E. coli virulence and pathotypes

Exercise - Interpretation of virulence gene results

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

DTU Food March 2026

‘Background’ of exercise

Background

Your laboratory has received 17 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. Several isolates from the same ward have been sequenced, and the clinicians want clarification on:

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

Data for the exercise

The lab has performed WGS on all 17 isolates (Illumina data).

You will be given:

Raw FASTQ files and assemblies (when available)

QC reports (FastQC)

Link to sciencedata: View it!

Here, you will additionally find:

VFDB resource file

Virulence&PlasmidDB results (FDA)

Tasks in 'homework'

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 - additional typing (MLST, rMLST, serotype)
- If presumptive EHEC/STEC - which predicted O:H serotype
- Contains important virulence genes
- Can be assigned to a pathotype (EHEC/STEC, EPEC, ETEC, EAEC, EIEC/Shigella, UPEC, ExPEC, etc.)

Suggested approach:

Start by evaluating the QC reports (FastQC) or use your own pipeline to assess the sequence quality.

Evaluate the sequence quality and determine if any isolates should be excluded from further analysis due to poor quality or contamination.

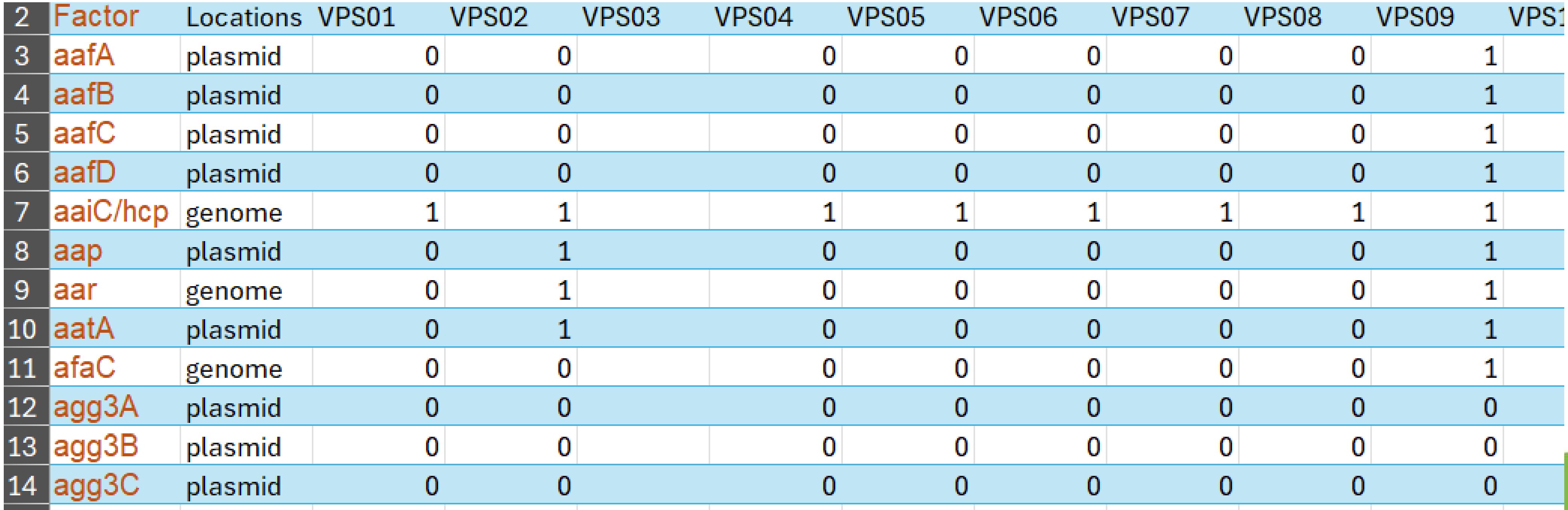
There are both raw fastq and fasta files available, as well as the QC reports. The fastq files can be used if participants want to do their own QC analysis but are otherwise not needed for the exercise.

Poor quality isolates should not be analysed further – but what is the cause?

Analyse virulence factors

Looking into virulence factors, the Virulence&PlasmidDB from FDA gives a good overview

- you can make a comparison with a batch upload
- we have provided an output on the exercise strains



2	Factor	Locations	VPS01	VPS02	VPS03	VPS04	VPS05	VPS06	VPS07	VPS08	VPS09	VPS1
3	aafA	plasmid	0	0		0	0	0	0	0	1	
4	aafB	plasmid	0	0		0	0	0	0	0	1	
5	aafC	plasmid	0	0		0	0	0	0	0	1	
6	aafD	plasmid	0	0		0	0	0	0	0	1	
7	aaiC/hcp	genome	1	1		1	1	1	1	1	1	
8	aap	plasmid	0	1		0	0	0	0	0	1	
9	aar	genome	0	1		0	0	0	0	0	1	
10	aatA	plasmid	0	1		0	0	0	0	0	1	
11	afaC	genome	0	0		0	0	0	0	0	1	
12	agg3A	plasmid	0	0		0	0	0	0	0	0	
13	agg3B	plasmid	0	0		0	0	0	0	0	0	
14	agg3C	plasmid	0	0		0	0	0	0	0	0	

Identification of notifiable strains

Group strains according to the FDA tool output; focus on the two groups **presumptive EIEC/Shigella** and **STEC/EHEC** first for further analysis

Perform serotyping, MLST and rMLST on EIEC/Shigella group to evaluate if this typing helps the **species interpretation**

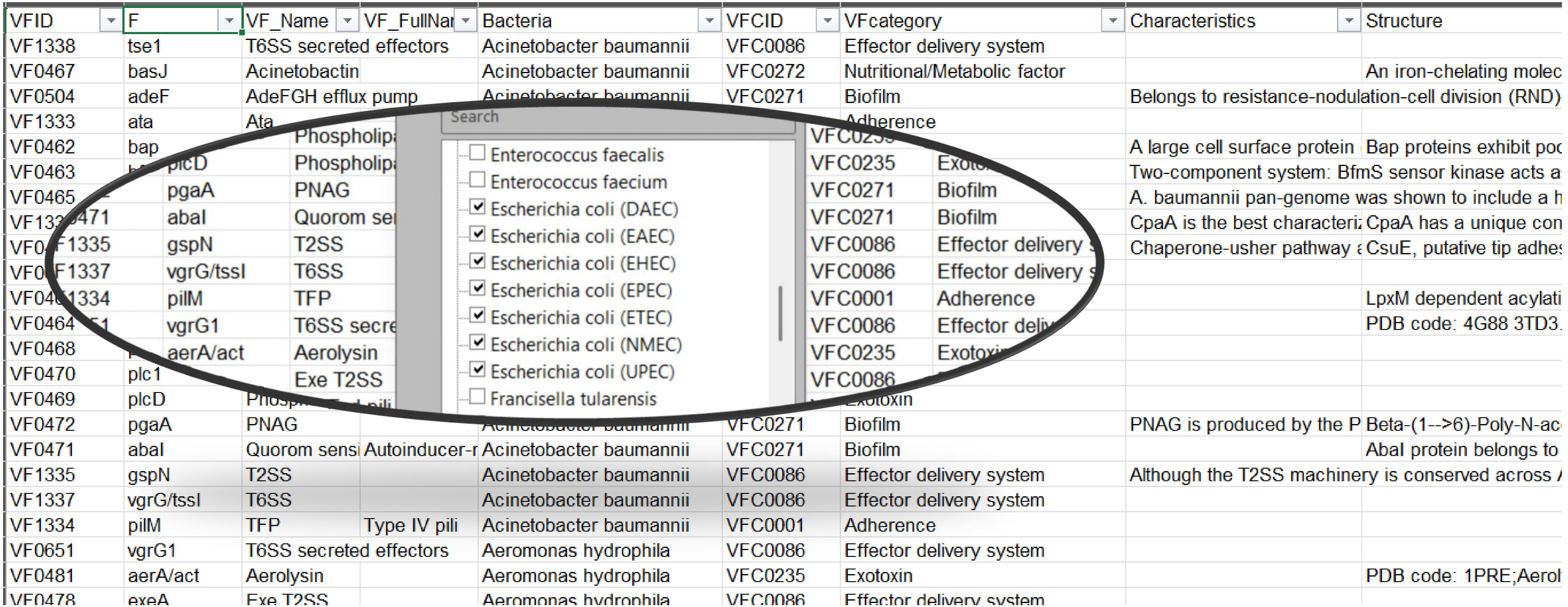
Perform serotyping of presumptive STEC/EHEC – are they common STEC serotypes?

Pathotyping of additional strains

Based on the analysis of virulence factors by FDA tool, VirulenceFinder or other tools, try to confirm the pathotypes of the remaining isolates

The VFDB table and comparisons can be used to confirm pathotypes?

VFDB outputs:



VFID	F	VF_Name	VF_FullNar	Bacteria	VFCID	VFcategory	Characteristics	Structure
VF1338	tse1	T6SS secreted effectors		Acinetobacter baumannii	VFC0086	Effector delivery system		
VF0467	basJ	Acinetobactin		Acinetobacter baumannii	VFC0272	Nutritional/Metabolic factor		An iron-chelating molec
VF0504	adeF	AdeFGH efflux pump		Acinetobacter baumannii	VFC0271	Biofilm	Belongs to resistance-nodulation-cell division (RND)	
VF1333	ata	Ata				Adherence		
VF0462	bap						A large cell surface protein	Bap proteins exhibit poo
VF0463							Two-component system: BfmS sensor kinase acts a	
VF0465							A. baumannii pan-genome was shown to include a h	
							CpaA is the best characteri	CpaA has a unique con
							Chaperone-usher pathway a	CsuE, putative tip adhes
								LpxM dependent acylati
VF0464								PDB code: 4G88 3TD3.
VF0468								
VF0470	plc1							
VF0469	plcD					Exotoxin		
VF0472	pgaA	PNAG			VFC0271	Biofilm	PNAG is produced by the P	Beta-(1-->6)-Poly-N-ac
VF0471	abaI	Quorom sensi	Autoinducer-r	Acinetobacter baumannii	VFC0271	Biofilm		AbaI protein belongs to
VF1335	gspN	T2SS		Acinetobacter baumannii	VFC0086	Effector delivery system	Although the T2SS machinery is conserved across	
VF1337	vgrG/tssI	T6SS		Acinetobacter baumannii	VFC0086	Effector delivery system		
VF1334	pilM	TFP	Type IV pili	Acinetobacter baumannii	VFC0001	Adherence		
VF0651	vgrG1	T6SS secreted effectors		Aeromonas hydrophila	VFC0086	Effector delivery system		
VF0481	aerA/act	Aerolysin		Aeromonas hydrophila	VFC0235	Exotoxin		PDB code: 1PRE;Aerol
VF0478	exeA	Exe T2SS		Aeromonas hydrophila	VFC0086	Effector delivery system		

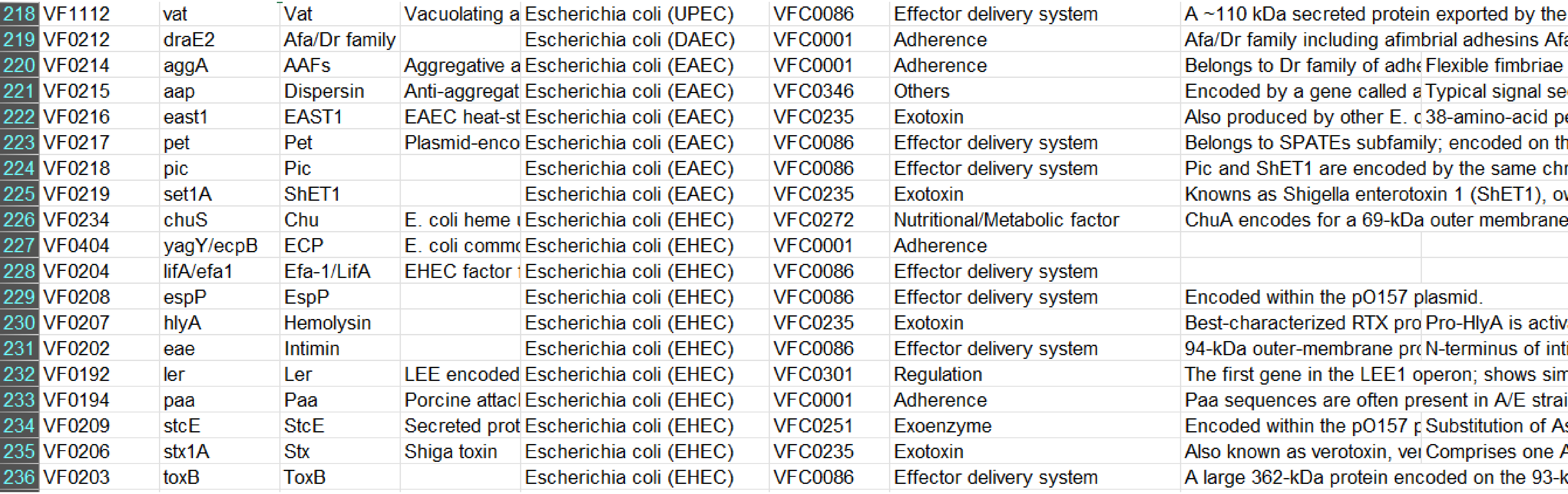
Search

- ☐ Enterococcus faecalis
- ☐ Enterococcus faecium
- ☑ Escherichia coli (DAEC)
- ☑ Escherichia coli (EAEC)
- ☑ Escherichia coli (EHEC)
- ☑ Escherichia coli (EPEC)
- ☑ Escherichia coli (ETEC)
- ☑ Escherichia coli (NMEC)
- ☑ Escherichia coli (UPEC)
- ☐ Francisella tularensis

Sort on *E. coli* and *Shigella*

VFDB output

Will it help in interpretation of pathotype?



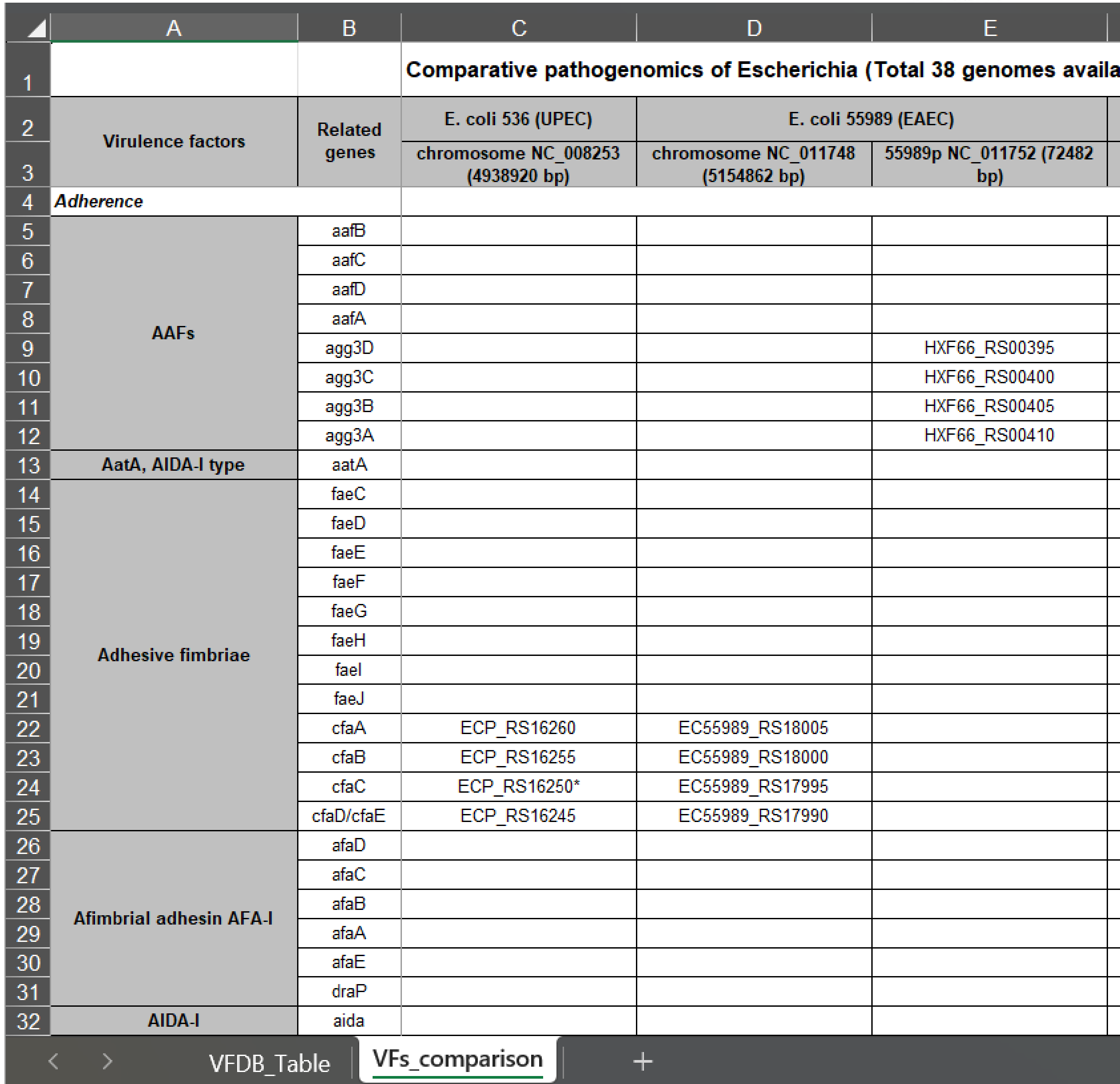
218	VF1112	vat	Vat	Vacuolating a	Escherichia coli (UPEC)	VFC0086	Effector delivery system	A ~110 kDa secreted protein exported by the	
219	VF0212	draE2	Afa/Dr family		Escherichia coli (DAEC)	VFC0001	Adherence	Afa/Dr family including afimbrial adhesins Afa	
220	VF0214	aggA	AAFs	Aggregative a	Escherichia coli (EAEC)	VFC0001	Adherence	Belongs to Dr family of adhe	Flexible fimbriae
221	VF0215	aap	Dispersin	Anti-aggregat	Escherichia coli (EAEC)	VFC0346	Others	Encoded by a gene called a	Typical signal se
222	VF0216	east1	EAST1	EAEC heat-st	Escherichia coli (EAEC)	VFC0235	Exotoxin	Also produced by other E. c	38-amino-acid pe
223	VF0217	pet	Pet	Plasmid-enco	Escherichia coli (EAEC)	VFC0086	Effector delivery system	Belongs to SPATEs subfamily; encoded on th	
224	VF0218	pic	Pic		Escherichia coli (EAEC)	VFC0086	Effector delivery system	Pic and ShET1 are encoded by the same chr	
225	VF0219	set1A	ShET1		Escherichia coli (EAEC)	VFC0235	Exotoxin	Knowns as Shigella enterotoxin 1 (ShET1), ov	
226	VF0234	chuS	Chu	E. coli heme	Escherichia coli (EHEC)	VFC0272	Nutritional/Metabolic factor	ChuA encodes for a 69-kDa outer membrane	
227	VF0404	yagY/ecpB	ECP	E. coli comm	Escherichia coli (EHEC)	VFC0001	Adherence		
228	VF0204	lifA/efa1	Efa-1/LifA	EHEC factor	Escherichia coli (EHEC)	VFC0086	Effector delivery system		
229	VF0208	espP	EspP		Escherichia coli (EHEC)	VFC0086	Effector delivery system	Encoded within the pO157 plasmid.	
230	VF0207	hlyA	Hemolysin		Escherichia coli (EHEC)	VFC0235	Exotoxin	Best-characterized RTX pro	Pro-HlyA is activ
231	VF0202	eae	Intimin		Escherichia coli (EHEC)	VFC0086	Effector delivery system	94-kDa outer-membrane pro	N-terminus of inti
232	VF0192	ler	Ler	LEE encoded	Escherichia coli (EHEC)	VFC0301	Regulation	The first gene in the LEE1 operon; shows sim	
233	VF0194	paa	Paa	Porcine attac	Escherichia coli (EHEC)	VFC0001	Adherence	Paa sequences are often present in A/E strai	
234	VF0209	stcE	StcE	Secreted prot	Escherichia coli (EHEC)	VFC0251	Exoenzyme	Encoded within the pO157 p	Substitution of As
235	VF0206	stx1A	Stx	Shiga toxin	Escherichia coli (EHEC)	VFC0235	Exotoxin	Also known as verotoxin, ver	Comprises one A
236	VF0203	toxB	ToxB		Escherichia coli (EHEC)	VFC0086	Effector delivery system	A large 362-kDa protein encoded on the 93-k	

Some genes can be found in several pathotypes!

Examples of strains of different pathotypes

In the second tab, you can see groups of virulence factors,

-And see the distribution in different types of strains

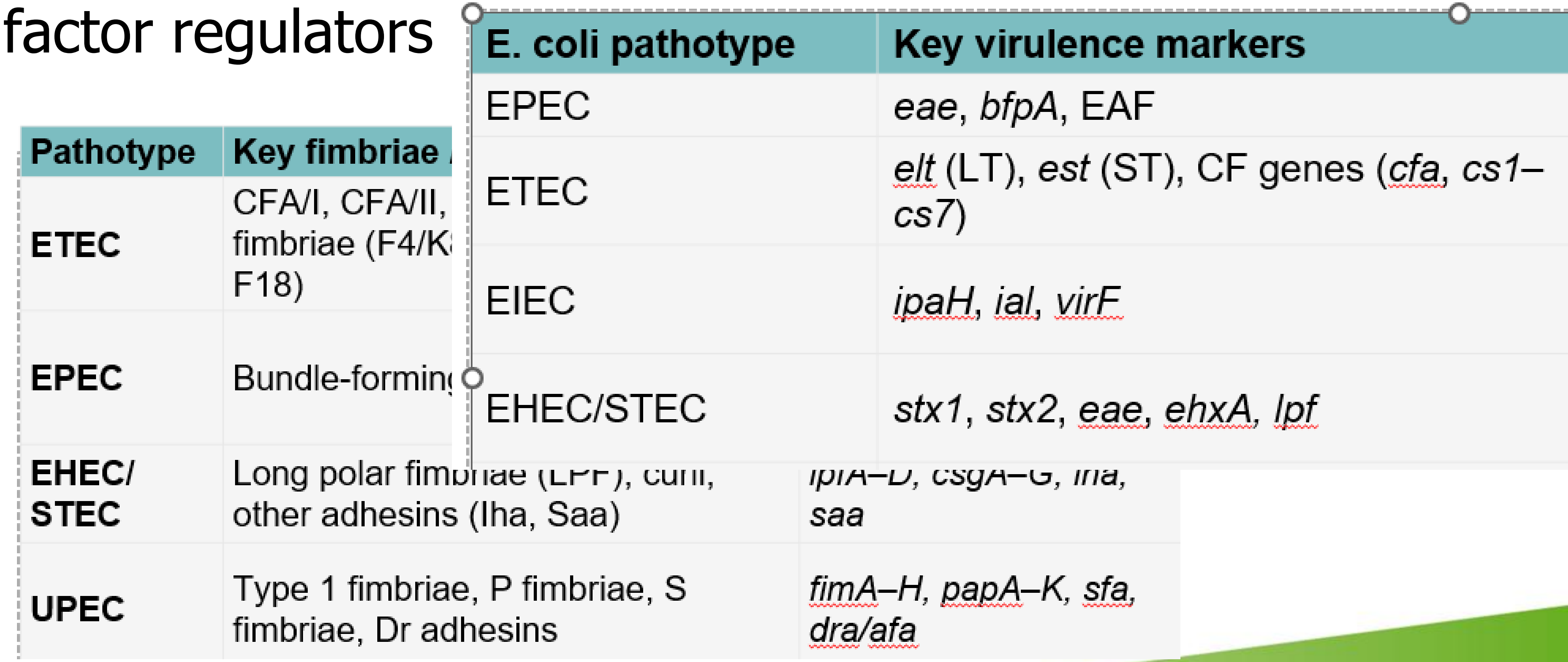


	A	B	C	D	E
1			**Comparative pathogenomics of Escherichia (Total 38 genomes availa**		
2	**Virulence factors**	**Related genes**	**E. coli 536 (UPEC)**	**E. coli 55989 (EAEC)**	
3			**chromosome NC_008253 (4938920 bp)**	**chromosome NC_011748 (5154862 bp)**	**55989p NC_011752 (72482 bp)**
4	***Adherence***				
5	**AAFs**	aafB			
6		aafC			
7		aafD			
8		aafA			
9		agg3D			HXF66_RS00395
10		agg3C			HXF66_RS00400
11		agg3B			HXF66_RS00405
12		agg3A			HXF66_RS00410
13	**AatA, AIDA-I type**	aatA			
14	**Adhesive fimbriae**	faeC			
15		faeD			
16		faeE			
17		faeF			
18		faeG			
19		faeH			
20		faeI			
21		faeJ			
22		cfaA	ECP_RS16260	EC55989_RS18005	
23		cfaB	ECP_RS16255	EC55989_RS18000	
24		cfaC	ECP_RS16250*	EC55989_RS17995	
25		cfaD/cfaE	ECP_RS16245	EC55989_RS17990	
26	**Afimbrial adhesin AFA-I**	afaD			
27		afaC			
28		afaB			
29		afaA			
30		afaE			
31		draP			
32	**AIDA-I**	aida			

Profiling?

Several groups of genes play an important role

- Fimbriae + adhesins
- Virulence factor regulators
- Toxins



E. coli pathotype	Key virulence markers
EPEC	*eae*, *bfpA*, EAF
ETEC	*elt* (LT), *est* (ST), CF genes (*cfa*, *cs1–cs7*)
EIEC	*ipaH*, *ial*, *virF*
EHEC/STEC	*stx1*, *stx2*, *eae*, *ehxA*, *lpf*

Pathotype	Key fimbriae	
ETEC	CFA/I, CFA/II, fimbriae (F4/K F18)	
EPEC	Bundle-formin	
EHEC/ STEC	Long polar fimbriae (LPF), curli, other adhesins (Iha, Saa)	*lpfA–D, csgA–G, iha, saa*
UPEC	Type 1 fimbriae, P fimbriae, S fimbriae, Dr adhesins	*fimA–H, papA–K, sfa, dra/afa*

Acknowledgements

The creation of this training material was commissioned by ECDC to Technical University of Denmark with the direct involvement of Jette Kjeldgaard.