

Introduction to whole genome sequencing and  
online sequence analysis by CGE tools

# De-mystifying WGS

Jette S. Kjeldgaard ([jetk@food.dtu.dk](mailto:jetk@food.dtu.dk))

DTU Food



# Introduction to sequencing

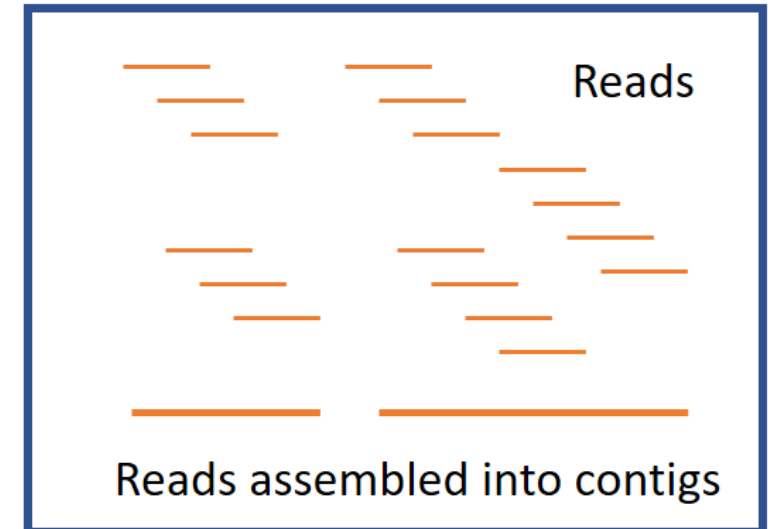
- One methodology with numerous applications
- Routine method in many laboratories
  - Relevant for fastidious or slow-growing microorganisms
  - Relevant for hard-to-cultivate microorganisms
- Still requires a pure culture for single isolate sequencing
  - Or requires skills to handle complex metagenomics samples

- ✓ Diagnostics
- ✓ Additional characterisation
- ✓ Outbreak detection



# Types of sequencers

- Short read technologies
  - **Illumina** has long been gold standard
  - Numerous types of equipment with various throughput
    - MiSeq, HighSeq, NextSeq, MiniSeq....
  - typical 150 -300 bp reads - low error rate
  - Most software accepts Illumina reads or contigs
- Long reads technologies
  - Oxford Nanopore Tech. (ONT) – **MinION**
  - **PacBio**
  - From 20 kbp to more than 200 kbp – higher error rate
  - Less software solutions available (yet)



# Flow of sequencing

- From pure culture or specimen
- DNA extraction (plus quality/quantity control)
- Library preparation
- Sequencing
- Data quality control

**Outsourcing of sequencing**



**DNA – largely same procedure for different types of organisms**

# Quality control aspects

- Sequence quality
  - Raw reads quality; length, quantity and quality scores (phred scores/fastq files)
  - Assembly; # of contigs, draft genome length, N50... (Assembly > fasta files)
- Sequence validation
  - Check for species ID
  - Check for contamination
  - (Metagenomics – check for target organism, abundance ...)

Sequence ok -> proceed to analysis

# 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

<https://cge.food.dtu.dk/>

## Overview of Services

### Phenotyping

#### [ResFinder](#)

Identification of acquired antibiotic resistance genes.

#### [ResFinderFG](#)

Identification of functional metagenomic antibiotic resistance determinants.

#### [LRE-finder](#)

Identification of genes and mutations leading to linezolid resistance.

#### [KmerResistance](#)

Identification 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.

#### [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](#)

# CGE Tools collection - overview

- Phenotyping
- Typing
- Phylogeny
- Metagenomics
- PCR tools
  
- Others:
  - User-defined databases
  - Identify bacteriophage sequences
  - Identify the bacterial host of phage genome
  - ...

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





# 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*, *Clostridium*, *E. coli*, *Listeria*, *Salmonella*, *Yersinia*
    - pMLST
    - **Serotyping** (*E. coli*, *P. aeruginosa*, *Salmonella*)
  - PlasmidFinder
  - MGE (Mobile genetic elements)

# Example: Kmer-finder (Species ID – and more!)

## KmerFinder 3.2

Service [Instructions](#) [Output](#) [Article abstract](#) [Citations](#)

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

### 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 address is https and not just http. Fix it by clicking [here](#).

Name	Size	Progress	Status

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

### Select database

- Archaea
- Archaea
- Bacteria organisms
- Fungi
- Protozoa
- Bacteria type strains
- Viral

# 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

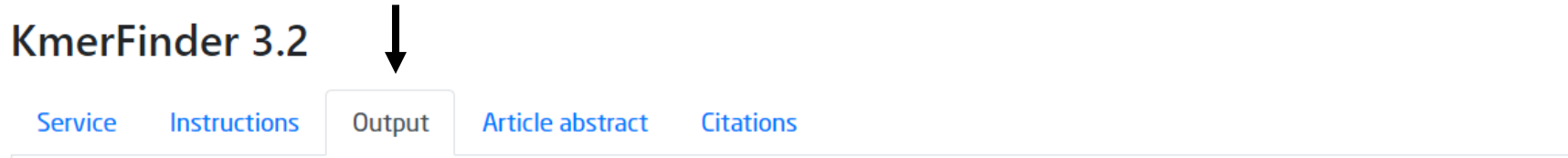
KmerFinder 3.2 results:

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 database 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



## 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	query_coverage	Coverage	Depth	tot_query_coverage
NZ_CP016952.1 Citrobacter freundii strain SL151 chromosome, complete genome	1723	127691	21	168352	71.33	76.91	0.76	71.33

## 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 template

**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 temp

**Template\_Coverage [%]:** is the template coverage.

**Depth:** is the number of matched kmers in the query sequence divided by the total sequencing depth.

# Example: Resfinder

## ResFinder 4.1

Service [Instructions](#) [Output](#) [Article abstract](#) [Citations](#) [Overview of genes](#) [Database history](#)

ResFinder identifies acquired genes and/or finds chromosomal mutations mediating antimicrobial resistance 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

\*Chromosomal point mutation database exists

Select type of your reads

**Chromosomal point mutations**

Select threshold for %ID

Select minimum length

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

Select minimum length

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

# Example: Resfinder

## ResFinder 4.1

Service **Instructions** Output Article abstract Citations Overview of genes Database history

ResFinder identifies acquired genes and/or finds chromosomal mutations mediating antimicrobial resistance 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.\*

Campylobacter spp.\*

Campylobacter jejuni\*

Campylobacter coli\*

Escherichia coli\*

Salmonella spp.\*

Plasmodium falciparum\*

Neisseria gonorrhoeae\*

Mycobacterium tuberculosis\*

Enterococcus faecalis\*

Enterococcus faecium\*

Klebsiella\*

Helicobacter pylori\*

Staphylococcus aureus\*

Other

Select type of your reads

Assembled Genome/Contigs

Assembled Genome/Contigs

454 - single end reads

454 - paired end reads

Illumina - single end reads

Illumina - paired end reads

Ion Torrent

SOLiD - single end reads

SOLiD - paired end reads

SOLiD - mate pair reads



# 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)

- Beta-lactam resistance - ampicillin
  - TEM-1B, TEM-1C, TEM-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, ceftioxiime, piperacillin, ticarcillin

- All the TEM- genes are in fact the same gene
- – placed on (contig/) 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/fastq

- Almost all the CGE tools are available for Illumina reads/fastq 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....	

# 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
MinTyper	Cluster detection and SNP calling

# CGE tools **Typing**

- **Species identification**
  - **KmerFinder** (kmer-based)
  - SpeciesFinder (16S ribosomal DNA)
- **Typing**
  - **MLST**
    - cgMLSTFinder
      - *Campylobacter*, *Clostridium*, *E. coli*, *Listeria*, *Salmonella*, *Yersinia*
    - pMLST
    - **Serotyping** (*E. coli*, *P. aeruginosa*, *Salmonella*)
  - PlasmidFinder
  - MGE (Mobile genetic elements)

# Serotyping

## SeqSero 1.2

Service

Instructions

Output

Article abstract

Citations

SeqSero predicts the Salmonella serotype of either the pre-assembled or raw read sequence data provided to the service.

**Note:** This service is hosted by CGE but all credit and scientific questions should be given to the original authors from [Deng Lab \(SeqSero\)](#).

### More info on Salmonella serotypes


From [Deng Labs website](#)

Software version: [Available on GitHub here](#)

Download the Salmonella determinants databases from: [Deng Labs website \(zip file\)](#)

### Select Data type

Assembled Genome/Contigs

 Choose File(s)

# Serotype result

**SeqSero-1.2 Server - Results**

**SeqSero-1.2 Server - Results**

**Serotype:**

**Illa 13,22:z4,z23:- or Ajiobo or Ried or Illa 13,23:z4,z23,[z32]:-\***

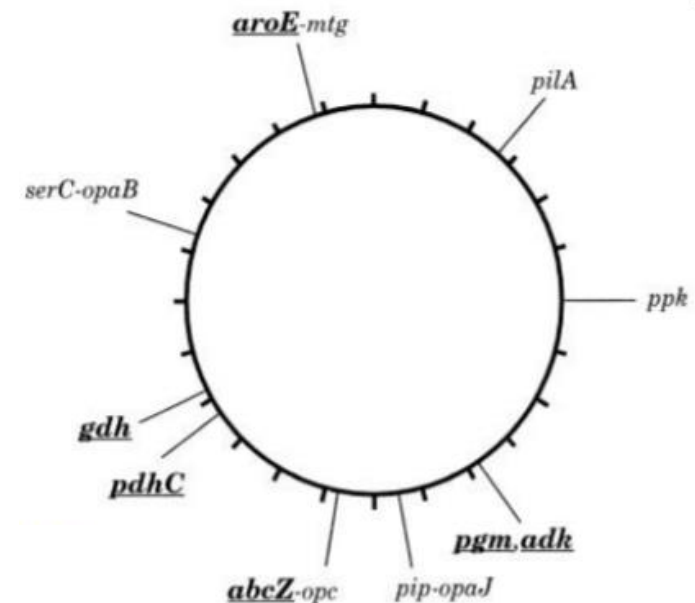
Antigenic profile	13:z4,z23:-
O antigen	O-13
H1 antigen	z4,z23
H2 antigen	-



# Multi-Locus Sequence Typing (MLST)

Classical MLST:

- The golden standard for typing
- First developed in 1998 for *Neisseria meningitidis* (Maiden et al. PNAS 1998. 95:3140-3145)
- The nucleotide sequence of internal regions of app. 7 housekeeping genes are determined by PCR followed by Sanger sequencing
- Different alleles are each assigned a random number
- The unique combination of alleles is the sequence type (ST)

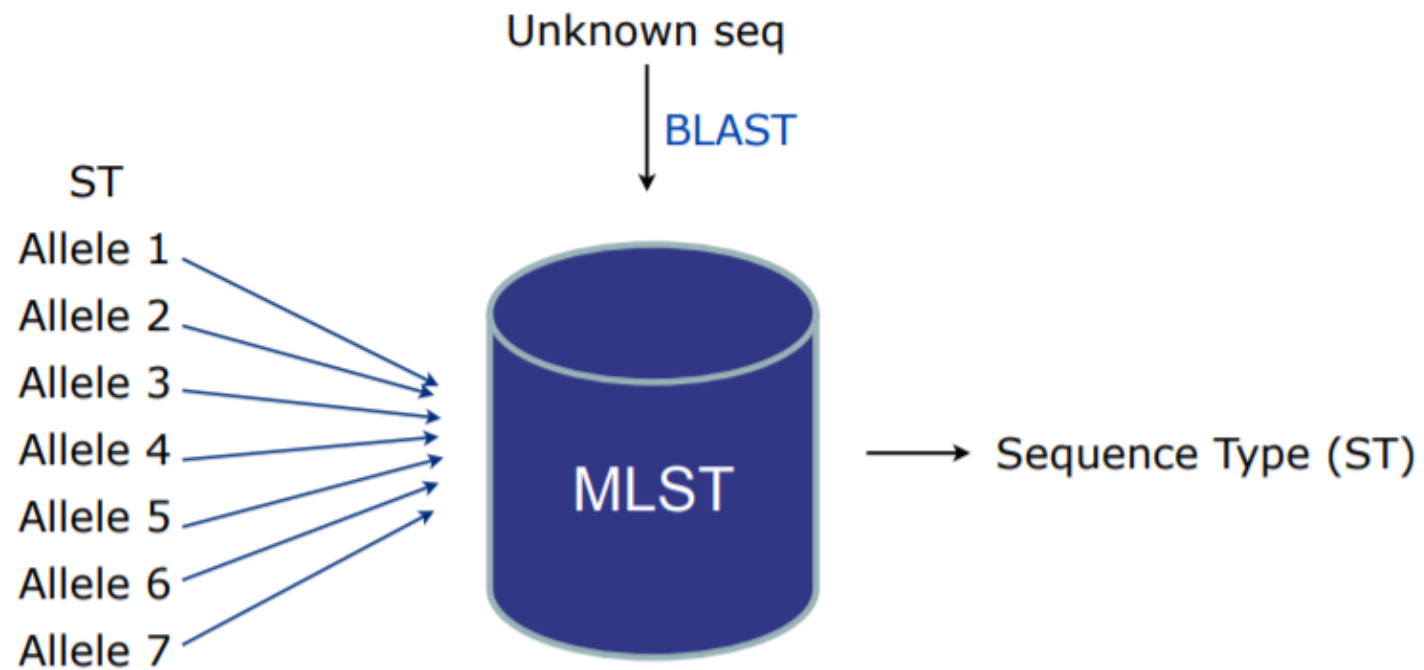


## MLST now

- For many bacterial species, MLST is considered the gold standard of typing
  - It is traditionally performed in an expensive and time-consuming way
- As the cost of WGS continue to decline, it becomes increasingly available to scientists and routine diagnostics laboratories
  - Currently, the WGS cost is typically below that of traditional MLST

**7 x PCR and sequencing vs. 1 x WGS**

# MLST Typing by WGS



# MLST result output

## MLST-2.0 Server - Results

mlst Profile: *Imonocytogenes*

Organism: *Listeria monocytogenes*

Sequence Type: 6

Locus	Identity	Coverage	Alignment Length	Allele Length	Gaps	Allele
abcZ	100	100	537	537	0	abcZ_3
bglA	100	100	399	399	0	bglA_9
cat	100	100	486	486	0	cat_9
dapE	100	100	462	462	0	dapE_3
dat	100	100	471	471	0	dat_3
ldh	100	100	453	453	0	ldh_1
lhkA	100	100	480	480	0	lhkA_5

extended output

Input Files: *Lm02.fa*

**One limitation: ONE variation in bases of one of the seven genes: new allele number = different ST**

**Why limit to SEVEN genes when we sequence the whole genome?  
-> core genome MLST**

# cgMLST – core genome

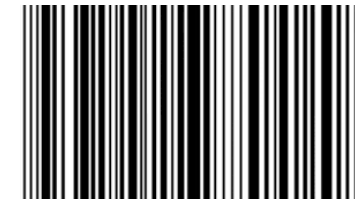
- Core genome = genes common for all (almost) within the species
  - *Salmonella* has approx. 5000 genes, hereof >3000 are selected for the cgMLST

Locus	Identity	Coverage	Alignment Length	Allele Length	Gaps	Allele
abcZ	100	100	537	537	0	abcZ_3
bglA	100	100	399	399	0	bglA_9
cat	100	100	486	486	0	cat_9
dapE	100	100	462	462	0	dapE_3
dat	100	100	471	471	0	dat_3
ldh	100	100	453	453	0	ldh_1
lhkA	100	100	480	480	0	lhkA_5

Gene08						
Gene09						
Gene10						
Gene11						
Gene12						
Gene13						
Gene14						
Gene15						
Gene16						
Gene17						
Gene18						
Gene19						
Gene20						

Each gene variant has an allele number

Each allele combination has a **cg ST** assigned based on the cgMLST scheme



By cgMLST very closely related genomes are 'lumped' together in a Complex Type (CT)

Can also be used to interpret clusters

# Whole genome based phylogeny

- Single Nucleotide Polymorphism (SNP)
  - Require reference genome
- Gene-by-gene approach
  - cgMLST – core genome MLST/wgMLST - whole genome MLST
  - No reference genome required
  - Require species specific cgMLST scheme
- What is phylogeny used for?
  - Classify taxonomy – the classic use
  - Outbreak detection – detection of clones – increasing with WGS data

# Focus on (CSI)phylogeny

- Phylogenetic comparisons allow for determining clusters and clonal spread of microorganisms
- SNP calling – to determine variants in the DNA (Single Nucleotide Polymorphism)
- Different sequencing technology has a systematic bias making integration of data generated from different platforms difficult.
  - CSIPhylogeny has incorporated two different procedures for identifying variable sites and inferring phylogenies in WGS data across multiple platforms

## CSI Phylogeny 1.4 (Call SNPs & Infer Phylogeny)

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

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

# Data quality and SNP calling

- Good data quality ensures reliability of your analysis
  - Poor quality sequences can rarely be used for SNP analysis
- For assembled contigs - good coverage is essential ( $\geq 30x$ )
- Consider the quality of your raw data (specifically phred scores)
- **CSI Phylogeny SNP filtering criteria:**
  - SNP quality:  $\geq 30$  (Phred score, base call accuracy: 99.9%)
  - SNPs with a sequence depth of  $< 10$  are removed
  - A SNP is removed if it is  $< 10$  bps from the nearest SNP (Pruning)(recombination do not reflect naturally evolved SNPs)

**Preferably analyse raw reads  
for better resolution!**



# SNPs detection (CSIPhylogeny)

- Calling of single nucleotide polymorphism
  - Variants in the DNA – compared to reference

....ATCGAATTCCGGGTTTTTAACCGGATCGTACGATCGGGAAAAA..

TTCCAGG

TTCCAGG

TTCCAGG

TTCCAGG

TTCCAGG

TTCCAGG

SNPs are called on **the nucleotides which all isolates in the analysis share** with the reference.

Higher variation between isolates = higher difference from reference

->

Decreasing amount of nucleotides to call SNPs from (Valid positions/ percentage of reference covered)

# Cluster analysis tools

- Upload of reference genome (fasta file)
- Change parameters
  - Or go with standard settings
- Upload of additional files
  - Either fasta or fastq files can be used
- Wait for analysis result
  - (or type in email address)

**Input data**

Upload reference genome (fasta format)  
Note: Reference genome must not be compressed.

Gennemse... Ingen fil valgt.  
 Include reference in final phylogeny.

Select min. depth at SNP positions  
10x

Select min. relative depth at SNP positions  
10 %

Select minimum distance between SNPs (prune)  
10 bp

Select min. SNP quality  
30

Select min. read mapping quality  
25

Select min. Z-score  
1.96

Ignore heterozygous SNPs

**Comment (to yourself)**  
This comment will appear unaltered on your output page. It has no effect on the analysis.

Use altered FastTree (more accurate)  
Note: Read more [here](#)

Upload read files and/or assembled genomes (fasta or fastq format)  
Note: Read files must be compressed with gzip (compressed files often ends with .gz).  
If you get an "Access forbidden. Error 403": Make sure the start of the web address is https and not just http. Fix it

Name

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

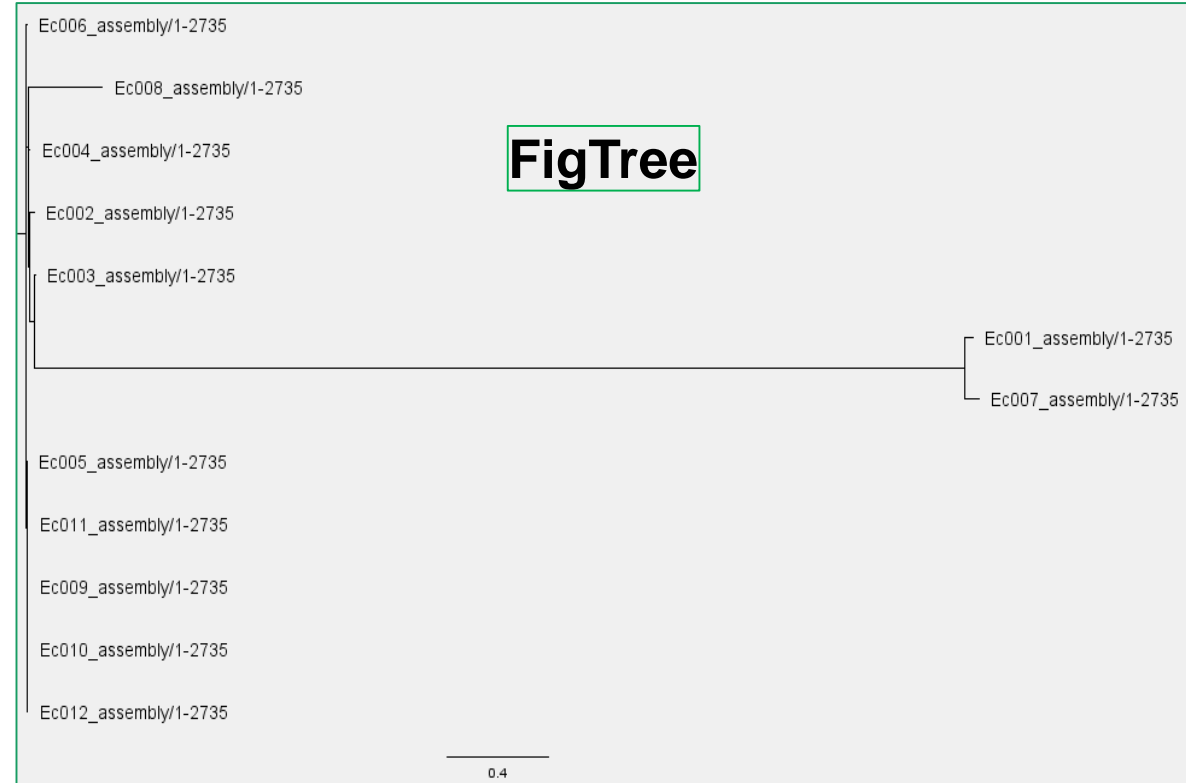
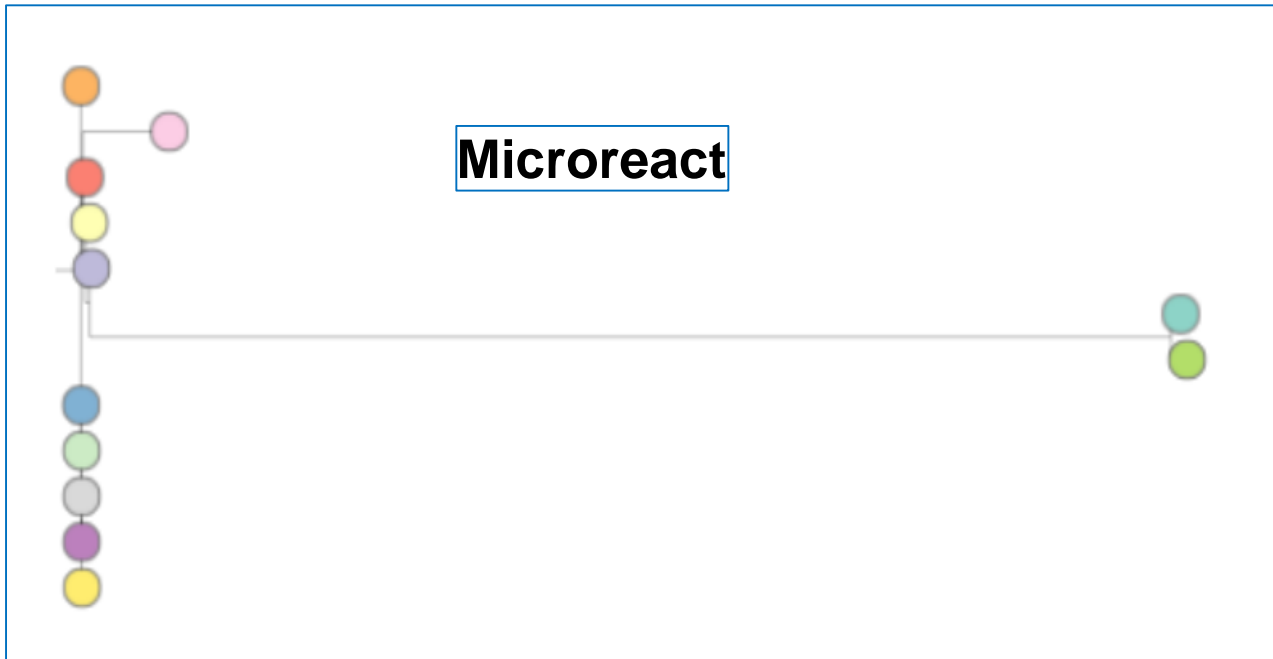
# Cluster analysis outputs

- SNP matrix
  - Shows pairwise variations between tested sequences

	EC1	EC2	EC3	EC4	EC5	EC6	EC7	EC8	EC9	EC10	EC11	EC12
EC1	0	2176	2122	2180	2176	2176	216	2280	2179	2179	2182	2184
EC2	2176	0	94	80	78	78	2212	644	81	81	84	86
EC3	2122	94	0	98	96	96	2170	662	99	99	102	104
EC4	2180	80	98	0	38	38	2222	604	41	41	44	46
EC5	2176	78	96	38	0	2	2218	598	5	5	8	10
EC6	2176	78	96	38	2	0	2218	598	5	5	8	10
EC7	216	2212	2170	2222	2218	2218	0	2322	2221	2221	2224	2226
EC8	2280	644	662	604	598	598	2322	0	601	601	604	606
EC9	2179	81	99	41	5	5	2221	601	0	0	3	5
EC10	2179	81	99	41	5	5	2221	601	0	0	3	5
EC11	2182	84	102	44	8	8	2224	604	3	3	0	8
EC12	2184	86	104	46	10	10	2226	606	5	5	8	0

# Cluster analysis output

- Newick file – distance matrix for visualisation with various tools



## Take home messages

- Numerous free, online tools available for diagnostics and comparison of microorganisms
- CGE tools offers a broad range of tools
- Different types of sequencers generate different types of data
- Sequence files generally available as fasta and/or fastq files
  - Some tools are restricted to one type

# Thank you for the attention!

- **Questions or comments?**
- **Jetk@food.dtu.dk**