Introduction to web-based bioinformatics tools - Resfinder, Virulence Finder using CGE

Dr Mabel Aworh-Ajumobi

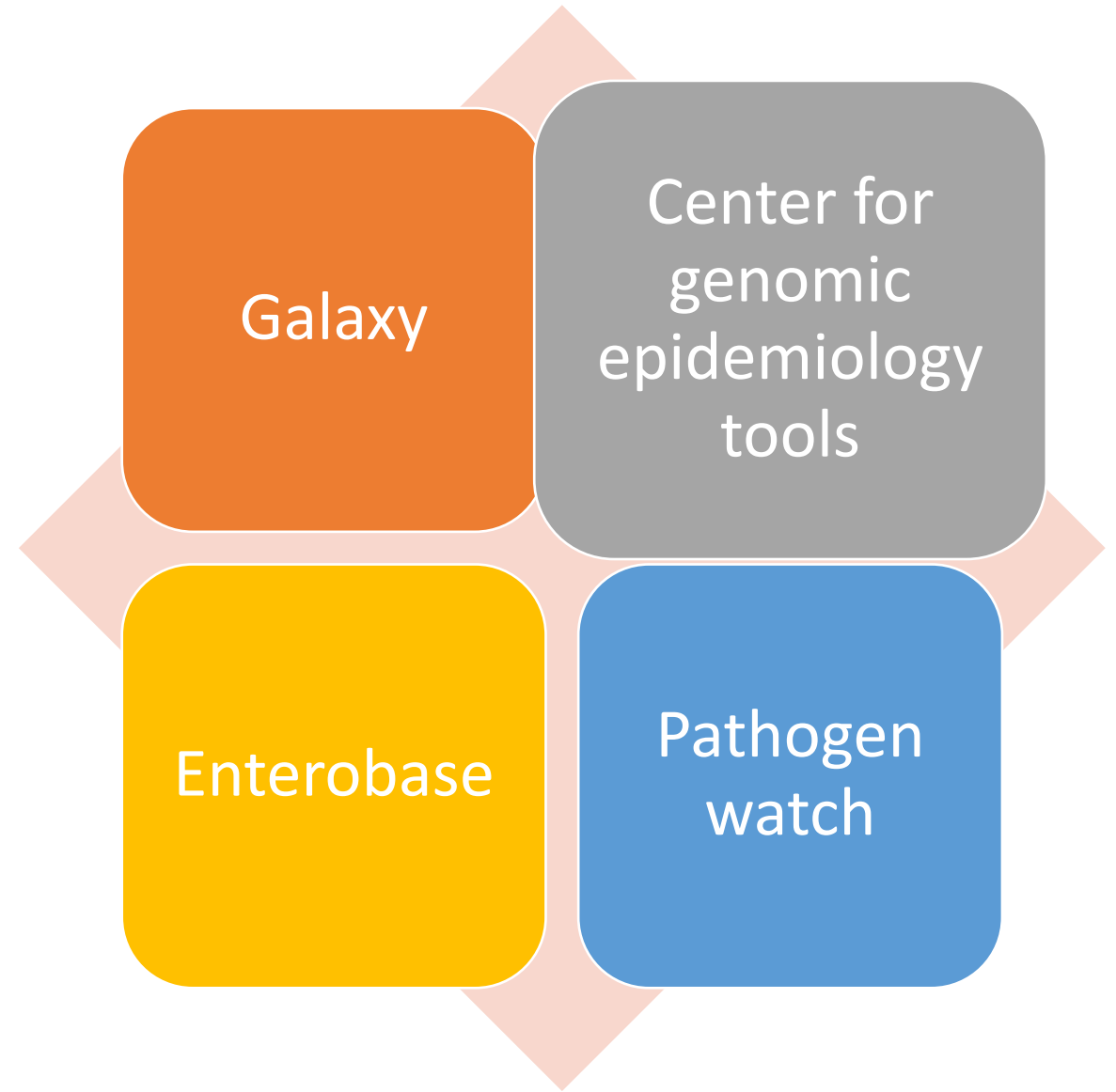**North Carolina State University, Raleigh**

Learning objectives:

- To learn how to access CGE Tools

- To use ResFinder in determination of AMR genes

- To use Virulence Finder in identification of virulence genes in bacteria species

- To retrieve and interpret outputs from ResFinder and Virulence Finder

Bioinformatics web-based tools

- Variety of methods and tools are available to analyze bacterial pathogens
- Most bioinformatics tools are implemented in Unix environments
- Require at least some bioinformatics expertise for usage
- Web-based bioinformatics tools
  - Often free for use
  - Do not require computational power from the user
  - Limited bioinformatics knowledge
  - In some cases demand that that users deposit the analyzed data in public repositories
  - Always make an effort to browse through the documentation of web based platforms
    - Helps with choice of parameters and interpretation of results

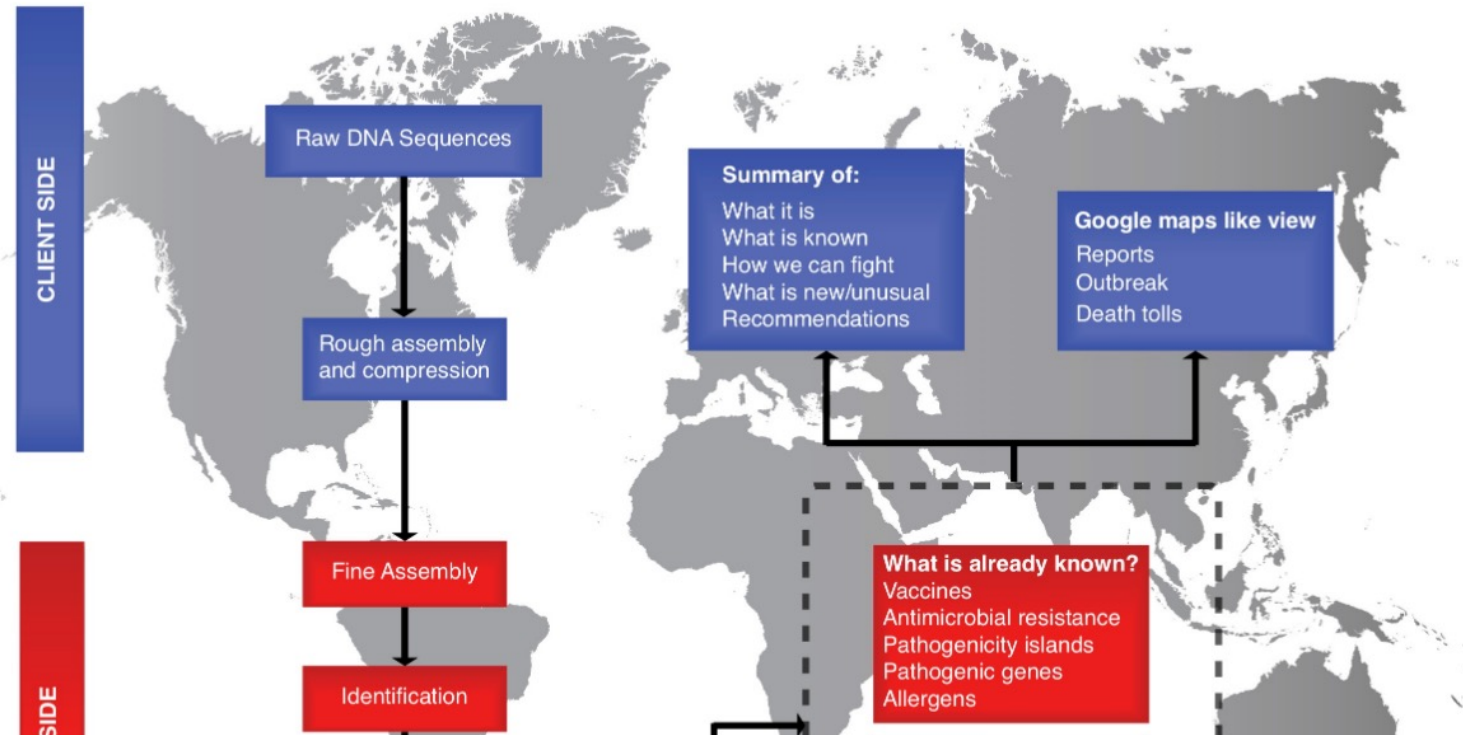# Examples of free online web-based tools

Galaxy

Center for genomic epidemiology tools

Enterobase

Pathogen watch

Accessing the CGE homepage

Overview of CGE Tools

Phenotyping

ResFinder
Identifcation of acquired antibiotic resistance genes.

ResFinderFG
Identifcation of functional metagenomic antibiotic resistance determinants.

LRE-finder
Identifcation of genes and mutations leading to linezolid resistance.

KmerResistance
Identifcation of acquired antibiotic resistance genes using Kmers.

PathogenFinder
Prediction of a bacteria's pathogenicity towards human hosts.

VirulenceFinder
Identifcation of acquired virulence genes.

Restriction-ModificationFinder
Determination of Restriction-Modification sites (based on REBASE.)

SPIFinder
SPIFinder identifies Salmonella Pathogenicity Islands.

ToxFinder
ToxFinder identifies genes involved in mycotoxin synthesis.

Phylogeny

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

CSIPhylogeny
CSI Phylogeny calls SNPs, filters the SNPs, does site validation and infers a phylogeny based on the concatenated alignment of the high quality* SNPs.

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

Evergreen
Evergreen generates a forest of constantly updated phylogenetic trees with publicly available whole-genome sequencing data from foodborne, bacterial isolates that were deposited in the short sequencing read archives (NCBI SRA/ENA).

TreeViewer
Phylogeny Tree Viewer.

Metagenomics

CCMetagen
CCMetagen: Comprehensive and accurate identification of eukaryotes and prokaryotes in metagenomic data.

ResFinder – for detecting resistance genes

- Is based on curated database, public databases as well as scientific papers
- Is web-based friendly interface and freely accessible tool
- Will detect the presence of whole genome resistance genes
- The database contains genes for 14 major antimicrobial drug classes

We want to learn how to detect AMR genes in bacteria genome

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: (2020-10-21)
ResFinder database: (2021-01-20)
PointFinder database: (2019-07-02)

Chromosomal point mutations ☐

Acquired antimicrobial resistance genes ☐

Select species

| Escherichia coli* | ⌄ |

*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

⊕ Upload   🗑 Remove

8

To upload your sequence file

Ensure to check all the important boxes has shown by the green tick marks

Assessing your results

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: [maworh@gmail.com] [Notify me via email]

This page will update itself automatically.

cgewebface@cbs.dtu.dk
To: maworh@gmail.com

Fri, Feb 3 at 5:55 PM

Dear sir/madam,
    Your job 63DDB57D0000051B70E4CBE5 has finished on our ResFinder-4.1 server.
    You may visit the following link to study the results:
    https://cge.food.dtu.dk//cgi-bin/webface.fcgi?jobid=63DDB57D0000051B70E4CBE5

    The data will expire at 2023-02-11 02:31:41.

Yours sincerely,
    CGE webserver

You can wait for the page to update after analysis is completed or choose to get notified by email

ResFinder-4.1 Server - Results

Input Files: *isolate2.fasta*

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	Resistant	blaTEM-1B (blaTEM-1B_AY458016)
ampicillin+clavulanic acid	beta-lactam	No resistance	
cefotaxime	beta-lactam	No resistance	
ciprofloxacin	quinolone	Resistant	gyrA (p.S83L)
colistin	polymyxin	No resistance	
sulfamethoxazole	folate pathway antagonist	Resistant	sul2 (sul2_AY034138)
imipenem	beta-lactam	No resistance	
trimethoprim	folate pathway antagonist	Resistant	dfrA14 (dfrA14_KF921535)
nalidixic acid	quinolone	Resistant	gyrA (p.S83L)
ertapenem	beta-lactam	No resistance	
tetracycline	tetracycline	Resistant	tet(A) (tet(A)_AJ517790)
fosfomycin	fosfomycin	No resistance	

11

Aminoglycoside

Resistance gene	Identity	Alignment Length/Gene Length	Position in reference	Contig or Depth	Position in contig	Phenotype	PMID	Accession no.	Notes
aph(6)-Id	100.0	837/837	1..837	NODE_79_length_7882_cov_4.782979	3194..4030	streptomycin	2653965	M28829	Alternative name strB
aph(3")-Ib	100.0	804/804						AF321551	Alternative name strA

Alignment length is the length between the best matching resistance gene and corresponding sequence in the genome while **Gene length** is the length of the best matching resistance gene in the database

Resistance gene	Identity	Alignment Length/Gene Length	Position in reference	Contig or Depth	Position in contig	Phenotype	PMID	Accession no.	Notes
tet(A)	100.0	1200/1200	1..1200	NODE_79_length_7882_cov_4.782979	387..1586	doxycycline,tetracycline	12654659	AJ517790	

The dark green colour indicates a perfect match for a given gene. The % ID is 100 and means that the sequence in the genome covers the entire length of the resistance gene in the database

Beta-lact

Resistance gene	Identity	Alignment Length/Gene Length	Position in reference	Contig or Depth	contig	Phenotype	PMID	Accession no.	Notes
blaTEM-1B	100.0	861/861	1..861	NODE_102_length_2700_cov_6.467548	1323..2183	amoxicillin,ampicillin,cephalothin,piperacillin,ticarcillin	15388431	AY458016	Class A

Other ways to access your results

Download acquired AMR gene results:

| Results as text | Hit in genome sequences | Resistance gene sequences | Results as tabseperated file |

Download Chromosomal point mutation results:

| Results as tabseperated file | Results as a text file |

Selected %ID threshold for ResFinder: *90 %*

Selected minimum length for ResFinder: *60 %*

Selected %ID threshold for PointFinder: *90 %*

Selected minimum length for PointFinder: *60 %*

| extended output |

Click on extended output for more information

extended output

dfrA14, ID: 100.00 %, Alignment Length/Gene Length: 474/474, Coverage: 100.0, Positions in reference: 1..474, Contig name: NODE_104_length_2160_cov_

```
Resistance gene seq: TTGAAAGTATCATTGATGGCTGCGAAAGCGAAAAACGGCGTGATTGGTTGCGGTCCAGAC
Hit in genome:       TTGAAAGTATCATTGATGGCTGCGAAAGCGAAAAACGGCGTGATTGGTTGCGGTCCAGAC

Resistance gene seq: ATACCCTGGTCCGCGAAAGGGGAGCAGCTACTTTTTAAAGCATTGACCTACAATCAGTGG
Hit in genome:       ATACCCTGGTCCGCGAAAGGGGAGCAGCTACTTTTTAAAGCATTGACCTACAATCAGTGG

Resistance gene seq: CTTCTGGTGGGTCGCAAGACGTTTGAATCTATGGGCGCACTCCCCAATAGGAAATACGCG
Hit in genome:       CTTCTGGTGGGTCGCAAGACGTTTGAATCTATGGGCGCACTCCCCAATAGGAAATACGCG

Resistance gene seq: GTCGTTACCCGCTCAGGTTGGACATCAAATGATGACAATGTAGTTGTATTTCAGTCAATC
Hit in genome:       GTCGTTACCCGCTCAGGTTGGACATCAAATGATGACAATGTAGTTGTATTTCAGTCAATC

Resistance gene seq: GAAGAGGCCATGGACAGGCTAGCTGAATTCACCGGTCACGTTATAGTGTCTGGTGGCGGA
Hit in genome:       GAAGAGGCCATGGACAGGCTAGCTGAATTCACCGGTCACGTTATAGTGTCTGGTGGCGGA

Resistance gene seq: GAAATTTACCGAGAAACATTACCCATGGCCTCTACGCTCCACTTATCGACGATCGACATC
Hit in genome:       GAAATTTACCGAGAAACATTACCCATGGCCTCTACGCTCCACTTATCGACGATCGACATC

Resistance gene seq: GAGCCAGAGGGGGATGTTTTCTTCCCGAGTATTCCAAATACCTTCGAAGTTGTTTTTGAG
```

This shows you the alignment between your sequence in the genome and the resistance genes. Green colour shows a perfect match between the sequence in the genome and resistance genes in the database

VirulenceFinder 2.0

| Service | Instructions | Output | Article abstract | Citations | Version history |

Software version: 2.0.3 (2020-05-21)
Database version: (2022-12-02)

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

Select species
Listeria
S. aureus
Escherichia coli
Enterococcus

Select threshold for %ID
90 %

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 |

This tool is used to analyse bacterial sequence data to detect the presence of virulence genes
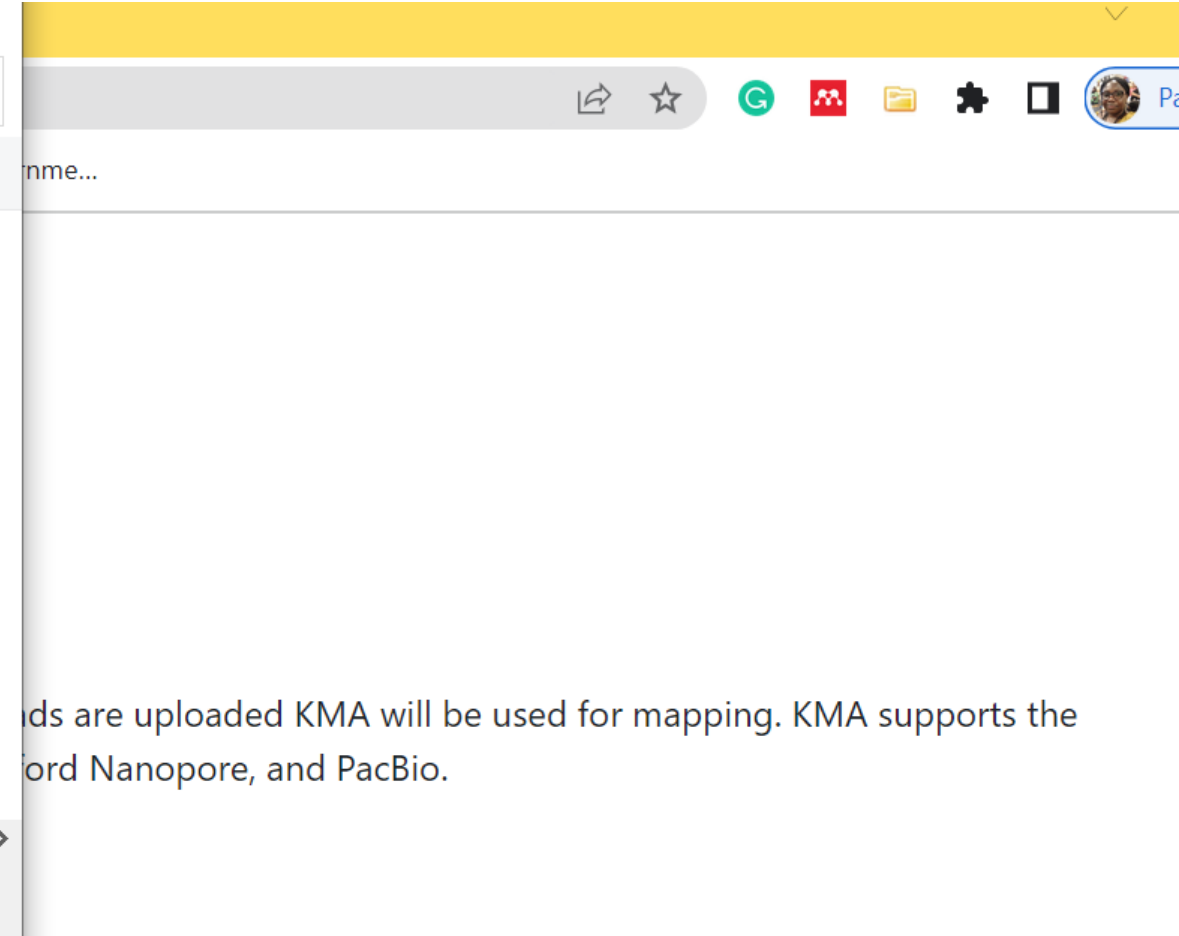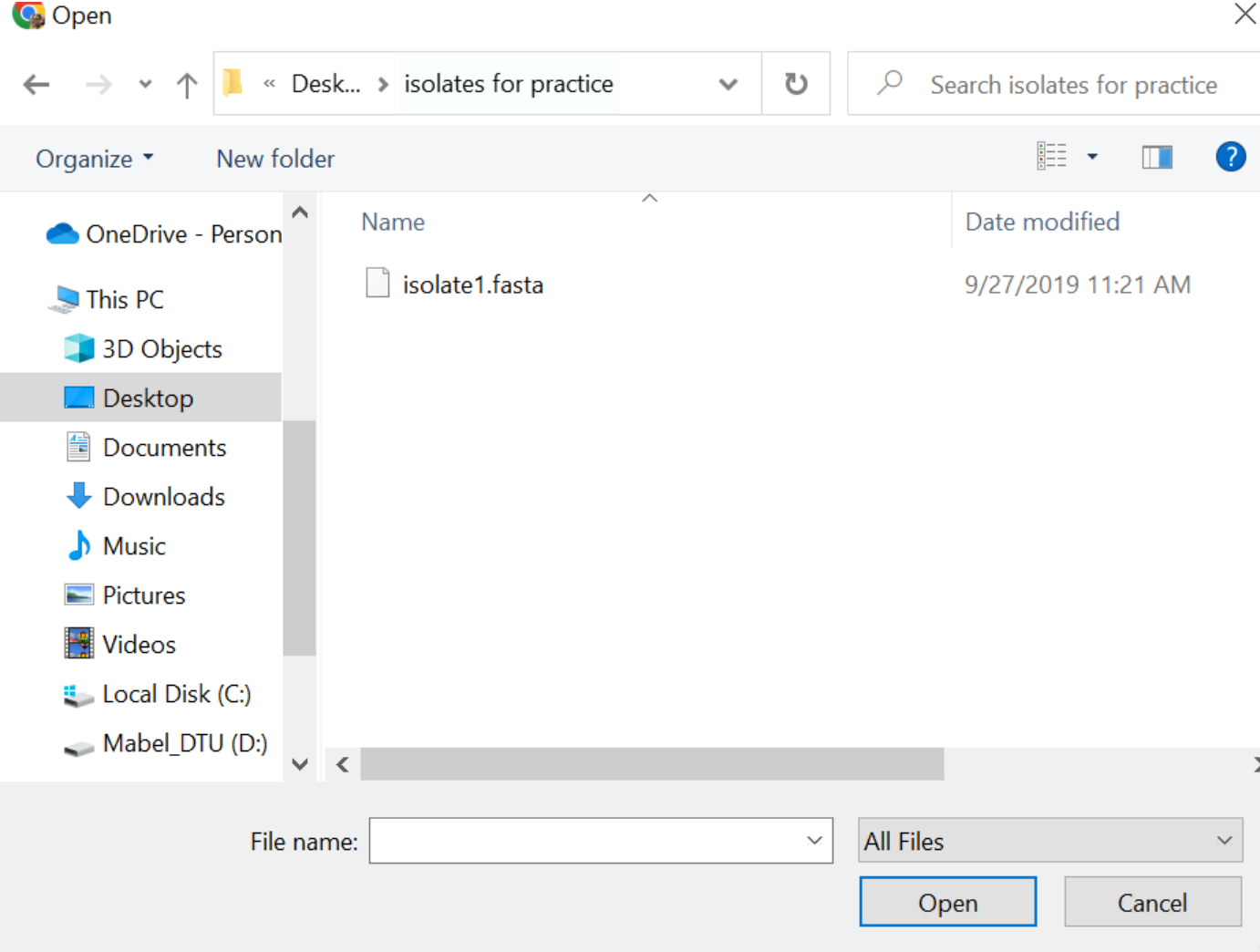
15

Desk... > isolates for practice

Search isolates for practice

Organize ▼   New folder

OneDrive - Person

This PC
3D Objects
Desktop
Documents
Downloads
Music
Pictures
Videos
Local Disk (C:)
Mabel_DTU (D:)

Name	Date modified
isolate1.fasta	9/27/2019 11:21 AM

...ds are uploaded KMA will be used for mapping. KMA supports the

...ord Nanopore, and PacBio.

File name: [   ]   All Files

Open   Cancel

Choose File(s)

Name	Size	Progress	Status

Upload   Remove

16

Assessing your results

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.
maworh@gmail.com

To get notified by email: [_____] [Notify me via email]

This page will update itself automatically

CGEwebface@cbs.dtu.dk <cgewebface@cbs.dtu.dk>   Fri, Feb 3 at 5:37 PM
To: maworh@gmail.com

Dear sir/madam,
    Your job 63DDB14700007D13AF2F51DF has finished on our VirulenceFinder-2.0 server.
    You may visit the following link to study the results:
    https://cge.food.dtu.dk//cgi-bin/webface.fcgi?jobid=63DDB14700007D13AF2F51DF

    The data will expire at 2023-02-11 02:13:43.

Yours sincerely,
    CGE webserver

VirulenceFinder-2.0 Server - Results

Organism(s): *Escherichia coli*

Virulence genes for Escherichia coli						
Virulence factor	Identity	Query / Template length	Contig	Position in contig	Protein function	Accession number
AslA	93.9	1657 / 1656	NODE_13_length_88665_cov_5.581852	37650..39305		CP054236
csgA	100	456 / 456	NODE_89_length_9640_cov_5.145275	294..749	curlin major subunit CsgA	CP069646
fdeC	92.88	4258 / 4254	NODE_14_length_87312_cov_5.365935	1478		
fimH	99.59	489 / 489	NODE_68_length_19341_cov_4.739565	6213		
hlyE	100	918 / 918	NODE_5_length_144015_cov_5.105777	104334..105251	Avian E. coli haemolysin	ECU57430
nlpI	99.77	885 / 885	NODE_12_length_94487_cov_5.441109	41575..42459	lipoprotein NlpI precursor	CP000243

Shiga-toxin genes						
Virulence factor	Identity	Query / Template length	Contig	Position in contig	Protein function	Accession number
No hit found						

The dark green color indicates a perfect match for a given gene. The % ID is 100 and means that the sequence in the genome covers the entire length of the virulence gene in the database

The light green color indicates a warning due a non-perfect match, ID < 100% although both query and template lengths are equal

The red color indicates that no virulence gene was detected for the shiga toxin

18

Click on extended output for more information

extended output

Results as text | Results tsv | Hits in genome seqs | Virulence factor seqs

Input Files: *ma_011.fasta*

CITATIONS

For publication of results, please cite:

- Real-time whole-genome sequencing for routine typing, surveillance, and outbreak detection of verotoxigenic Escherichia coli.
  Joensen KG, Scheutz F, Lund O, Hasman H, Kaas RS, Nielsen EM, Aarestrup FM.
  J. Clin. Micobiol. 2014. 52(5): 1501-1510.
  View the abstract

fimH_NA

```
template   TTCGCCTGTAAAACCGCCAATGGTACAGCTATCCCTATTGGCGGTGGCAGCGCTAATGTT
query      TTCGCCTGTAAAACCGCCAATGGTACAGCTATCCCTATTGGCGGTGGCAGCGCTAATGTT

template   TATGTAAACCTTGGCGCCCGCCGTGAATGTGGGGCAAAACCTGGTCGTAGATCTTTCGACG
query      TATGTAAACCTTTCGCCCGCCGTGAATGTGGGGCAAAACCTGGTCGTAGATCTTTCGACG

template   CAAATCTTTTGCCATAACGATTATCCGGAAACCATTACAGACTATGTCACACTGCAACGA
query      CAAATCTTTTGCCATAACGATTATCCGGAAACCATTGCAGACTATGTCACACTGCAACGA

template   GGCTCGGCTTATGGCGGCGTGTTATCTAATTTTTCCGGGACCGTAAAATATAGTGGCAGT
query      GGCTCGGCTTATGGCGGCGTGTTATCTAATTTTTCCGGGACCGTAAAATATAGTGGCAGT

template   AGCTATCCATTTCCTACCACCAGCGAAACGCCGCGCGTTGTTTATAATTCGAGAACGGAT
query      AGCTATCCATTTCCTACCACCAGCGAAACGCCGCGCGTTGTTTATAATTCGAGAACGGAT

template   AAGCCGTGGCCGGTGGCGCTTTATTTGACGCCTGTGAGCAGTGCGGGCGGGGTGGCGATT
query      AAGCCGTGGCCGGTGGCGCTTTATTTGACGCCTGTGAGCAGTGCGGGCGGGGTGGCGATT

template   AAAGCTGGCTCATTAATTGCCGTGCTTATTTTGCGACAGACCAACAACTATAACAGCGAT
query      AAAGCTGGCTCATTAATTGCCGTGCTTATTTTGCGACAGACCAACAACTATAACAGCGAT

template   GATTTCCAGTTTGTGTGGAATATTTACGCCAATAATGATGTGGTGGTGCCCACTGGCGGC
query      GATTTCCAGTTTGTGTGGAATATTTACGCCAATAATGATGTGGTGGTGCCCACTGGCGGC

template   TGTGATGTT
query      TGTGATGTT
```

This shows you the alignment between your sequence in the genome & the virulence gene. The red colour shows a variation has occurred e.g instead of a base "G" there is a "T" and instead of base "A" we have a "G"

20

hlyE_ECU57430

Now you should be able:

- To access ResFinder & VirulenceFinder and upload your sequences

- To access your results or outputs from ResFinder & VirulenceFinder

- To interpret the outputs from ResFinder & VirulenceFinder

Thank you!

Happy to take any questions.
You can contact me:
maworh@gmail.com