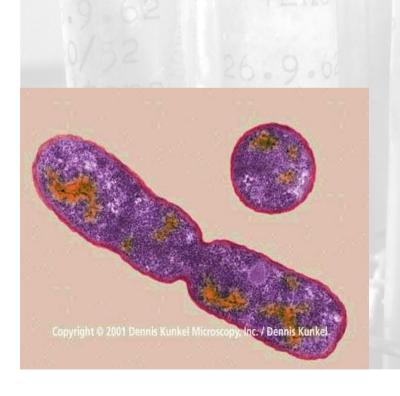


Whole genome sequencing (WGS) applications to VTEC surveillance and outbreak investigations

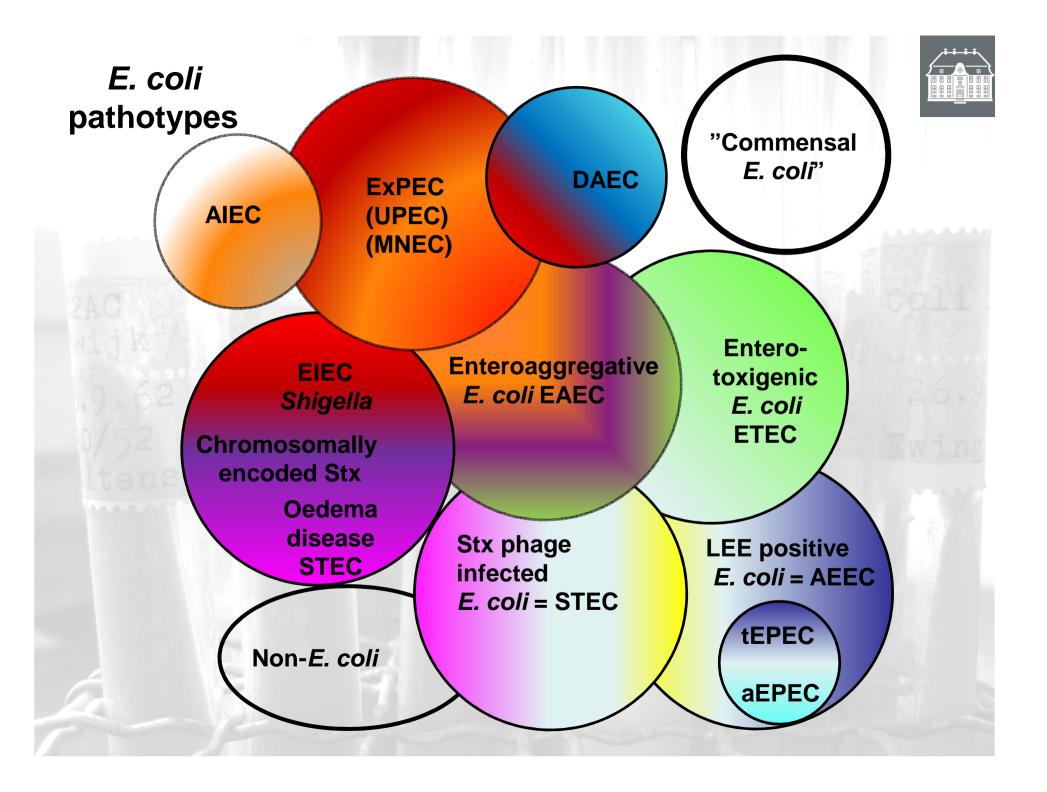




Katrine G. Joensen Ph.d student Flemming Scheutz

WHO Collaborating Centre for Reference and Research on Escherichia and Klebsiella

Foodborne Infections
Department for Microbiology
& Infection Control
STATENS SERUM INSTITUT



Real-Time Surveillance of **VTEC using WGS**





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

Katrine Grimstrup Joensen An Flemming Scheutz 2 Oke Lund 4 Henrik Hasman 4 Rolf S. Kaas 44 Eva M. Nielsen 2 Frank M. Aarestrup National Local Institute, Distant for Epidemiology and Microbial Generalics, Technical University of Denmark, Kongress Langby, Denmark's Department of Microbial Congress Cong and infection Control, Daterra Semen Institut, Copenhagen, Dinamedii (Control for Biological Sequence Analysis, Department of System Biology, Technical University of Denmark Songens Langby Denmark

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 Escherichta coli (VTEC). In Denmark, the Statens Serum Institut (SSI) routinely receives all suspected VTEC isolates. During a 7-week period in the full 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. colt strulence 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 strulence genes included in routine typing, explicitly serocytotoxin 1 (vex1), verocytotoxin 2 (vex2), and intimin (nar), and also detected additional virulence penes. VirulenceFinder is also a robust method for assigning verocytotoxin (vex) 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 pathosens.

Bacterial pathogens still pose a major threat to public health, matter skills. To achieve this, the Center for Genomic Epidemiological in order to limit their spread and provent infectious discounting only (CGE) provides public, user-friendly web-tools for rapid hanease outbreaks, accurate and rapid diagnostics and classification of isolates are of great importance. In current routine practice, for diagnostics, surveillance, and outbreak investigations for the isolation and identification are mostly performed at clinical microbiological laboratories, and verification and further charactertration are performed for a few selected pathogens at national, or regional, reference laboratories, using a variety of species-specific methods. Typing and surveillance of bacterial pathogens rely mainly on well-established, standardized phenotypic and molectalar typing methods, such as serotyping and ptalsed-field gel electrophoresis (PPGE) (1, 2). However, to obtain sufficient discrimtriation between isolates, it is typically necessary to combine typing results from several different typing techniques, both phenotypic and genetypic. As a result, it is laborious, time-consuming, and expensive to perform proper typing for surveillance and outbreak detection.

As the cost of whole-genome segmenting (WGS) has decreased and benchtop sequencing machines enable fast turnaround. It has become increasingly attractive for use in routine diagnostics and typing, and the approach has already been found useful in retrospective outbreak investigations (3, 4).

Although WGS provides detailed information that will, in theory, enable diagnostics and typing solely on the basis of the features in the bacterial genome, it is a challenge to define and extract the appropriate information from the large amount of sequence data that is generated. Thus, to facilitate the use of WGS data for routine diagnostics, typing, and surveillance, it is important that the sequence data can be automatically and quickly converted to clinically relevant information that can be easily interpreted by obwicters and proble health professionals with limited biotnior

dling of WGS data and extraction of relevant information, useful global medical society (www.genomicepidemiology.org),

In this study, as a proof of concept, we demonstrate the usefulness of WGS for routine typing, surveillance, and outbreak detection of verocytotoxin-producing Escherichia call (VTEC) infections in Denmark, VTEC, also known as Shipp toxin-producing E. cult (STEC), is a gastrointestinal pathogen, which is hypically spread by injestion of contaminated food or water or person-toperson contact. Rapid and reliable diagnostics and detection of outbreak clusters are of utmost importance for control. VTEC infections cause bloody diarrhea and in some cases hemolyticuremic syndrome (HUS), which is characterized by kidney fathure. thrombocytopenia, and microangiopathic hemolytic anemia, and can be fatal to young children. VI'BC pathogenicity is facili-

Replaced 3 Services 2014 Statuened for modification 5 Editors (2014) Accepted 181 streamy 2014

Published should of print 26 Telescory 2014

Address principle of the Committee Symposis Suppliffeed the di-Supplemental number for this article may be based at http://michalog/101136

Copyright & 2014, American Society for Informatickapy, All Rights Reserved.

The authors have publicative to allow immediate flow accounts this article









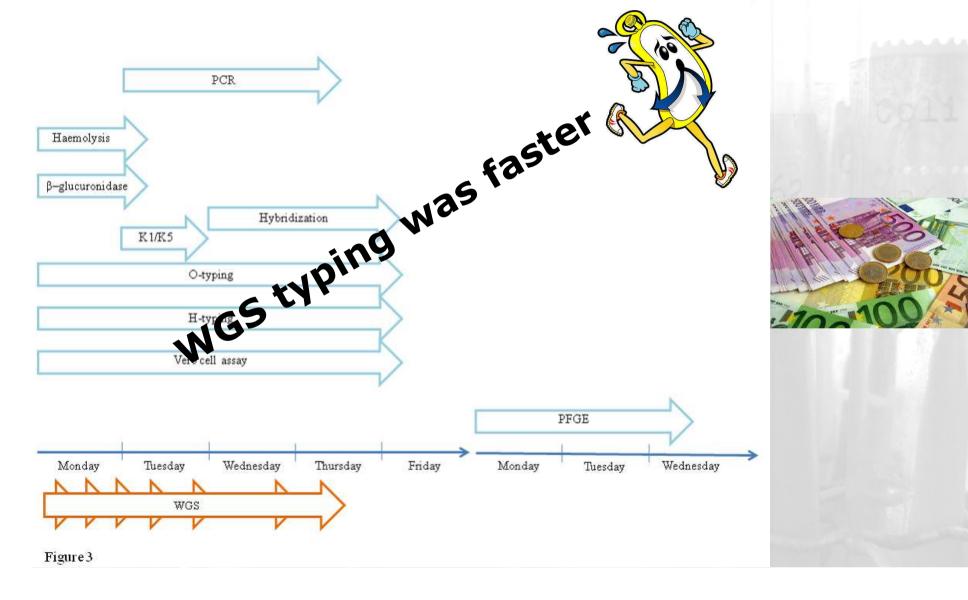






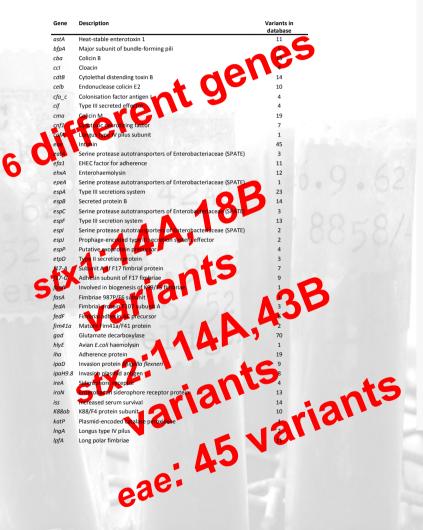
Real-Time Surveillance of VTEC using WGS

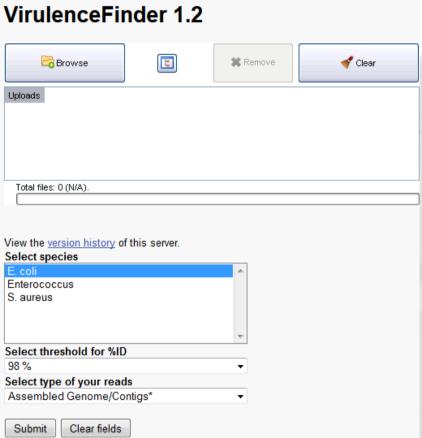












Virulence genes and vtx subtyping

SETTINGS:

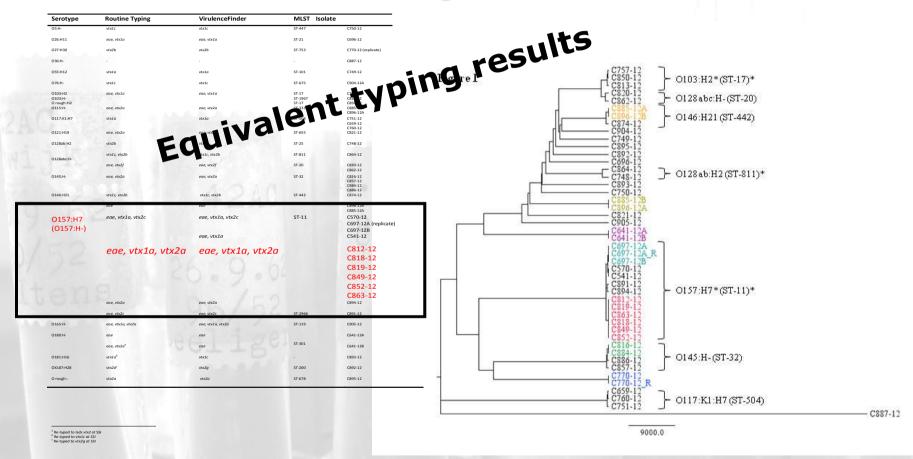
Selected %ID threshold: 85.00

Virulence - E. coli								
Virulence factor	%ldentity	Query/HSP length	Contig	Position in contig	Protein function	Accession number		
iroN	99.68	2178 / 2178	NODE_206_length_14143_cov_15.210423	1183814015	Enterobactin siderophore receptor protein	CP000034		
cnf1	99.97	3045 / 3045	NODE_395_length_28846_cov_16.705471	1764820692	Cytotoxic necrotizing factor	<u>CP002167</u>		
iss	97.28	294 / 294	NODE_325_length_6410_cov_14.389079	20552348	Increased serum survival	CP001509		
vat	100.00	4131 / 4131	NODE_52_length_14997_cov_16.533173	3594489	Vacuolating autotransporter toxin	CU651637		
stx2A	98.96	960 / 960	NODE_161_length_64579_cov_15.699051	5552656485	Shiga toxin 2, subunit A, variant b	<u>AJ313015</u>		
stx2B	96.93	270 / 261	NODE_161_length_64579_cov_15.699051	5649856758	Shiga toxin 2, subunit B	FN252458		
iss	100.00	342 / 342	NODE_259_length_12930_cov_15.107579	57806121	Increased serum survival	AE014075		
sfaS	100.00	492 / 492	NODE_65_length_8417_cov_17.987049	59296420	S-fimbriae minor subunit	CP000243		
prfB	99.89	882 / 882	NODE_168_length_118197_cov_16.083166	7635377234	P-related fimbriae regulatory gene	<u>CU651637</u>		

	stx - Holotoxin								
Virulence % Query/HSP Contig				Position in contig	Protein function	Accession number			
stx2	99.19	1236 / 1236	NODE_161_length_64579_cov_15.699051	5552656761	O22 3143-97, variant b	<u>AJ313015</u>			



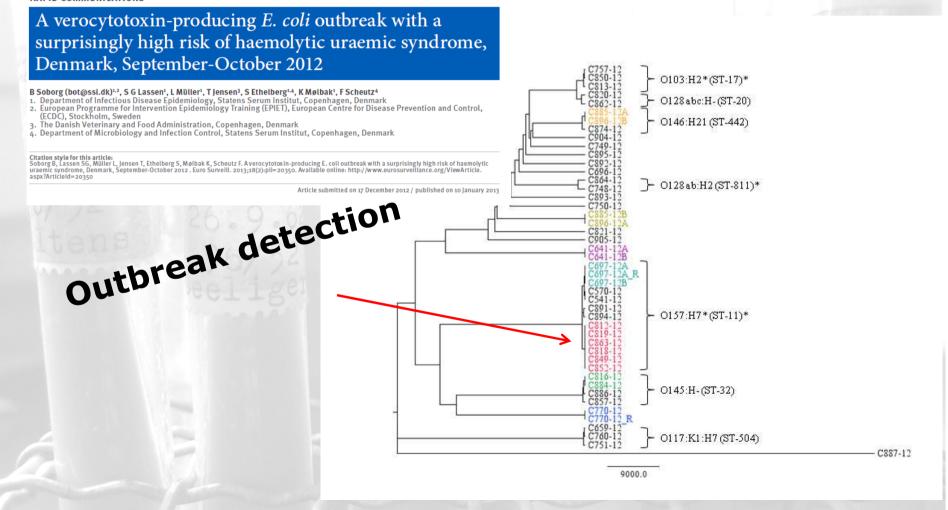
Real-Time Surveillance of VTEC using WGS





Real-Time Surveillance of VTEC using WGS

RAPID COMMUNICATIONS





Center for Genomic Epidemiology



Home Services Datasets

Overview of Services

Phenotyping

ResFinder
PathogenFinder
VirulenceFinder
Restriction-ModificationFinder

Typing

MLST
pMLST *
PlasmidFinder *
KmerFinder
SpeciesFinder *
Reads2Type
Tapir (OBS! Only works for Firefox)
TaxonomyFinder *
SeroTypeFinder

Phylogeny

snpTree CSI Phylogeny NDtree

Other

PanFunPro MyDbFinder



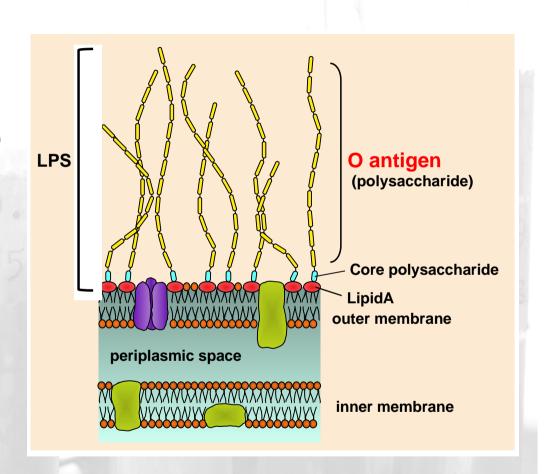
SerotypeFinder: O

- WZX and WZY

 (O-antigen flippase and polymerase)
- wzm and wzt (ABC transporter)

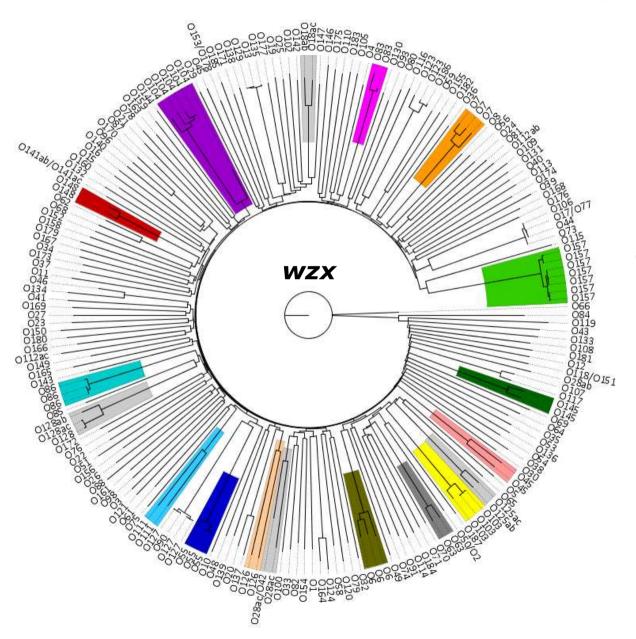


Anna Maria Malberg TetzschnerMaster student



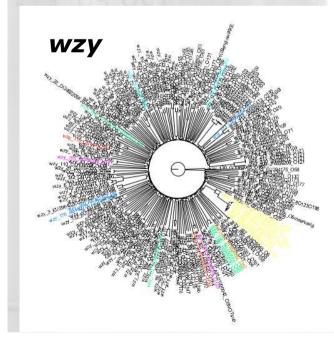
SerotypeFinder: O-Typing



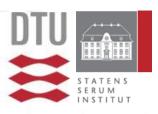


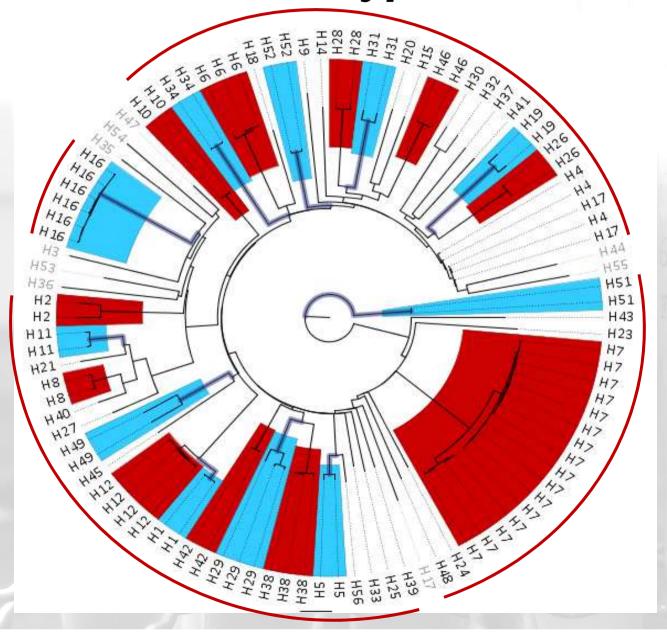
Two types are not in database: 057 014

Probably because of a deletion of these genes



SerotypeFinder: -H





All 53 H-types represented

fliC genes

O:H serotyping



Center for Genomic Epidemiology

Home Services Instructions Output Overview of genes Article abstract

SerotypeFinder-1.0 Server - Results

	H type							
Serotype gene %Identity Query/HSP Contig Position in Prediction Contig Serotype						Accession number		
fliC	99.94	1647 / 1647	NODE_2_length_16121_cov_14.234663	49796625	H6	AY249991		

	O type								
Serotype gene	%ldentity	Query/HSP length	Contig	Position in contig	Predicted serotype	Accession number			
WZX	98.67	975 / 1263	NODE_60_length_82759_cov_13.704238	1975	05002	EU549863			
wzy	97.93	723 / 963	NODE_64_length_2739_cov_9.434465	20912813	02	EU549863			

Predicted Serotype: O50O2:H6

extended output

Results as text Resul

Results tab separated

Hit in genome sequences

Serotype gene sequences

Selected %ID threshold: 85.00 %

Selected minimum length: 60 %





Number of genomes

2AC	for validation	with detected genes	with consistent WGS -and conventional results		
O-typing	601	569 (~95%)	560° (~98%)		
H-typing	509	508 (~100%)	503 (~100%)		

a) 12 predictions were ambiguous between the two O-processing genes:

O17/O44, O118/O151 (7), O164/O124, O134/O46, O90/O127, O162/O101

Co-existence of wzx/wzy and wzt/wzm (and fliC and non-fliC) genes as detected using the SerotypeFinder



wzx/wzy fliC	<i>wzm wzt</i> non- <i>fliC</i>	Serotype by phenotypic
wzx/wzyO128abc	wzm/wztO60	O20:K67:H17
wzyO138	wzm/wztO60	O138:K81:H48
wzx/wzyO20 fliC4	wzm/wztO60 fliC48	O 20:H 4
wzyO20 fliC4	wzm/wztO60 fliC48	O 20:H 4

Total of 45,947 strains (> 10)



The prevalence of the different O-groups for *E. coli* strains within the WHOCC *E. coli* database since 1963. Number of strains and rank is listed.

O-group	No. of strains	Rank	O-group	No. of strains	Rank	O-group	No. of strains	Rank
O26	3236	10	O149	251	42	019	105	83
O6	2520	2	O27	235	43	0104	104	84
O8	1871	3	0141	224	44	040	101	85
O2	1760	4	O125ac	218	45	O123	101	86
O55	1619	5	014	213	46	O89	93	87
O157	1452	6	0174	213	47	O112ab	92	88
O172	57	124	O168	36	152	3	0 40	
018,23	56	125	034	34	156	033,04	0 16	184
O116	56	126	O165	32	157	04,018		
O65	53	127	093	31	158	O23	16	184
O108	53	128	077,010			O50	16	184
O136	52	129	6	30	159	066	16	184
O178	52	130	O170	28	160	08,075	16	184
O18abe	50	131	O181	28	160	012.01	6 15	100
O38	50	132	O52	26	162	O60	15	190
O87	50	133	O164	26	162	O62	15	190
O130	50	134	OX186	26	162	018,06	8 14	193
O131	50	135	O61	25	165	042	14	193
O69	49	136	O144	25	165	05,070	14	195
OX182	49	137	O112ac	24	167	05,071		193
O28ac	48	138	O176	24	167	O18ac,0		
O80	48	139	OX187	23	169	23	13	197

Co-existence of *fliC* and *non-fliC* genes as detected using the *SerotypeFinder* in 23 strains

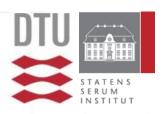


fliC	non- <i>fliC</i>	Serotype by	y phenotypic tests: All non-fliC
fliC2	flkA35	3 strains	
fliC2	flkA47	2 strains	
fliC4	fIIA44	2 strains	
fliC4	fIIA55	1 strain	
fliC11	flkA3	1 strain	
fliC11	flkA47	1 strain	* Drevieusly, described
fliC11	flmA54	1 strain	* Previously described combinations
fliC16	flkA53	1 strain	Combinations
fliC21	flkA36	1 strain	
fliC21	flkA47*	1 strain	
fliC21	flmA54*	1 strain	
fliC25	fIIA55	1 strain	
fliC27	flkA3	1 strain	
fliC27	flkA36	2 strains	
fliC27	flmA54	1 strain	
∠ fliC38	fIIA44*	1 strain	
fliC38	fIIA55*	1 strain	
fliC40	flkA53	1 strain	

EIEC H types H antigen



	ID	fliC gene found in
H18	96.6%	S. dysenteriae 1
H16	99.8%	S. sonnei
H45	97.9%	S. boydii 5
H14	96.7%	S. flexneri F2a



Conclusion

- In combination with other available
 WGS-typing tools, *E. coli* O:H serotyping can be performed solely from WGS-data
- Faster and cheaper than current routine typing procedures
- WGS-typing is already a superior alternative to conventional typing strategies.





http://vbs.psu.edu/events/EcoliWorkshop

Thank yous and acknowledgements!



The lab



Susanne Jespersen



Hansen



Pia Møller Christian Vråby Pedersen



Eva Møller Nielsen (The boss)



Katrine G. Joensen (The hard work)



Flemming Scheutz



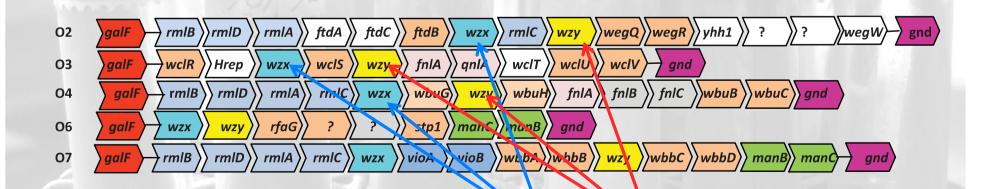
Anna Maria Malberg Tetzschner (more hard work)

SPECIAL THANKS TO: CDC, Atlanta, USA Atsushi Iguchi, Miyazaki **University, Japan**

O-antigen gene cluster



- Cluster on chromosome
 - between galF & gnd, (his & gnd for O8 & O9)
- Three gene classes:
 - Nucleotide sugar genes
 - Sugar transferase genes
 - O-unit processing genes: wzx, wzy, wzm & wzt



Virulence genes and vtx subtyping



stx2: WARNING, Identity: 99.19%, Query/HSP: 1236/1236, Contig name: NODE 9 length 64883 cov 15.699967, Position: 55830..57065 Virulence gene seq: atgaagtgtatattgttaaaatgggtactgtgcctgttactgggcttttcttcggtatcc Hit in genome: atqaaqtqtatattqttaaaatqq<mark>a</mark>tactqtqcctqtt<mark>q</mark>ctqqqcttttcttcqqtatco Virulence gene seg: tattcccgggaatttacgatagacttttcgactcaacaaagttatgtctcttcgttaaat tattcccgggaatttacgatagacttttcgactcaacaaagttatgtctcttcgttaaat Hit in genome: Virulence gene seg: agtatacggacagaatatcgactcctcttgagcatatatctcaggggaccacatcggtg Hit in genome: agtatacggacagaaatatcgactcctcttgagcatatatctcaggggaccacatcggto Virulence gene seg: tctgttattaaccacacccgggcagttattttgctgtgggatatacgagggcttgat Hit in genome: tctgttattaaccacccccaccgggcagttattttgctgtggatatacgagggcttgat Virulence gene seg: gtctatcaggcgcgttttgaccatcttcgtctgattattqagcaaaataatttatatgtc Hit in genome: gtctatcaggcgcgttttgaccatcttcgtctgatt<mark>g</mark>ttgagcaaaataatttatatgtg Virulence gene seg: gccqqatttqttaatacqqcaacaaatactttctacaqattttcaqattttqcacatata Hit in genome: gccggatt<mark>e</mark>gttaatacggcaacaaatactttctacagattttc<mark>t</mark>gattttgcacatata Virulence gene seg: tcaqtqcccqqtqtqacaactqtttccatqacaacqqacaqcaqttataccactctqcaa Hit in genome: tcagtgcccggtgtgacaactgtttccatgacaacggacagcagttataccactctgca Virulence gene seg: cgtgtcgcaqcgctggaacqttccggaatqcaaatcagtcgtcactcactggtttcatca Hit in genome: Virulence gene seg: tatctggcgttaatggagtttagtggaaatgccatgaccagagatgcatccagagcagtt Hit in genome: tatctqqcqttaatqqaqtttaqtqqaaatqccatqaccaqaqatqcatccaqaqcaqt Virulence gene seg: ctgcgttttgtcactgtcacagcagaagccttacggttcaggcaaatacagagagaattt Hit in genome: tgcgttttgt<mark>t</mark>actgtcacagcagaagccttacggttcaggcaaatacagagagaattt Virulence gene seg: cgtctggcactgtctgaaactgctcctgtttatacgatgacaccggaagaaatggacctc Hit in genome: cqtctqqcactqtctqaaactqctcctqtttatacqatqacaccqqaaqaa Virulence gene seq: acactgaactgggggaqaatcagcaatgtgcttccggagtttcggggagagggtggtgtc acactgaactggggggggaatcagcaatgtgcttccggagtttcggggagagggtggtgt Hit in genome: Virulence gene seq: agagtgggggaatatcctttaataatatatcagcgatactgggcacagtggcggttata Hit in genome: agagtggggggaatatcctttaataatatatcagcgatactgggcacagtggcggttata Virulence gene seg: Hit in genome: ctgaattgccatcatcagggggc<mark>a</mark>cgttccgttcgcgccgtgaatgaagagatacaacca Virulence gene seg: gaatgtcagataactggcgacaggccagttataaggataaacaatactttatgggaaagt Hit in genome: gaatgtcagataactggcgacaggccagttataaggataaacaatactttatgggaaagt



EAggECFinder

Center for Genomic Epidemiology

Home Services Instructions Output

VirulenceFinder-1.3 Server - Results

SETTINGS:

Selected %ID threshold: 85.00

	EAEC database								
Virulence factor	% Identity	Query/HSP length	Contig	Position in contig	Protein function	Accession number			
aaiC	99.71	684 / 684	NODE_251_length_69439_cov_17.582914	1711317796	Type VI secretion protein	<u>cp003301</u>			
capU	99.82	1089 / 1089	NODE_252_length_27346_cov_21.620201	50626150	Hexosyltransferase homolog	AFRH01000016			

extended output

Results as text Results tab separated

Hit in genome sequences

Virulence gene sequences

Input Files: PRI-HRD4-77-11_S12_L001_R1_001.fastq PRI-HRD4-77-11_S12_L001_R2_001.fastq

CONTIGS INFO

Technology: Illumina Paired Ends Reads

N50: 51806

DOWNLOAD CONTIGS

MLST typing



Center for Genomic Epidemiology

Home Services Instructions Output

MLST-1.7 Server - Typing Results

Sequence Type: ST-141

Locus	% Identity	HSP Length	Allele Length	Gaps	Allele
adk	100.00	536	536	0	adk-13
fumc	100.00	469	469	0	fumc-52
gyrb	100.00	460	460	0	gyrb-10
icd	100.00	518	518	0	icd-14
mdh	100.00	452	452	0	mdh-17
pura	100.00	478	478	0	pura-25
reca	100.00	510	510	0	reca-17

extended output

MLST Profile: ecoli

Organism: Escherichia coli#1

Input Files: C532-11 S1 L001 R1 001.fastq C532-11 S1 L001 R2 001.fastq

CONTIGS INFO

Technology: Illumina Paired Ends Reads

N50: 64374

DOWNLOAD CONTIGS