



GenEpi-BioTrain Virtual training 23

Virulence Profiling of *E. coli* Pathotypes – Day 2

Dr. Jette Kjeldgaard (jetk@food.dtu.dk)
Technical University of Denmark (DTU), National Food Institute

<March 2026>

Intended Learning Objectives

Specific objectives of this session:

Learn about the biological diversity within the *E. coli* pathotypes (incl. *Shigella*)

Learn about the challenges in quality control of WGS *E. coli* data

Analyse and interpret the different sub-typing methods, including MLST, rMLST, and serotyping of *E. coli*.

Apply virulence profiling and typing using simple bioinformatics tools and suggest the resulting pathotype

Learn about real-life challenges in tracing emerging clones of pathogenic *E. coli*.

Outline

This session consists of the following elements

1. Discuss questions and tools used for QC and check for contamination
2. Discuss results of the exercise and interpretation of results
3. Learn about *E. coli* outbreaks

Agenda

AGENDA FOR SESSION 2 – Tuesday March 10, 2026

Agenda	Speakers	Hours
ECDC introduction and rules of conduct		09:00 – 09:10
1. Guided discussion using interactive polls: - Participants will engage in a structured discussion of their analysis results using live polls and questions. - Topics covered include: QC of isolates – discussion of species-specific parameters; typing and discriminatory power of different methods (species, serotyping, MLST), PCR targets etc., application of virulence databases (fastq or fasta; online or command line), interpretation of virulence profiles	Dr Jette Kjeldgaard and João Cardoso	09:10 – 10:20
Break		10:20 – 10:30
2. Comparative and critical analysis: - Facilitators will explore different analytical paths using VFDB and VirulenceFinder and how this may impact conclusions. - Live discussion of the advantages and limitations of typing methods and virulence profiling, incl. file input format (reads vs. fasta) and command line vs. user-friendly tools	Dr Jette Kjeldgaard and João Cardoso	10:30 – 10:50
3. Troubleshooting: - Discussion of common challenges encountered during the analysis, including ambiguities and tool limitations. - Strategies for dealing with atypical profiles and possible contamination	Dr Jette Kjeldgaard and João Cardoso	10:50 – 11:20
Break		11:20 -11.30
4. Reflections: - Real-life examples of WGS-based pathogen characterisation and tracing of outbreaks - Danish EIEC outbreak, 2021 - Atypical pathotypes? – German O104:H4 outbreak, 2011 - Open Q&A session for participants to raise questions and share reflections on integrating virulence profiling	Emily Dibba White (SSI, Denmark) Dr Jette Kjeldgaard All speakers	11:30 – 12:00 12.00 - 12:10 12:10-12:30

Exercise background

Your laboratory has received 18 bacterial isolates from a regional hospital following an increase in cases of acute gastroenteritis and bloody diarrhea. The hospital suspects that multiple *E. coli* pathotypes may be involved. Isolates have been sequenced, and the clinicians want clarification on if:

- the sequencing data is of sufficient quality, or if mixed cultures or contamination occurred during processing
- the isolates are *E. coli* or *Shigella* spp.
- they represent multiple pathotypes and which
- they carry (important) virulence factors
- any isolates require urgent public health notification (e.g., STEC, EIEC/Shigella)

Exercise tasks

Your task is to analyse the genomic data and report whether each isolate:

- has adequate sequence quality, or is contaminated or mixed
- is *E. coli* or *Shigella spp.*
- If presumptive EIEC or Shigella - which sequence type (MLST/rMLST)
- If presumptive EHEC/STEC - which predicted O:H serotype
- contains which virulence genes
- can be assigned to a pathotype (EHEC/STEC, EPEC, ETEC, EAEC, EIEC/Shigella, UPEC, ExPEC, etc.)

But first:

Sequence quality and contamination

Virulence factor overview

Batch upload to [Virulence&PlasmidDB](#) from FDA – for comparison

- Presence/absence of genes based on their database
- Prediction on plasmid/genome presence of gene

Column	Column	Column	Column	Column	Column	Column	Column	Column
Factor	Locations	VFS01	VFS02	VFS03	VFS04	VFS05	VFS06	VFS07
aafA	plasmid	0	0		0	0	0	0
aafB	plasmid	0	0		0	0	0	0
aafC	plasmid	0	0		0	0	0	0
aafD	plasmid	0	0		0	0	0	0
aaiC/hcp	genome	1	1		1	1	1	1
aap	plasmid	0	1		0	0	0	0
aar	genome	0	1		0	0	0	0
aatA	plasmid	0	1		0	0	0	0
afaC	genome	0	0		0	0	0	0
agg3A	plasmid	0	0		0	0	0	0
agg3B	plasmid	0	0		0	0	0	0
agg3C	plasmid	0	0		0	0	0	0
agg3D	plasmid	0	0		0	0	0	0
agg5A	genome	0	0		0	0	0	0



What did you overall think of the FDA virulence output?

Virulence factor overview

Batch upload to [Virulence&PlasmidDB](#) from FDA – for comparison

Limitations:

- Lists only gene name
 - no explanation in case of nomenclature/function ambiguity
- No contig/base position information to elucidate issues around # of variants
- No score on ID/coverage

Factor	Locations	VFS01	VFS02	VFS03	VFS04	VFS05	VFS06	VFS07
aafA	plasmid	0	0		0	0	0	0
aafB	plasmid	0	0		0	0	0	0
aafC	plasmid	0	0		0	0	0	0
aafD	plasmid	0	0		0	0	0	0
aaiC/hcp	genome	1	1		1	1	1	1
aap	plasmid	0	1		0	0	0	0
aar	genome	0	1		0	0	0	0
aatA	plasmid	0	1		0	0	0	0
afaC	genome	0	0		0	0	0	0
agg3A	plasmid	0	0		0	0	0	0
agg3B	plasmid	0	0		0	0	0	0
agg3C	plasmid	0	0		0	0	0	0
agg3D	plasmid	0	0		0	0	0	0
agg5A	genome	0	0		0	0	0	0



**Could you assign a (presumptive)
pathotype based on the FDA tool?**

Output table from FDA

Apply the key virulence markers on the FDA output

E. coli pathotype	Key virulence markers
EPEC	<i>eae</i> , <i>bfpA</i> , EAF
ETEC	<i>elt</i> (LT), <i>est</i> (ST), CF genes (<i>cfa</i> , <i>cs1–cs7</i>)
EIEC	<i>ipaH</i> , <i>ial</i> , <i>virF</i>
EHEC/STEC	<i>stx1</i> , <i>stx2</i> , <i>eae</i> , <i>ehxA</i> , <i>lpf</i>
EAEC	<i>aggR</i> , <i>aatA/aap</i> , <i>aaiC</i>
DAEC	<i>daa</i> , <i>afa/dra</i> genes
AIEC	No single marker; sometimes <i>lpfA</i> , <i>fimH</i> variants used

Factor	Locations	VFS01	VFS02	VFS03	VFS04	VFS05	VFS06	VFS07
<i>aafA</i>	plasmid	0	0		0	0	0	0
<i>aafB</i>	plasmid	0	0		0	0	0	0
<i>aafC</i>	plasmid	0	0		0	0	0	0
<i>aafD</i>	plasmid	0	0		0	0	0	0
<i>aaiC/hcp</i>	genome	1	1		1	1	1	1
<i>aar</i>	plasmid	0	1		0	0	0	0
<i>aatA</i>	plasmid	0	1		0	0	0	0
<i>afaC</i>	genome	0	0		0	0	0	0
<i>agg3A</i>	plasmid	0	0		0	0	0	0
<i>agg3B</i>	plasmid	0	0		0	0	0	0
<i>agg3C</i>	plasmid	0	0		0	0	0	0
<i>agg3D</i>	plasmid	0	0		0	0	0	0
<i>agg5A</i>	genome	0	0		0	0	0	0



Output table from FDA

Common pathotype markers - FDA Output		
Strain	EPEC	
	eae	bfpA
VFS01		
VFS02		
VFS03		
VFS04		
VFS05	eae	
VFS06		
VFS07		
VFS08		
VFS09	eae	
VFS10		
VFS11		
VFS12	eae	
VFS13		
VFS14		
VFS15		
VFS16		
VFS17		

E. coli pathotype	Key virulence markers
EPEC	<i>eae</i> , <i>bfpA</i> , EAF
ETEC	<i>elt</i> (LT), <i>est</i> (ST), CF genes (<i>cfa</i> , <i>cs1–cs7</i>)
EIEC	<i>ipaH</i> , <i>ial</i> , <i>virF</i>
EHEC/STEC	<i>stx1</i> , <i>stx2</i> , <i>eae</i> , <i>ehxA</i> , <i>lpf</i>
EAEC	<i>aggR</i> , <i>aatA/aap</i> , <i>aaiC</i>
DAEC	<i>daa</i> , <i>afa/dra</i> genes
AIEC	No single marker; sometimes <i>lpfA</i> , <i>fimH</i> variants used



EPEC:

Indication that some strains could be EPEC, but intimin (*eae*) is found in several pathotypes (EPEC/EHEC-STEC)

ETEC:

The nomenclature of the Heat-Stable/Labile genes in the output is not clear!?

Virulence&PlasmidDB database genes?

<https://virulence.fda.gov/>



Virulence&PlasmidDB

Salmonella Virulence Plasmid Transfer Factors **Ecoli Virulence** Tools Contact

Show 10 entries

Search: heat

ID	Virulence Gene	Locus Tag	Note	Product	DB_Xref	Location
5223	hra1	None	heat resistant agglutinin 1	Hra1	None	genome
5139	stb	None	None	heat stable enterotoxin II (S Tb)	HSSP:P22542	genome
5138	sta1	None	heat-stable enterotoxin I	None	None	genome
5021	astA	None	EAST1	heat-stable enterotoxin I	None	genome

Show 10 entries

Search: labile

1

Next

ID	Virulence Gene	Locus Tag	Note	Product	DB_Xref	Location
----	----------------	-----------	------	---------	---------	----------

No matching records found

Showing 0 to 0 of 0 entries (filtered from 235 total entries)

Previous Next

Ecoli Virulence Tools

Virulence Genes and Products

Heat-Stable/Heat-Labile toxins – gene names?

NCBI (and FDA): sta1, stb /ltcA

VFDB: estIa/eltA

Ge...	Product name
astA	heat-stable enterotoxin EAST1
astA	heat-stable enterotoxin EAST1
astA	heat-stable enterotoxin EAST1
astA	heat-stable enterotoxin EAST1
astA	heat-stable enterotoxin EAST1
stb	heat-stable enterotoxin ST-II
stb	heat-stable enterotoxin ST-II
stb	heat-stable enterotoxin ST-II
ystC	ST-I family heat-stable enterotoxin YstC
sta1	heat-stable enterotoxin ST-I group a
ystA	ST-I family heat-stable enterotoxin YstA
ystB	ST-I family heat-stable enterotoxin YstB
sta1	heat-stable enterotoxin ST-I group a

VF0210	eltA	Heat-labile toxin (LT)
VF0211	estla	Heat-stable toxin (ST)

VirulenceFinder: estb-STb1/eltIAB-4

Virulence genes for Escherichia coli					
Virulence factor	Identity	Query / Template length	Contig	Position in contig	Protein function
astA	100	117 / 117	SRR8901388_00148 len=6501 cov=13.2	5991..6107	Heat-stable enterotoxin EAST-1
eltIAB-4	100	1148 / 1148	SRR8901388_00148 len=6501 cov=13.2	913..2060	Heat-labile enterotoxin LTlh-4
estb-STb1	99.53	214 / 216	SRR8901388_00122 len=8747 cov=19.9	8503..8716	Heat-stable enterotoxin STb1 porcine variant

astA: EAST-1 Enteroaggregative *E. coli* heat stable enterotoxin

Output table from FDA

Common pathotype markers - FDA Output								
Strain	EPEC		ETEC			EIEC		
	eae	bfpA	elt	est (sta1/stb)	CFs	ipaH	ial	virF
VFS01						ipaH3/ipaH9.8		
VFS02								
VFS03								
VFS04						ipaH3/ipaH9.8		
VFS05	eae							
VFS06						ipaH3/ipaH9.8		
VFS07				stb				
VFS08						ipaH3/ipaH9.8		
VFS09	eae							
VFS10								
VFS11						ipaH3/ipaH9.8		
VFS12	eae							
VFS13								
VFS14						ipaH3/ipaH9.8		
VFS15						ipaH3/ipaH9.8		
VFS16								
VFS17								

E. coli pathotype	Key virulence markers
EPEC	<i>eae</i> , <i>bfpA</i> , EAF
ETEC	<i>elt</i> (LT), <i>est</i> (ST), CF genes (<i>cfa</i> , <i>cs1–cs7</i>)
EIEC	<i>ipaH</i> , <i>ial</i> , <i>virF</i>
EHEC/STEC	<i>stx1</i> , <i>stx2</i> , <i>eae</i> , <i>ehxA</i> , <i>lpf</i>
EAEC	<i>aggR</i> , <i>aatA/aap</i> , <i>aaiC</i>
DAEC	<i>daa</i> , <i>afa/dra</i> genes
AIEC	No single marker; sometimes <i>lpfA</i> , <i>fimH</i> variants used



EPEC:

Indication that some strains could be EPEC, but intimin (*eae*) is found in several pathotypes (EPEC/EHEC-STEC)

ETEC:

Heat-Stable gene (*stb*) identified in one isolate

EIEC:

ipaH-gene variants found in seven isolates. No other markers found

Output table from FDA

Status?



E. coli pathotype	Key virulence markers
EPEC	<i>eae</i> , <i>bfpA</i> , EAF
ETEC	<i>elt</i> (LT), <i>est</i> (ST), CF genes (<i>cfa</i> , <i>cs1</i> – <i>cs7</i>)
EIEC	<i>ipaH</i> , <i>ial</i> , <i>virF</i>
EHEC/STEC	<i>stx1</i> , <i>stx2</i> , <i>eae</i> , <i>ehxA</i> , <i>lpf</i>
EAEC	<i>aggR</i> , <i>aatA/aap</i> , <i>aaiC</i>
DAEC	<i>daa</i> , <i>afa/dra</i> genes
AIEC	No single marker; sometimes <i>lpfA</i> , <i>fimH</i> variants used



Common pathotype markers - FDA Output								
Strain	EPEC		ETEC			EIEC		
	<i>eae</i>	<i>bfpA</i>	<i>elt</i>	<i>est</i> (<i>sta1/stb</i>)	CFs	<i>ipaH</i>	<i>ial</i>	<i>virF</i>
VFS01						<i>ipaH3/ipaH9.8</i>		
VFS02								
VFS03								
VFS04						<i>ipaH3/ipaH9.8</i>		
VFS05	<i>eae</i>							
VFS06						<i>ipaH3/ipaH9.8</i>		
VFS07				<i>stb</i>				
VFS08						<i>ipaH3/ipaH9.8</i>		
VFS09	<i>eae</i>							
VFS10								
VFS11						<i>ipaH3/ipaH9.8</i>		
VFS12	<i>eae</i>							
VFS13								
VFS14						<i>ipaH3/ipaH9.8</i>		
VFS15						<i>ipaH3/ipaH9.8</i>		
VFS16								
VFS17								

Strain	EPEC	ETEC	EIEC
VFS01			X
VFS02			
VFS03			
VFS04			X
VFS05	X		
VFS06			X
VFS07		X	
VFS08			X
VFS09	X		
VFS10			
VFS11			X
VFS12	X		
VFS13			
VFS14			X
VFS15			X
VFS16			
VFS17			

Status:
Some presumptive
pathotypes,
Not excluding
possible
STEC/EHEC and
Shigella spp.

Output table from FDA

Resource file from VFDB
(excel list provided on Sciencedata)

Look for additional *Shigella*-specific virulence genes?

Common pathotype markers - FDA Output								
Strain	EPEC		ETEC			EIEC		
	eae	bfpA	elt	est (sta1/stb)	CFs	ipaH	ial	virF
VFS01						ipaH3/ipaH9.8		
VFS02								
VFS03								
VFS04						ipaH3/ipaH9.8		
VFS05	eae							
VFS06						ipaH3/ipaH9.8		
VFS07				stb				
VFS08						ipaH3/ipaH9.8		
VFS09	eae							
VFS10								
VFS11						ipaH3/ipaH9.8		
VFS12	eae							
VFS13								
VFS14						ipaH3/ipaH9.8		
VFS15						ipaH3/ipaH9.8		
VFS16								
VFS17								

	DsbA		<i>Shigella flexneri</i>
ipaA	TTSS secreted effectors		<i>Shigella flexneri</i>
virF	VirF		<i>Shigella flexneri</i>
shuS	Shu		<i>Shigella dysenteriae</i>
gspC	T2SS	Type II secret	<i>Shigella dysenteriae</i>
stxA	Shiga toxin		<i>Shigella dysenteriae</i>
iucA	Aerobactin		<i>Shigella flexneri</i>
icsA/virG	IcsA (VirG)	Ics for interce	<i>Shigella flexneri</i>
icsP/sopA	IcsP (SopA)		<i>Shigella flexneri</i>
gtrA	LPS		<i>Shigella flexneri</i>
pic	Pic	Protein involv	<i>Shigella flexneri</i>
set1B	ShET1	<i>Shigella enter</i>	<i>Shigella flexneri</i>
senB	ShET2	<i>Shigella enter</i>	<i>Shigella flexneri</i>
sigA	SigA	<i>Shigella IgA-li</i>	<i>Shigella flexneri</i>
ipaD	TTSS	Type III secre	<i>Shigella flexneri</i>
hcp1/tssD1	T6SS	Type VI secre	<i>Shigella sonnei</i>

Not a good strategy - It can be difficult to find good indicators,
Even in this small dataset, there are exceptions where this 'specific' list is not specific!

Look for additional *Shigella*-specific virulence genes?

Factor	Locations	VFS01	VFS04	VFS08	VFS14
aaiC/hcp	genome	1	1	1	1
capU	plasmid	1	1	1	1
csgD	genome	1	1	1	1
csgE	genome	1	1	1	1
csgF	genome	1	1	1	1
csgG	genome	1	1	1	1
entA	genome	1	1	1	1
espX1	genome	1	1	1	0
feoB	genome	1	1	1	1
feoC	genome	1	1	1	1
fhuA	genome	1	1	1	1
flgB	genome	1	1	1	1
gad	genome	1	1	1	1
gsp	genome	1	1	1	1
iap	genome	1	1	1	1
ipaD	plasmid	1	1	1	1
ipaH3	plasmid	1	1	1	1
ipaH9.8	genome	1	1	1	1
lpfA	genome	1	1	1	1
tcdA	genome	1	1	1	1
uspB	genome	1	1	1	1
yfcV	genome	1	1	1	1
yfeA	plasmid	1	1	1	1
yfeC	genome	1	1	1	1
yfeD	genome	1	1	1	1
iucA	plasmid	1	1	0	1
iucB	plasmid	1	1	0	1
iucC	plasmid	1	1	0	1

	DsbA		<i>Shigella flexneri</i>
ipaA	TTSS secreted effectors		<i>Shigella flexneri</i>
virF	VirF		<i>Shigella flexneri</i>
shuS	Shu		<i>Shigella dysenteriae</i>
gspC	T2SS	Type II secre	<i>Shigella dysenteriae</i>
stxA	Shiga toxin		<i>Shigella dysenteriae</i>
iucA	Aerobactin		<i>Shigella flexneri</i>
icsA/virG	IcsA (VirG)	Ics for interce	<i>Shigella flexneri</i>
icsP/sopA	IcsP (SopA)		<i>Shigella flexneri</i>
gtrA	LPS		<i>Shigella flexneri</i>
pic	Pic	Protein involv	<i>Shigella flexneri</i>
set1B	ShET1	<i>Shigella enter</i>	<i>Shigella flexneri</i>
senB	ShET2	<i>Shigella enter</i>	<i>Shigella flexneri</i>
sigA	SigA	<i>Shigella IgA-li</i>	<i>Shigella flexneri</i>
ipaD	TTSS	Type III secre	<i>Shigella flexneri</i>
hcp1/tssD1	T6SS	Type VI secre	<i>Shigella sonnei</i>

Output table from FDA

	Common pathotype markers - FDA Output					
	EPEC		STEC/EHEC			
Strain	eae	bfpA	stx	eae	ehxA	lpf
VFS01						
VFS02						
VFS03						
VFS04						
VFS05	eae		multiple	eae	ehxA	lpf1/2
VFS06						
VFS07						
VFS08						
VFS09	eae		multiple	eae	ehxA	
VFS10						
VFS11						
VFS12	eae			eae		
VFS13						
VFS14						
VFS15						
VFS16						
VFS17						

E. coli pathotype	Key virulence markers
EPEC	<i>eae</i> , <i>bfpA</i> , EAF
ETEC	<i>elt</i> (LT), <i>est</i> (ST), CF genes (<i>cfa</i> , <i>cs1–cs7</i>)
EIEC	<i>ipaH</i> , <i>ial</i> , <i>virF</i>
EHEC/STEC	<i>stx1</i> , <i>stx2</i> , <i>eae</i> , <i>ehxA</i> , <i>lpf</i>
EAEC	<i>aggR</i> , <i>aatA/aap</i> , <i>aaiC</i>
DAEC	<i>daa</i> , <i>afa/dra</i> genes
AIEC	No single marker; sometimes <i>lpfA</i> , <i>fimH</i> variants used



EHEC/STEC:

Two STEC strains, carrying multiple *stx* genes!?

Clear Filter From "Column1"

Filter by Color >

Text Filters >

stx
 ×

☒ (Select All Search Results)
☐ Add current selection to filter
☒ stx1dB
☒ stx2dA
☒ stx2dB
☒ stx2fA
☒ stx2fB
☒ stx2gA
☒ stx2gB
☒ stxA1a
☒ stxA1b
☒ stxA1c
☒ stxA1d
☒ stxA2a

OK
Cancel

	Column1	Column2	Column3	Column4	Column5	Column6	Column7	Column8	Column9	Column10	Column11	Column12
0												
0												
2	stx1dB	genome	0	0	0	0	0	0	0	0	1	0
3	stx2dA	genome	0	0	0	1	0	0	0	0	1	0
4	stx2dB	genome	0	0	0	1	0	0	0	0	1	0
5	stx2fA	genome	0	0	0	0	0	0	0	0	0	0
6	stx2fB	genome	0	0	0	1	0	0	0	0	1	0
7	stx2gA	genome	0	0	0	1	0	0	0	0	1	0
8	stx2gB	genome	0	0	0	1	0	0	0	0	1	0
9	stxA1a	genome	0	0	0	0	0	0	0	0	1	0
0	stxA1b	genome	0	0	0	0	0	0	0	0	1	0
1	stxA1c	genome	0	0	0	0	0	0	0	0	1	0
2	stxA1d	genome	0	0	0	0	0	0	0	0	1	0
3	stxA2a	genome	0	0	0	1	0	0	0	0	1	0
4	stxA2b	genome	0	0	0	1	0	0	0	0	1	0
5	stxA2c	genome	0	0	0	1	0	0	0	0	1	0
6	stxA2e	genome	0	0	0	1	0	0	0	0	1	0
7	stxB1a	genome	0	0	0	0	0	0	0	0	1	0
8	stxB1c	genome	0	0	0	0	0	0	0	0	1	0
9	stxB1d	genome	0	0	0	0	0	0	0	0	1	0
0	stxB2a	genome	0	0	0	1	0	0	0	0	1	0
1	stxB2b	genome	0	0	0	1	0	0	0	0	1	0
2	stxB2c	genome	0	0	0	1	0	0	0	0	1	0
3	stxB2dA	genome	0	0	0	1	0	0	0	0	1	0
4	stxB2dB	genome	0	0	0	1	0	0	0	0	1	0
5	stxB2eA	genome	0	0	0	1	0	0	0	0	1	0
6	stxB2eB	genome	0	0	0	1	0	0	0	0	1	0
7	stxB2fA	genome	0	0	0	0	0	0	0	0	0	0
8	stxB2fB	genome	0	0	0	1	0	0	0	0	1	0
9	stxB2gA	genome	0	0	0	1	1	0	0	0	1	0
0	stxB2gB	genome	0	0	0	1	1	0	0	0	1	0



Did you try other tools, which were able to differentiate the stx genes?

Output table from FDA

Common pathotype markers - FDA Output								
	STEC/EHEC				EAEC			
Strain	stx	eae	ehxA	lpf	aggR	aatA/aap	aaiC	astA
VFS01							aaiC/hcp	
VFS02					aggR	aatA/aap	aaiC/hcp	
VFS03								
VFS04							aaiC/hcp	
VFS05	multiple	eae	ehxA	lpf1/2			aaiC/hcp	astA
VFS06							aaiC/hcp	
VFS07							aaiC/hcp	astA
VFS08							aaiC/hcp	
VFS09	multiple	eae	ehxA		aggR	aatA/aap	aaiC/hcp	astA
VFS10								
VFS11							aaiC/hcp	
VFS12		eae					aaiC/hcp	
VFS13					aggA-D/R	aatA/aap	aaiC/hcp	
VFS14							aaiC/hcp	
VFS15							aaiC/hcp	
VFS16					aggR	aatA/aap	aaiC/hcp	
VFS17								

E. coli pathotype	Key virulence markers
EPEC	<i>eae</i> , <i>bfpA</i> , EAF
ETEC	<i>elt</i> (LT), <i>est</i> (ST), CF genes (<i>cfa</i> , <i>cs1-cs7</i>)
EIEC	<i>ipaH</i> , <i>ial</i> , <i>virF</i>
EHEC/STEC	<i>stx1</i> , <i>stx2</i> , <i>eae</i> , <i>ehxA</i> , <i>lpf</i>
EAEC	<i>aggR</i> , <i>aatA/aap</i> , <i>aaiC</i>
DAEC	<i>daa</i> , <i>afa/dra</i> genes
AIEC	No single marker; sometimes <i>lpfA</i> , <i>fimH</i> variants used



EHEC/STEC:

Two STEC strains, carrying multiple *stx* genes

EAEC:

Four strains carry markers for the EAEC type incl. the *aggR* regulon and dispersin (*aap/aat*).

The enterotoxin EAST-1 (*astA*) is present in one of these (plus the two STECs)

Some adhesins are found in all strains!

-> Not EAEC-specific

Output table from FDA



E. coli pathotype	Key virulence markers
EPEC	<i>eae</i> , <i>bfpA</i> , EAF
ETEC	<i>elt</i> (LT), <i>est</i> (ST), CF genes (<i>cfa</i> , <i>cs1</i> – <i>cs7</i>)
EIEC	<i>ipaH</i> , <i>ial</i> , <i>virF</i>
EHEC/STEC	<i>stx1</i> , <i>stx2</i> , <i>eae</i> , <i>ehxA</i> , <i>lpf</i>
EAEC	<i>aggR</i> , <i>aatA/aap</i> , <i>aaiC</i>
DAEC	<i>daa</i> , <i>afa/dra</i> genes
AIEC	No single marker; sometimes <i>lpfA</i> , <i>fimH</i> variants used

Common pathotype markers - FDA Output

	STEC/EHEC				EAEC				DAEC	
Strain	stx	eae	ehxA	lpf	aggR	aatA/aap	aaiC	astA	daa	afa/dra
VFS01							aaiC/hcp			
VFS02					aggR	aatA/aap	aaiC/hcp			
VFS03										
VFS04							aaiC/hcp			
VFS05	multiple	eae	ehxA	lpf1/2			aaiC/hcp	astA		
VFS06							aaiC/hcp			
VFS07							aaiC/hcp	astA		
VFS08							aaiC/hcp			
VFS09	multiple	eae	ehxA		aggR	aatA/aap	aaiC/hcp	astA		afaC
VFS10										
VFS11							aaiC/hcp			
VFS12		eae					aaiC/hcp			
VFS13					aggA-D/R	aatA/aap	aaiC/hcp			
VFS14							aaiC/hcp			
VFS15							aaiC/hcp			
VFS16					aggR	aatA/aap	aaiC/hcp			
VFS17										

EHEC/STEC:

Two STEC strains, carrying *stx* genes

EAEC:

Four strains carry markers for the EAEC type incl. the *aggR* regulon and dispersin (*aap/aat*)

DAEC:

One strain carry typical adhesin (*afa*)

Output table from FDA

Status?



	Common pathotype markers - FDA Output										
	STEC/EHEC				EAEC				DAEC		
Strain	stx	eae	ehxA	lpf	aggR	aatA/aap	aaiC	astA	daa	afa/dra	
VFS01							aaiC/hcp				
VFS02					aggR	aatA/aap	aaiC/hcp				
VFS03											
VFS04							aaiC/hcp				
VFS05	multiple	eae	ehxA	lpf1/2			aaiC/hcp	astA			
VFS06							aaiC/hcp				
VFS07							aaiC/hcp	astA			
VFS08							aaiC/hcp				
VFS09	multiple	eae	ehxA		aggR	aatA/aap	aaiC/hcp	astA		afaC	
VFS10											
VFS11							aaiC/hcp				
VFS12		eae					aaiC/hcp				
VFS13					aggA-D/R	aatA/aap	aaiC/hcp				
VFS14							aaiC/hcp				
VFS15							aaiC/hcp				
VFS16					aggR	aatA/aap	aaiC/hcp				
VFS17											

E. coli pathotype	Key virulence markers
EPEC	<i>eae</i> , <i>bfpA</i> , EAF
ETEC	<i>elt</i> (LT), <i>est</i> (ST), CF genes (<i>cfa</i> , <i>cs1</i> – <i>cs7</i>)
EIEC	<i>ipaH</i> , <i>ial</i> , <i>virF</i>
EHEC/STEC	<i>stx1</i> , <i>stx2</i> , <i>eae</i> , <i>ehxA</i> , <i>lpf</i>
EAEC	<i>aggR</i> , <i>aatA/aap</i> , <i>aaiC</i>
DAEC	<i>daa</i> , <i>afa/dra</i> genes
AIEC	No single marker; sometimes <i>lpfA</i> , <i>fimH</i> variants used



Strain	EPEC	ETEC	EIEC	STEC/EHEC	EAEC	DAEC
VFS01			X			
VFS02					X	
VFS03						
VFS04			X			
VFS05	X			X		
VFS06			X			
VFS07		X				
VFS08			X			
VFS09	X			X	X	X
VFS10						
VFS11			X			
VFS12	X					
VFS13					X	
VFS14			X			
VFS15			X			
VFS16					X	
VFS17						



Did you find any genomic tools able to differentiate between *E. coli* and *Shigella* spp.?



Did you find any genomic tools (not in the previous list), which could differentiate between *E. coli* and *Shigella* spp.?