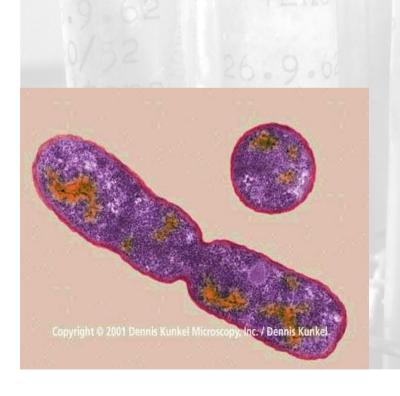# Whole genome sequencing (WGS) applications to VTEC surveillance and outbreak investigations
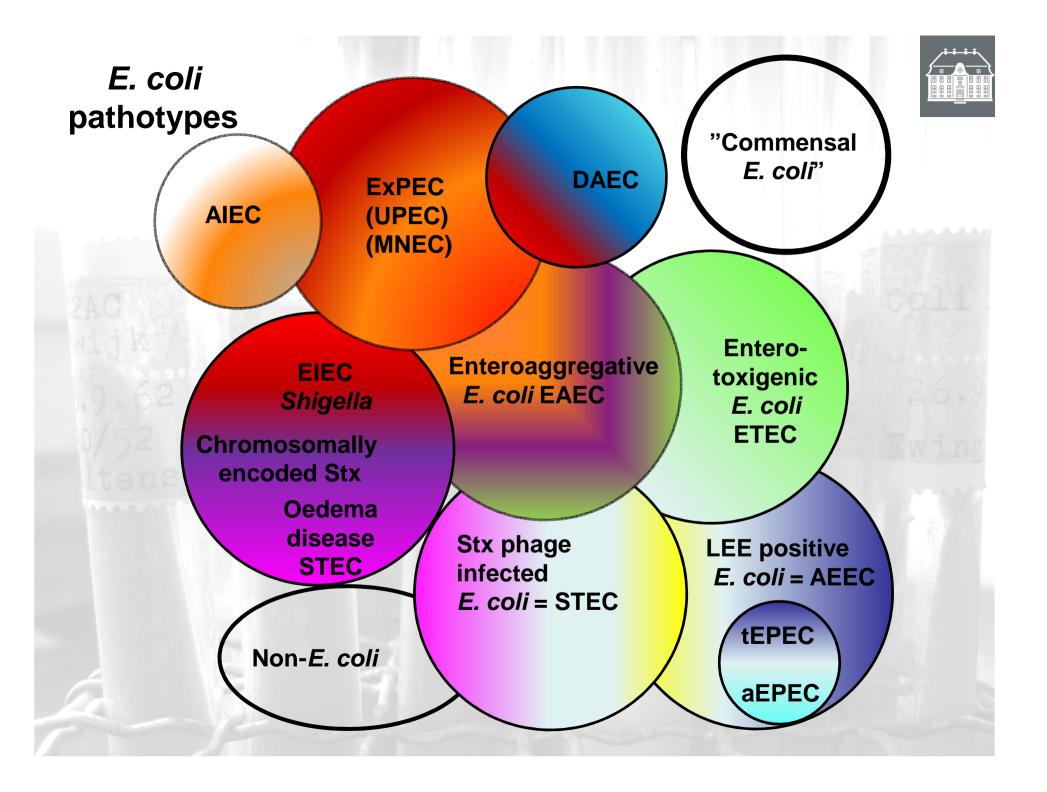
Katrine G. Joensen
Ph.d student

Flemming Scheutz
WHO Collaborating Centre for Reference and Research on *Escherichia* and *Klebsiella*

**Foodborne Infections
Department for Microbiology
& Infection Control
STATENS SERUM INSTITUT**

Copyright © 2001 Dennis Kunkel Microscopy, Inc. / Dennis Kunkel

E. coli pathotypes

AIEC

ExPEC (UPEC) (MNEC)

DAEC

"Commensal *E. coli*"

EIEC *Shigella*

Chromosomally encoded Stx

Oedema disease STEC

Enteroaggregative *E. coli* EAEC

Entero-toxigenic *E. coli* ETEC

Stx phage infected *E. coli* = STEC

LEE positive *E. coli* = AEEC

tEPEC

aEPEC

Non-*E. coli*

Real-Time Surveillance of VTEC using WGS

Real-Time Surveillance of VTEC using WGS

Figure 3

Real-Time Surveillance of VTEC using WGS

Gene	Description	Variants in database
astA	Heat-stable enterotoxin 1	11
bfpA	Major subunit of bundle-forming pili	
cba	Colicin B	
ccl	Cloacin	
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	Serine protease autotransporters of Enterobacteriaceae (SPATE)	3
efa1	EHEC factor for adherence	11
ehxA	Enterohaemolysin	12
epeA	Serine protease autotransporters of Enterobacteriaceae (SPATE)	1
espA	Type III secretions system	23
espB	Secreted protein B	14
espC	Serine protease autotransporters of Enterobacteriaceae (SPATE)	3
espF	Type III secretion system	13
espI	Serine protease autotransporters of Enterobacteriaceae (SPATE)	2
espJ	Prophage-encoded type III secretion system 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	
fedA	Fimbrial protein F107 subunit A	3
fedF	Fimbrial adhesin AC precursor	
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	
ireA	Siderophore receptor	4
iroN	Enterobactin siderophore receptor protein	13
iss	Increased serum survival	14
K88ab	K88/F4 protein subunit	10
katP	Plasmid-encoded catalase peroxidase	1
lngA	Longus type IV pilus	
lpfA	Long polar fimbriae	

76 different genes

stx:114A,18B variants

stx2:114A,43B variants

eae: 45 variants

VirulenceFinder 1.2

Browse | | Remove | Clear

Uploads

Total files: 0 (N/A).

View the version history of this server.

Select species

E. coli
Enterococcus
S. aureus

Select threshold for %ID

98 %

Select type of your reads

Assembled Genome/Contigs*

Submit | Clear fields

Virulence genes and vtx *subtyping*

SETTINGS:
Selected %ID threshold: *85.00*
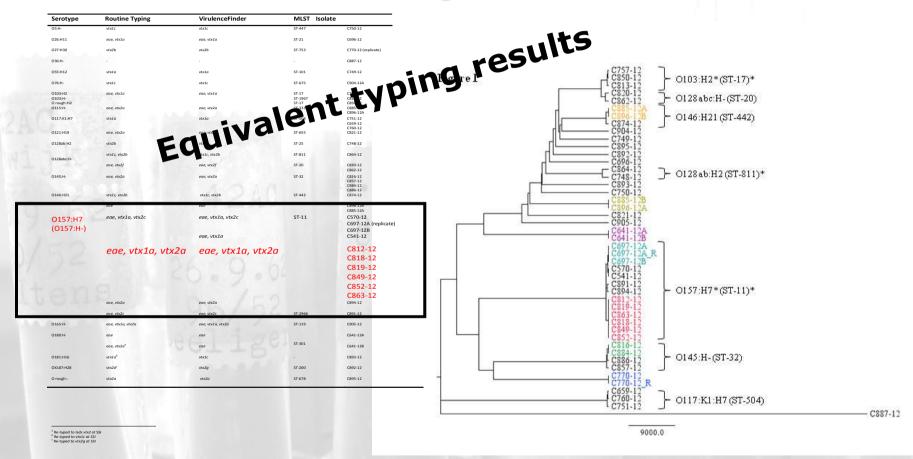
Virulence - E. coli						
Virulence factor	%Identity	Query/HSP length	Contig	Position in contig	Protein function	Accession number
iroN	99.68	2178 / 2178	NODE_206_length_14143_cov_15.210423	11838..14015	Enterobactin siderophore receptor protein	CP000034
cnf1	99.97	3045 / 3045	NODE_395_length_28846_cov_16.705471	17648..20692	Cytotoxic necrotizing factor	CP002167
iss	97.28	294 / 294	NODE_325_length_6410_cov_14.389079	2055..2348	Increased serum survival	CP001509
vat	100.00	4131 / 4131	NODE_52_length_14997_cov_16.533173	359..4489	Vacuolating autotransporter toxin	CU651637
stx2A	98.96	960 / 960	NODE_161_length_64579_cov_15.699051	55526..56485	Shiga toxin 2, subunit A, variant b	AJ313015
stx2B	96.93	270 / 261	NODE_161_length_64579_cov_15.699051	56498..56758	Shiga toxin 2, subunit B	FN252458
iss	100.00	342 / 342	NODE_259_length_12930_cov_15.107579	5780..6121	Increased serum survival	AE014075
sfaS	100.00	492 / 492	NODE_65_length_8417_cov_17.987049	5929..6420	S-fimbriae minor subunit	CP000243
prfB	99.89	882 / 882	NODE_168_length_118197_cov_16.083166	76353..77234	P-related fimbriae regulatory gene	CU651637

stx - Holotoxin						
Virulence factor	% Identity	Query/HSP length	Contig	Position in contig	Protein function	Accession number
stx2	99.19	1236 / 1236	NODE_161_length_64579_cov_15.699051	55526..56761	O22 3143-97, variant b	AJ313015

Real-Time Surveillance of VTEC using WGS

Equivalent typing results

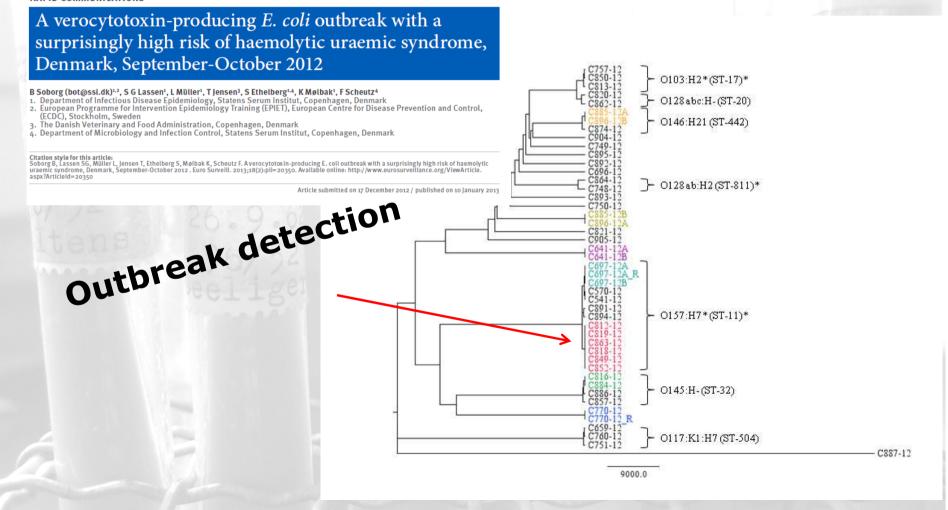| Serotype	Routine Typing	VirulenceFinder	MLST	Isolate
O5:H-	vtx1c	vtx1c	ST-447	C750-12
O26:H11	eae, vtx1a	eae, vtx1a	ST-21	C696-12
O27:H30	vtx2b	vtx2b	ST-753	C770-12 (replicate)
O36:H-	-	-	-	C887-12
O55:H12	vtx1a	vtx1a	ST-101	C749-12
O76:H-	vtx1c	vtx1c	ST-675	C904-12A
O103:H2	eae, vtx1a	eae, vtx1a	ST-17	
O103:H-			ST-1967	
O rough:H2			ST-17	
O115:H-	eae, vtx2a	eae, vtx2a		C896-12A
O117:K1:H7	vtx1a	vtx1a		C751-12
				C659-12
				C760-12
O121:H19	eae, vtx2a		ST-655	C821-12
O128ab:H2	vtx2b		ST-25	C748-12
O128abc:H-	vtx1c, vtx2b	vtx1c, vtx2b	ST-811	C864-12
	eae, vtx2f	eae, vtx2f	ST-20	C820-12
				C862-12
O145:H-	eae, vtx2a	eae, vtx2a	ST-32	C816-12
				C857-12
				C884-12
				C886-12
O146:H21	vtx1c, vtx2b	vtx1c, vtx2b	ST-442	C874-12
	eae	eae		C896-12B
				C885-12A
O157:H7	eae, vtx1a, vtx2c	eae, vtx1a, vtx2c	ST-11	C570-12
(O157:H-)				C697-12A (replicate)
		eae, vtx1a		C697-12B
				C541-12
	eae, vtx1a, vtx2a	**eae, vtx1a, vtx2a**		**C812-12**
				C818-12
				C819-12
				C849-12
				C852-12
				C863-12
	eae, vtx2a	eae, vtx2a		C894-12
	eae, vtx2c	eae, vtx2c	ST-2966	C891-12
O165:H-	eae, vtx1a, vtx2a	eae, vtx1a, vtx2a	ST-119	C905-12
O180:H-	eae	eae		C641-12A
	eae, vtx2a[c]	eae	ST-301	C641-12B
O181:H16	vtx1a[b]	vtx1c	-	C893-12
OX187:H28	vtx2d[c]	vtx2g	ST-200	C892-12
O rough:-	vtx2a	vtx2a	ST-678	C895-12

[a] Re-typed to lack vtx2 at SSI
[b] Re-typed to vtx1c at SSI
[c] Re-typed to vtx2g at SSI

Real-Time Surveillance of VTEC using WGS

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
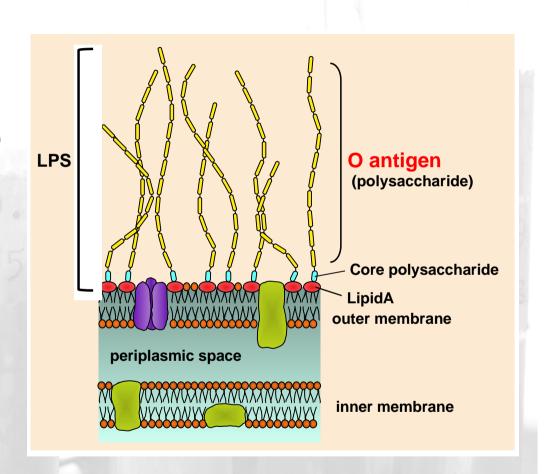CSI Phylogeny
NDtree

Other

PanFunPro
MyDbFinder

Support   Technical problems

SerotypeFinder: O

- ## *wzx* and *wzy*
  (O-antigen flippase and polymerase)

- ## *wzm* and *wzt*
  (ABC transporter)
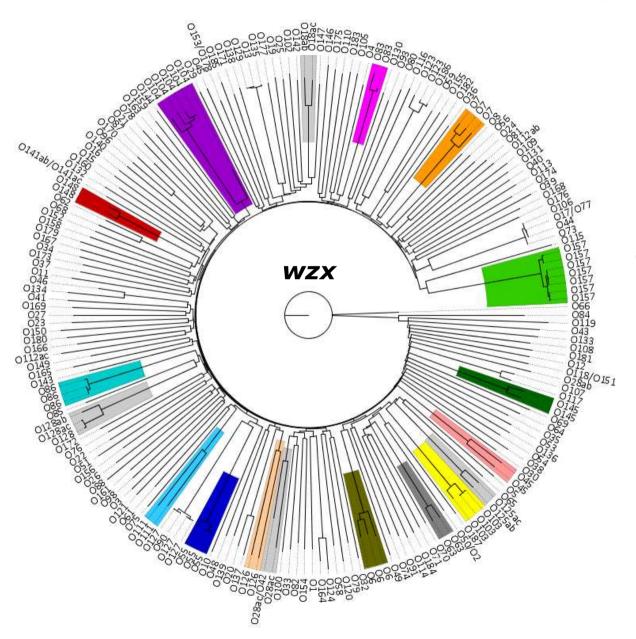
Anna Maria Malberg Tetzschner
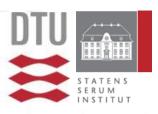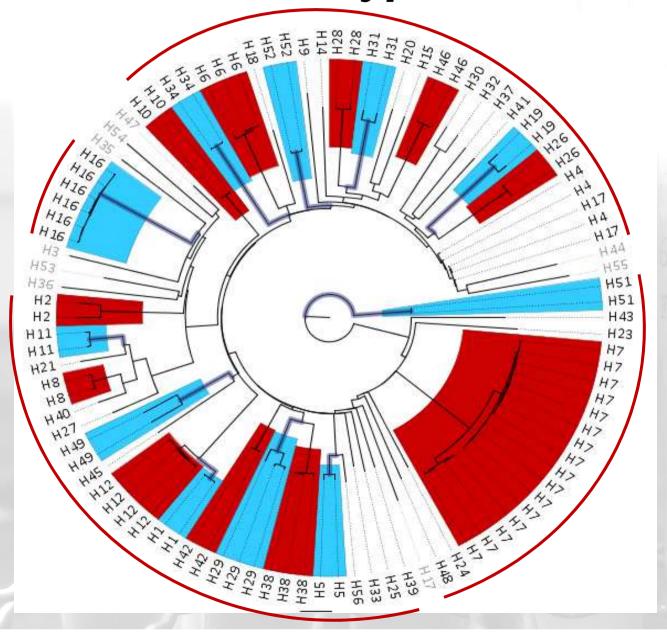Master student
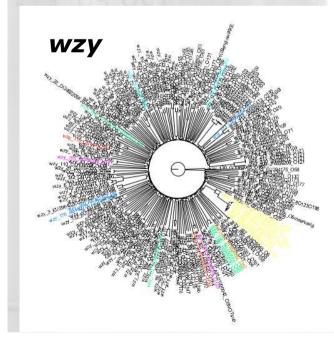
SerotypeFinder: O-Typing

Two types are not in database:
O57
O14

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 length	Contig	Position in contig	Predicted serotype	Accession number
fliC	99.94	1647 / 1647	NODE_2_length_16121_cov_14.234663	4979..6625	H6	AY249991

O type						
Serotype gene	%Identity	Query/HSP length	Contig	Position in contig	Predicted serotype	Accession number
wzx	98.67	975 / 1263	NODE_60_length_82759_cov_13.704238	1..975	O50O2	EU549863
wzy	97.93	723 / 963	NODE_64_length_2739_cov_9.434465	2091..2813	O2	EU549863

Predicted Serotype: O50O2:H6

extended output

Results as text | Results tab separated | Hit in genome sequences | Serotype gene sequences

Selected %ID threshold: *85.00 %*

Selected minimum length: *60 %*

SerotypeFinder:
Preliminary validation of O:H

Number of genomes

	for validation	with detected genes	with consistent WGS -and conventional results
O-typing	601	569 (~95%)	560[a] (~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*	*wzm*/*wzt* non-*fliC*	Serotype by phenotypic
wzx/*wzy*O128abc	*wzm*/*wzt*O60	O20:K67:H17
*wzy*O138	*wzm*/*wzt*O60	O138:K81:H48
wzx/*wzy*O20 *fliC*4	*wzm*/*wzt*O60 *fliC*48	O 20:H 4
*wzy*O20 *fliC*4	*wzm*/*wzt*O60 *fliC*48	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	1	O149	251	42	O19	105	83
O6	2520	2	O27	235	43	O104	104	84
O8	1871	3	O141	224	44	O40	101	85
O2	1760	4	O125ac	218	45	O123	101	86
O55	1619	5	O14	213	46	O89	93	87
O157	1452	6	O174	213	47	O112ab	92	88

O-group	No. of strains	Rank	O-group	No. of strains	Rank	O-group	No. of strains	Rank
O172	57	124	O168	36	152	3		
O18,23	56	125	O34	34	156	O33,O40	16	184
O116	56	126	O165	32	157	O4,O18,		
O65	53	127	O93	31	158	O23	16	184
O108	53	128	O77,O10			O50	16	184
O136	52	129	6	30	159	O66	16	184
O178	52	130	O170	28	160	O8,O75	16	184
O18abc	50	131	O181	28	160	O12,O16	15	190
O38	50	132	O52	26	162	O60	15	190
O87	50	133	O164	26	162	O62	15	190
O130	50	134	OX186	26	162	O18,O68	14	193
O131	50	135	O61	25	165	O42	14	193
O69	49	136	O144	25	165	O5,O70	14	193
OX182	49	137	O112ac	24	167	O5,O71	14	193
O28ac	48	138	O176	24	167	O18ac,O		
O80	48	139	OX187	23	169	23	13	197

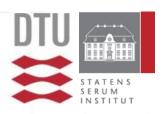# Co-existence of *fliC* and *non-fliC* genes as detected using the *SerotypeFinder* in 23 strains

fliC	non-*fliC*	Serotype by phenotypic tests: All non-*fliC*
fliC2	flkA35	3 strains
fliC2	flkA47	2 strains
fliC4	fllA44	2 strains
fliC4	fllA55	1 strain
fliC11	flkA3	1 strain
fliC11	flkA47	1 strain
fliC11	flmA54	1 strain
fliC16	flkA53	1 strain
fliC21	flkA36	1 strain
fliC21	flkA47*	1 strain
fliC21	flmA54*	1 strain
fliC25	fllA55	1 strain
fliC27	flkA3	1 strain
fliC27	flkA36	2 strains
fliC27	flmA54	1 strain
fliC38	fllA44*	1 strain
fliC38	fllA55*	1 strain
fliC40	flkA53	1 strain

* Previously described combinations

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   **Pia Møller Hansen**   **Christian Vråby Pedersen**

Eva Møller Nielsen (The boss)

Katrine G. Joensen (The hard work)

Flemming Scheutz
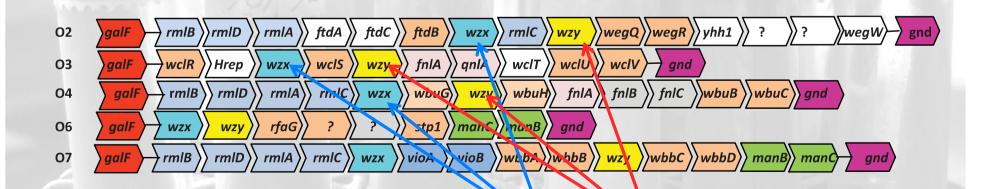
Anna Maria Malberg Tetzschner (more hard work)

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*

O2	galF	rmlB	rmlD	rmlA	ftdA	ftdC	ftdB	wzx	rmlC	wzy	wegQ	wegR	yhh1	?	?	wegW	gnd
O3	galF	wclR	Hrep	wzx	wclS	wzy	fnlA	qnlA	wclT	wclU	wclV	gnd
O4	galF	rmlB	rmlD	rmlA	rmlC	wzx	wbuG	wzy	wbuH	fnlA	fnlB	fnlC	wbuB	wbuC	gnd
O6	galF	wzx	wzy	rfaG	?	?	stp1	manC	manB	gnd
O7	galF	rmlB	rmlD	rmlA	rmlC	wzx	vioA	vioB	wbbA	wbbB	wzy	wbbC	wbbD	manB	manC	gnd

wzx   *wzy*

22

Virulence genes and vtx *subtyping*

EAggECFinder

MLST typing