Outbreaks, typing & AMR/Day 7

Serotyping and virulence typing of *Salmonella* and *E. coli*

Section of Foodborne infections

Dep. of Bacteria, Parasites and Fungi, Statens Serum Institut

Katrine Grimstrup Joensen and Pernille Gymoese, April 2024

Objectives

Specific objectives of this session:

1. Learn about *in silico* serotyping of *Salmonella* and *E. coli* isolates
2. Learn about *in silico* virulence profiling of *E. coli*
3. Learn how to apply SerotypeFinder, SeqSero and VirulenceFinder
4. Interpret typing results

Related to the course objectives:

A. Process sequencing data (from wgs data to results)
B. Extract relevant information from processed data
C. Perform basic analysis supporting epidemiological investigations, including interaction with public databases

Outline

This session consists of the following elements

1. Introduction to *in silico* serotyping of *Salmonella*
2. SeqSero
3. Introduction to *in silico* serotyping and virulence profiling of *E. coli*
4. SerotypeFinder
5. VirulenceFinder
6. Issues regarding *in silico* typing
7. Introduction to exercise
8. Group exercise integrating *in silico* tools for serotyping of *E. coli* and *Salmonella* and virulence profiling of *E. coli*
9. Group discussion of results

Surveillance at SSI

Surveillance flow from hospital to SSI and back

[redacted]

Serotypes
ST
Species
Virulence

AMR (genotype and phenotype)

Cluster analysis

Results

Reports and trends

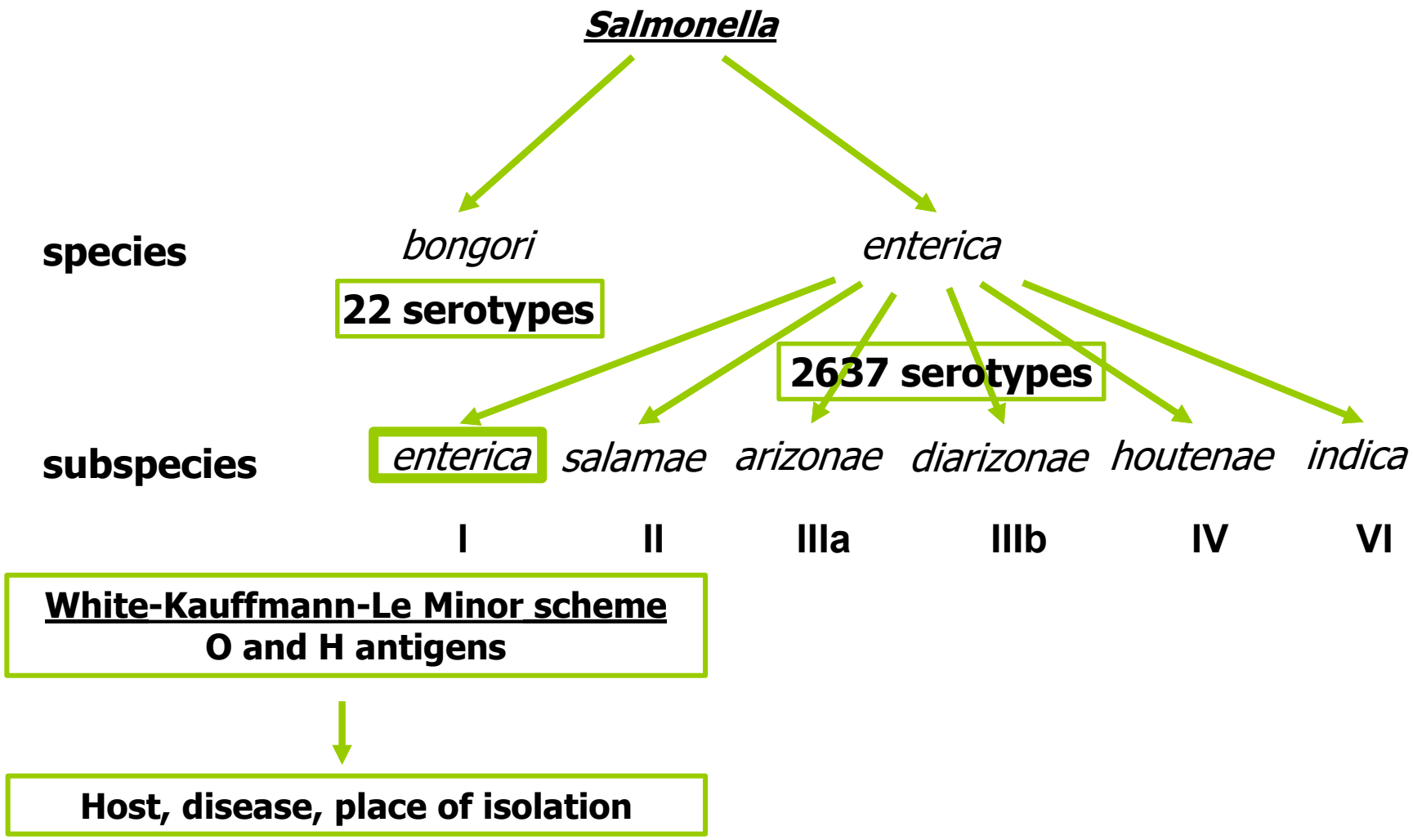
Outbreak investigation

[redacted]

[redacted]

[redacted]

Serotyping of *Salmonella*



- Based on O and H antigens
- **O:H phase 1:H phase 2**
- Subspecies *enterica* is named after host, disease or geographical place of isolation
- Other subspecies after antigen formula

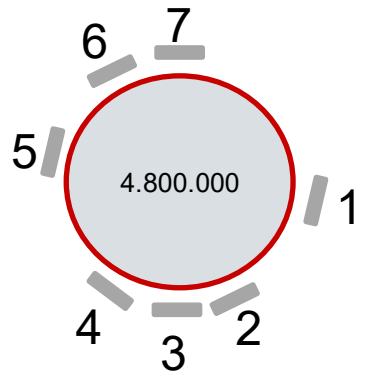
Examples:

- **Serotype:** *Salmonella enterica* subspecies *enterica* serotype London (*S.* London)
- **Antigen profile:** 3,10:l,v:1,6
- **Serotype:** *Salmonella enterica* subspecies *salamae* serotype
- II 58:a:z_6

Sequence typing – new nomenclature

- **Multi Locus Sequence Typing - MLST**
  - Seven conserved genes
  - Genetic relatedness

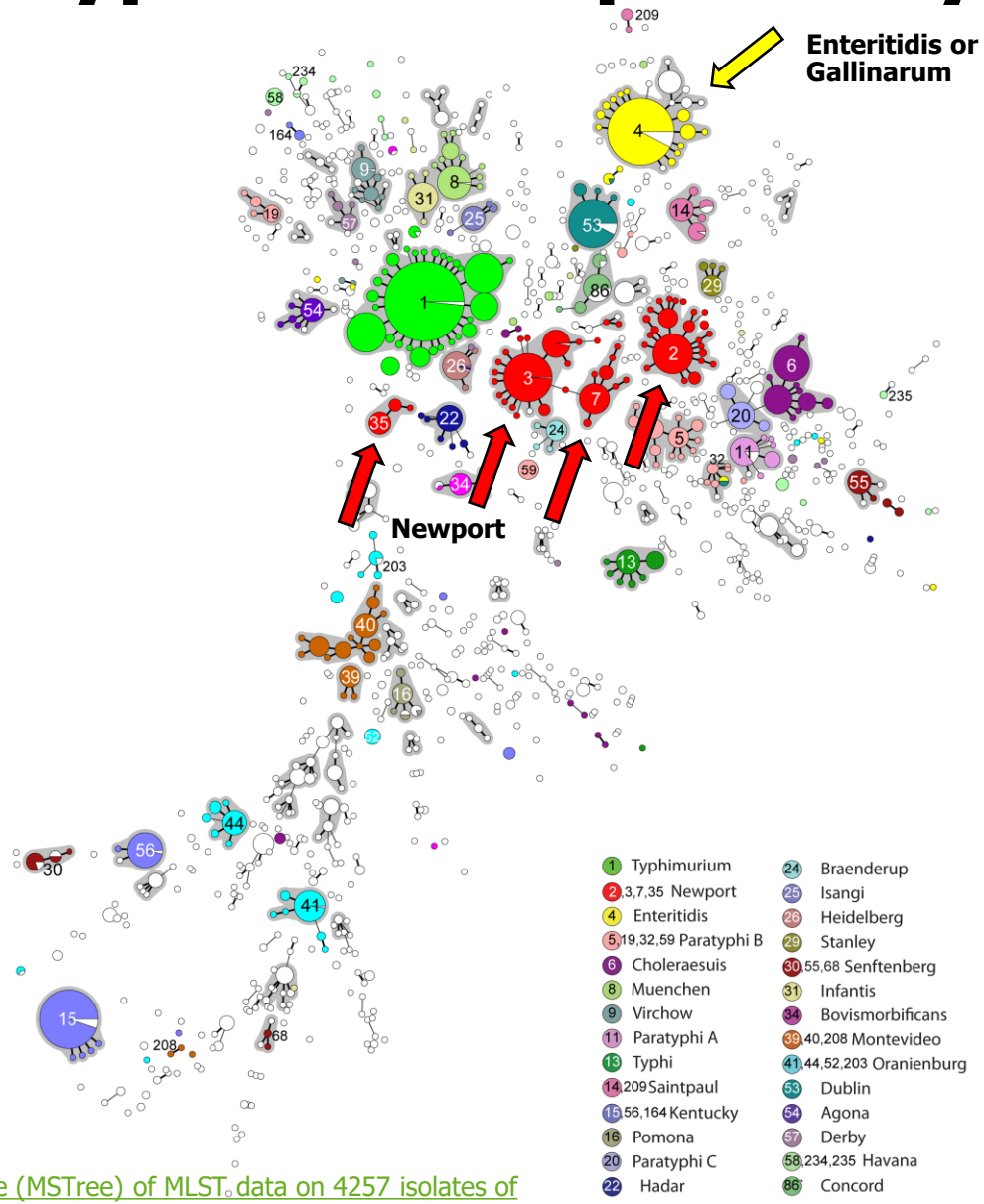
MLST - 7 genes



- **eBG – eBurstGroup**
  - Closely related ST´s
  - Correlates to serotypes

aroC	dnaN	hemD	hisD	purE	sucA	thrA		ST	eBG	Serotype
17	18	22	17	5	21	19		32	31	Infantis
17	341	22	17	5	21	19		1032	31	Infantis

Serotypes and sequence types



Minimal spanning tree (MSTree) of MLST data on 4257 isolates of S. enterica subspecies enterica by Achtman et al. (2012)

Serotypes

- Pros
  - Well known nomenclature
  - Can provide information on properties like disease and host specificity
- Cons
  - Not necessarily genetic related

ST and eBG

- Pros
  - Information on genetic relatedness
- Cons
  - No information on properties

Replacement for serotyping? – or a combination of both

Interpretation of *Salmonella* serotype results

- **Same antigen formula - different subspecies**
- **Biotypes/variants** – same antigen formula – different phenotypes
  - Additional biochemical test
- **Paratyphi B and Paratyphi B var. Java**
  - Differentiated by fermentation of d-tartrate
- **Enteritidis/Gallinarum**
  - Genotypic serotype looks like Enteritidis – non-expressed H phase
  - Phenotypic serotype looks like non-motile Enteritidis
  - Additional biochemical test
  - Gene *sdf* missing in Gallinarum and present i common Enteritidis – but is not exclusive….
- See http://denglab.info/SeqSero2/supple or https://www.pasteur.fr/sites/default/files/veng_0.pdf for further details

SeqSero2

- **Input**
  - Raw reads
  - Assemblies

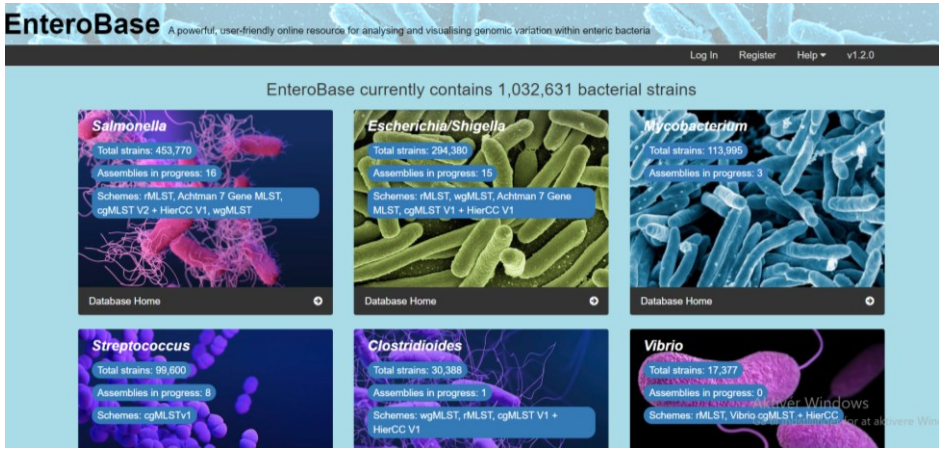
- **Output**
  - Detected O and H antigenes
  - Predicted antigen formula
  - Predicted serotype
  - Predicted species and subspecies
  - Potential inter-serotype contamination
  - Notes

SeqSero2 vs SeqSero1

- **Faster – k-mer workflow**
- **Identify species and subspecies**
- **Additional genetic markers**
  - Differentiate Paratyphi B and variant Java (D-tartrate fermentation)
  - +/- *sdf* gene in Enteritidis – indication of variant Gallinarum (*sdf* negative)
  - O5 negative mutation (-O5) – variant of Typhimurium
  - Differentiation of O22 and O23 in the O13 group
- **Inter-serotype contamination**
  - Can determine more than one O antigen and more than 2 H antigens – indication of contamination

Alternatives

- **SISTR**
  - Serotype prediction from MLST and extended MLST schemes (rMLST and cgMLST)
  - https://doi.org/10.1371/journal.pone.0147101
- **Enterobase**
  - Seqsero2, SISTR and predicted type after ST/eBG
  - https://enterobase.warwick.ac.uk/



Serotyping of *E. coli*

- Based on O and H antigens
- Typical single O and H antigens
- Named according to antigen formula

Example:

- O157:H7

E. coli pathotypes

- **Diarrheagenic *E. coli* (DEC)**
  - **STEC - Shiga toxin-producing**
  - EPEC - Enteropathogenic
  - ETEC - Enterotoxigenic
  - EIEC - Enteroinvasive
  - EAEC - Enteroaggregative
- **Extraintestinal *E. coli* (ExPEC)**
  - UPEC - Uropathogenic

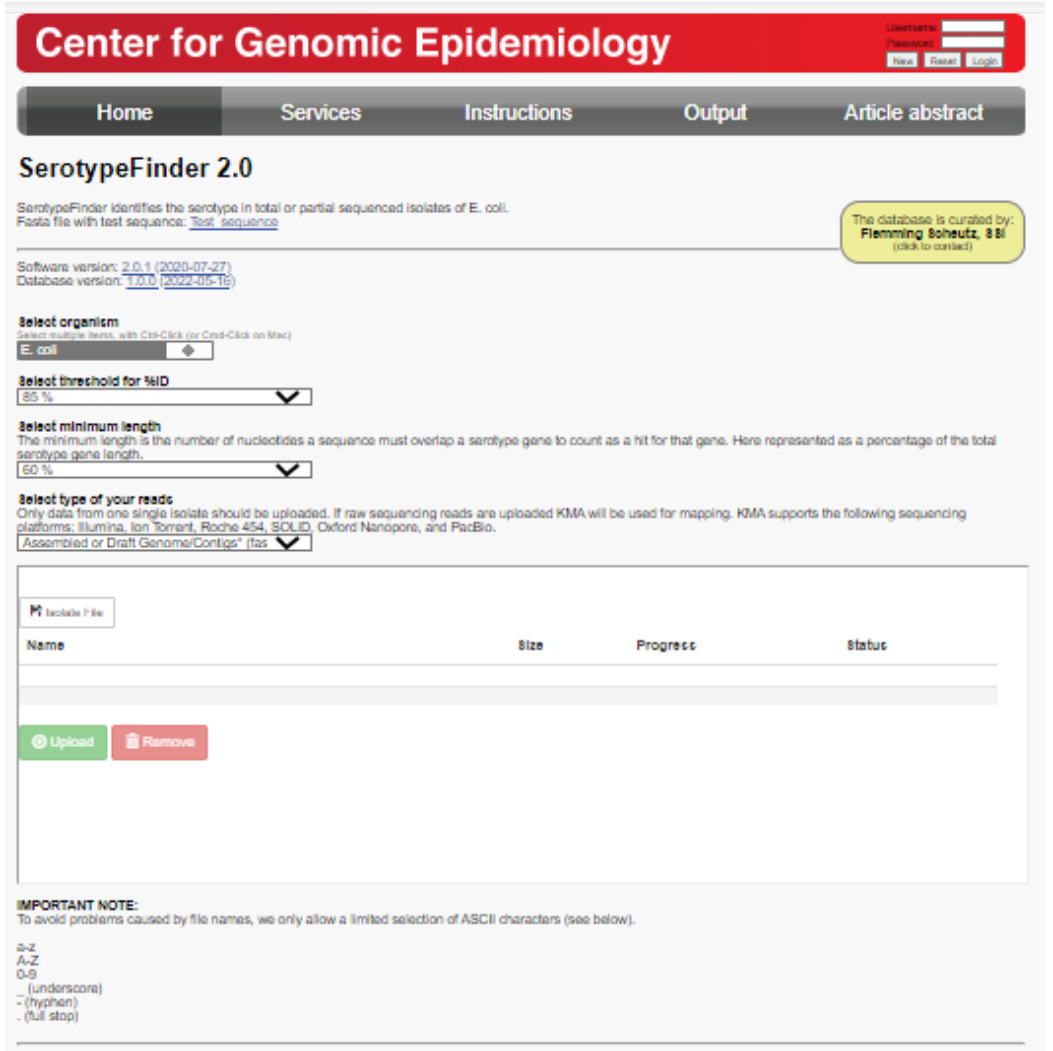
Characterized by virulence genes

E. coli typing

- **Previous STEC typing at SSI:**
  - Serotyping
  - Resistance profiling
  - Test for hemolysin, β-glucuronidase, indole production
  - PCR and Hybridization to detect virulence genes (*eae, vtx1*, *vtx2*…)
  - PFGE for suspected outbreak isolates
- **Replaced by WGS analysis**

SerotypeFinder & VirulenceFinder

- **Developed as online CGE tools**
  - Select database
  - Select ID/Coverage (default 85%, 60%)
  - Select data type
  - Upload data
- **Blast-based, KMA for raw data**
- **Also available command-line**
- **At SSI - integrated databases (not the whole tool) in our pipeline (with mapping)**



SerotypeFinder: *in silico* O and H-typing

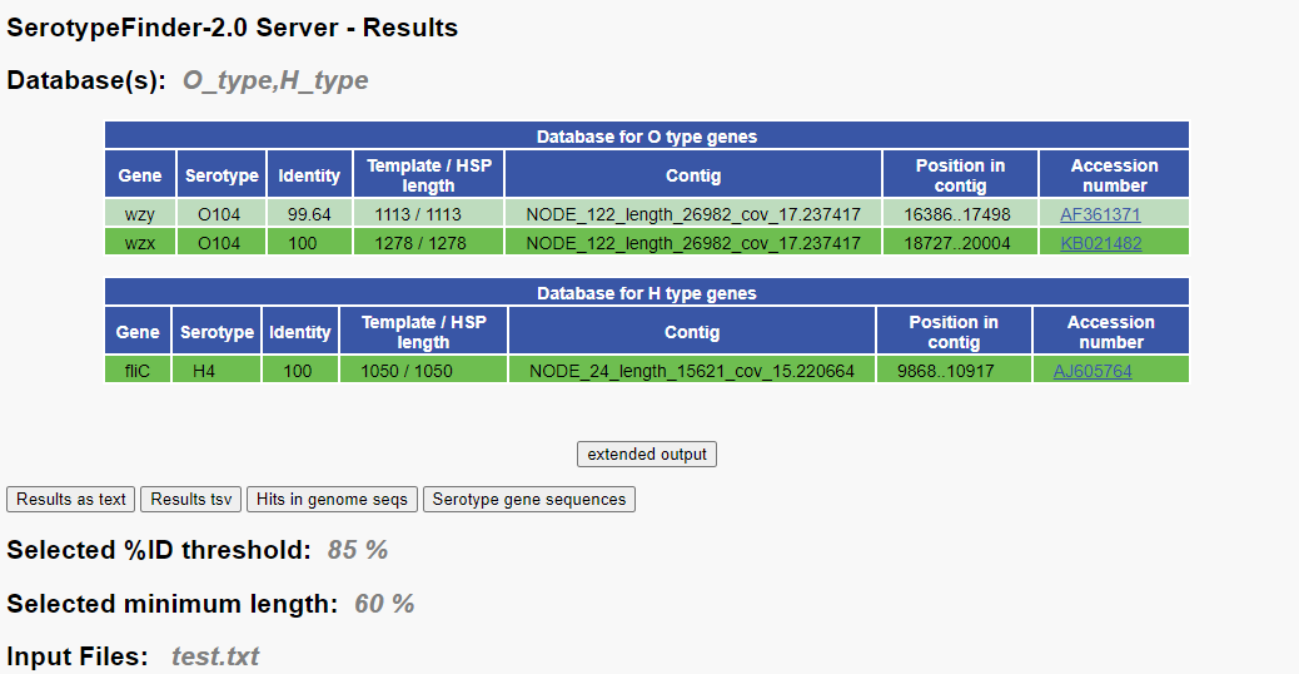
- **187 known *E. coli* O-antigens (O1-O187)**
  - More added now
  - ***wzx*** and ***wzy*** (O-antigen flippase and polymerase)
  - ***wzm*** and ***wzt*** (ABC transporter)

- **All 53 known H-types**
  - *fliC, flkA, flnA, flmA, fllA*

SerotypeFinder results

- **Input**
  - Raw reads
  - Assemblies

- **Output**
  - O - combined wzx/wzy
  - H - flagellin genes
  - Sequence identity, coverage and positions
  - Reference sequence
  - Examine alignments
  - Command line output more messy and needs to be sorted



SerotypeFinder-2.0 Server - Results

Database(s): *O_type,H_type*

Database for O type genes						
Gene	Serotype	Identity	Template / HSP length	Contig	Position in contig	Accession number
wzy	O104	99.64	1113 / 1113	NODE_122_length_26982_cov_17.237417	16386..17498	AF361371
wzx	O104	100	1278 / 1278	NODE_122_length_26982_cov_17.237417	18727..20004	KB021482

Database for H type genes						
Gene	Serotype	Identity	Template / HSP length	Contig	Position in contig	Accession number
fliC	H4	100	1050 / 1050	NODE_24_length_15621_cov_15.220664	9868..10917	AJ605764

extended output

Results as text | Results tsv | Hits in genome seqs | Serotype gene sequences

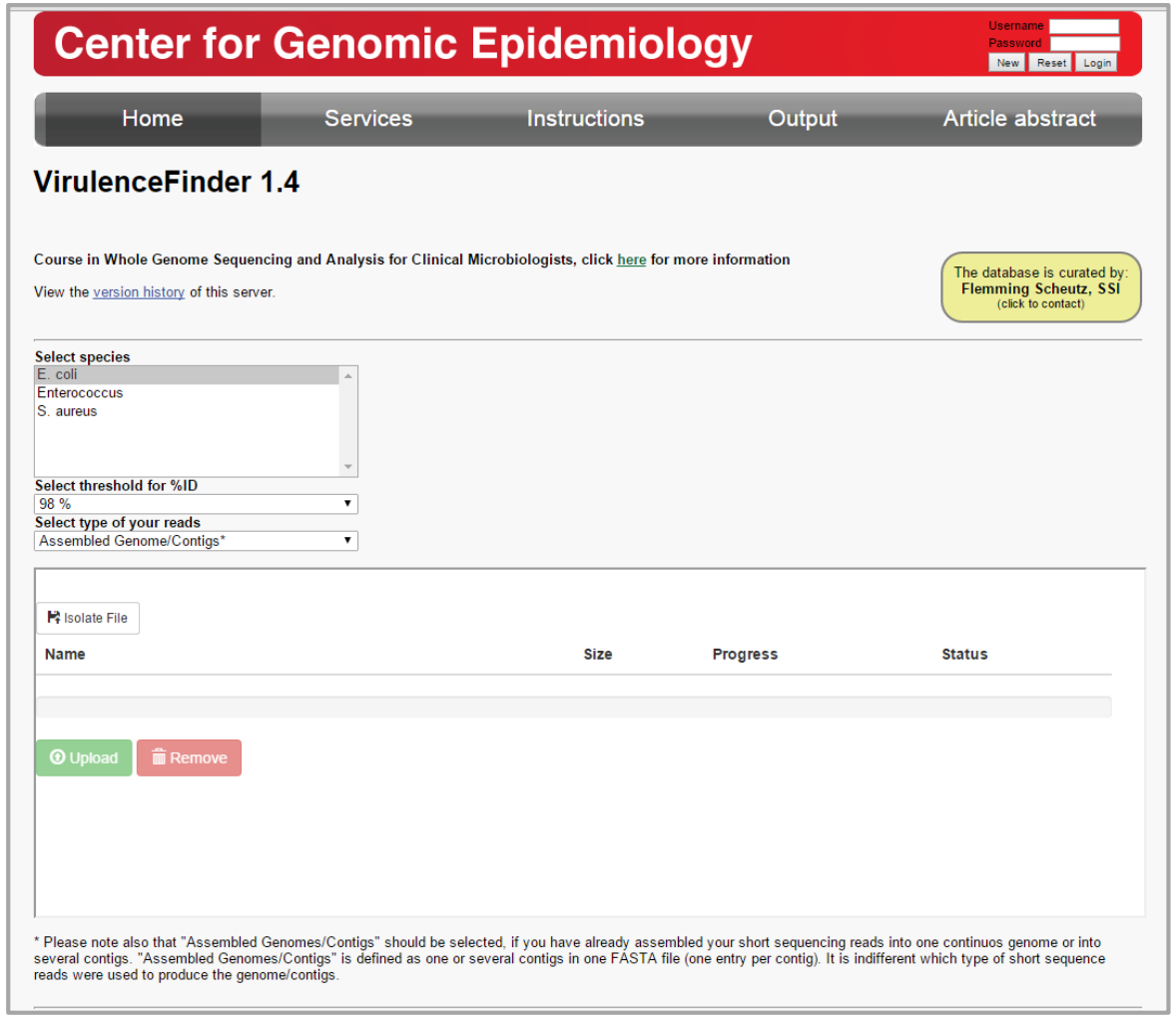
Selected %ID threshold: *85 %*

Selected minimum length: *60 %*

Input Files: *test.txt*

VirulenceFinder

- Initially 76 virulence genes – and growing….



JCM
Journals.ASM.org

Real-Time Whole-Genome Sequencing for Routine Typing, Surveillance, and Outbreak Detection of Verotoxigenic *Escherichia coli*

Katrine Grimstrup Joensen,[a,b] Flemming Scheutz,[b] Ole Lund,[c] Henrik Hasman,[a] Rolf S. Kaas,[a,c] Eva M. Nielsen,[b] Frank M. Aarestrup[a]

National Food Institute, Division for Epidemiology and Microbial Genomics, Technical University of Denmark, Kongens Lyngby, Denmark[a]; Department of Microbiology and Infection Control, Statens Serum Institut, Copenhagen, Denmark[b]; Center for Biological Sequence Analysis, Department of System Biology, Technical University of Denmark, Kongens Lyngby, Denmark[c]

Fast and accurate identification and typing of pathogens are essential for effective surveillance and outbreak detection. The current routine procedure is based on a variety of techniques, making the procedure laborious, time-consuming, and expensive. With whole-genome sequencing (WGS) becoming cheaper, it has huge potential in both diagnostics and routine surveillance. The aim of this study was to perform a real-time evaluation of WGS for routine typing and surveillance of verocytotoxin-producing *Escherichia coli* (VTEC). In Denmark, the Statens Serum Institut (SSI) routinely receives all suspected VTEC isolates. During a 7-week period in the fall of 2012, all incoming isolates were concurrently subjected to WGS using IonTorrent PGM. Real-time bioinformatics analysis was performed using web-tools (www.genomicepidemiology.org) for species determination, multilocus sequence type (MLST) typing, and determination of phylogenetic relationship, and a specific VirulenceFinder for detection of *E. coli* virulence genes was developed as part of this study. In total, 46 suspected VTEC isolates were characterized in parallel during the study. VirulenceFinder proved successful in detecting virulence genes included in routine typing, explicitly verocytotoxin 1 (*vtx1*), verocytotoxin 2 (*vtx2*), and intimin (*eae*), and also detected additional virulence genes. VirulenceFinder is also a robust method for assigning verocytotoxin (*vtx*) subtypes. A real-time clustering of isolates in agreement with the epidemiology was established from WGS, enabling discrimination between sporadic and outbreak isolates. Overall, WGS typing produced results faster and at a lower cost than the current routine. Therefore, WGS typing is a superior alternative to conventional typing strategies. This approach may also be applied to typing and surveillance of other pathogens.

VirulenceFinder

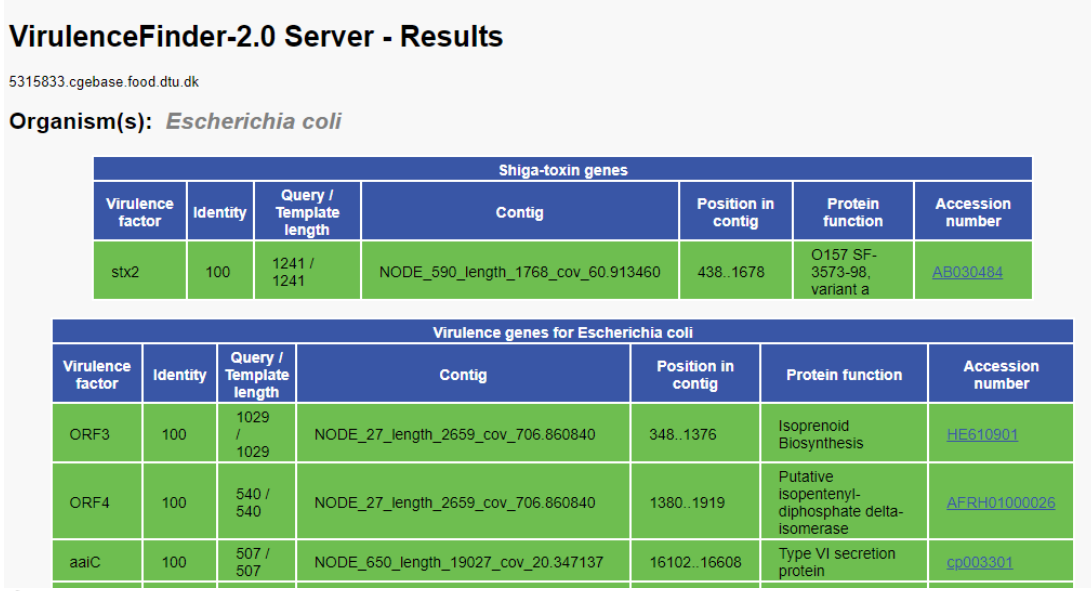
- **Detection of pathotype-specific genes**

- **STEC**
  - *eae - 45 variants*
  - *stx1* - 14A,18B variants
  - *stx2* - 114A,43B variants
  - *stx* subtypes
    - Important for HUS associated types

Gene	Description	Variants in database
astA	Heat-stable enterotoxin 1	11
bfpA	Major subunit of bundle-forming pili	5
cba	Colicin B	15
ccl	Cloacin	4
cdtB	Cytolethal distending toxin B	14
celb	Endonuclease colicin E2	10
cfa_c	Colonisation factor antigen I	4
cif	Type III secreted effector	4
cma	Colicin M	19
cnf1	Cytotoxic necrotizing factor	7
cofA	Longus type IV pilus subunit	1
eae	Intimin	45
eatA	SPATE	3
efa1	EHEC factor for adherence	11
ehxA	Enterohaemolysin	12
epeA	SPATE	1
espA	Type III secretions system	23
espB	Secreted protein B	14
espC	SPATE	3
espF	Type III secretion system	13
espI	SPATE	2
espJ	Prophage-encoded t3SS effector	2
espP	Putative exoprotein precursor	4
etpD	Type II secretion protein	3
f17-A	Subunit A of F17 fimbrial protein	7
f17-G	Adhesin subunit of F17 fimbriae	9
fanA	Involved in biogenesis of K99/F5 fimbriae	1
fasA	Fimbriae 987P/F6 subunit	1
fedA	Fimbrial protein F107 subunit A	3
fedF	Fimbrial adhesin AC precursor	6
fim41a	Mature Fim41a/F41 protein	2
gad	Glutamate decarboxylase	70
hlyE	Avian *E.coli* haemolysin	1
iha	Adherence protein	19
ipaD	Invasion protein *Shigella flexneri*	9
ipaH9.8	Invasion plasmid antigen	8
ireA	Siderophore receptor	4
iroN	Enterobactin siderophore receptor	13

VirulenceFinder results

- **Input**
  - Raw reads
  - Assemblies
- **Output**
  - Sequence identity, coverage and positions
  - Reference sequence
  - Examine alignments
  - Sort in relevance of genes



VirulenceFinder-2.0 Server - Results

5315833.cgebase.food.dtu.dk

Organism(s): *Escherichia coli*

Shiga-toxin genes						
Virulence factor	Identity	Query / Template length	Contig	Position in contig	Protein function	Accession number
stx2	100	1241 / 1241	NODE_590_length_1768_cov_60.913460	438..1678	O157 SF-3573-98, variant a	AB030484

Virulence genes for Escherichia coli						
Virulence factor	Identity	Query / Template length	Contig	Position in contig	Protein function	Accession number
ORF3	100	1029 / 1029	NODE_27_length_2659_cov_706.860840	348..1376	Isoprenoid Biosynthesis	HE610901
ORF4	100	540 / 540	NODE_27_length_2659_cov_706.860840	1380..1919	Putative isopentenyl-diphosphate delta-isomerase	AFRH01000026
aaiC	100	507 / 507	NODE_650_length_19027_cov_20.347137	16102..16608	Type VI secretion protein	cp003301

Issues to be aware of

- **Data quality**
- **Sequence bias**
  - Undetected/partly detected genes
- **Non-expressed genes**
  - Phenotype vs genotype problem
- **New variants and multiple variants of same gene**
- **Serotypes do not always correlate with phylogeny**
  - Combine with MLST
  - Enterobase is good for overview of serotypes and ST´s
- **Consider coverage, ID, and positions**
- **Much output for VirulenceFinder**
  - Sort and select relevant genes

Introduction to exercise

- ***Salmonella***
  - MLST
  - Serotype – Seqsero2
- ***E. coli***
  - Serotype – SerotypeFinder
  - Virulence - VirulenceFinder
- Note the results in sheets and discuss in group
- Any problems?

Results - *Salmonella*

Strain ID	ST	Species	Subspecies	Serotype	O antigen	H1 antigen	H2 antigen	Antigen profile	Comments
SRR27241771	50	enterica	enterica	Saintpaul	4	e,h	1,2	4:e,h:1,2	No notes
SRR27241772	198	enterica	enterica	Kentucky	8	i	z6	8:i:z6	No notes
SRR27944993	3226	enterica	enterica	Paratyphi C or Choleraesuis or Typhisuis	7	c	1,5	7:c:1,5	**The predicted serotypes share the same general formula: 7:c:1,5 and can be differentiated by additional analysis** Biochemical test necessary for differentiation – important to determine since Paratyphi C can cause invasive disease (enteric fewer) and Choleraesuis is host adapted to pigs and Typhisuis is host restricted to pigs
SRR27944994	86	enterica	enterica	Paratyphi B	4	b	1,2	4:b:1,2	**Detected the SNP in gene STM3356 that is associated with the d-tartrate nonfermenting phenotype characteristic of the typhoidal pathotype** Paratyphi B (d-tartrate negative) can cause enteric fewer and the Java variant (d-tartrate fermenting/positive) is less pathogenic and cause gastroenteritis

Results – *E. coli*

Strain ID	O	H	Serotype	eae genes	stx genes	stx subtype	Repeated genes	Comments
STEC-EQA-11-BB-10	O187	H28	O187:H28	none	stx2g-Out-S-8	stx2g	**gad, terC**, traT	traT – multiple variants in database and/or multiple copies of genes
STEC-EQA-11-BH-1	O80	H2	O80:H2	eae-e07-xi	stx2d-O55-5905 stx2d-OR-TS06-08	stx2d	afaB, cia, etsC, **gad**, iss, mchF, nleB, nleC, ompT, shiA, **terC**, traJ	**O80 is only detected in wzx – and with low coverage (812/1221) and ID% (96,43%).** Additional analysis needed to determine the O-type. afaB – bug (same reference sequence) cia, etsC, iss, mchF nleB, nleC, ompT, shiA - multiple variants in database and/or multiple copies of genes traJ – same gene with different reference sequences – clean up in database.
STEC-EQA-11-strain0017	O157	H7	O157:H7	eae-g01-gamma	stx1a-O157-FLY16, stx2c-O157-FLY16	stx1a, stx2c	astA, **gad**, nleB, tccP, **terC**	astA, nleB, tccP - multiple variants in database and/or multiple copies of genes

gad and terC – found in almost all *E. coli* - disregard

Further reading and references

- ***Salmonella* serotyping and MLST**
  - White-Kauffmann-Le Minor serotype classification scheme White-Kauffmann-Le Minor Scheme (https://www.pasteur.fr/sites/default/files/veng_0.pdf)
  - SeqSero2 – http://denglab.info/SeqSero2, https://doi.org/10.1128/AEM.01746-19
  - SeqSero1 – http://www.denglab.info/SeqSero, https://doi.org/10.1128/jcm.00323-15
  - Sequence bias SeqSero - https://doi.org/10.1128/AEM.00614-20
  - MLST and eBG - https://doi.org/10.1371/journal.ppat.1002776
  - Enterobase - https://enterobase.warwick.ac.uk/
  - SISTR - https://doi.org/10.1371/journal.pone.0147101

Further reading and references

- ***E. coli* serotyping and virulence**
  - Pathotypes review - https://doi.org/10.1093%2Ffemsre%2Ffuac031
  - SerotypeFinder - https://cge.food.dtu.dk/services/SerotypeFinder/, https://journals.asm.org/doi/10.1128/jcm.00008-15
  - VirulenceFinder - https://cge.food.dtu.dk/services/VirulenceFinder/, https://doi.org/10.1128/jcm.03617-13, https://doi.org/10.1128/jcm.01269-20

Acknowledgements

The creation of this training material was commissioned by ECDC to Statens Serum Institut (SSI) and The Danish Technical University (DTU) with the direct involvement of Egle Kudirkiene (SSI), Sofie Holtsmark Nielsen (SSI), Katrine Grimstrup Joensen (SSI) and Pernille Gymoese (SSI).