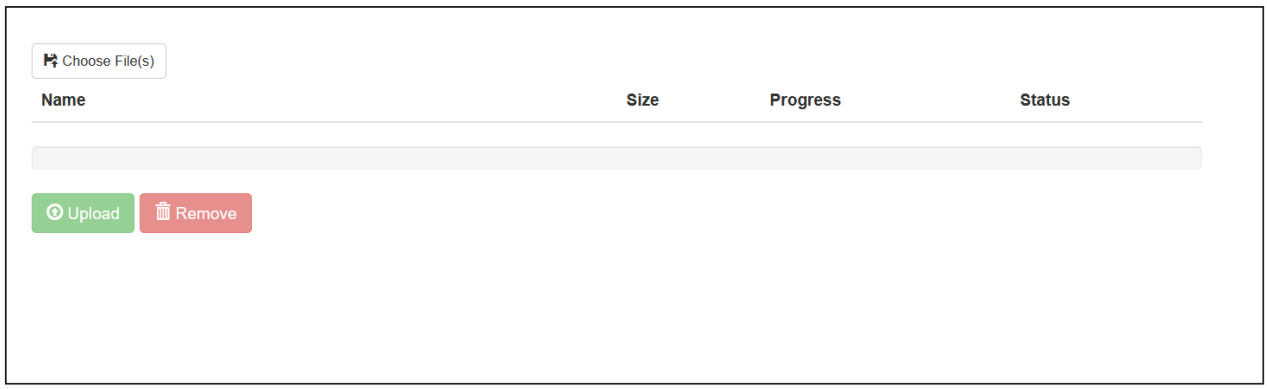



**Step 1a:** Select species.

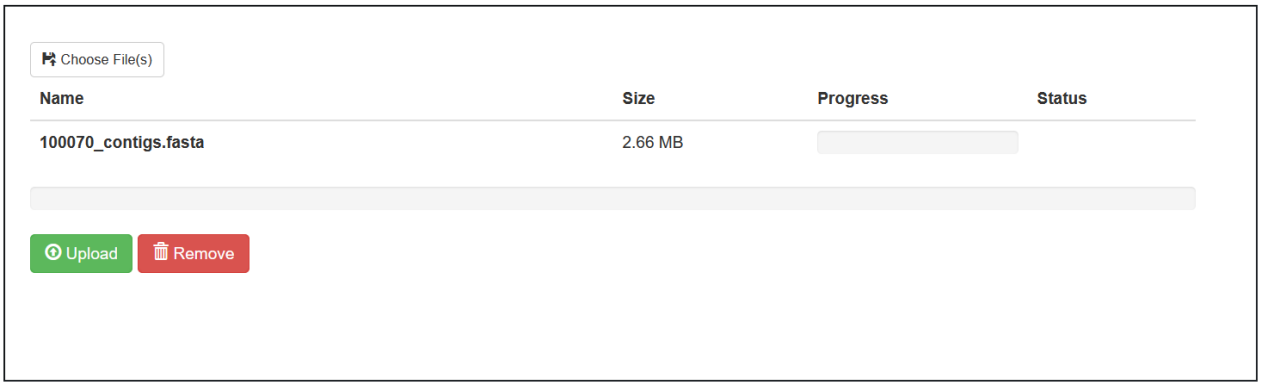
**Step 1b:** Select threshold for %ID (default value will do).

**Step 1c:** Select minimum length (default value will do).

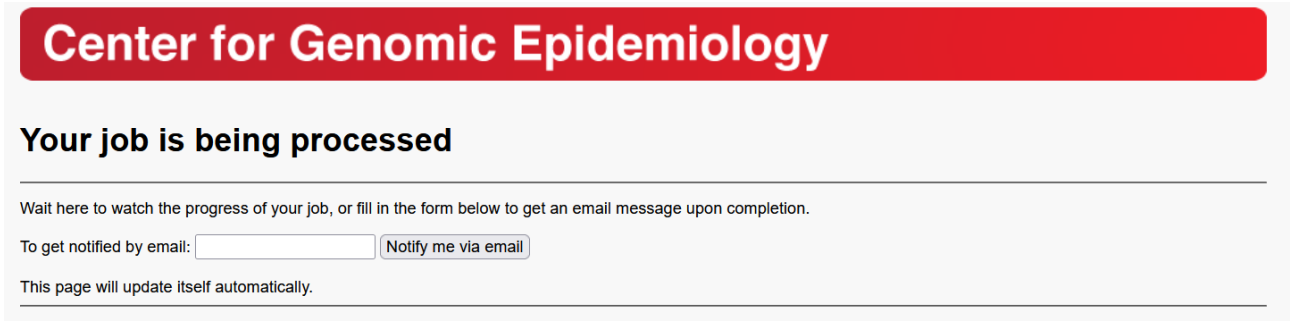
**Step 1d:** Select type of reads. Use Assembled or Draft Genome/Contigs for this exercise.




**Step 2:** Choose file.




**Step 3:** Press upload.




**Step 4:** Enter email for link to results.

Unix command:

Use following command to explore all options:
python virulencefinder.py -h

Example on running VirulenceFinder on S.aureus exoenzyme DB (fill the red text with our own paths):

python [/path/to/virulencefinder.py] -i [path/to/input_file] -p [path/to/virulencefinder_db] [-d] s.aureus_exoenzyme

## VirulenceFinder-2.0 Server - Results

**Organism(s):** *S. aureus*

| Hostimm genes for S. aureus | | | | | | |
|---|---|---|---|---|---|---|
| Virulence factor | Identity | Query / Template length | Contig | Position in contig | Protein function | Accession number |
| sak | 99.8 | 492 / 492 | 100070_NODE_11_length_103052_bpcov_120.1 | 85501..85992 | staphylokinase | CP000253.1 |
| sak | 99.8 | 492 / 492 | 100070_NODE_11_length_103052_bpcov_120.1 | 85501..85992 | staphylokinase | CP003979.1 |
| sak | 99.8 | 492 / 492 | 100070_NODE_11_length_103052_bpcov_120.1 | 85501..85992 | staphylokinase | HE681097.1 |
| scn | 100 | 351 / 351 | 100070_NODE_11_length_103052_bpcov_120.1 | 88206..88556 | staphylococcal complement inhibitor | AP009351.1 |

| Exoenzyme genes for S. aureus | | | | | | |
|---|---|---|---|---|---|---|
| Virulence factor | Identity | Query / Template length | Contig | Position in contig | Protein function | Accession number |
| aur | 99.93 | 1530 / 1530 | 100070_NODE_5_length_180116_bpcov_164.5 | 9186..10715 | aureolysin | BA000033.2 |
| splA | 98.02 | 708 / 708 | 100070_NODE_28_length_14173_bpcov_72.3 | 6296..7003 | serine protease splA | AP014653.1 |
| splB | 99.59 | 723 / 723 | 100070_NODE_28_length_14173_bpcov_72.3 | 5449..6171 | serine protease splB | AP014942.1 |

| Toxin genes for S. aureus | | | | | | |
|---|---|---|---|---|---|---|
| Virulence factor | Identity | Query / Template length | Contig | Position in contig | Protein function | Accession number |
| hlgA | 100 | 930 / 930 | 100070_NODE_7_length_137991_bpcov_134.8 | 16239..17168 | gamma-hemolysin chain II precursor | CP001781.1 |
| hlgB | 99.8 | 977 / 977 | 100070_NODE_7_length_137991_bpcov_134.8 | 18684..19660 | gamma-hemolysin component B precursor | BA000018.3 |
| hlgB | 99.8 | 977 / 977 | 100070_NODE_7_length_137991_bpcov_134.8 | 18684..19660 | gamma-hemolysin component B precursor | BA000033.2 |
| hlgC | 99.89 | 948 / 948 | 100070_NODE_7_length_137991_bpcov_134.8 | 17735..18682 | gamma-hemolysin component C | BA000018.3 |
| lukD | 99.8 | 984 / 984 | 100070_NODE_28_length_14173_bpcov_72.3 | 11119..12102 | leukocidin D component | AP014653.1 |
| lukE | 99.79 | 936 / 936 | 100070_NODE_28_length_14173_bpcov_72.3 | 12104..13039 | leukocidin E component | AP014942.1 |

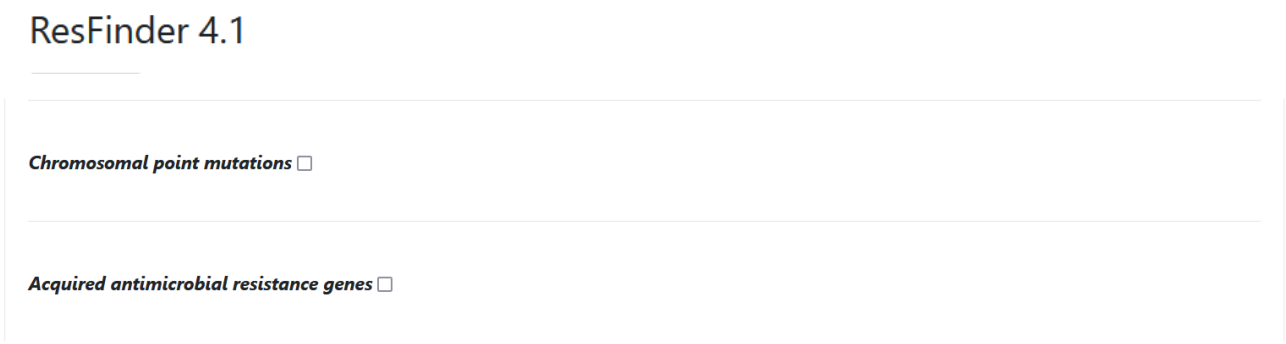
extended output

**Results**. The Finder will look for many virulence genes and report those detected. In this exercise we want to know if the isolates contain the host immune evasion gene *scn*, and the genes encoding the toxin PVL (*lukF*/*lukS*). Register +/- in the result table.

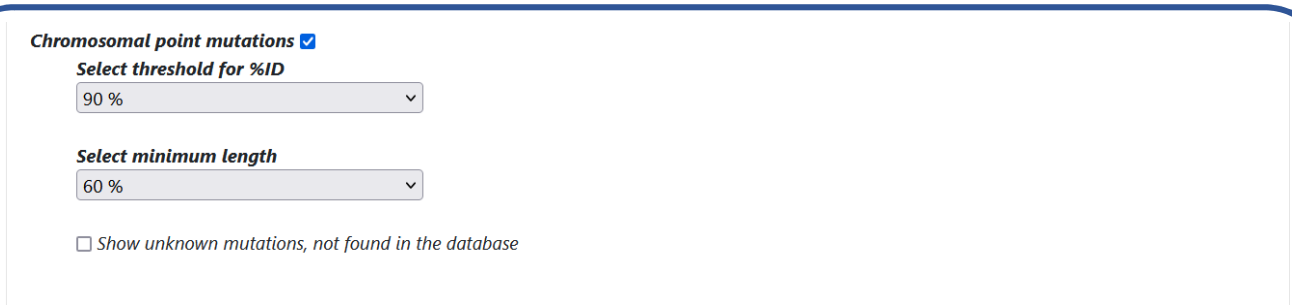
## Exercise 5: ResFinder

In the microbiological lab, the phenotypic resistance is determined by different methods. From WGS data we can extract resistance genes to find the genotypic marker for detected phenotypic resistances.

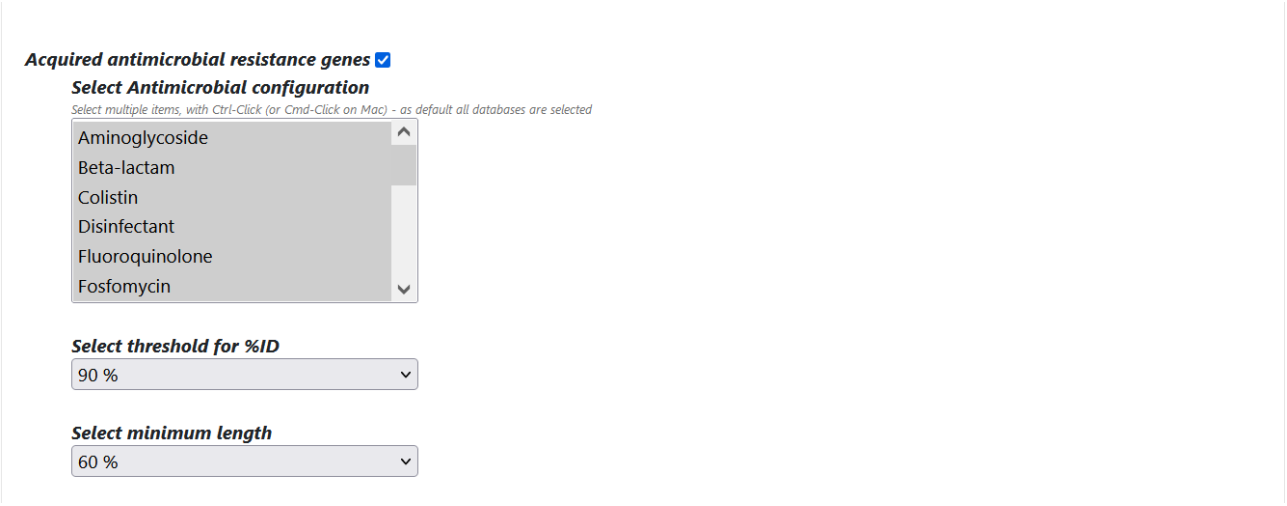
Find the service at: https://cge.food.dtu.dk/services/ResFinder/




**Step 1:** Select types of resistance mechanisms (chromosomal point mutations and/or acquired antimicrobial resistance genes). Tick both boxes in this exercise.

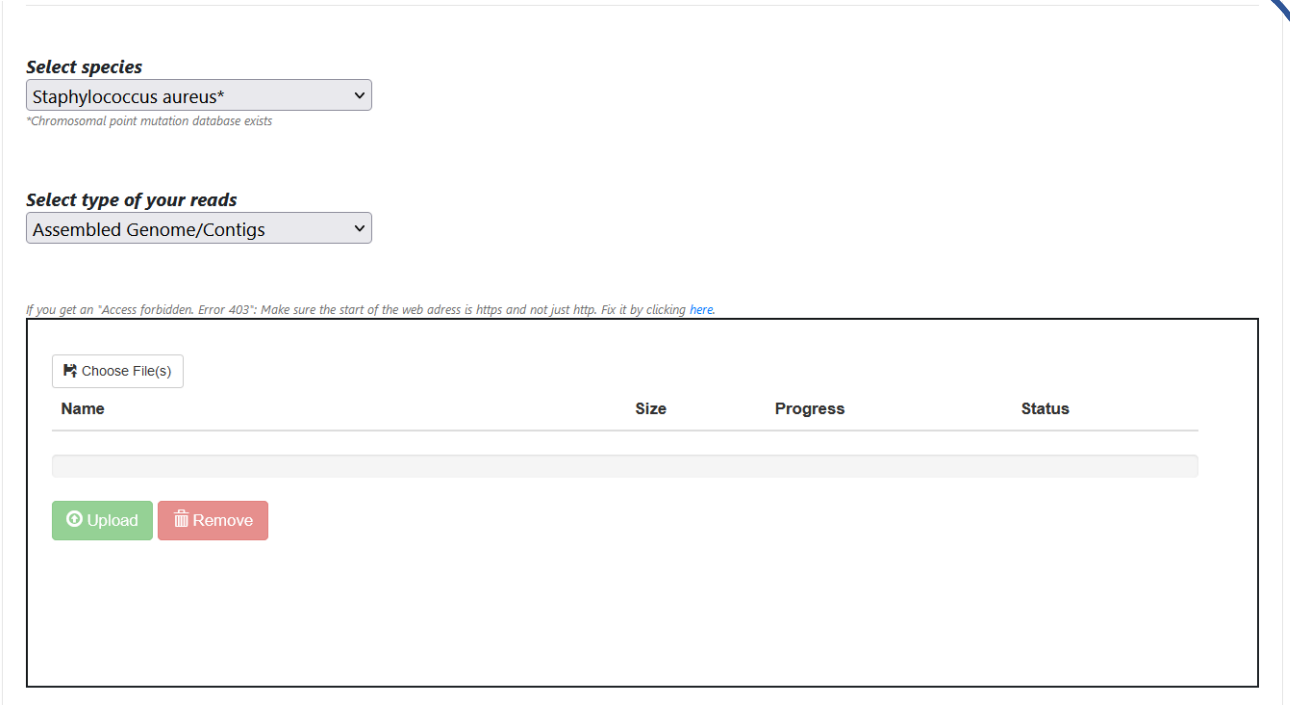



**Step 2:** Select threshold and minimum length for chromosomal point mutations. Use default values in this exercise.




**Step 3a:** It is possible to choose one or more antimicrobial classes to be included in the search. Select all in this exercise.

**Step 3b:** Select threshold and minimum length for acquired antimicrobial resistance genes. Use default values in this exercise.

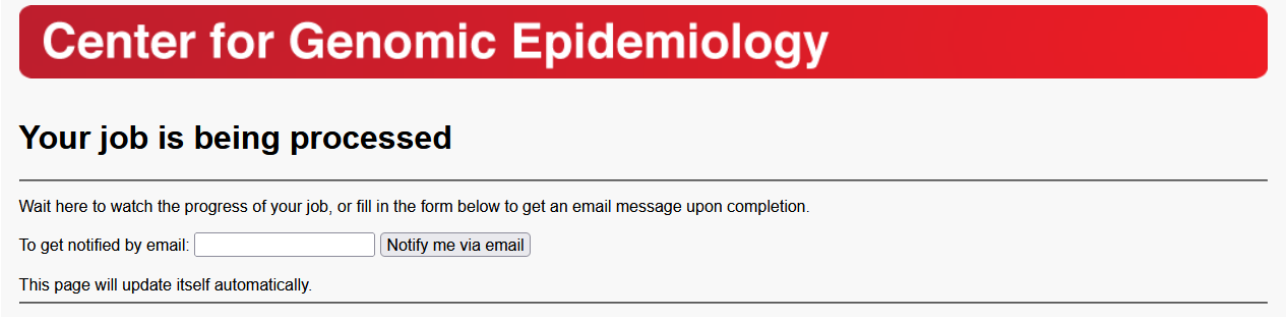



**Step 4a:** Select species. Use *Staphylococcus aureus* in this exercise.

**Step 4b:** Select type of reads. Use Assembled Genome/Contigs in this exercise.

**Step 4c:** Choose file.


**Step 5:** Press upload.




**Step 6:** Enter email to get link to results.

Unix command:

Example on running ResFinder on S. aureus (fill the red text with our own paths):

python -m resfinder -o [/path/to/outdir] -i “Staphylococcus aureus” -l 0.6 -t 0.8 –acquired –point -ifq [path/to/readsfiles.*]

| Beta-lactam | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Resistance gene | Identity | Alignment Length/Gene Length | Position in reference | Contig or Depth | Position in contig | Phenotype | PMID | Accession no. | Notes |
| mecA | 99.9501743896 | 2007/2007 | 1..2007 | 100070_NODE_5_length_180116_bpcov_164.5 | 139006..141012 | amoxicillin,amoxicillin+clavulanic acid,ampicillin,ampicillin+clavulanic acid,cefepime,cefixime,cefotaxime,cefoxitin,ceftazidime,ertapenem,imipenem,meropenem,piperacillin,piperacillin+tazobactam | 15774886 | NC_002951 | |
| blaZ | 100.0 | 846/846 | 1..846 | 100070_NODE_24_length_25510_bpcov_459.2 | 12363..13208 | amoxicillin,ampicillin,penicillin,piperacillin | 12044378 | AP004832 | Class A |

| Detection PointFinder Genes | |
|---|---|
| dfrB | Gene found without known mutations |
| pbp2 | Gene found without known mutations |
| gyrA | Gene found without known mutations |
| ileS | Gene found without known mutations |
| grlB | Gene found without known mutations |
| grlA | Gene found without known mutations |
| rpoB | Gene found without mutations |
| pbp4 | Gene found without known mutations |
| fusA | Gene found without mutations |
| 23S | Gene found with coverage, 0.016079, below minimum coverage threshold: 0.6 |

Extract of the results. Many genotypic markers are investigated. In this exercise, look for the ones given in the result table and mark either + for detected, or – for not detected. If the genotype is a mutation, register the predicted amino acid changes

Appendix 1 Result table

| **Isolate** | ***spa* type** | ***spa* type WGS** | ***spa* repeats** | **MLST** | ***arcC*** | ***aroE*** | ***glpF*** | ***gmk*** | ***pta*** | ***tpi*** | ***yqiL*** |
|---|---|---|---|---|---|---|---|---|---|---|---|
| MRSA1 | t008 | | | | | | | | | | |
| MRSA2 | t019 | | | | | | | | | | |
| MRSA3 | t044 | | | | | | | | | | |
| MRSA4 | t2872 | | | | | | | | | | |
| MRSA5 | t223 | | | | | | | | | | |
| MRSA6 | t034 | | | | | | | | | | |
| MRSA7 | t843 | | | | | | | | | | |
| MRSA8 | t041 | | | | | | | | | | |
| MRSA9 | t030 | | | | | | | | | | |

| **Isolate** | **SCC*mec* gene (BLAST)** | **SCC*mec* cassette (k-mer)** | ***scn*** | **PVL** | ***mecA*** | ***mecC*** |
|---|---|---|---|---|---|---|
| MRSA1 | | | | | | |
| MRSA2 | | | | | | |
| MRSA3 | | | | | | |
| MRSA4 | | | | | | |
| MRSA5 | | | | | | |
| MRSA6 | | | | | | |
| MRSA7 | | | | | | |
| MRSA8 | | | | | | |
| MRSA9 | | | | | | |

| Isolate | Aph(3´)-III | aac(6´)-aph(2´´) | ant(9)-1a | blaZ | mph(C) | tet(K) | tet(M) | msr(A) | fusA | fusB | ermA | ermC | dfrG | lnuB | lsa(E) | gyrA* | grlA* | ileS* | rpoB* |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| MRSA1 | | | | | | | | | | | | | | | | | | | |
| MRSA2 | | | | | | | | | | | | | | | | | | | |
| MRSA3 | | | | | | | | | | | | | | | | | | | |
| MRSA4 | | | | | | | | | | | | | | | | | | | |
| MRSA5 | | | | | | | | | | | | | | | | | | | |
| MRSA6 | | | | | | | | | | | | | | | | | | | |
| MRSA7 | | | | | | | | | | | | | | | | | | | |
| MRSA8 | | | | | | | | | | | | | | | | | | | |
| MRSA9 | | | | | | | | | | | | | | | | | | | |

*register the predicted amino acid substitutions, eg. S84L

STATENS SERUM INSTITUT

The exercise was developed by Hülya Kaya and Andreas Petersen, National Reference Laboratory for Antimicrobial Resistance, Statens Serum Institut.