



Norges miljø- og
biovitenskapelige
universitet

Masteroppgave 2020 30 stp

Fakultet for kjemi, bioteknologi og matvitenskap

Mattrygghet – Helgenomsekvensering og karakterisering av hybrid-patogene *E. coli* (ST28) varianter

Food Safety – Whole genome sequencing and
characterization of hybrid pathogenic *E. coli* (ST28)
variants

Audun Solberg Kristensen

Matvitenskap: Matvaretrygghet, -kvalitet og -hygiene

Forord

Denne 30 studiepoengs masteroppgaven ble skrevet vårsemester 2020 som siste del av en mastergrad i matvitenskap med studieretningen matvaretrygghet-, kvalitet og -hygiene ved Fakultetet for kjemi, bioteknologi og matvitenskap (KBM) ved Norges miljø- og biovitenskapelige universitet (NMBU).

Jeg er takknemlig for å ha hatt mulighet til å arbeide med svært spennende og relevante temaer innenfor bioinformatikk, molekylærbiologi og mikrobiologi som er fremtidsrettede, samt få innsikt i metoder knyttet til disse med denne masteroppgaven. For denne muligheten vil jeg takke tilleggsveileder Lin Brandal og hovedveileder Bjørn-Arne Lindstedt, som har bidratt med god veiledning og oppklarende innsikt innenfor temaer som har vært mer eller mindre kjent fra før.

Til slutt vil jeg takke mine foreldre som alltid er støttende og viser mye omtanke, noe jeg setter stor pris på.

Ås, mai 2020

Audun Solberg Kristensen

Sammendrag

Bakgrunn: *Escherichia coli* (*E. coli*) er en vanlig tarmbakterie som kan isoleres både fra mennesker og dyr. Denne bakterien har et plastisk genom, som er årsaken til at vi i dag finner mange genetiske ulike stammer, både patogene og apatogene. Patogen *E. coli* kan forårsake både intestinal (IPEC) og ekstraintestinal (ExPEC) sykdom. Den kontinuerlige evolusjonen av patogen *E. coli* er en utfordring for klassifisering av bakterien i ulike patogruyper. Nye patogene hybride stammer med virulensfaktorer tilhørende både IPEC og ExPEC oppdages stadig, og det er viktig å karakterisere og følge med på utviklingen av disse stammene da de kan utgjøre en global trussel.

Hensikt: I denne masteroppgaven hadde vi til hensikt å karakterisere og undersøke forskjellige patogene *E. coli* (ST28) stammer på genomnivå, som har forårsaket intestinal sykdom i lang tid i Norge.

Metode: Helgenomsekvenser av 7 ulike ST28 stammer isolert fra pasienter med diarè (6 isolert fra Norge og en fra Italia) ble analysert og sammenliknet ved bruk av forskjellige bioinformatiske verktøy med hensikt å karakterisere og undersøke forskjellige virulensassosierte gener (VAGs).

Resultater: Alle stammene som ble undersøkt hadde både VAGs assosiert med IPEC og ExPEC. Forskjellige stammer hadde også kombinasjoner av VAGs assosiert med patotypene Enteropatogen *Escherichia coli* (EPEC), Enterotoksisk *Escherichia coli* (ETEC), Meningitt-assosiert *Escherichia coli* (MNEC) og Uropatogen *Escherichia coli* (UPEC). Stammen M884Italy hadde samme EPEC assosierte VAGs som de fleste stammene, men det ble også indentifisert shigatoksin type 2 (*stx2*), som er spesifikt assosiert med Enterohemoragisk *Escherichia coli* (EHEC).

Konklusjon: Funnene viser at kombinasjoner av VAGs assosiert med IPEC og ExPEC finnes i ST28 stammer isolert fra norske pasienter med diarè. Dette kan utgjøre en trussel for den norske befolkningen, særlig for VAGs som kan bidra til å forårsake systemisk i utsatte pasientgrupper. Den høye forekomsten av høyst prediktive UPEC VAGs hos flere av stammene er bekymringsfullt, særlig med hensyn til hvor vanlig denne typen infeksjon er både globalt og nasjonalt i Norge.

Abstract

Background: *Escherichia coli* (*E. coli*) is a common intestinal bacterium found in both animals and humans. This bacterium has a plastic genome, which is the reason we today find many genetically different strains, both pathogenic and apathogenic. Pathogenic *E. coli* is able to cause both intestinal (IPEC) and extraintestinal (ExPEC) disease. The continuous evolution of pathogenic *E. coli* is a challenge to the classification of this bacteria into different pathogroups. With novel pathogenic hybrid strains emerging, harboring possible virulence factors associated with both IPEC and ExPEC, it is important that these strains are monitored and characterized as they may become a global threat.

Purpose: In this master thesis, we wanted to characterize and investigate on the genomic level different pathogenic *E. coli* (ST28) strains which have caused intestinal disease in Norway for a long period of time.

Methods: Whole genome sequences of 7 different ST28 strains isolated from patients with diarrhea (6 isolated from Norway and 1 from Italy) were analyzed and compared using different bioinformatic tools with the aim of characterization and investigation of different virulence associated genes (VAGs).

Results: All the strains investigated were shown to have VAGs associated with both IPEC and ExPEC. Different strains also had combinations of VAGs associated with the pathotypes Enteropathogenic *Escherichia coli* (EPEC), Enterotoxigenic *Escherichia coli* (ETEC), Meningitis associated *Escherichia coli* (MNEC) and Uropathogenic *Escherichia coli* (UPEC). The strain M884Italy had the same EPEC associated VAGs as most of the other strains, but also harbored the shigatoxin type 2 (*stx2*) gene specifically associated with Enterohemorrhagic *Escherichia coli* (EHEC).

Conclusion: The findings show that combinations of VAGs associated with ExPEC and IPEC are found in ST28 strains isolated from Norwegian patients with diarrhea. This could be a potential threat for the Norwegian population, especially because of VAGs that could contribute to systemic infections in susceptible patients. The common occurrence of highly predictive VAGs of the UPEC pathotype among the strains is alarming, considering how common this type of infection is both globally and nationally in Norway.

Forkortelser

AAF/I	Aggregative adherence fimbriae
A/E-lesjoner	Attaching and effacing lesjoner
aEPEC	Atypisk Enteropatogen <i>Escherichia coli</i>
BFP	Bundle forming pili
CFAs	Colonizing factor antigens
EAF	<i>Escherichia coli</i> adherence factor
EAST-1	Enteroaggregative heat-stable enterotoxin 1
EHEC	Enterohemoragisk <i>Escherichia coli</i>
EIEC	Enteroinvasiv <i>Escherichia coli</i>
EPEC	Enteropatogen <i>Escherichia coli</i>
ETEC	Enterotoksisk <i>Escherichia coli</i>
ExPEC	Ekstraintestinal patogen <i>Escherichia coli</i>
HAI	Helsetjensteassosierte infeksjoner
HAIU	Helsetjensteassosierte urinveisinfeksjoner
HUS	Hemolytisk uremisk syndrom
IPEC	Intestinal patogen <i>Escherichia coli</i>
LEE	Locus of enterocyte effacement
LT enterotoksin	Varmelabilt enterotoksin
MNEC	Meningitt-assosiert <i>Escherichia coli</i>
MLST	Multilocus sequence typing
NGS	Neste generasjons sekvensering
Nles	Non-LEE encoded effectors
ORF	Åpne leserammer
PCR	Polymerase chain reaction
PE sekvensering	Paired-end sekvensering
SE sekvensering	Single-end sekvensering
SPATEs	Serin protease autotransporters of <i>Enterobacteriaceae</i>
ssDNA	Enkeltrådet DNA
ST enterotoksin	Varmestabilt enterotoksin
STEC	Shigatoksinproduserende <i>Escherichia coli</i>
ST28	Sekvenstype 28
tEPEC	typisk Enteropatogen <i>Escherichia coli</i>
TTP	Trombotisk trombocytopenisk purpura
T3SS	Type III sekresjonssystem
UPEC	Uropatogen <i>Escherichia coli</i>
UTI	Urinveisinfeksjoner
VAGs	Virulensassosierte gener
Vat	Vacuolating autotransporter toxin

Innholdsfortegnelse

1.0 Innledning	7
1.1 Mikrobiologisk teori	7
1.1.1 E. coli patotyper	7
1.1.2 Virulensgener hos EPEC og EHEC	12
1.1.3 Hvorfor hybrider er interessante	16
1.1.4 Fenotyper og serotyping	19
1.1.5 ST28 og EPEC	19
1.2 Molekylærbiologisk- og bioinformatisk teori	20
1.2.1 Neste generasjons sekvensering: Illumina- og nanopore sekvensering.	20
1.2.2 Adaptersekvenser	25
1.3 Bioinformatiske verktøy	25
1.3.1 Trimmomatic	25
1.3.2 Shovill (assembly)	25
1.3.3 Prokka (annotation)	26
2.0 Metode	27
3.0 Resultater	29
3.1 Serotyper	29
3.2 MLST	29
3.3 Plasmider	30
3.4 Antibiotikaresistens	30
3.5 Virulensgener med VirulenceFinder	31
3.6 Virulensgener med egendefinert søkt	33
3.7 Fordeling av stammene innenfor IPEC, ExPEC og patotyper basert på VAGs	41
3.8 Protein BLAST-søk for virulensgenene <i>eae</i> , <i>tir</i> og <i>vat</i>	43
3.9 ClustalOmega sequence alignment for virulensgenene <i>eae</i> , <i>tir</i> og <i>vat</i> på nukleinsyrenivå	45
3.10 ClustalOmega sequence alignment for virulensgenene <i>eae</i> , <i>tir</i> og <i>vat</i> på aminosyrenivå.	48
3.11 SNP-analyse av de ulike stammene	51
4.0 Diskusjon	52
5.0 Konklusjon	57
6.0 Referanser	58
7.0 Vedlegg	65
7.1 Matrix benyttet i SNP-analysene	65
7.2 Tilleggsfil 1: Egendefinert virulenssøk	65

1.0 Innledning

1.1 Mikrobiologisk teori

1.1.1 *E. coli* patotyper

Escherichia coli (*E. coli*) er Gramnegativ, ikke-sporedannende fakultativ anaerob stavbakterie som tilhører familien *Enterobacteriaceae* (Granum, 2015). *E. coli* forekommer normalt i tarmen hos både mennesker og varmblodige dyr, og flestparten regnes derfor ikke som patogener. Enkelte varianter har derimot spesifikke gener tilknyttet virulens (VAGs) som gjør at de har egenskaper hvor de kan forårsake sykdom. Det er vanlig å skille mellom *E. coli* som kan forårsake intestinal sykdom (IPEC) og de som forårsaker ekstraintestinal sykdom (ExPEC), og videre dele disse inn i subgrupper (patotyper) basert på karakteristiske VAGs (Köhler & Dobrindt, 2011) (Lindstedt et al., 2018). IPEC inkluderer patotypene ETEC (Enterotoksisk *E. coli*), EPEC (Enteropatogen *E. coli*), EHEC (Enterohemoragisk *E. coli*), EIEC (Enteroinvasiv *E. coli*) og EAEC (Enteroaggregativ *E. coli*), hvor disse er obligate patogener (Köhler & Dobrindt, 2011). ExPEC derimot er fakultative patogener, og inkluderer patotypene UPEC (Uropatogen *E. coli*), MNEC (Meningitt-assosiert *E. coli*) og SEPEC (Sepsis-forårsakende *E. coli*). ExPEC kan forårsake en rekke sykdommer, hvor enkelte stammer vil kunne gi svært alvorlige infeksjoner. De ekstraintestinale sykdommene kan være urinveisinfeksjon, sårinfeksjon, sepsis, meningitt og hemolytisk-uremisk syndrom (HUS) (Granum, 2015).

E. coli har et plastisk genom, som består av opp mot 5000 gener, men hvor kun rundt 2000 av disse er konservert og utgjør kjernegenomet til alle stammene av *E. coli* (Granum, 2015). Denne plastisiteten gjør at det finnes mange ulike stammer av *E. coli*, både patogener av ulik grad og apatogene. For å skille mellom de ulike patogener og apatogene *E. coli* benyttes det serotyping basert på O-, H- K-antigener, men også ved bruk av ulike molekylærbiologiske typingsmetoder som f.eks. multilocus sequence typing (MLST) og patotyping basert på spesifikke VAGs. En oversikt over de ulike humane patotypene tilhørende *E. coli* er presentert i tabell 1.

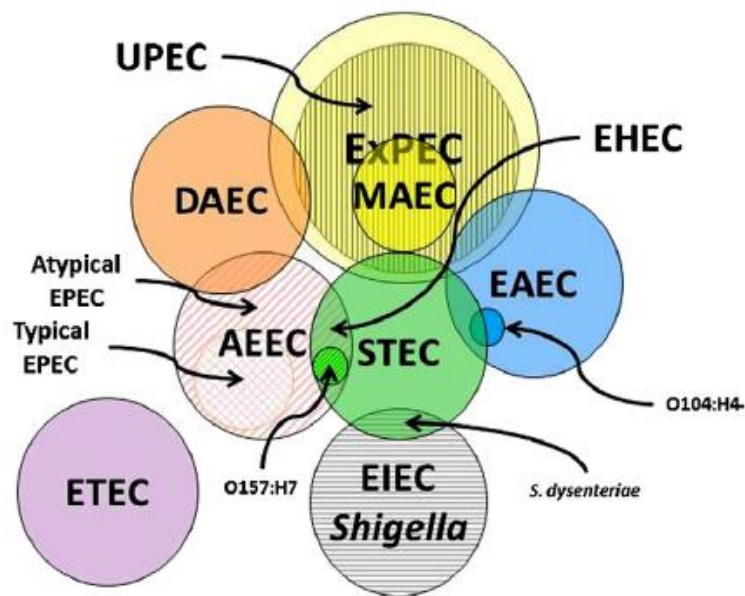
Tabell 1. Oversikt over de seks *E. coli* humane patotypene. Tabellen er reprodusert fra (Granum, 2015).

Patotype	Hovedresorvar	Viktige egenskaper	Kjennetegn for patogruppe
Enterohemorragisk <i>E. coli</i> (EHEC)	Drøvtyggere	Produserer shigatoksiner og har evne til AE (attaching and effacing) adhesjon ved tir, intimin og type 3 sekresjonssystem T3SS i LEE	Påvisning av <i>stx</i> og <i>eae</i> . Tilhører spesifikke serogrupper
Enterotoksiske <i>E. coli</i> (ETEC)	Smittede mennesker	Produserer varmelabile (LT) og varmestabile (ST) toksiner. Adherer til tarmepitel ved spesifikke fimbrier på bakterieoverflaten	LT og/eller ST. Fimbrier. Tilhører spesifikke serogrupper
Enteropatogene <i>E. coli</i> (EPEC)	Smittede mennesker	Evne til AE adhesjon (har LEE med <i>tir</i> og <i>eae</i> gener og T3SS)	Påvist <i>eae</i> , <i>tir</i> . Fravær av <i>stx</i> . tEPEC har bfp
Enteraggregative <i>E. coli</i> (EAEC)	Drøvtyggere	Adherer til celler i cellekultur etter et spesielt aggregerende mønster. Kan produsere ulike entertoksiner og cytotoksiner	Adhesjonsmønster i cellekulturer. Påvisning av shigatoksiner og adhesjongsener assosiert med stammer av EAEC
Enteroinvasive <i>E. coli</i> (EIEC)	Mennesker	Har evne til å invandere tarmepitel, svært lik <i>Shigella</i> . Stammene mangler ofte evnen til å fermentere laktose og er ofte ubevegelige	Biokjemiske og fenotypiske egenskaper som laktosenegative og ubevegelige. Evne til å invandere cellekultur
Diffus adherente <i>E. coli</i> (DAEC)	Mennesker		Adhesjonsmønster i cellekulturer av Hep-2 og HeLa celler

MLST er en typingsmetode hvor en kan skille mellom ulike stammer av bakterier (Maiden et al., 2013). Denne metoden ble utviklet da metoder som 16S rRNA ikke klarer å skille nært beslektede bakterier, hvor forandringer som følge av vertikal og horisontal genoverføring ikke vil fanges godt nok opp. MLST sammenlikner alleler (genvarianter) av typisk 7 loci («housekeeping genes»), og fanger derfor enkelt og effektivt opp mer vanlige genetiske forandringer som via horisontal genoverføring. Ved å sammenlikne sekvensene til de 7

forskjellige loci med databaser av andre alleler, kan man raskt finne ut om man har samme sekvenstype eller ikke. Basert på sekvensene til de 7 loci, gis bakterien en spesifikk profil med et unikt nummer for hvert sekvensert loci som dermed blir bakteriens sekvenstype.

De ulike patotypene av *E. coli* må sees i sammenheng med denne bakterieartens plastiske genom, som stadig i er forandring via horisontal genoverføring (Donnenberg, 2013). Nomenklaturen og definisjonene forandrer seg og utvides ettersom vår forståelse av evolusjonær utvikling og patogenese utvikler seg, samtidig som nye patotyper oppdages. I figur 1 er de ulike patotypene presentert hvor en kan se hvordan enkelte av patotypene overlapper hverandre, ettersom det har blitt oppdaget patovarianter (hybrider) med VAGs på tvers av de ulike patotypene.



Figur 1. En fremstilling av kompleksiteten, diversiteten og sammenvevingen av *E. coli* patotyper som er et resultat av denne slektens genomiske plastisitet. Innenfor ExPEC finner man blant annet «meningitis-associated» *E. coli* (MAEC), og innenfor «attaching and effacing» (AEEC) finner man blant annet EPEC. Figuren er hentet fra (Donnenberg, 2013).

Det finnes VAGs som er assosiert med de ulike patotypene, men for ExPEC eksiterer det ikke en universell akseptert liste med VAGs som definerer en *E. coli* som en ExPEC (Lindstedt et al., 2018). For å sikkert klassifisere ExPEC er det derfor nødvendig å identifisere hvor bakterien har blitt isolert fra.

Vanlige VAGs assosiert med IPEC og ExPEC er presentert i tabell 2, som er basert på informasjon fra (Lindstedt et al., 2018).

Tabell 2. Oversikt over enkelte vanlige IPEC og ExPEC assosierte VAGs.

IPEC										
IPEC VAGs	<i>stx1</i>	<i>stx2</i>	<i>eaeA</i>	<i>ipaH</i>	<i>LTI</i>	<i>STIa</i>	<i>STIb</i>	<i>aggR</i>	<i>ehxA</i>	<i>bfp</i>
ExPEC										
Cytotoxic necrotising factors	<i>cnf1</i>	<i>cnf2</i>	<i>cnf3</i>							
Autotransporters	<i>sat</i>	<i>tsh</i>	<i>vat</i>	<i>ehaA</i>	<i>ehaG</i>					
Iron aquisition	<i>iutA</i>	<i>sitA</i>	<i>iucD</i>	<i>iroC</i>	<i>fbpB</i>	<i>fyuA</i>				
Adhesins	<i>sfaS</i>	<i>papC</i>	<i>tosA</i>							
Protectins	<i>kpsS</i>	<i>traT</i>	<i>iss</i>							
Invasive gene	<i>ibeA</i>									

Virulensfaktorer assosiert med de ulike *E. coli* patotypene er presentert i tabell 3, som er basert på informasjon fra (Kaper et al., 2004).

Tabell 3. Oversikt over virulensfaktorer assosiert med de ulike *E. coli* patotypene.

Patotyper								
	EPEC	EHEC	EAEC	DAEC	EIEC	ETEC	MNEC	UPEC
V. faktorer	<i>eae</i> (intimin)	<i>eae</i> (intimin)	<i>aaf</i> (aggregative adherence fimbriae)	Dr adhesins	<i>icsA</i> (<i>virG</i>)	CFAs (colonizing factor antigens)	S fimbriae (adhesin)	Dr adhesins
	<i>bfp</i> (bundle forming pilus)	<i>paa</i> (adhesin)	<i>shET1</i> (shigella enterotoxin 1)		Aerobactin (iron acquisition, siderophore receptor)	LT (heat-labile enterotoxin)	<i>ompA</i> (adhesin)	<i>pap</i> fimbriae (adhesin)
	<i>paa</i> (adhesin)	<i>toxB</i> (adhesin)	<i>pic</i> (autotransporter)		ShET1* (shigella enterotoxin 1)	STa (heat-stable enterotoxin a)	<i>lbeA, B, C</i> (promotes invasion)	F1C fimbriae (adhesin)
	<i>lpf</i> (long polar fimbriae)	<i>efa-1/lifa</i> (adhesin)	<i>pet</i> (autotransporter)		<i>pic</i> * (autotransporter)	STb (heat-stable enterotoxin b)	<i>aslA</i> (promotes invasion)	S fimbriae (adhesin)
	<i>espC</i> (autotransporter)	Lpf (long polar fimbriae)			<i>sepA</i> * (autotransporter)		K antigen capsules (antiphagocytic)	Type 1 fimbriae (adhesin)
	<i>cif</i> (cycle-inhibiting factor)	<i>saa</i> (adhesin)			<i>sigA</i> * (autotransporter)			<i>ireA</i> (iron acquisition, siderophore receptor)
	<i>espF</i> (type III effector)	<i>ompA</i> (adhesin)			<i>ipaA</i> (type III effector)			<i>iroN</i> (iron acquisition, siderophore receptor)
	<i>espH</i> (type III effector)	<i>stx</i> (shiga toxin)			<i>ipaB</i> (type III effector)			<i>pic</i> (autotransporter)
	<i>map</i> (type III effector)	Urease (ABC subunit, cleaves urea)			<i>ipaC</i> (type III effector)			<i>sat</i> (autotransporter)
	<i>tir</i> (type III effector)	<i>espP</i> (autotransporter)			<i>ipaH</i> (type III effector)			<i>hylA</i> (RTX toxin, cell lysis)
	<i>lifa/efa</i> (inhibits lymphocyte activation)	<i>cif</i> (cycle-inhibiting factor)			<i>ipgD</i> (type III effector)			CNF 1, 2 (cytotoxic necrotizing factors, altered cytoskeleton, necrosis)
		<i>espF</i> (type III effector)			<i>virA</i> (type III effector)			<i>usp</i> (uropathogenic Specific protein)
		<i>espH</i> (type III effector)						

	<i>map</i> (type III effector)						
	<i>tir</i> (type III effector)						
	<i>stcE</i> (type II effector)						
	<i>Ehx</i> (RTX toxin)						
	<i>lifA/efa</i> (inhibits lymphocyte activation)						

*Disse faktorene har blitt karakterisert i *Shigella* arter, men enda ikke blitt bestemt for EIEC.

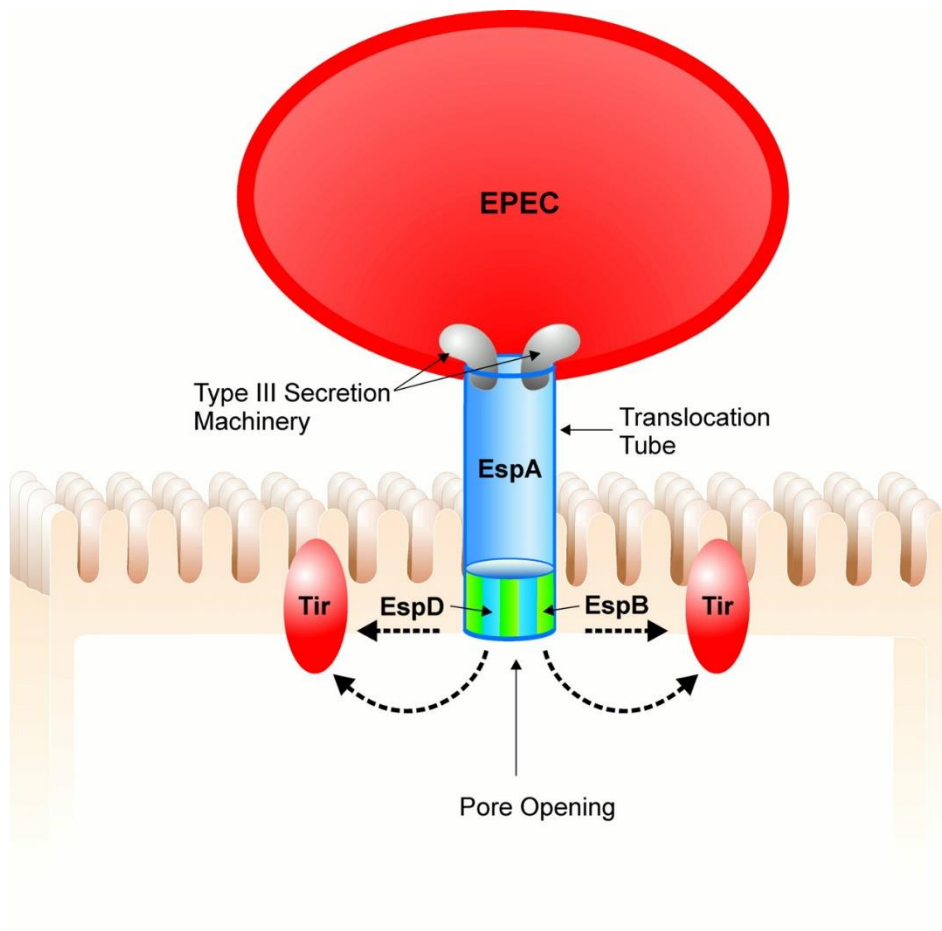
1.1.2 Virulensgener hos EPEC og EHEC

Virulensfaktorer tilknyttet EPEC befinner seg på et stort plasmid kalt for EAF («*E. coli* adherence factor») og kromosomalt på patogenisitetsoyene LEE («locus of enterocyte effacement»), i tillegg til *bfp* og *per* operonet (Donnenberg, 2013). Disse virulensfaktorene øker bakteriens evne til overlevelse i møte med vertens immunsystem samt gir evne til å forårsake en infeksjon og sykdom hos verten.

EAF plasmidet har en størrelse som varierer mellom 50-70 MDa, finnes kun hos tEPEC (typisk EPEC) og kan i noen tilfeller overføres mellom *E. coli* stammer via konjugasjon, avhengig av om det inneholder nødvendige overføringsgener (Donnenberg, 2013). *Bfp* og *per* er to viktige loci på kromosomet som koder for transkripsjonsfaktorer tilknyttet virulens. EAF plasmidet spiller også en viktig rolle i uttrykk av *eae* ved å regulere aktiviteten til genet.

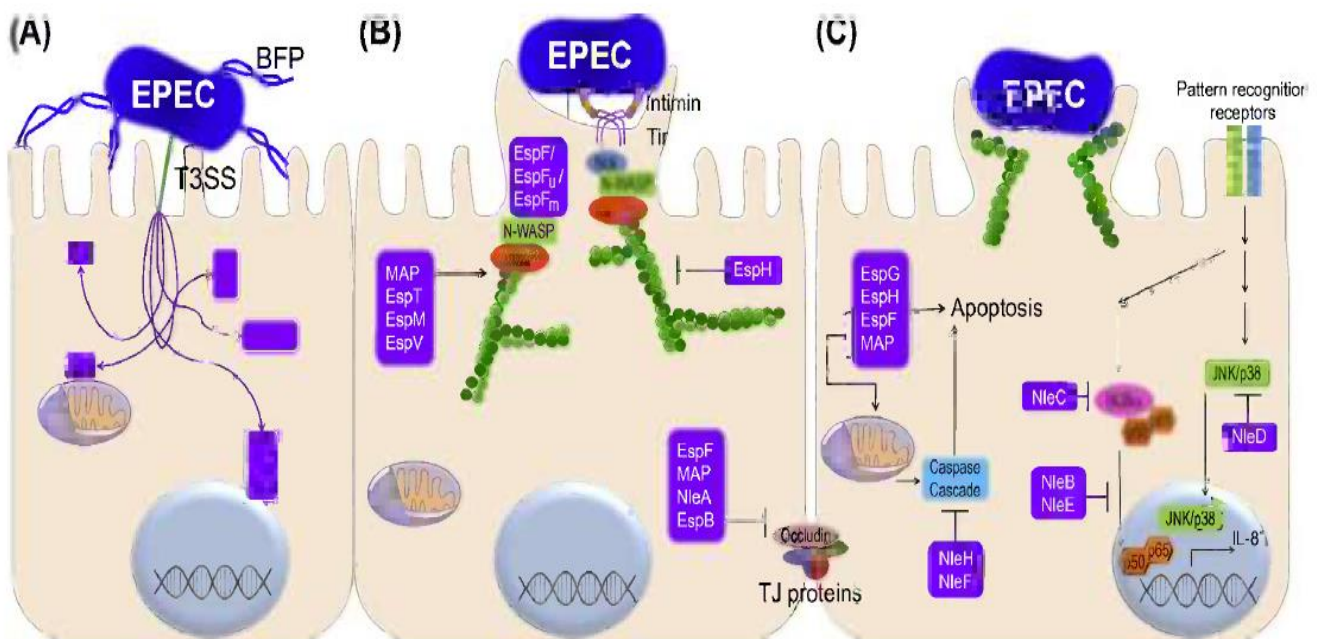
Patogenisitetsoyene LEE har består av 35.6 kb med genetisk informasjon, og spiller en viktig rolle i *E. coli* sin evne til å danne A/E-lesjoner («attaching and effacing») (Donnenberg, 2013). Hos EPEC består LEE av 41 åpne leserammer (ORFs) som koder for transkripsjonsfaktorer, et adhesin, type 3 sekresjonssystem (T3SS), effektor proteiner, type 3 spesifikke chaperoner. T3SS er et transportsystem som gjør at EPEC kan transportere proteiner over bakteriens indre og ytre membran, samt over vertens cytoplasmatiske membran. Totalt består LEE av 41 ORFs, som blant annet koder for mer enn 20 proteiner som inngår i T3SS. Translokator proteinene EspA, EspB og EspD er viktige for translokasjon av proteiner inn i vertscellen, og uten disse vil

EPEC ikke kunne danne A/E-lesjon. EspA inngår som en strukturell komponent i et rørignende kompleks som danner koblingen mellom EPEC og vertscellen. EspB og EspD vil etter at bakterien har festet seg til vertscellen, danne en porestruktur hvor translokasjonen av proteiner kan foregå.



Figur 2. Viser translokasjonen av effektor proteiner som blant annet Tir gjennom T3SS via translokator proteinene proteinene EspA, EspB og EspD. Figuren er hentet fra (Vallance & Finlay, 2000).

Effektor proteinene som translokeres gjennom T3SS resulterer i at proteinene intimin og Tir vil danne en interaksjon som gjør at EPEC får feste til epitelcellen og danner A/E-lesjonen (Donnenberg, 2013). Andre sentrale effektor proteiner kodet av LEE som blir translokert via T3SS er blant annet EspF, EspH, EspG, EspZ og Map. Effektor proteiner translokert via T3SS som ikke kodes av LEE er kjent som Nles («non-LEE encoded effectors»).



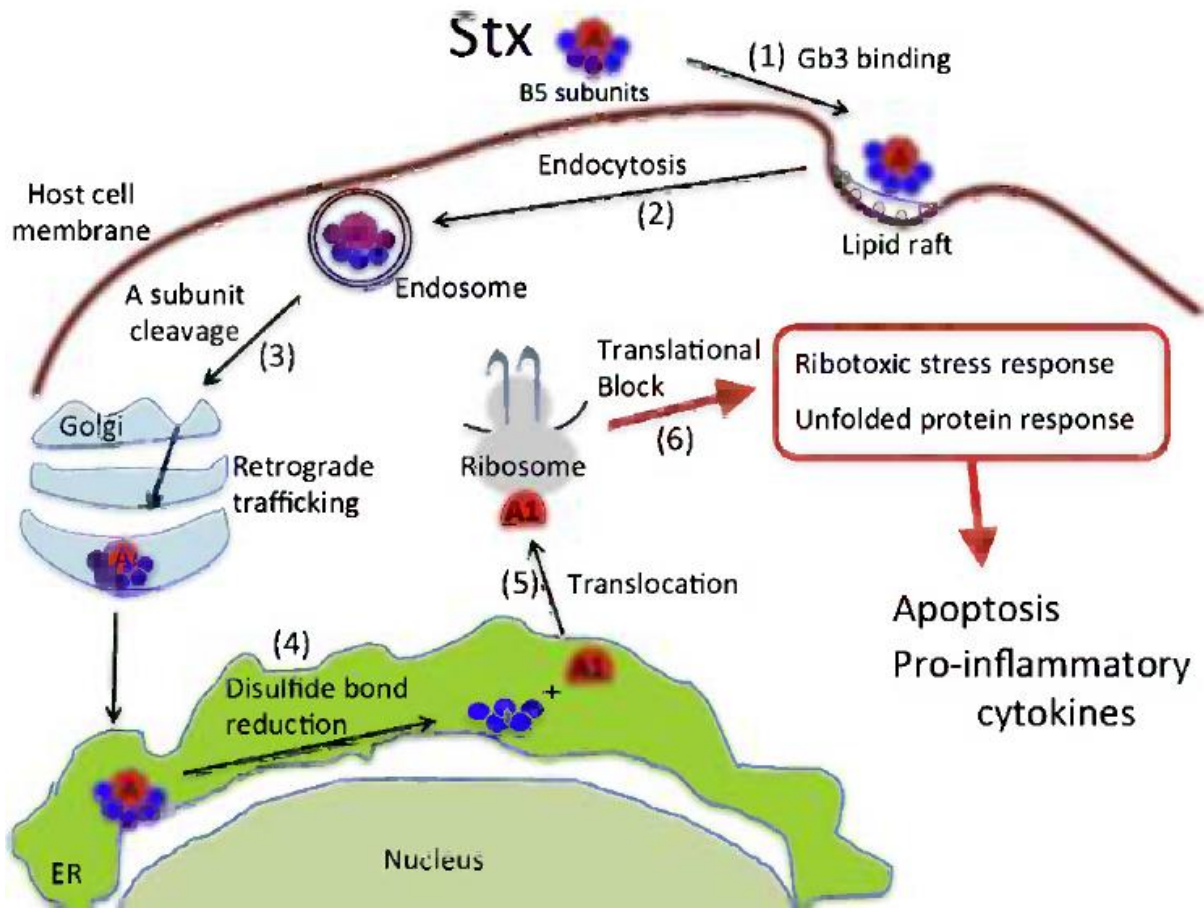
Figur 3. Viser hvordan EPEC først fester seg til epitelceller via BFP, deretter benyttes T3SS til translokere LEE-kodene og ikke-LEE kodene effektor proteiner inn i vertscellen (A). Effektor proteinene sammen med intimin gjør at det dannes epitelcelleforandringer (B), som resulterer i at det dannes en pidestall-lignende struktur hvor EPEC kan feste seg (C). Effektor proteinene vil også påvirke apoptose og inhibering av vertscellens immunrespons. Figuren er hentet fra (Donnenberg, 2013).

«Enteroaggregative heat-stable enterotoxin 1» (EAST-1) er et enterotoksin som kodes av genet *astA* og produseres av EAEC stammer, men har også blitt påvist hos EPEC stammer (Donnenberg, 2013). Dette toksinet peptid bestående av 38-aminosyrer og en molekylærvækt på 41 kDa, og er 50 % homologt med varmestabile STa enterotoksiner fra ETEC med lignende reseptor interaksjoner (Dubreuil, 2019). Toksinet produseres av stammer isolert fra mennesker og dyr. Toksinet er assosiert med diarè hos mennesker, men i ulik grad, da det blant annet har blitt vist at frivillige infisert med EAST-1 produserende EAEC stammer ikke utviklet diarè selv ved effektiv tarmkolonisering (Nataro et al., 1995). Det er også isolert *astA*-positive stammer i betydelige mengder fra feces hos barn og voksne uten at disse har hatt symptomer på diarè, noe som indikerer at mennesker kan være asymptomatiske bærere av disse stammene. Toksinet virker på samme måte som STa enterotoksin produsert fra ETEC ved å stimulere enzymet guanylat cyklase i tarmcellene. Aktivering av guanylat cyklase gir økt nivå av cyklisk GMP

med påfølgende lekkasje av elektrolytter og vann ut i tarmen som resulterer i diarè (Dubreuil, 2019; Granum, 2015).

For å overleve den lave pHen i magesekken må tarmpatogene bakterier ha ulike forsvarssystemer, hvor for EPEC spiller glutamat decarboxylase en rolle i å motvirke den lave pHen forårsaket av magesyren (Shin et al., 2001). Transkripsjonen av glutamat decarboxylase styres av tre gener, *gadA*, *gadB* og *gadC* som igjen reguleres av ulike faktorer. GAD-systemet har vist seg å ha en felles regulator kalt for GadX med *per* operonet som styrer transkripsjonen av LEE-gener viktig for type 3 sekresjonssystemet. GAD-systemet som aktiveres ved lav pH kan derfor regnes en virulensfaktor ved at aktivering av dette systemet også fører til aktivering av *per* operonet.

EHEC som er en undergruppe av STEC, som i tillegg til shigatoksin produksjon har virulensfaktorer som gjør at den kan feste seg til og invadere vertscellene via T3SS (Donnenberg, 2013). Produksjonen av shigatoksiner gjør EHEC til en fryktet patogen som kan gi svært alvorlige systemiske infeksjoner. Shigatoksinene deles inn i to grupper, *stx1* og *stx2* hvor hver gruppe har flere undergrupper. Shigatoksinene har til felles at de er proteiner bestående av en A-subenhet (StxA) og fem B-subenheter (StxB) i en kvartærstruktur. StxA har en *N*-glycosidase aktivitet, mens StxB binder seg til glykolipid reseptorer på celleoverflater. Shigatoksinene er styrt av operoner, hvor disse i stor grad er like strukturelt for de ulike toksin typene. Derimot er de ulike gruppene og undergruppene ulike i hvilken grad de kan fremkalle sykdom. For O157:H7 *E. coli* stammer er Stx2 produksjon i større grad assosiert med utvikling av HUS og alvorlige komplikasjoner hos mennesker, sammenliknet med Stx1. Genene som koder for shigatoksinene er et resultat av lambdaoide profager inkorporert i *E. coli* sitt genom. Ved aktivering av SOS responssystemet hos *E. coli* som for eksempel ved bruk av enkelte antibiotika, vil lytisk syklus hos profagene promoteres som kan gi økt transkripsjon av shigatoksinene.



Figur 4. Viser hvordan shigatoksinerne promoterer cellulær intoksikasjon ved å skade ribosomene som resulterer i inhibering av proteinsyntese, ribotoksisk stress responser og respons knyttet til protein utkveiling. Patologisk vil dette gi økt apoptose og inflammasjon. Figuren er hentet fra (Donnenberg, 2013).

Ved cellulær intoksikasjon vil pentameren StxB bindes vertens celleoverflate via en glykosid reseptor (Donnenberg, 2013). Toksinet fraktes inne i cellen via endocytose, og fraktes gjennom golgiapparatet til endoplasmatisk retikulum. Proteolytisk spalting av StxA, resulter i en fragmentering hvor StxA1 til slutt dissosieres fra StxA2 og StxB via en reduksjon av en intramolekylær disulfidbinding. StxA1 fraktes deretter videre til ribosomet hvor toksinet forårsaker strukturelle forandringer (depurinering av «alpha sarcin loop» i 28S ribosomalt RNA).

1.1.3 Hvorfor hybrider er interessante

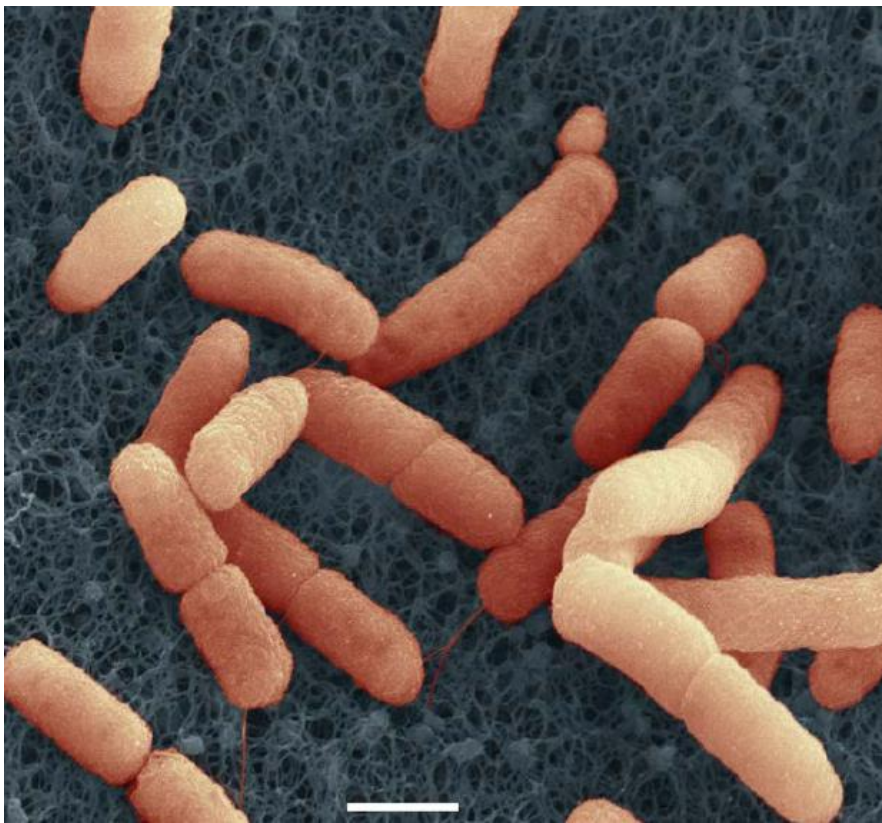
Shigatoksin produserende *E. coli* (STEC) og andre *E. coli* som gir lokal tarminfeksjon (IPEC) og systematisk infeksjon (ExPEC) kan tilegne seg virulensgener via horisontal genoverføring

fra andre patogener, og kan dermed føre til utvikling av patogene hybridstammer (Nyholm et al., 2015).

Hybridstammer kan derfor teoretisk både klassifiseres som ExPEC og IPEC, noe som vil kunne være relevant å ta hensyn til ved identifisering av patogener både i rutinekontroller innenfor næringsmiddelindustri og i klinisk setting. En stor metastudie har vist at stammer med VAGs som gjør at de kan klassifiseres som ExPEC utgjør omtrent 10 % av fekale isolater hos friske individer (Starčič Erjavec & Žgur-Bertok, 2015). En annen studie har vist at en *stx*-kodende bakteriofag via lysogenisk konvertering kan tilegne *stx*-gener til en atypisk enteropatogen *E. coli* (aEPEC) (Eichhorn et al., 2018). aEPEC, i motsetning til EPEC, regnes som både å ha reservoar hos mennesker og dyr, og har derfor potensielle for zoonose. Den påviste lysogeniske konverteringen av en aEPEC til EHEC viser oss at utvikling av mer aggressive hybridstammer fra mindre alvorlige patotyper kan oppstå, og kan derfor være relevant å ta stilling til ved ulike screeninger for bakterier og ved infeksjoner av denne typen.

Konsekvensene av høypatogene hybridstammer har allerede rammet Europa, kjent ved et stort utbrudd i Tyskland og et mindre i Frankrike i mai til juli 2011 (Donnenberg, 2013). I Tyskland utviklet omkring 4000 smittede blodig diarè, med 855 tilfeller av HUS og 53 døde som følge av komplikasjonene (Institute, 2011). Smittekilden til de Europeiske utbruddene ble lenket til importerte kontaminerte sporer av bukkehornkløver fra Egypt. I slike tilfeller er det sannsynlig at vanningsvann kontaminert med dyrefeces er årsaken til kontaminasjonen. Begge utbruddene var forårsaket av en STEC serotype O104:H4. Videre viste det seg at dette ikke var en vanlig STEC, men den hadde også VAGs tilknyttet patogruppen EAEC. Denne stammen viste seg i tillegg å være multiresistent mot flere antibiotika, inkludert tredje generasjons cefalosporiner. Denne stammen ble trolig til via tap og opptak av kromosomale- og plasmidkodende VAGs via horisontal genoverføring, hvor en modell foreslår profag mediert overføring av shigatoksin 2 (Rasko et al., 2011). Slike hybridstammer er bekymringsfulle da kombinasjonen av flere VAGs fra ulike patogrupper vil kunne skape stammer med svært effektive infeksjonsevne, samtidig som de vil kunne ha en høyere enn tradisjonell mortalitet eller føre til utvikling av alvorlige komplikasjoner. For stammen som forårsaket utbruddet i Tyskland og Frankrike, spilte trolig bakterienes evne til å effektivt kolonisere tarmepitelet og svekke tarmbarrieren, en viktig rolle i å gjøre pasientene mer utsatt for shigatoksinet og den høye forekomsten av HUS blant de

smittede. Blant annet ble det påvist gjennom PCR (polymerase chain reaction) *aggA* genet hos stammen, som koder for en viktig komponent i AAF/I («aggregative adherence fimbriae») som er assosiert med dannelse av biofilm og hemagglutinasjon av humane erythrocytter (Donnenberg, 2013; Harrington et al., 2006). I tillegg ble det påvist virulens regulator genet *aggR* som er typisk for EAEC (Donnenberg, 2013). Et annet viktig funn hos O104:H4 stammen var gener knyttet til jernopptak vanligvis assosiert med EAEC stammer, mer spesifikt «aerobactin transport system», som igjen er nært knyttet bakteriens evne til virulens og overlevelse ved infeksjon (Donnenberg, 2013).



Figur 5. Viser EHEC O104:H4, bakteriestammen som forårsaket de alvorlige utbruddene i Europa i 2011. Bilde er tatt med skanning elektronmikroskop, hvor skalaen (hvite horisontale linjen) representerer 1 μ m. Bilde er hentet fra (Institute, 2011).

1.1.4 Feno- og serotyping

Fenotype til en bakterie er dens genetiske uttrykk, og er et resultat av hvilke gener som blir transkribert (Tronsmo, 2016). Virulensgener er i denne sammenhengen viktige potensielle fenotypiske trekk hos *E. coli* som gir et uttrykk for hvilke evner bakterien har til å gi sykdom. En måte å skille bakterier innenfor samme art med ulik genetisk sammensetning er ved å benytte seg av serotyping. Ulike serotyper innenfor samme bakterieart vil ha forskjellige antigener med sine komplementære antistoffer. For eksempel serotypes *E. coli* basert på O- og H-antigener, hvor disse representerer respektivt lipopolysakkarid og flagelle antigener (Adams & Moss, 2008). O-antigenene vil for *E. coli* definere serogrupper som videre deles inn i serotyper basert på H-antigenene. Feno- og serotyping gir nyttige informasjon om karakteristiske trekk som kan benyttes til å identifisere og vurdere virulens hos ulike patogener.

1.1.5 ST28 og EPEC

Sekvenstype 28 (ST28) er en klassifisering av *E. coli* etter MLST hvor denne klassen er blitt vist å tilhøre patotype EPEC (Yu et al., 2018). Andre funn viser også at ST28 kan tilhøre patotypen EHEC (Wirth et al., 2006). Dette samsvarer også med funn, hvor det har blitt vist at aEPEC kan tilegne seg *stx* gener gjennom lysogenisk konvertering og dermed bli en EHEC (Eichhorn et al., 2018). Samtidig har det også blitt vist at EHEC igjen kan tape *stx* gener og bli konvertert tilbake til en aEPEC (Bielaszewska et al., 2007). Slike aEPEC stammer som har mistet tidligere *stx* gener betegnes i smittevernveilederen som EHEC-LST (Folkehelseinstituttet, 2019). Spesifikt står det i definisjonen fra smitteveilederen «EHEC-LST er en *eae*-positiv *E. coli* som kan være HUS-assosiert EHEC som har mistet sine *stx* gener.»

EPEC stammer av *E. coli* har evne til å gi diarè til verten, men produserer ikke shiga-toksiner (Donnenberg, 2013). I tynntarmen har EPEC evne til å feste seg til å tarmepitelceller (mikrovilli) som resulterer i strukturelle forandringer i tarmepitelet, kjent som A/E-lesjoner. De genkodende sekvensene for A/E-lesjon virulens befinner seg på patogenistetsøyen referert til som LEE. EPEC klassifiseres videre inn i to underklasser, tEPEC og aEPEC *E. coli* avhengig av om bakterien har plasmidet EAF (Granum, 2015). EAF koder for BFP («bundle-forming pili»), som spiller en rolle i bakterien sin evne til adhesjon til tarmepitel. BFP er assosiert med

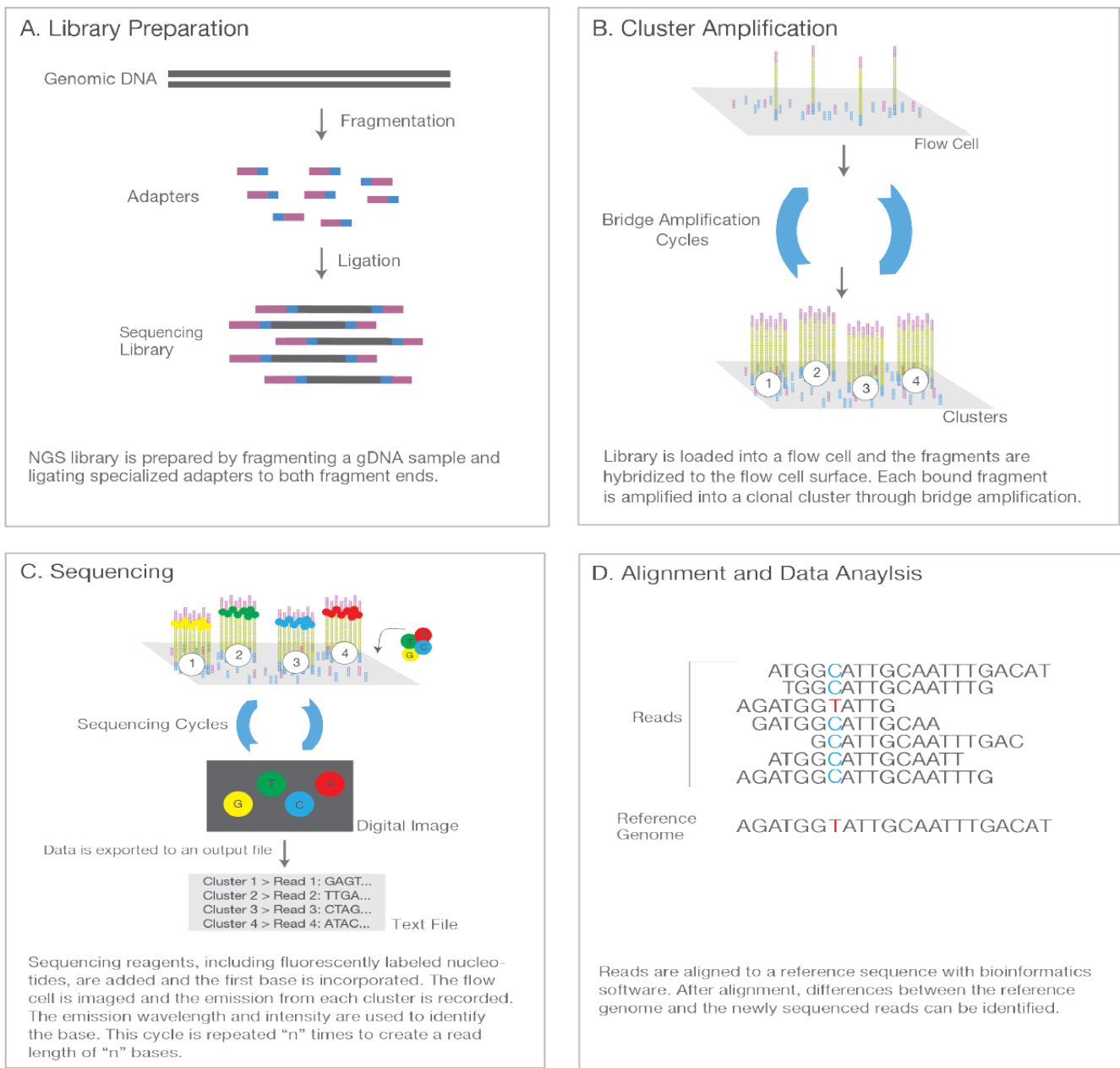
tEPEC, men ikke aEPEC. EPEC smitter via fekal-oral rute via smittekilder som kontaminert mat og vann, men også direkte kontakt med smittebærere (Donnenberg, 2013). Det er også blitt beskrevet et tilfelle med smitte mellom en hund og et barn med tEPEC, noe som også kan bety at kjæledyr er potensielle bærere og smittekilder (Rodrigues et al., 2004). Infeksjonsdosen beskrives som 10^8 - 10^{11} KDE for voksne, men som er sannsynligvis lavere for barn (Donnenberg, 2013). Diarè er det vanligste symptomet på EPEC infeksjon, hvor diarèn kan være slimete. En høy andel smittede med diarè vil også rapportere andre symptomer som magesmerter, feber, muskelsmerter, oppkast og kvalme. Hos spedbarn er dehydrering en alvorlig komplikasjon som følge av diarè forårsaket av EPEC, og kan være dødelig.

1.2 Molekylærbiologisk- og bioinformatisk teori

1.2.1 Neste generasjons sekvensering: Illumina- og nanopore sekvensering.

Sekvensering har hatt en rask utvikling de siste tiårene. Den første helgenom sekvenseringen av et menneske ble publisert i Nature og Science i 2001, og tok 15 år å utføre med en prislapp på nærmere 3 milliarder dollar (Illumina, 2017). Til sammenlikning kan neste generasjons sekvenserings (NGS) teknologi, slik som HiSeqX Ten utgitt i 2014 av Illumina, utføre 45 helgenom sekvenser av mennesket på en dag med en prislapp på omtrent 1000 dollar for hvert genom. Helgenom sekvensering gir en komplett kartlegging av genomer til bakterier og er derfor svært nyttig i studier av patogene bakterier og virulensfaktorer over tid. For eksempel hadde det svært alvorlige tyske EHEC utbruddet i 2011 51 dødsfall og 4397 rapporterte tilfeller med infeksjon (Folkehelseinstituttet, 2019). I dette utbruddet ble det benyttet NGS teknologi (Illumina, 2017). Med effektiv helgenomsekvensering kunne de raskt finne smitekilden samt identifisere genetiske endringer hos *E. coli* som gav opphav til det fatale utbruddet (Grad et al., 2012).

Illumina NGS sekvenserings teknologi baserer seg på inkorporering av fluoriserende nukleotider når DNA syntetiseres, bølgelengden som avgis kan detekteres og dermed kan DNA sekvenseres (Illumina, 2017). Med NGS teknologi kan dette utføres parallelt for millioner av DNA fragmenter som gir en svært effektiv sekvensering med høy pålitelighet. En illustrasjon av arbeidsflyten i en Illumina NGS sekvensering er fremstilt i figur 6.



Figur 6: Arbeidsflyten i en Illumina NGS sekvensering innebærer 4 grunnleggende trinn: (A) forberedelser for konstruksjon av bibliotek med adapter ligerte DNA-fragmenter (B) amplifisering av biblioteket festet til flytcellen (C) sekvensering (D) alignment og data analyser. Figuren er hentet fra (Illumina, 2017).

Det første trinnet i en NGS sekvensering består av tilfeldig fragmentering av DNA ved bruk av enzymer (Illumina, 2017). DNA fragmentene liggeres med adaptersekvenser i 5' og 3' endene. De ligerte fragmentene amplifiseres deretter ved bruk av PCR og renses med gel elektroforese. I trinn nummer to blir fragmentene påført flytcellen som har en overflate med oligonukleotider som er komplementære til adaptersekvensene. Fragmentene i flytcellen amplifiseres videre slik at hvert fragment danner en klynge av identiske fragmenter gjennom

en teknikk kalt for «bridge amplification». Biblioteket (fragmentene) er nå klart for sekvensering. I trinn nummer tre blir DNA fragmentene sekvensert via deteksjon av merkede fluoriserende nukleotider som blir inkorporert under amplifiseringen. Bølgelengde og intensitet på strålingen fra den spesifikke nukleotidene som blir inkorporert, benyttes til å identifisere og skille mellom de 4 ulike nukleotidene. Det siste trinnet i en NGS sekvensering er basert på bioinformatisk dataanalyse. Dataanalysen går i hovedsak ut på å sette sammen de sekvensdataene fra DNA fragmentene tilbake til slik som det originale komplett genomet ville vært konstruert, referert til som sekvens assembly. Etterfulgt av alignment hvor man sidestiller sekvensdataene med et referansegеноm for å finne områder med høyest grad av likhet. Videre vil man utføre annotering for å finne kodene områder og ikke-kodende områder, samt forsøke å forutsi hvilke funksjoner de ulike genene har.

En fordel med NGS sekvensering er at man kan utføre paired-end (PE) sekvensering (Illumina, 2017).

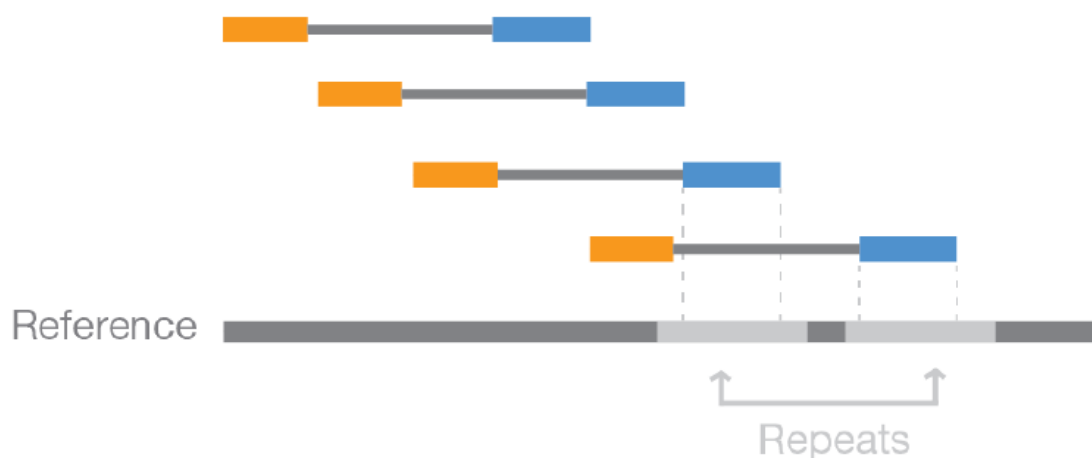
Paired-End Reads



Figur 7. Med paired-end sekvensering blir biblioteket sekvensert en bestemt lengde fra hver ende, som gir oppgav til to avlesninger (read 1 og read 2). Figuren er hentet og modifisert fra (Illumina, 2017).

Med denne typen sekvensering blir DNA fragmentene sekvensert en bestemt lengde i begge ender. Dette gir mer sekvensdata for hvert fragment som gjør at alignment blir mer presis (Illumina, 2017; Nakazato et al., 2013). En annen fordel er at metoden også gir mulighet for deteksjon av feil under amplifisering hvor delesjon eller feil inkorporasjon av en base kan oppstå (Illumina, 2017). Slike feil kan ikke oppdages med single-end (SE) sekvensering.

Alignment to the Reference Sequence

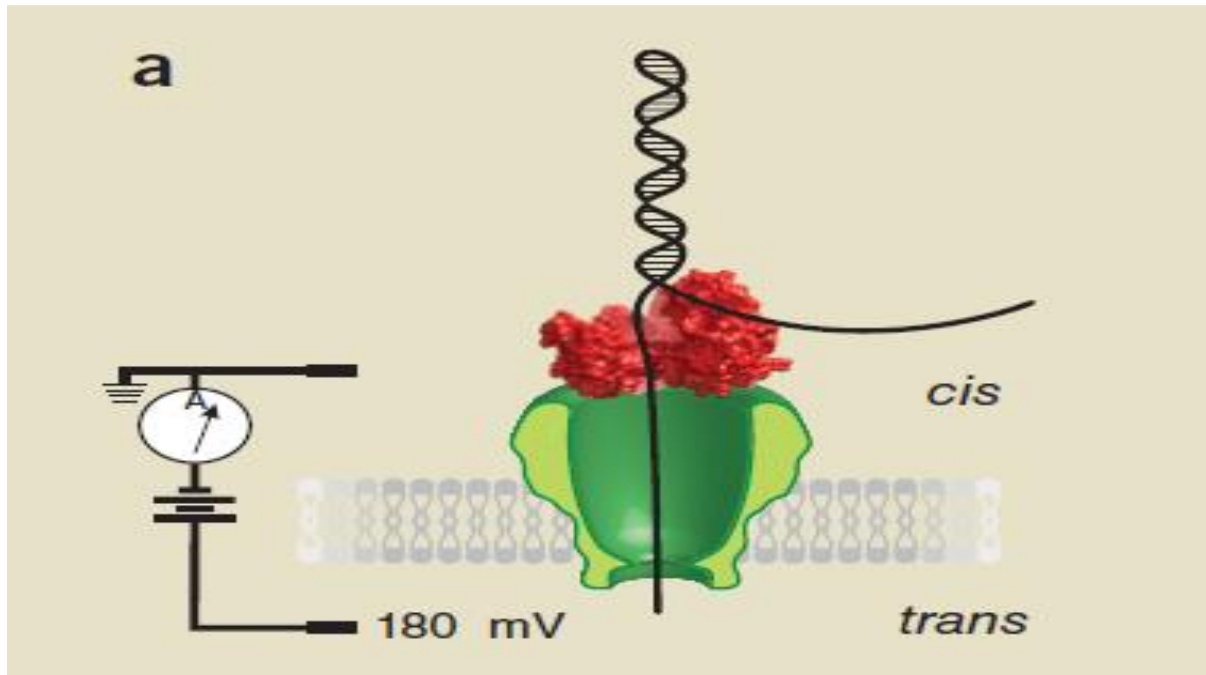


Figur 8. Bibliotek sekvensert med paired-end sekvensering gir mer informasjon når alignment med et referansegenom utføres, slik at en får mer presise sammenstillinger. Særlig gjelder dette for områder som er vanskelig å sammenstille, slik som repetitive områder i genomet. Figuren er hentet og modifisert fra: (Illumina, 2017).

Under alignment av paired-end sekvenseringer benyttes det algoritmer som utnytter informasjonen fra sekvensering av begge endene sammen med sekvenslengden til DNA-fragmentene, og derfor kan man med høyere presisjon sammenstille repetitive områder i genomet (Illumina, 2017).

Nanopore sekvensering er en svært effektiv fremtredende tredje generasjons sekvenseringsteknologi, hvor det benyttes strukturer på størrelsesnivå mindre enn 100 nanometer (Helmer Fjellvåg et al., 2019). Denne sekvenseringsteknologien skiller seg fra tradisjonelle sekvenseringsteknologier ved at sekvenseringen kan utføres uten aktiv DNA

syntese (Kono & Arakawa, 2019). Nanopore sekvensering utføres ved å benytte membran proteiner (nanopore) til å avlese enkelttrådet DNA (ssDNA) (Deamer et al., 2016).



Figur 9. Viser hvordan ssDNA beveger seg igjennom nanoporen (grønn) fra cis-side til trans-side av membranen via elektroporese. Dobbeltrådet DNA kveiles opp og bevegelsen til nukleotidene kontrolleres av enzymet (rødt) festet til nanoporen. Figuren er hentet og modifisert fra: (Deamer et al., 2016).

Nanoporen befinner seg i mellom to separate kammer med forskjellige ioniske løsninger som skaper et potensiale over nanoporen (Deamer et al., 2016). Til nanoporen er det tilkoblet et enzym (et helikase for eksempel) som kontrollerer bevegelsen til ssDNA igjennom membranen. Ved å benytte seg av elektroforese vil ssDNA beveges seg fra cis-side av membranen til trans-side, nukleotid for nukleotid. Hvert nukleotid som passerer nanoporen vil skape et unikt spenningsnivå respektivt med type nukleotid, som kan måles og benyttes til å bestemme sekvensen. Slik kan DNA sekvensen til et ssDNA molekyl avlese direkte uten noen form for aktiv DNA syntese.

1.2.2 Adaptersekvenser

Adaptersekvenser er kjemisk syntetiserte DNA sekvenser, også kjent som oligonukleotider (Watson, 2014). Disse kan bindes til 3' og 5' ender av DNA fragmenter som man ønsker å sekvensere. Adapterne gjør at man kan identifisere DNA fragmentene (barcoding), introduksjon av primere samt som de utgjør bindende sekvenser for å immobilisere DNA fragmentene til flytcelle (Momand et al., 2016).

1.3 Bioinformatiske verktøy

1.3.1 Trimmomatic

Trimmomatic er et bioinformatisk preprosesserings program som gjør at man kan endre sekvensdata fra illumina-sekvensering slik at man får fjernet overflødige adaptersekvenser (Bolger et al., 2014). Før nedstrømsanalyser av sekvensdata kan utføres må adaptersekvenser fjernes, da adaptersekvensene vil kunne forstyrre og gi feilaktige analyser.

1.3.2 Shovill (assembly)

Assembly er et viktig post sekvensering dataanalyse trinn hvor man forsøker å sette sammen de sekvenserte DNA fragmentene til originalsekvensen (Claverie & Notredame, 2006). Fragmentering av DNA er ofte nødvendig ved sekvensering av lange sekvenser, på grunn av begrensninger i sekvenseringsteknologi, hvor man blir nødt til å sekvensere flere kortere sekvenser (ofte kalt for «shotgun sekvensering»). Ulike programmer basert på assembly algoritmer er utviklet for å forsøke å rekonstruere originalsekvenser. Disse programmene vil produsere en contigs-fil som er konsensus sekvens bestemt av programmet i sitt forsøk på å rekonstruere originalsekvensen. Shovill er et program for sekvens assembly som baserer seg på algoritmer fra SPAdes, og er utviklet av Torsten Seemann (Carriço et al., 2018; Seemann, 2020). Shovill egner seg for haploide genomer av mindre størrelse, slik som bakterier, og støtter kun PE-sekvenseringer.

1.3.3 Prokka (annotation)

Når en konsensus sekvens har blitt konstruert fra assembly er det nødvendig å identifisere de proteinkodene områdene i sekvensen (genene) (Watson, 2014). Dette dataanalysetrinnet kalles for annotering hvor man i hovedsak søker etter ORFs som indikerer genkodede sekvenser. Videre kan de genkodende sekvensene sammenliknes med kjente sekvenserte modell organismer for å identifisere funksjonalitet til genene. Databaser med kjente protein-kodene sekvenser eller aminosyresekvenser kan søkes i ved bruk av algoritmer hvor man sammenlikner sekvensdataene med disse, for å bestemme funksjonalitet til gener, proteiner eller proteinstruktur.

2.0 Metode

Totalt ble sekvensdataene til 7 *E. coli* stammer analysert i denne oppgaven: ECA, ECB, ECC, ECD, FHI_NMBU_03, M844Italy og Trh49. Alle sekvensdataene bortsett fra ECA og ECB ble mottatt av hovedveileder Bjørn Arne Lindstedt, hvor ECA og ECB ble mottatt av tilleggsveileder Lin Brandal. Stammene ECA og ECB er hentet fra det nasjonale referanselaboratoriet for enteropatogene bakterier ved folkehelseinstituttet (FHI), hvor disse er nyere stammer som ble isolert respektivt fra 2019 og 2020. ECB og ECD også hentet fra FHI, og er eldre stammer fra som ble isolert respektivt 2012 og 2013. M844Italy er en stamme isolert fra Italia 2011 og som tidligere har blitt beskrevet i (Grande et al., 2016). Trh49 og FHI_NMBU_03 er stammer utlevert respektivt fra St. Olavs hospital og FHI, hvor FHI_NMBU_03 tidligere har blitt beskrevet i Lindstedt et al. og Trh49 er en stamme isolert i løpet av en studie som gikk fra 2002-2003, publisert i Afset et al (2004, 2008) (Afset et al., 2004; Afset et al., 2008; Lindstedt et al., 2018) . Alle stammene er fra sykehusprøver med personer som har hatt diaré, hvor spesifikt Trh49 er fra et barn <5 år som hadde diaré.

Filene med sekvensdata ble lastet opp til Galaxy serveren via nettstedet www.usegalaxy.eu, som er et åpent kilde prosjekt med en akademisk fri lisens som vedlikeholdes av Freiburg Galaxy Team samt kollektivt av frivillig grupper og personer omkring i Europa. Dette prosjektet tilbyr lagring, prosessorkraft og bioinformatiske verktøy til bioinformatiske analyser uten at det er nødvendig med programmeringskunnskaper. Enkelte av sekvensene var ikke ferdig assemblert og det var nødvendig å benytte verktøyet Trimmomatic i Galaxy for å fjerne adaptersekvenser, og deretter benytte verktøyet Shovill for å produsere assembly-filer.

Assembly-filer av sekvensdataene ble deretter analysert med verktøyet Prokka for å lage annoterte filer. Det ble også benyttet annoteringsverktøyet VirulenceFinder via nettstedet www.genomicepidemiology.org for søke spesifikt etter virulensgener. Dette nettstedet ble også benyttet til å søke etter antibiotikaresistens via ResFinder, MLST-typing via MLST, serotyping via SeroTypeFinder og plasmider via PlasmidFinder. For virulensgenet var identifisert med VirulenceFinder, ble dette identifisert som hbp (hemoglobin-binding protease hbp autotransporter) med Prokka, altså ble sekvensene ulikt annotert selv om de hadde samme genkodende region i genomet angitt av CDS-nummerene (coding sequence).

Virulensgenene *eae*, *tir* og *vat* som ble identifisert for de ulike stammene ved bruk av VirulenceFinder, ble videre undersøkt på aminosyre- og nukleinsyrenivå ved å sammenlikne de for forskjeller. Dette ble gjort ved å benytte det nettbaserte verktøyet Clustal Omega Multiple Sequence Alignment (www.ebi.ac.uk/Tools/msa/clustalo/). Aminosyresekvensene for disse virulensgenene ble deretter analysert med verktøyet BLAST (Basic Local Alignment Search Tool) via www.blast.ncbi.nlm.nih.gov/Blast.cgi, for å finne mest lignende proteinsekvenser i deres databaser med tilhørende organisme.

Hovedveileder utførte en SNP analyse av stammene ved å benytte verktøyet CSIPhylogeny via www.cge.cbs.dtu.dk/services/CSIPhylogeny/. SNP analysen gir en oversikt over slektskap mellom stammene, hvor *E. coli* K-12 MG1655 ble benyttet som referansestamme. Det ble også utført et nukleotid BLAST-søk med hovedveileders egendefinert virulensdatabase (upublisert) for alle stammene, med hensikt å identifisere flere virulensgener. Fra det egendefinerte virulenssøket ble det plukket ut VAGs som kun var unike for en stamme, eller som var tilstede for flere stammer, men som ikke fantes hos alle. Det ble også plukket ut VAGs som var tilstede hos alle stammene, men hvor det var ulike genvarianter. Gener som også indikerte antibiotika resistens ble også plukket ut. Hensikten med utvalget fra det egendefinerte søket var å velge ut VAGs etter oppgavens hovedmål som er å sammenlikne de ulike stammene med hensyn til forskjeller på genomnivå.

Virulensgenene identifisert for de ulike stammene ved bruk av VirulenceFinder og det egendefinerte søket, ble sammenliknet med definerte VAGs assosiert med IPEC, ExPEC og 7 ulike patotyper (EPEC, EHEC, EAEC, DAEC, EIEC, ETEC, MNEC, UPEC) slik som definert i tabell 2 og 3. Dette ble gjort med hensikt å undersøke fordeling av VAGs assosiert med IPEC, ExpEC og ulike patotyper for de ulike stammene, som deretter kunne benyttes til å sammenlikne forskjeller mellom stammene.

3.0 Resultater

Det ble benyttet SeroTypeFinder, MLST, PlasmidFinder og ResFinder for alle de ulike stammene, deretter ble funnene sammenliknet slik som presentert i tabell 4, 5, 6 og 7.

3.1 Serotyper

Tabell 4. Sammenlikning av serotyper funnet for de ulike stammene ved å benytte verktøyet SeroTypeFinder. Accession number, nummeret som angir databasereferanse som er benyttet er oppgitt i parentes.

Serotyper							
Gener	ECA	ECB	ECC	ECD	M884Italy	FHI_NMBU_03	Trh49
wzx	O33 (AB811611)	-	O81 (CU928162)	O84 (AB812036)	O96 (AB812043)	-	O81 (CU928162)
wzy	O33 (AB811611)	-	O81 (CU928162)	O84 (AB812036)	O96 (AB812043)	-	O81 (CU928162)
fliC	H6 (AIEY01000041)	H6 (AIEY01000041)	H6 (AY249991)	H7 (AF228492)	H7 (AF228492)	H6 (AIFA01000047)	H6 (AIFA01000047)

Det ble funnet ulike O antigener (O33, O81, O84 og O96) og H antigener (H6 og H7) for de ulike stammene. For stammene ECB og FHI_NMBU_03 ble det ikke funnet noen O antigener. I følge hovedveileder er FHI_NMBU_03 en O81 serotype, basert på serologiske analyser gjort ved FHI.

3.2 MLST

Tabell 5. Sammenlikning av MLST funn for de ulike stammene ved å benytte verktøyet MLST. Identity query (prosent) for de ulike stammene sammenliknet med databasereferanse.

MLST							
Sekvenstype	ECA	ECB	ECC	ECD	M884Italy	FHI_NMBU_03	Trh49
ST28	100 %	100 %	100 %	100 %	100 %	100 %	100 %

Sekvenstype 28 (ST28) ble funnet for alle de ulike stammene med 100 % match.

3.3 Plasmider

Tabell 6. Sammenlikning av funn av plasmider for de ulike stammene ved å benytte verktøyet PlasmidFinder. Accession number er oppgitt i parentens angitt med AP/Y/B-nummer og identity query (prosent) for de ulike stammene sammenliknet med databasereferanse.

Plasmider							
	ECA	ECB	ECC	ECD	M884Italy	FHI_NMBU_03	Trh49
IncFIB (AP001918)	98.24 %	96.63 %	96.63 %	-	-	96.63 %	-
IncFIC (FII) (AP001918)	-	95.79 %	-	-	-	-	-
Col (KPHS6) (016841)	-	99.44 %	-	-	-	-	-
IncFII (AY458016)	-		96.55 %	96.55 %	-	96.55 %	96.55 %
IncFIA (AY458016)	-	-	-	-	99.04 %	-	-
IncFIB (pB171) (AB024946)	-	-	-	-	99.38 %	-	-
IncFIA (AP001918)	-	-	-	-	99.74 %	-	96.63 %

Det ble funnet plasmider hos alle de ulike stammene. For ECA og ECD ble det kun funnet ett plasmid, IncFIB og IncFII respektivt. For ECC og FHI_NMBU_03 ble det funnet to plasmider, mens for ECB, M884Italy og Trh49 ble det funnet tre plasmider.

3.4 Antibiotikaresistens

Tabell 7. Sammenligning av antibiotikaresistensgenerfunnet for de ulike stammene ved å benytte verktøyet ResFinder. Accesion number er oppgitt i parents og identity query (prosent) for de ulike stammene sammenliknet med databasereferanse.

Antibiotikaresistens							
Resistensgener	ECA	ECB	ECC	ECD	FHI_NMBU_03	M884Italy	Trh49
<i>mdf(A)</i> (macrolide) (Y08743)	98.3 %	98.3 %	98.3 %	98.3 %	98.3 %	98.3 %	98.3 %

Det ble kun funnet antibiotikaresistensgenet som koder for resistens mot antibiotikagruppen makrolider hos alle de ulike stammene.

3.5 Virulensgener med VirulenceFinder

Alle stammene ble annotert spesifikt for VAGs ved å benytte VirulenceFinder, deretter ble alle funnene sammenliknet med hverandre slik som det er presentert i tabell 8 og 9.

Tabell 8. Sammenlikning av VAGs funnet for de ulike stammene ved å benytte verktøyet VirulenceFinder. I rutene for de ulike stammene er det angitt identity query (prosent) for de ulike VAGs sammenliknet med databasereferanse.

Virulens							
Virulensfaktorer	ECA	ECB	ECC	ECD	FHI_NMBU_03	M884Italy	Trh49
<i>astA</i> (EAST-1 heat-stable toxin)	-	100 %	-	-	-	99.15 %	-
<i>cif</i> (type III secreted effector)	100 %	100 %	100 %	100 %	100 %	100 %	100 %
<i>eae</i> (intimin)	90.86 %	90.82 %	90.82 %	90.86 %	90.82 %	90.86 %	90.82 %
<i>espA</i> (type III secretions system)	100 %	100 %	100 %	100 %	100 %	100 %	100 %
<i>espF</i> (type III secretions system)	96.62 %	96.62 %	96.62 %	-	-	96.62 %	96.62 %
<i>espJ</i> (prophage-encoded type III secretions system) effector	95.11 %	-	-	-	-	98.17 %	-
<i>gad</i> (glutamate decarboxylase)	99.91 %	99.82 %	99.91 %	99.93 %	99.79 %	99.64 %	99.91 %
<i>iroN</i> (enterobactin siderophore receptor protein)	98.94 %	98.9 %	98.94 %	98.94 %	98.94 %	-	98.94 %
<i>iss</i> (increased serum survival)	99.35 %	99.35 %	99.35 %	99.35 %	99.35 %	-	99.35 %
<i>nleA</i> (non-LEE encoded effector A)	-	100 %	99.84 %	100 %	99.84 %	100 %	99.84 %
<i>nleB</i> (non-LEE encoded effector B)	-	100 %	100 %	100 %	100 %	99.9 %	100 %
<i>stx2A</i> (shiga toxin 2, subunit A, variant f)	-	-	-	-	-	100 %	-
<i>stx2B</i> (shiga toxin 2, subunit B, variant f)	-	-	-	-	-	100 %	-
<i>tccP</i> (tir-cytoskeleton coupling protein)	-	-	-	-	-	98.59 %	-
<i>tir</i> (translocated intimin receptor protein)	99.98 %	99.88 %	99.98 %	99.88 %	99.88 %	99.94 %	99.88 %
<i>vat</i> (vacuolating autotransporter toxin)	99.95 %	99.93 %	99.95 %	99.95 %	99.95 %	99.95 %	99.95 %

Tabell 9. Angir accession number for alle VAGs funnet for de ulike stammene ved bruk av VirulenceFinder.

Virulensfaktorer	Accession number						
	ECA	ECB	ECC	ECD	FHI_NMBU_03	M884Italy	Trh49
<i>astA</i>	-	AB042002	-	-	-	AB042005	-
<i>cif</i>	AY128535	AF497476	AF497476	AF497476	AF497476	AF497476	AF497476
<i>eae</i>	EUCU59503	ECU59503	ECU59503	ECU59503	ECU59503	ECU59503	ECU59503
<i>espA</i>	AJ225016	AJ225016	AJ225016	AJ225016	AJ225016	AJ225016	AJ225016
<i>espF</i>	AE005174	AE005174	AE005174	-	-	AE005174	AE005174
<i>espJ</i>	AB303060	-	-	-	-	AE005174	-
<i>gad</i>	CP001671	FM180568	CP001671	CP002167	FM180568	CP001671	CP001671
<i>iroN</i>	AF449498	AF449498	AF449498	AF449498	AF449498	-	AF449498
<i>iss</i>	DQ381420	DQ381420	DQ381420	DQ381420	DQ381420	-	DQ381420
<i>nleA</i>	-	AB303062	AB303062	AB303062	AB303062	AB303062	AB303062
<i>nleB</i>	-	AB303062	AB303062	AB303062	AB303062	AB303062	AB303062
<i>stx2A</i>	-	-	-	-	-	AB232172	-
<i>stx2B</i>	-	-	-	-	-	AB232172	-
<i>tccP</i>	-	-	-	-	-	AB355995	-
<i>tir</i>	DQ206455	DQ206455	DQ206455	DQ206455	DQ206455	DQ206455	DQ206455
<i>vat</i>	AE014075	AE014075	AE014075	AE014075	AE014075	AE014075	AE014075

Virulensgenene *cif*, *eae*, *espA*, *gad*, *tir* og *vat* ble funnet hos alle de ulike stammene. Virulensgenene som kun ble funnet hos to stammer var *astA* og *espJ*, hvor disse ble funnet hos ECB og M884Italy, og ECA og M884Italy, respektivt. M884Italy var den eneste stammen hvor virulensgener var unike, dette gjaldt genene *tccP*, *stx2A* og *stx2B*. Resterende virulensgener ble funnet hos 5 stammer, med ulike kombinasjoner av stammer for hvert enkelt gen.

3.6 Virulensgener med egendefinert søkt

Alle stammene ble annotert spesifikt for VAGs ved å benytte et nukleotid BLAST-søk med en egendefinert virulensdatabase. Funn fra søket for de ulike stammene ble plukket ut og er presentert i tabell 10 og 11.

Tabell 10. Sammenlikning av VAGs funnet for de ulike stammene ved å benytte et nukleotid BLAST-søk med en egendefinert virulensdatabase. Tabellen viser VAGs fra søket som enten er unike for kun en stamme (markert med gult), eller for flere enn en stamme men som ikke finnes hos alle.

Virulensfaktor	Kommentar	Accession number	Identity query	Coverage	Stammer
A31E_00140	(autotransporter gene)	ASVR01000002	99.72 %	3909 / 3909	ECA
<i>aatA</i>	(APEC autotransporter adhesin A)		98.86 %	3498 / 3498	M884Italy
antigen-43	A self-recognizing adhesin that is associated with cell aggregation and biofilm formation in E. coli.	NZ_BGJU0100062	96.38 %	2847 / 2847	ECB, ECC, FHI_NMBU_03, Trh49
autotransporter FHI_NMBU_03 plasmid	Autotransporter (putative virulence factor) from the virulence plasmid of strain FHI_NMBU_03	CP019456	99.97 %, 100 %, 99.97 %, 100 %, 100 %	3909 / 3909	ECB, ECC, ECD, FHI_NMBU_03, Trh49
Bor lipoprotein	Serum resistance lipoprotein Bor (iss homolog)	AWDN01000013	100 %	294 / 294	ECB, ECC, ECD
BXO92_12800	DUF4917-domain Protein from FHI_NMBU_03	CP019455	100 %	1011 / 1011	ECB, ECC, ECD, M884Italy, FHI_NMBU_03, Trh49
<i>cdtA</i>	Cytolethal distending toxin, subunit A	NZ_NMIQ0100039	98.46 %, 97.06 %, 98.46 %, 99.02 %	714 / 714	ECB, ECC, ECD, M884Italy
<i>cdtB</i>	Cytolethal distending toxin, subunit B	NC_008563	99.76 %, 99.88 %, 99.76 %, 98.42 %	822 / 822	ECB, ECC, ECD, M884Italy
<i>cdtC</i>	Cytolethal distending toxin, subunit C	NZ_CP013025	99.83 %, 100 %, 99.83 %, 100 %	573 / 573	ECB, ECC, ECD, M884Italy
<i>cesAB</i>	Chaperone for the type-III translocator proteins EspA and EspB	FM986651	100 %	324 / 324	ECA, ECC, ECD, M884Italy, FHI_NMBU_03, Trh49
<i>cesF</i>	A 14-kDa protein which specifically interacts with EspF	AJ633129	99.45 %	363 / 363	ECA, ECC, ECD, M884Italy, FHI_NMBU_03, Trh49
<i>cesT</i>	CesT is a bivalent enteropathogenic Escherichia coli chaperone required for translocation of both Tir and Map	LT903847	100 %	471 / 471	ECA, ECC, ECD, M884Italy, FHI_NMBU_03, Trh49
Cyclic β -1,2-glucan synthase	Cyclic β -1,2-glucan is a brucella virulence factor required for intracellular survival. https://doi.org/10.1038/ni1202	NZ_NMDZ0100025	99.85 %, 99.87 %, 99.85 %, 99.86 %, 99.85 %, 99.85 %	8583 / 8583	ECA, ECB, ECC, ECD, FHI_NMBU_03, Trh49
EAST1-toxin (astA-allele2)	Enteraggative Escherichia coli heat-stable enterotoxin 1 (EAST1)	AF143819	100 %, 98.29 %	117 / 117	ECB, M884Italy
ECED1_RS12835	Tellurite resistance	NC_011745	93.53 %	433 / 432	M884Italy
<i>ecf1</i>	The <i>ecf</i> operon has previously been identified in the large enterohemolysin-encoding plasmid of <i>eae</i> -positive Shiga toxin-producing E. coli (STEC). The <i>ecf1</i> gene has been proposed as a marker gene to detect Shiga toxin-producing Escherichia coli in beef samples. doi: 10.4315/0362-028X.JFP-14-417.	AP018692	95.89 %	754 / 822	ECB
ECSE_RS24525	Colicin 1A immunity protein	NC_011419	99.7 %, 91.67 %	336 / 336	ECA, ECB
<i>escC</i>	Outer membrane ring protein	LT903847	99.94 %, 99.94 %, 99.87 %, 99.94 %, 99.94 %, 99.94 %	1539 / 1539	ECA, ECC, ECD, M884Italy, FHI_NMBU_03, Trh49

<i>escD</i>	The proteins forming the inner-membrane T3SS rings, EscJ and EscD, are crucial for the oligomerization of EscC.	BA000007	99.02 %, 99.1 %, 99.1 %, 99.1 %, 99.1 %	1221 / 1221	ECA, ECC, ECD, M884Italy, FHI_NMBU_03, Trh49
<i>escF</i>	Type III secretion system LEE needle major subunit	NC_002695	100 %	222 / 222	ECA, ECC, ECD, M884Italy, FHI_NMBU_03, Trh49
<i>escJ</i>	The proteins forming the inner-membrane T3SS rings, EscJ and EscD, are crucial for the oligomerization of EscC.	CP027103	100 %	573 / 573	ECA, ECC, ECD, M884Italy, FHI_NMBU_03, Trh49
<i>escN</i>	T3SS ATPase that is essential for functional T3 translocation	NC_002695	100 %, 100 %, 100 %, 99.93 %, 100 %, 100 %	1341 / 1341	ECA, ECC, ECD, M884Italy, FHI_NMBU_03, Trh49
<i>escO</i>	Positive Regulator of EscN ATPase Activity of the Enteropathogenic Escherichia coli Injectisome	BA000007	100 %	378 / 378	ECA, ECC, ECD, M884Italy, FHI_NMBU_03, Trh49
<i>escR</i>	Suggested to be the first T3SS protein to localize to the bacterial inner membrane, self-oligomerize, and initiate the assembly of the T3SS complex by interacting with additional export apparatus proteins.	BA000007	99.69 %	654 / 654	ECA, ECC, ECD, M884Italy, FHI_NMBU_03, Trh49
<i>escS</i>	The smallest T3SS export apparatus protein	BA000007	100 %	270 / 270	ECA, ECC, ECD, M884Italy, FHI_NMBU_03, Trh49
<i>escT</i>	LEE1 operon T3SS gene	BA000007	99.61 %	777 / 777	ECA, ECC, ECD, M884Italy, FHI_NMBU_03, Trh49
<i>escU</i>	EscU auto-cleavage in EPEC promotes effector protein translocation into host cells during infection	AP010958	96.15 %	1038 / 1038	ECA, ECC, ECD, M884Italy, FHI_NMBU_03, Trh49
<i>escV</i>	Exported substrates are chaperone-delivered to the translocase, EscV in enteropathogenic Escherichia coli.	NC_002695	99.70 %	2028 / 2028	ECA, ECC, ECD, M884Italy, FHI_NMBU_03, Trh49
<i>espF</i>	Cooperative actions of the EPEC effectors EspF and Map have been reported to mediate tight junction disruption.	BA000007	96.62 %	473 / 747	ECA, ECB, ECC, ECD, M884Italy, Trh49
<i>espJ</i>	The Escherichia coli effector EspJ blocks Src kinase activity via amidation and ADP ribosylation. EspJ effector in enterohemorrhagic E. coli translocates into host mitochondria via an atypical mitochondrial targeting signal.	CP028379	95.11 %	654 / 654	ECA

<i>etgA</i>	The LEE pathogenicity island encodes a lytic transglycosylase (LT) homologue named EtgA. The muramidase EtgA from enteropathogenic Escherichia coli is required for efficient type III secretion.	FM986650	100 %	459 / 459	ECA, ECC, ECD, M884Italy, FHI_NMBU_03, Trh49
<i>etsC</i>	Outer membrane efflux protein EtsC	CP010316	91.51 %, 91.35 %, 91.43 %, 91.43 %, 91.64 %, 91.64 %	1307 / 1308, 1307 / 1308, 1307 / 1308, 1292 / 1308, 1292 / 1308	ECA, ECB, ECC, ECD, FHI_NMBU_03, Trh49
<i>faeC</i>	K88 fimbrial protein A. The K88 fimbrial adhesin enables certain strains of enterotoxigenic Escherichia coli to adhere to the porcine small intestine.	CP012499	95.87 %, 95.39 %, 93.06 %, 93.06 %, 93.85 %, 93.85 %	460 / 561, 456 / 561, 461 / 561, 461 / 561, 439 / 561, 439 / 561	ECA, ECB, ECC, ECD, FHI_NMBU_03, Trh49
<i>faeD</i>	Outer membrane usher protein FaeD. Involved in the export and assembly of K88ab fimbrial subunits across the outer membrane.	CP012499	92.68 %, 93.38 %, 92.97 %, 92.97 %	2432 / 2445, 2448 / 2445, 2432 / 2445, 2432 / 2445	ECA, ECB, ECC, ECD
<i>faeH</i>	K88 minor fimbrial subunit FaeH. K88 minor fimbrial subunit, plays an essential role in the biogenesis of the K88 fimbriae. They enable bacteria to colonize the epithelium of specific host organs.	AWDN01000013	99.25 %, 99.13 %, 98.00 %, 98.00 %, 98.00 %, 98.00 %	801 / 801	ECA, ECB, ECC, ECD, FHI_NMBU_03, Trh49
<i>faeI</i>	K88 minor fimbrial subunit FaeI. F4 or K88 fimbriae are long filamentous polymeric surface proteins of enterotoxigenic Escherichia coli (EPEC).	AWDN01000013	97.39 %, 98.69 %, 94.90 %, 94.90 %, 94.90 %, 94.90 %	765 / 765	ECA, ECB, ECC, ECD, FHI_NMBU_03, Trh49
Ferritin-B	Ferritin is a universal intracellular protein that stores iron and releases it in a controlled fashion	CP019777	100 %	504 / 504	ECD, M884Italy
gene BB546_13115	Locus tag BB546_13115 from NTES01000019	NTES01000019	99.59 %	1461 / 1461	ECC, FHI_NMBU_03
<i>grlA</i>	Transcription of the locus of enterocyte effacement (LEE) genes in enterohemorrhagic Escherichia coli (EHEC) is regulated by the LEE-encoded Ler and GrlR/GrlA proteins.		97.34 %, 97.34 %, 97.34 %, 97.34 %, 97.56 %, 97.56 %	413 / 414, 413 / 414, 413 / 414, 413 / 414, 409 / 414, 409 / 414	ECA, ECC, ECD, M884Italy, FHI_NMBU_03, Trh49
<i>gtrA</i>	A type-IV O-antigen modification gene	AF288197	90.08 %	363 / 363	ECC, FHI_NMBU_03
<i>hbp</i>	Escherichia coli Hemoglobin Protease Autotransporter. Released from E. coli, this serine protease autotransporter degrades hemoglobin and delivers heme to the bacteria.	CP009072	99.95 %, 99.93 %, 99.95 %, 99.95 %, 99.95 %, 99.95 %	4131 / 4131	ECA, ECB, ECC, ECD, M884Italy, FHI_NMBU_03
Hemolysin	A hemolysin from Escherichia Albertii strain B156	ADJH01000002	99.58 %, 99.31 %, 99.27 %, 99.27 %	2607 / 3394	ECA, ECB, ECC, ECD

Hypothetical protein	Hypothetical protein from E. coli strain 317	CP023357	99.90 %	1023 / 1023	ECB, ECC, FHI_NMBU_03, Trh49
<i>ibrA</i>	Immunoglobulin-binding regulator	CP003034	98.91 %, 98.93 %, 98.91 %	1196 / 1185	ECA, ECD, M884Italy
<i>iia</i>	Colicin 1A immunity protein	NC_025198	99.7 %, 91.67 %	336 / 336	ECA, ECB
<i>iroB</i>	Salmochelin siderophore gene	NC_017659	99.73 %, 99.82 %, 99.82 %, 99.82 %, 99.82 %	1115 / 1116	ECA, ECB, ECC, ECD, FHI_NMBU_03, Trh49
<i>iroC</i>	Salmochelin siderophore gene	AF449500	99.73 %, 99.82 %, 99.82 %, 99.82 %, 99.82 %	1115 / 1116	ECA, ECB, ECC, ECD, FHI_NMBU_03, Trh49
<i>iroD</i>	Salmochelin siderophore gene	CP030791	99.59 %, 99.67 %, 99.67 %, 99.67 %, 99.67 %	1230 / 1230	ECA, ECB, ECC, ECD, FHI_NMBU_03, Trh49
<i>iroE</i>	Salmochelin siderophore gene	LT985319	99.69 %	957 / 957	ECA, ECB, ECC, ECD, FHI_NMBU_03, Trh49
<i>iroN</i>	Salmochelin siderophore gene	JX077110	98.94 %, 98.90 %, 98.94 %, 98.94 %, 99.12 %, 99.12 %	2162 / 2178	ECA, ECB, ECC, ECD, FHI_NMBU_03, Trh49
<i>iss</i>	The increased serum survival gene	CP030791	99.35 %	309 / 309	ECA, FHI_NMBU_03, Trh49
<i>iss</i>	The increased serum survival gene (O55:H7 variant)	NC_013941	98.30	294 / 294	ECA
<i>ler</i>	Type III secretion system LEE master regulator	BA000007	99.49 %	390 / 390	ECA, ECC, ECD, M884Italy, FHI_NMBU_03, Trh49
<i>lktB</i>	Leukotoxin translocation ATP-binding protein	AWDN01000013	99.76 %, 99.95 %, 100 %, 100 %, 100 %, 100 %	2112 / 2112	ECA, ECB, ECC, ECD, FHI_NMBU_03, Trh49
<i>matD</i>	Predicted outer membrane protein associated with E. coli common pilus (ECP) formation in pathogenic E. coli strains	HM102365	99.72 %, 99.76 %, 99.76 %, 99.72 %, 99.76 %, 99.76 %	2526 / 2526	ECA, ECC, ECD, M884Italy, FHI_NMBU_03, Trh49
Mpc	Mpc (multiple point controller) protein, also known as CesL. Modulates LEE expression in EHEC.	BA000007	99.72 %	354 / 354	ECA, ECC, ECD, M884Italy, FHI_NMBU_03, Trh49
<i>nleA</i>	NleA plays a role in controlling the host immune response through targeting of NLRP3.	AB303062	100 %, 99.84 %, 100 %, 100 %, 99.94 %, 99.84 %	1282 / 1282	ECB, ECC, ECD, M884Italy, FHI_NMBU_03, Trh49
<i>nleB</i>	The Type III Effectors NleE and NleB from Enteropathogenic E. coli and OspZ from Shigella Block Nuclear Translocation of NF-κB p65	AB303062	100 %, 100 %, 100 %, 99.9 %, 100 %, 100 %	981 / 981	ECB, ECC, ECD, M884Italy, FHI_NMBU_03, Trh49

<i>nleC</i> -like	NleC like gene from CYEL01000033	CYEL01000033	100 %	264 / 264	ECB, ECC, ECD, M884Italy, FHI_NMBU_03, Trh49
<i>nleG</i>	NleG Type 3 effectors from enterohaemorrhagic Escherichia coli are U-Box E3 ubiquitin ligases.	AB303062	100 %, 100 %, 100 %, 99.83 %, 100 %, 100 %	576 / 576	ECB, ECC, ECD, M884Italy, FHI_NMBU_03, Trh49
<i>nleH</i>	NleHs and their homologous effector kinases form a new kinase family within the cluster of eukaryotic-like kinases that includes also Rio, Bud32, and KdoK families.	AB303062	99.75 %	812 / 812	ECB, ECC, ECD, M884Italy, FHI_NMBU_03, Trh49
<i>ompW</i>	Outer membrane protein W	CP027060	90.75 %	400 / 639	FHI_NMBU_03, Trh49
Orf80	Open Reading Frame from a Salmonella Phage	EU570103	96.68 %	241 / 243	ECC, FHI_NMBU_03
<i>paa</i>	Adhesin involved in the early steps of the adhesion mechanism of attaching and effacing Escherichia coli.	BA000007	99.74 %	759 / 759	M884Italy
<i>papC</i>	Part of the pyelonephritis-associated pilus (pap) operon	AWDN01000013	98.43 %, 99.88 %, 99.92 %, 99.92 %, 99.92 %, 99.92 %	2548 / 2547, 2460 / 2460, 2460 / 2460, 2460 / 2460, 2460 / 2460, 2460 / 2460	ECA, ECB, ECC, ECD, FHI_NMBU_03, Trh49
<i>papD</i>	Part of the pyelonephritis-associated pilus (pap) operon	AWDN01000013	95.77 %, 100 %, 100 %, 100 %, 100 %	732 / 732	ECA, ECB, ECC, ECD, FHI_NMBU_03, Trh49
protease-7	Protease-7 from Escherichia species KTE159. Protease that can cleave T7 RNA polymerase, ferric enterobactin receptor protein (FEP), antimicrobial peptide protamine and other proteins.	ASVR01000002	100 %, 99.89 %, 99.79 %, 99.97 %, 99.97 %, 99.97 %	951 / 951	ECA, ECB, ECC, ECD, FHI_NMBU_03, Trh49
<i>rafD</i>	Part of raf operon. The plasmid-borne raf operon of Escherichia coli encodes proteins involved in the uptake and utilisation of the trisaccharide raffinose.	NC_010558	91.51 %	1425 / 1431	M884Italy
<i>rafY</i>	The porin RafY encoded by the raffinose plasmid pRSD2 of Escherichia coli forms a general diffusion pore and not a carbohydrate-specific porin.	U82290	90.37 %	924 / 1395	FHI_NMBU_03, Trh49
Rhs-family protein pseudogene	Rhs-family protein pseudogene from E. coli strain EPEC E2348/69	LT827011	99.09 %, 98.98 %, 98.98 %, 99.09 %, 98.98 %, 98.98 %	883 / 883	ECA, ECB, ECC, M884Italy, FHI_NMBU_03, Trh49
<i>sepQ</i>	Is essential for T3S, interacts with EscN and is associated with the bacterial membrane	CP003109	95.42 %, 95.53 %, 95.53 %, 95.53 %, 95.53 %, 95.53 %	918 / 918	ECA, ECC, ECD, M884Italy, FHI_NMBU_03, Trh49

<i>sfpC</i> (pSFO157)	The <i>sfp</i> cluster, encoding Sfp fimbriae and located in the large plasmid of sorbitol-fermenting (SF) enterohemorrhagic <i>Escherichia coli</i> (EHEC) O157 (pSFO157), has been considered a unique characteristic of this organism.	AF401292	95.85 %	2508 / 2508	M884Italy
<i>stx2f</i>	Shiga-toxin 2, variant f	LN997803	100 %	960 / 960	M884Italy
<i>Tccp2</i>	Tir-cytoskeleton coupling protein	AB253581	95.07 %	507 / 753	M884Italy
<i>traJ</i>	<i>Escherichia coli</i> is a major cause of neonatal bacterial sepsis and meningitis. We recently identified a gene, <i>traJ</i> , which contributes to the ability of <i>E. coli</i> K1 to penetrate the blood-brain barrier in the neonatal rat. DOI: 10.1128/iai.72.1.478-488.2004	DQ381420	100 %, 99.71 %, 98.86 %	701 / 701, 700 / 701, 701 / 701	ECA, ECB, M884Italy
<i>vat</i>	The vacuolating autotransporter toxin (Vat) contributes to uropathogenic <i>Escherichia coli</i> (UPEC) fitness during systemic infection.	KR094935	99.50 %	5651 / 5631	Trh49
XhIA-hemolysin	XhIA is a cell-surface associated haemolysin that lyse the two most prevalent types of insect immune cells (granulocytes and plasmatocytes) as well as rabbit and horse erythrocytes.	LDCR0100004 6	99.73 %	372 / 372	ECC, FHI_NMBU_03
<i>yciR</i>	Hypothetical protein from <i>E. coli</i> FHI_NMBU_03	CP019456	95.61 %, 100 %, 100 %, 100 %, 100 %	638 / 639, 639 / 639, 639 / 639, 639 / 639, 639 / 639	ECA, ECB, ECD, FHI_NMBU_03, Trh49
<i>yecl</i> (ferritin-like protein 2)	Gene encoding a ferritin-like protein (FtnB). Ferritin Mutants of <i>Escherichia coli</i> Are Iron Deficient and Growth Impaired	AE014075	99.80 %, 100 %, 100 %, 100 %	504 / 504	ECA, ECB, ECC, FHI_NMBU_03, Trh49
<i>yfaL</i>	Autotransporter in the AIDA-I family. Its overexpression induce aggregation of cells without a detrimental effect.	CP019455	99.97 %, 99.97 %, 100 %, 100 %, 100 %	3765 / 3765	ECA, ECC, ECD, M884Italy, FHI_NMBU_03, Trh49

11 av de totalt 80 utvalgte virulensfaktorene i tabell 8 var unike for kun en stamme, mens resterende 69 virulensfaktorer var tilstede hos ≥ 2 men < 7 av de totalt 7 ulike stammene. For genet *iss* ble det funnet to genvarianter hos ECA stammen, en variant med 99.35 % match med CP030791 og en variant med 98.35 % match med NC_013941.

Tabell 11. Sammenlikning av VAGs funnet for de ulike stammene ved å benytte et nukleotid BLAST-søk med en egendefinert virulensdatabase. Tabellen viser VAGs fra søket som er til stede hos alle stammene, men hvor det ble oppdaget ulike genvarianter (angitt av ulikt accession number). Virulensfaktorer knyttet til antibiotika resistens er også inkludert.

Virulensfaktorer	Kommentar	Accession number	Identity query	Coverage	Stammer ¹
<i>AcrA</i>	Multidrug efflux pump subunit <i>AcrA</i>	NC_000913	99.16 %	1194 / 1194	
<i>ampC</i>	<i>AmpC</i> β -lactamases (also termed class C or group 1) are typically encoded on the chromosome of many Gram-negative bacteria	CP009072	99.91 %	1134 / 1134	
<i>bcr</i>	Bicyclomycin resistance protein	CP027060	96.39 %, 96.47 %, 96.47 %, 96.47 %, 96.47 %	1191 / 1191	
<i>BlaAMPH</i>	A weak beta-lactamase	AP012030	97.50 %	1158 / 1158	
<i>eae</i> -Beta2	<i>eae</i> subtype Beta2	AJ715407, AJ715407, AJ715407, AB647562, AB647562, AB647562, AB647562	100 %	2820 / 2820, 2524 / 2820, 2820 / 2820, 2820 / 2820, 2820 / 2820, 2820 / 2820, 2820 / 2820	
<i>fsr</i> (Fosmidomycin resistance)	Fosmidomycin resistance protein	CP027060	97.95 %, 98.03 %, 98.03 %, 98.03 %, 98.03 %, 98.03 %	1221 / 1221	
<i>mdfA</i>	Multidrug transporter	Y08743	98.30 %	1233 / 1233	
<i>mdtH</i>	Multidrug resistance protein	CP019777	99.17 %, 99.83 %, 99.83 %, 99.83 %, 99.83 %, 99.83 %	1209 / 1209	

¹ Stammerekkefølgen er: ECA, ECB, ECC, ECD, M884Italy, FHI_NMBU_03, Trh49. Alle verdier oppgitt følger denne rekkefølgen der det er relevant.

Det ble funnet to varianter av *eae*-Beta2 virulensgenet som var tilstede hos alle de 7 ulike stammene (AJ715407 og AB647562). De resterende 6 genene av totalt 7 i tabell 9 var knyttet til antibiotikaresistens, og fantes hos alle stammene.

3.7 Fordeling av stammene innenfor IPEC, ExPEC og patotyper basert på VAGs

VAGs identifisert med VirulenceFinder og egendefinert søk for de ulike stammene, ble benyttet til å undersøke fordelingen av IPEC og ExPEC assosierte VAGs.

Tabell 12. Viser fordeling av IPEC og ExPEC assosierte VAGs for de ulike stammene som ble funnet via virulenssøk (VirulenceFinder og egendefinert søk). IPEC og ExPEC assosierte VAGs er definert i tabell 2. Parentesene angir hvor mange ExPEC/IPEC VAGs som ble funnet.

	ECA	ECB	ECC	ECD	FHI_NMBU_03	M884Italy	Trh49
Fordeling	IPEC (1) ExPEC (6)	IPEC (1) ExPEC (6)	IPEC (1) ExPEC (6)	IPEC (1) ExPEC (6)	IPEC (1) ExPEC (6)	IPEC (2) ExPEC (3)	IPEC (1) ExPEC (7)
IPEC VAGs	<i>eae</i>	<i>eae</i>	<i>eae</i>	<i>eae</i>	<i>eae</i>	<i>eae, stx2</i>	<i>eae</i>
ExPEC VAGs	<i>iroC, fyuA*, papC, traT*, iss, ibeA</i>	<i>iroC, fyuA, papC, traT, iss, ibeA</i>	<i>iroC, fyuA, papC, traT, iss, ibeA</i>	<i>iroC, fyuA, papC, traT, iss, ibeA</i>	<i>iroC, fyuA, papC, traT, iss, ibeA</i>	<i>fyuA, traT, ibeA</i>	<i>vat, iroC, fyuA, papC, traT, iss, ibeA</i>

* VAGs *fyuA* og *traT* er ikke presentert i resultatkapittelet, da disse genene var tilstede hos alle stammene, men finnes i egen Excel-fil (se vedlegg) som omfatter alle dataene fra det egendefinerte virulenssøket. Gjelder for alle stammene.

For ECA, ECB, ECC, ECD og FHI_NMBU_03 ble det funnet 1 IPEC assosiert VAG og 6 ExPEC assosierte VAGs. For M884Italy ble det funnet 2 IPEC assosierte VAGs og 3 ExPEC assosierte VAGs. For Trh49 ble det funnet 1 IPEC assosiert VAG og flest ExPEC assosierte VAGs med 7 ulike.

VAGs identifisert med VirulenceFinder og egendefinert søk for de ulike stammene, ble benyttet til å undersøke fordelingen av VAGs assosiert med patotypene (EPEC, EHEC, EAEC, DAEC, EIEC, ETEC, MNEC, UPEC).

Tabell 13. Viser fordeling av VAGs assosiert med ulike *E. coli* patotyper for de ulike stammene som ble funnet via virulenssøk (VirulenceFinder og egendefinert søk). VAGs assosiert med patotypene er definert i tabell 3. Parentes angir hvor mange VAGs for den spesifikke patotypen som ble funnet.

	ECA	ECB	ECC	ECD	FHI_NMBU_03	M884Italy	Trh49
Fordeling	EPEC (5) EHEC (5) ETEC (4) MNEC (5) UPEC (5)	EPEC (5) EHEC (5) ETEC (4) MNEC (5) UPEC (5)	EPEC (5) EHEC (5) ETEC (4) MNEC (5) UPEC (5)	EPEC (4) EHEC (4) ETEC (4) MNEC (5) UPEC (5)	EPEC (4) EHEC (4) ETEC (3) MNEC (5) UPEC (5)	EPEC (6) EHEC (7) MNEC (5) UPEC (2)	EPEC (5) EHEC (5) ETEC (3) MNEC (5) UPEC (5)
EPEC/EHEC VAGs	<i>eaec, cif, espF, tir, map*</i>	<i>eaec, cif, espF, tir, map</i>	<i>eaec, cif, espF, tir, map</i>	<i>eaec, cif, tir, map</i>	<i>eaec, cif, tir, map</i>	<i>eaec, cif, stx2, espF, tir, paa, map</i>	<i>eaec, cif, espF, tir, map</i>
EAEC VAGs							
DAEC VAGs							
EIEC VAGs							
ETEC VAGs	<i>faec, faeD, faeH, fael</i>	<i>faec, faeD, faeH, fael</i>	<i>faec, faeD, faeH, fael</i>	<i>FaeC, faeD, faeH, fael</i>	<i>faec, faeH, fael</i>		<i>faec, faeH, fael</i>
MNEC VAGs	<i>ompA*, ibeA*, ibeB*, ibeC*, asIA*</i>	<i>ompA, ibeA, ibeB, ibeC, asIA</i>	<i>ompA, ibeA, ibeB, ibeC, asIA</i>	<i>ompA, ibeA, ibeB, ibeC, asIA</i>	<i>ompA, ibeA, ibeB, ibeC, asIA</i>	<i>ompA, ibeA, ibeB, ibeC, asIA</i>	<i>ompA, ibeA, ibeB, ibeC, asIA</i>
UPEC VAGs	<i>iroN, papC, papD, yfcV*, usp*</i>	<i>iroN, papC, papD, yfcV, usp</i>	<i>iroN, papC, papD, yfcV, usp</i>	<i>iroN, papC, papD, yfcV, usp</i>	<i>iroN, papC, papD yfcV, usp</i>	<i>yfcV, usp</i>	<i>iroN, papC, papD, yfcV, usp</i>

* VAGs *map, ompA, ibeA, ibeB, ibeC, asIA, yfcV, usp* er ikke presentert i resultatkapittelet, da disse genene var tilstede hos alle stammene, men finnes i egen Excel-fil som omfatter alle dataene fra det egendefinerte virulenssøket. Gjelder for alle stammene.

For ECA, ECB, ECC og Trh49 ble det funnet 5 VAGs assosiert med EPEC, for ECD og FHI_NMBU_03 ble det funnet 4, og for M884Italy ble det funnet 6. EHEC assosierte VAGs var samme som for EPEC for stammene, med unntak av for M884Italy hvor også shigatoksin type 2 (*stx2*) ble funnet. ETEC assosierte VAGs som ble funnet var CFAs *faec, faeD, faeH* og *fael* for ECA, ECB, ECC og ECD. For FHI_NMBU_03 og Trh49 ble det funnet *faec, faeH* og *fael*, mens for M884Italy ble det ikke funnet ETEC assosierte VAGs. Det ble funnet 5 MNEC assosierte VAGs for alle stammene. 5 UPEC assosierte VAGs ble funnet for alle stammene, med unntak av M884Italy hvor det kun ble funnet *usp* og *yfcv* (type 1 fimbriae). Det ble ikke funnet noen VAGs assosiert med EAEC og DAEC.

3.8 Protein BLAST-søk for virulensgenene *eae*, *tir* og *vat*

Aminosyresekvensene for virulensgenene *eae*, *tir* og *vat* funnet hos de ulike stammene ved å benytte VirulenceFinder, ble analysert med verktøyet BLAST for å finne lignende proteiner med best match i databasene.

Tabell 14. Viser funn fra protein BLAST-søk for virulensgene *eae*, *tir* og *vat* funnet ved bruk av VirulenceFinder.

BlastP							
	Virulensfaktorer	Coverage	Identity	Positives	Gaps	Kommentar	Sequence ID
ECA	<i>eae</i>	939/939	100 %	100 %	0 %	intimin type beta	WP_024223565.1
	<i>tir</i>	549/549	100 %	100 %	0 %	type III secretion system LEE translocated intimin receptor Tir	WP_024240669.1
	<i>vat</i>	1376/1376	100 %	100 %	0 %	vacuolating autotransporter toxin Vat	WP_033552377.1
ECB	<i>eae</i>	839/839	100 %	100 %	0 %	intimin type beta	WP_024223565.1
	<i>tir</i>	549/549	100 %	100 %	0 %	type III secretion system LEE translocated intimin receptor Tir	WP_024240669.1
	<i>vat</i>	1375/1376	99 %	100 %	0 %	vacuolating autotransporter toxin Vat	WP_024223493.1
ECC	<i>eae</i>	939/939	100 %	100 %	0 %	intimin type beta	WP_024223565.1
	<i>tir</i>	549/549	100 %	100 %	0 %	type III secretion system LEE translocated intimin receptor Tir	WP_024240669.1
	<i>vat</i>	1376/1376	100 %	100 %	0 %	vacuolating autotransporter toxin Vat	WP_024223493.1
ECD	<i>eae</i>	939/939	100 %	100 %	0 %	intimin type beta	WP_024223565.1
	<i>tir</i>	549/549	100 %	100 %	0 %	type III secretion system LEE translocated intimin receptor Tir	WP_024240669.1
	<i>vat</i>	1376/1376	100 %	100 %	0 %	vacuolating autotransporter toxin Vat	WP_024223493.1
FHI_NMBU_03	<i>eae</i>	939/939	100 %	100 %	0 %	intimin type beta	WP_024223565.1

	<i>tir</i>	549/549	100 %	100 %	0 %	type III secretion system LEE translocated intimin receptor Tir	WP_024240669.1
	<i>vat</i>	1376/1376	100 %	100 %	0 %	vacuolating autotransporter toxin Vat	WP_024223493.1
M884Italy	<i>eae</i>	939/939	100 %	100 %	0 %	intimin type beta	WP_024223565.1
	<i>tir</i>	549/549	100 %	100 %	0 %	type III secretion system LEE translocated intimin receptor Tir	WP_024223567.1
	<i>vat</i>	1376/1376	100 %	100 %	0 %	vacuolating autotransporter toxin Vat	WP_024223493.1
Trh49	<i>eae</i>	939/939	100 %	100 %	0 %	intimin type beta	WP_024223565.1
	<i>tir</i>	549/549	100 %	100 %	0 %	type III secretion system LEE translocated intimin receptor Tir	WP_024240669.1
	<i>vat</i>	1376/1376	100 %	100 %	0 %	vacuolating autotransporter toxin Vat	WP_024223493.1

Alle de tre virulente proteinene som ble funnet matchet 100 % med den beste matchen i databasereferansen, bortsett fra Vat-proteinet for ECB. For M884Italy var Tir-proteinet av en ulik variant sammenlikning med de andre stammene med en sequence id som matchet 100 % med WP_024223567.1. For ECA var Vat-proteinet av en ulik variant sammenliknet med de andre stammene med en sequence id som matchet 100 % med WP_033552377.1.

3.9 ClustalOmega sequence alignment for virulensgenene *eae*, *tir* og *vat* på nukleinsyrenivå

Virulensgenene *eae*, *tir* og *vat* som ble identifisert for de ulike stammene ved bruk av VirulenceFinder, ble undersøkt på nukleinsyrenivå med hensikt å sammenlikne de for forskjeller ved å benytte Clustal Omega Multiple Sequence Alignment.

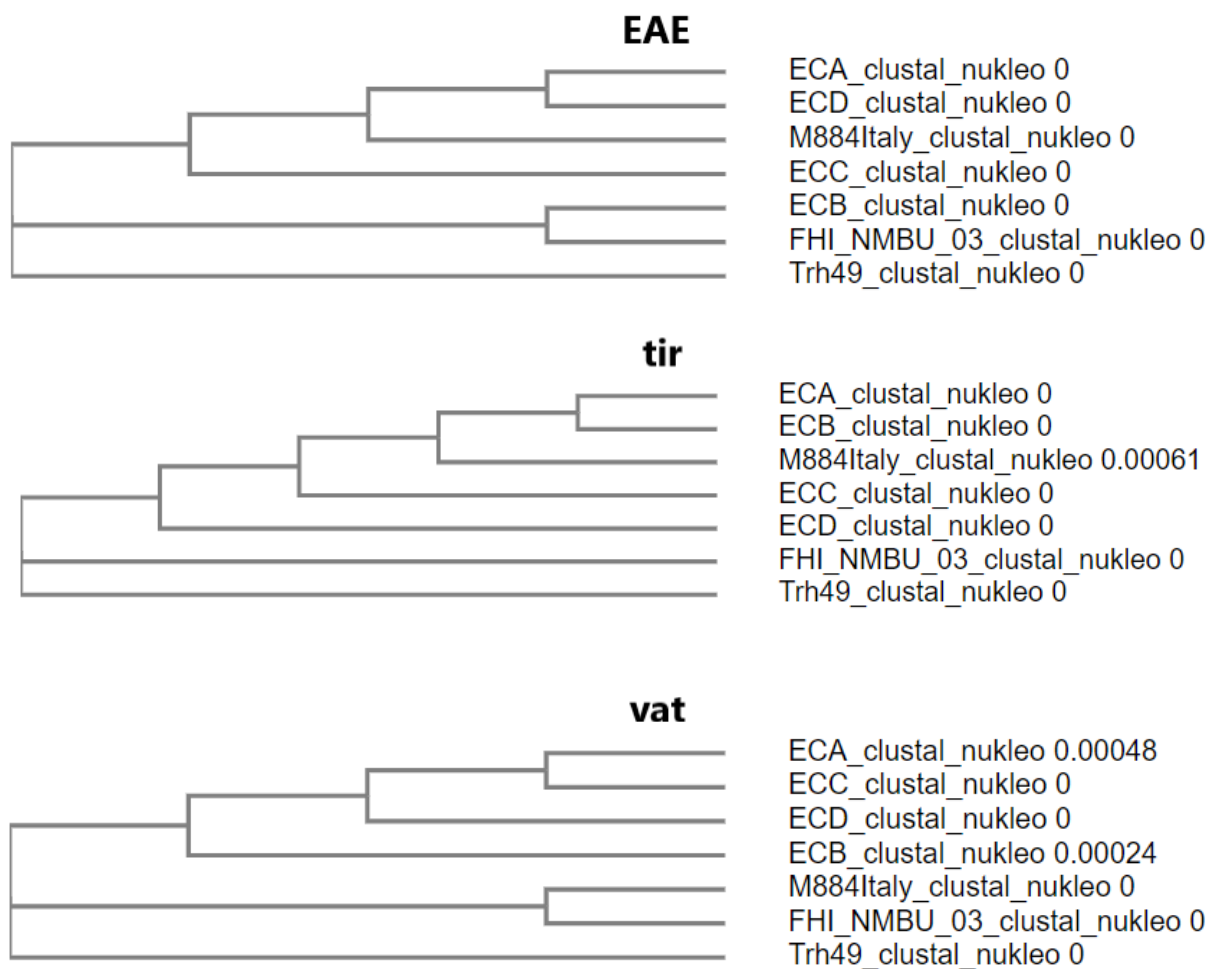
EAE		
ECA_clustal_nukleo	ATGATTACTCATGGTTGTTATTCCCGGACCCGGCACAAGCATAAGCTAAAAAAACATTT	60
ECD_clustal_nukleo	ATGATTACTCATGGTTGTTATTCCCGGACCCGGCACAAGCATAAGCTAAAAAAACATTT	60
M884Italy_clustal_nukleo	ATGATTACTCATGGTTGTTATTCCCGGACCCGGCACAAGCATAAGCTAAAAAAACATTT	60
ECB_clustal_nukleo	-----	0
ECC_clustal_nukleo	ATGATTACTCATGGTTGTTATTCCCGGACCCGGCACAAGCATAAGCTAAAAAAACATTT	60
FHI_NMBU_03_clustal_nukleo	ATGATTACTCATGGTTGTTATTCCCGGACCCGGCACAAGCATAAGCTAAAAAAACATTT	60
Trh49_clustal_nukleo	ATGATTACTCATGGTTGTTATTCCCGGACCCGGCACAAGCATAAGCTAAAAAAACATTT	60
ECA_clustal_nukleo	ATTATGCTTAGTGCTGGTTTAGGATTGTTTTTTATGTTAATCAGAACTCATTGCAAAAT	120
ECD_clustal_nukleo	ATTATGCTTAGTGCTGGTTTAGGATTGTTTTTTATGTTAATCAGAACTCATTGCAAAAT	120
M884Italy_clustal_nukleo	ATTATGCTTAGTGCTGGTTTAGGATTGTTTTTTATGTTAATCAGAACTCATTGCAAAAT	120
ECB_clustal_nukleo	-----	0
ECC_clustal_nukleo	ATTATGCTTAGTGCTGGTTTAGGATTGTTTTTTATGTTAATCAGAACTCATTGCAAAAT	120
FHI_NMBU_03_clustal_nukleo	ATTATGCTTAGTGCTGGTTTAGGATTGTTTTTTATGTTAATCAGAACTCATTGCAAAAT	120
Trh49_clustal_nukleo	ATTATGCTTAGTGCTGGTTTAGGATTGTTTTTTATGTTAATCAGAACTCATTGCAAAAT	120
ECA_clustal_nukleo	GGTGAAAAATTTTAAATGGGTTCCGGATTCAAACGTGTTAACTCATAATAGCTATCAG	180
ECD_clustal_nukleo	GGTGAAAAATTTTAAATGGGTTCCGGATTCAAACGTGTTAACTCATAATAGCTATCAG	180
M884Italy_clustal_nukleo	GGTGAAAAATTTTAAATGGGTTCCGGATTCAAACGTGTTAACTCATAATAGCTATCAG	180
ECB_clustal_nukleo	-----	0
ECC_clustal_nukleo	GGTGAAAAATTTTAAATGGGTTCCGGATTCAAACGTGTTAACTCATAATAGCTATCAG	180
FHI_NMBU_03_clustal_nukleo	GGTGAAAAATTTTAAATGGGTTCCGGATTCAAACGTGTTAACTCATAATAGCTATCAG	180
Trh49_clustal_nukleo	GGTGAAAAATTTTAAATGGGTTCCGGATTCAAACGTGTTAACTCATAATAGCTATCAG	180
ECA_clustal_nukleo	AATCGCCTTTTTTATACGTTGAAAAACAGGTGAAACTGTTGCCGATCTTCTAAATCGCAA	240
ECD_clustal_nukleo	AATCGCCTTTTTTATACGTTGAAAAACAGGTGAAACTGTTGCCGATCTTCTAAATCGCAA	240
M884Italy_clustal_nukleo	AATCGCCTTTTTTATACGTTGAAAAACAGGTGAAACTGTTGCCGATCTTCTAAATCGCAA	240
ECB_clustal_nukleo	-----	0
ECC_clustal_nukleo	AATCGCCTTTTTTATACGTTGAAAAACAGGTGAAACTGTTGCCGATCTTCTAAATCGCAA	240
FHI_NMBU_03_clustal_nukleo	AATCGCCTTTTTTATACGTTGAAAAACAGGTGAAACTGTTGCCGATCTTCTAAATCGCAA	240
Trh49_clustal_nukleo	AATCGCCTTTTTTATACGTTGAAAAACAGGTGAAACTGTTGCCGATCTTCTAAATCGCAA	240
ECA_clustal_nukleo	GATATTAATTTATCGACGATTTGGTTCGTTGAATAAGCATTATACAGTTCTGAAAGCGAA	300
ECD_clustal_nukleo	GATATTAATTTATCGACGATTTGGTTCGTTGAATAAGCATTATACAGTTCTGAAAGCGAA	300
M884Italy_clustal_nukleo	GATATTAATTTATCGACGATTTGGTTCGTTGAATAAGCATTATACAGTTCTGAAAGCGAA	300
ECB_clustal_nukleo	-----	0
ECC_clustal_nukleo	GATATTAATTTATCGACGATTTGGTTCGTTGAATAAGCATTATACAGTTCTGAAAGCGAA	300
FHI_NMBU_03_clustal_nukleo	GATATTAATTTATCGACGATTTGGTTCGTTGAATAAGCATTATACAGTTCTGAAAGCGAA	300

tir		
ECA_clustal_nukleo	GCAACAAAAGATCAGTTAACGCAAGAAGCATTCAAGGACCC TGAGAACCAGAAAGTTAAC	900
ECB_clustal_nukleo	GCAACAAAAGATCAGTTAACGCAAGAAGCATTCAAGGACCC TGAGAACCAGAAAGTTAAC	900
ECC_clustal_nukleo	GCAACAAAAGATCAGTTAACGCAAGAAGCATTCAAGGACCC TGAGAACCAGAAAGTTAAC	900
ECD_clustal_nukleo	GCAACAAAAGATCAGTTAACGCAAGAAGCATTCAAGGACCC TGAGAACCAGAAAGTTAAC	900
FHI_NMBU_03_clustal_nukleo	GCAACAAAAGATCAGTTAACGCAAGAAGCATTCAAGGACCC TGAGAACCAGAAAGTTAAC	900
Trh49_clustal_nukleo	GCAACAAAAGATCAGTTAACGCAAGAAGCATTCAAGGACCC TGAGAACCAGAAAGTTAAC	900
M884Italy_clustal_nukleo★	GCAACAAAAGATCAGTTAACGCAAGAAGCATTCAAGGACTCTGAGAACCAGAAAGTTAAC	900

vat		
ECA_clustal_nukleo★	AGTGATAAAACGGCTAATATCACTCTGGATTATCAGACGCATCCGGCAGACGTAAGTT	1620
ECB_clustal_nukleo	AGTGATAAAACGGCTAATATCACTCTGGATTATCAGACGCATCCGGCAGACGTAAGTT	1620
ECC_clustal_nukleo	AGTGATAAAACGGCTAATATCACTCTGGATTATCAGACGCATCCGGCAGACGTAAGTT	1620
ECD_clustal_nukleo	AGTGATAAAACGGCTAATATCACTCTGGATTATCAGACGCATCCGGCAGACGTAAGTT	1620
M884Italy_clustal_nukleo	AGTGATAAAACGGCTAATATCACTCTGGATTATCAGACGCATCCGGCAGACGTAAGTT	1620
FHI_NMBU_03_clustal_nukleo	AGTGATAAAACGGCTAATATCACTCTGGATTATCAGACGCATCCGGCAGACGTAAGTT	1620
Trh49_clustal_nukleo	AGTGATAAAACGGCTAATATCACTCTGGATTATCAGACGCATCCGGCAGACGTAAGTT	1620

vat		
ECA_clustal_nukleo	CCCATGGTTGATTTTAGCAGTGTGAATGTAAGCTCAAATCCCGGCGTTGCAACATTAATT	360
ECB_clustal_nukleo★	CCCATGGTTGATTTTAGCAGTGTGAATGTAAGCTCAAATCCCGGCGTTGCAACATTAATT	360
ECC_clustal_nukleo	CCCATGGTTGATTTTAGCAGTGTGAATGTAAGCTCAAATCCCGGCGTTGCAACATTAATT	360
ECD_clustal_nukleo	CCCATGGTTGATTTTAGCAGTGTGAATGTAAGCTCAAATCCCGGCGTTGCAACATTAATT	360
M884Italy_clustal_nukleo	CCCATGGTTGATTTTAGCAGTGTGAATGTAAGCTCAAATCCCGGCGTTGCAACATTAATT	360
FHI_NMBU_03_clustal_nukleo	CCCATGGTTGATTTTAGCAGTGTGAATGTAAGCTCAAATCCCGGCGTTGCAACATTAATT	360
Trh49_clustal_nukleo	CCCATGGTTGATTTTAGCAGTGTGAATGTAAGCTCAAATCCCGGCGTTGCAACATTAATT	360

Figur 10. Viser utdrag av resultatene fra ClustalOmega sequence alignment, hvor virulensgenene eae, tir og vat ble analysert på nukleinsyrenivå. Figuren viser kun områder hvor det ble funnet en forskjell angitt med rød pil (sekvensen), og sort stjerne for stammen forskjellen ble funnet for. For resten av sekvensene (ikke inkludert) hadde disse en likhet på 100 %.



Figur 11. Fylogenetiske trær produsert ved å bruke verktøyet *ClustalOmega sequence alignment*, hvor virulensgenene *eae*, *tir* og *vat* ble analysert på nukleinsyrenivå for de ulike stammene.

ClustalOmega analysene viser at *eae*, *tir* og *vat* sekvensene på nukleinsyrenivå var svært like hverandre mellom de ulike stammene. En markant forskjell var ECB stammen med hensyn til *eae*, hvor de første 300 nukleotidene i genet manglende sammenliknet med sekvensene for de andre stammene. Andre ulikheter var for enkeltbaser i genene; *tir* hvor M884Italy hadde en tyrosin istedenfor en cytosin, genet *vat* hvor ECA hadde en guanin istedenfor adenin og ECB en adenin istedenfor guanin.

3.10 ClustalOmega sequence alignment for virulensgenene *eae*, *tir* og *vat* på aminosyrenivå.

Virulensgenene *eae*, *tir* og *vat* som ble identifisert for de ulike stammene ved bruk av VirulenceFinder, ble undersøkt på aminosyrenivå med hensikt å sammenlikne de for forskjeller ved å benytte Clustal Omega Multiple Sequence Alignment.

EAE

Trh49_clustal_aa	MITHGCYSRTRHKHKLKKTFFIMLSAGLGLFFYVQNQSFANGENYFKLGSDSKLLTHNSYQ	60
FHI_NMBU_03_clustal_aa	MITHGCYSRTRHKHKLKKTFFIMLSAGLGLFFYVQNQSFANGENYFKLGSDSKLLTHNSYQ	60
M884Italy_clustal_aa	MITHGCYSRTRHKHKLKKTFFIMLSAGLGLFFYVQNQSFANGENYFKLGSDSKLLTHNSYQ	60
ECD_clustal_aa	MITHGCYSRTRHKHKLKKTFFIMLSAGLGLFFYVQNQSFANGENYFKLGSDSKLLTHNSYQ	60
ECC_clustal_aa	MITHGCYSRTRHKHKLKKTFFIMLSAGLGLFFYVQNQSFANGENYFKLGSDSKLLTHNSYQ	60
ECA_clustal_aa	MITHGCYSRTRHKHKLKKTFFIMLSAGLGLFFYVQNQSFANGENYFKLGSDSKLLTHNSYQ	60
ECB_clustal_aa	-----	0

Trh49_clustal_aa	NRLFYTLKTGETVADLSKSDINLSTIWSLNKHLYSSEEMMKAAPGQQIILPLKLLPFE	120
FHI_NMBU_03_clustal_aa	NRLFYTLKTGETVADLSKSDINLSTIWSLNKHLYSSEEMMKAAPGQQIILPLKLLPFE	120
M884Italy_clustal_aa	NRLFYTLKTGETVADLSKSDINLSTIWSLNKHLYSSEEMMKAAPGQQIILPLKLLPFE	120
ECD_clustal_aa	NRLFYTLKTGETVADLSKSDINLSTIWSLNKHLYSSEEMMKAAPGQQIILPLKLLPFE	120
ECC_clustal_aa	NRLFYTLKTGETVADLSKSDINLSTIWSLNKHLYSSEEMMKAAPGQQIILPLKLLPFE	120
ECA_clustal_aa	NRLFYTLKTGETVADLSKSDINLSTIWSLNKHLYSSEEMMKAAPGQQIILPLKLLPFE	120
ECB_clustal_aa	-----MMKAAPGQQIILPLKLLPFE	20

tir

ECA_clustal_aa	IAAGLAGLAATGIAQALAL TPEPDDPTTTDPDQAANAASATKDQLTQEAFKDPENQKVN	300
ECB_clustal_aa	IAAGLAGLAATGIAQALAL TPEPDDPTTTDPDQAANAASATKDQLTQEAFKDPENQKVN	300
ECC_clustal_aa	IAAGLAGLAATGIAQALAL TPEPDDPTTTDPDQAANAASATKDQLTQEAFKDPENQKVN	300
ECD_clustal_aa	IAAGLAGLAATGIAQALAL TPEPDDPTTTDPDQAANAASATKDQLTQEAFKDPENQKVN	300
FHI_NMBU_03_clustal_aa	IAAGLAGLAATGIAQALAL TPEPDDPTTTDPDQAANAASATKDQLTQEAFKDPENQKVN	300
Trh49_clustal_aa	IAAGLAGLAATGIAQALAL TPEPDDPTTTDPDQAANAASATKDQLTQEAFKDPENQKVN	300
M884Italy_clustal_aa	IAAGLAGLAATGIAQALAL TPEPDDPTTTDPDQAANAASATKDQLTQEAFKDPENQKVN	300

vat

ECA_clustal_aa	NNPYQTYRDF AENKGFQAGATNIP IFNNGELVGHLDKAPMVD FSSVNVSSNPGVATLI	120
ECB_clustal_aa	NNPYQTYRDF AENKGFQAGATNIP IFNNGELVGHLDKAPMVD FSSVNVSSNPGVATLI	120
ECC_clustal_aa	NNPYQTYRDF AENKGFQAGATNIP IFNNGELVGHLDKAPMVD FSSVNVSSNPGVATLI	120
ECD_clustal_aa	NNPYQTYRDF AENKGFQAGATNIP IFNNGELVGHLDKAPMVD FSSVNVSSNPGVATLI	120
M884Italy_clustal_aa	NNPYQTYRDF AENKGFQAGATNIP IFNNGELVGHLDKAPMVD FSSVNVSSNPGVATLI	120
FHI_NMBU_03_clustal_aa	NNPYQTYRDF AENKGFQAGATNIP IFNNGELVGHLDKAPMVD FSSVNVSSNPGVATLI	120
Trh49_clustal_aa	NNPYQTYRDF AENKGFQAGATNIP IFNNGELVGHLDKAPMVD FSSVNVSSNPGVATLI	120

vat

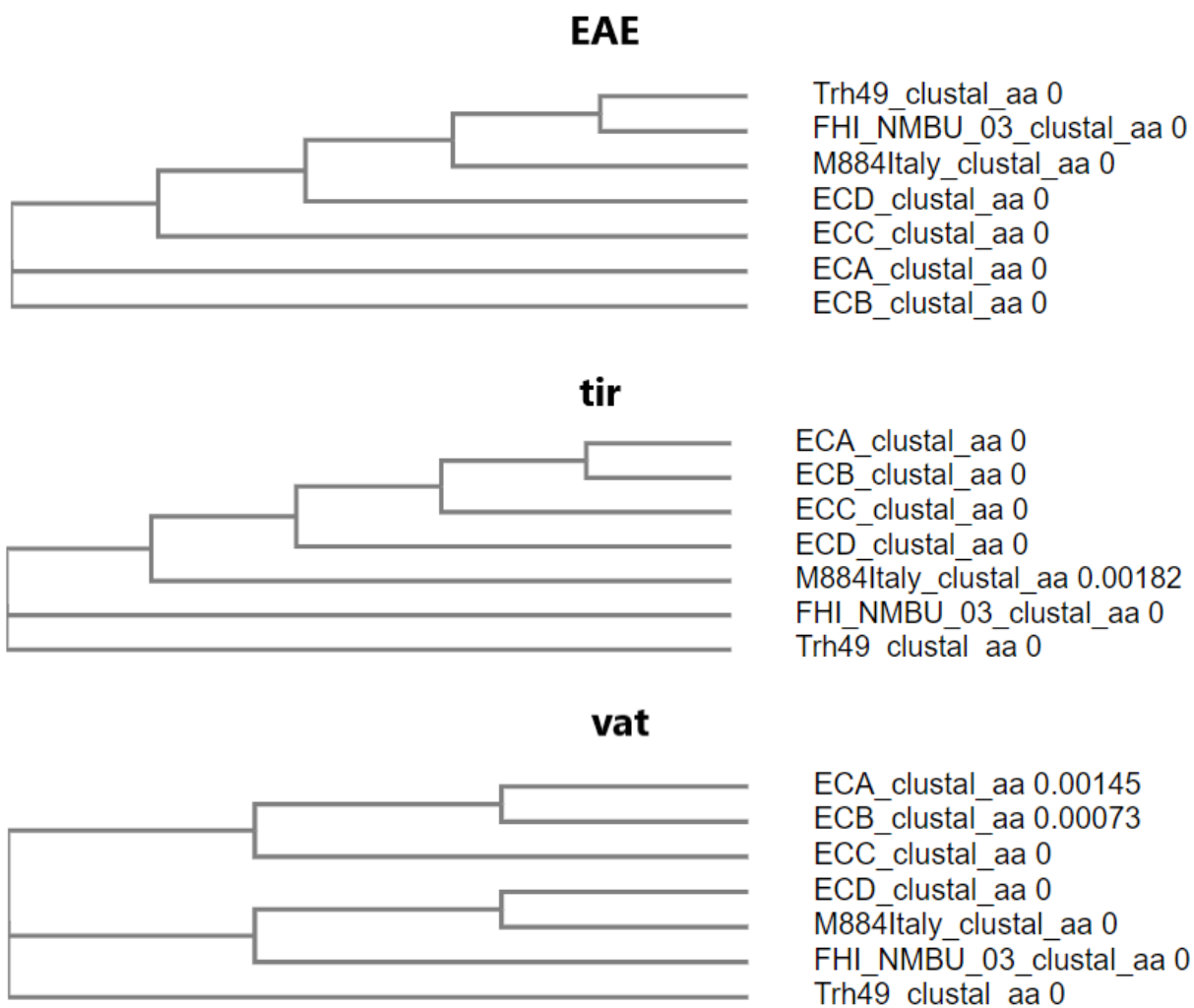
ECA_clustal_aa	QVNPNDISWGYRGGVLDVNGNDLTFHKLNAADYGATLGNSSDKTANITLDYQTRPADVKV	540
ECB_clustal_aa	QVNPNDISWGYRGGVLDVNGNDLTFHKLNAADYGATLGNSSDKTANITLDYQTRPADVKV	540
ECC_clustal_aa	QVNPNDISWGYRGGVLDVNGNDLTFHKLNAADYGATLGNSSDKTANITLDYQTRPADVKV	540
ECD_clustal_aa	QVNPNDISWGYRGGVLDVNGNDLTFHKLNAADYGATLGNSSDKTANITLDYQTRPADVKV	540
M884Italy_clustal_aa	QVNPNDISWGYRGGVLDVNGNDLTFHKLNAADYGATLGNSSDKTANITLDYQTRPADVKV	540
FHI_NMBU_03_clustal_aa	QVNPNDISWGYRGGVLDVNGNDLTFHKLNAADYGATLGNSSDKTANITLDYQTRPADVKV	540
Trh49_clustal_aa	QVNPNDISWGYRGGVLDVNGNDLTFHKLNAADYGATLGNSSDKTANITLDYQTRPADVKV	540

vat

ECA_clustal_aa	WDSGMDVSMRRNSVNPLVGRGTGVVSGKTFSGKDWSLTARAGLHYEFDLTDSADVHLKDA	1320
ECB_clustal_aa	WDSGMDVSMRRNSVNPLVGRGTGVVSGKTFSGKDWSLTARAGLHYEFDLTDSADVHLKDA	1320
ECC_clustal_aa	WDSGMDVSMRRNSVNPLVGRGTGVVSGKTFSGKDWSLTARAGLHYEFDLTDSADVHLKDA	1320
ECD_clustal_aa	WDSGMDVSMRRNSVNPLVGRGTGVVSGKTFSGKDWSLTARAGLHYEFDLTDSADVHLKDA	1320
M884Italy_clustal_aa	WDSGMDVSMRRNSVNPLVGRGTGVVSGKTFSGKDWSLTARAGLHYEFDLTDSADVHLKDA	1320
FHI_NMBU_03_clustal_aa	WDSGMDVSMRRNSVNPLVGRGTGVVSGKTFSGKDWSLTARAGLHYEFDLTDSADVHLKDA	1320
Trh49_clustal_aa	WDSGMDVSMRRNSVNPLVGRGTGVVSGKTFSGKDWSLTARAGLHYEFDLTDSADVHLKDA	1320

* *****

Figur 12. Viser utdrag av resultatene fra ClustalOmega sequence alignment, hvor virulensgenene eae, tir og vat ble analysert på aminosyrenivå. Figuren viser kun områder hvor det ble funnet en forskjell angitt med rød pil (sekvensen), og sort stjerne for stammen forskjellen ble funnet for. For resten av sekvensene (ikke inkludert) hadde disse en likhet på 100 %.

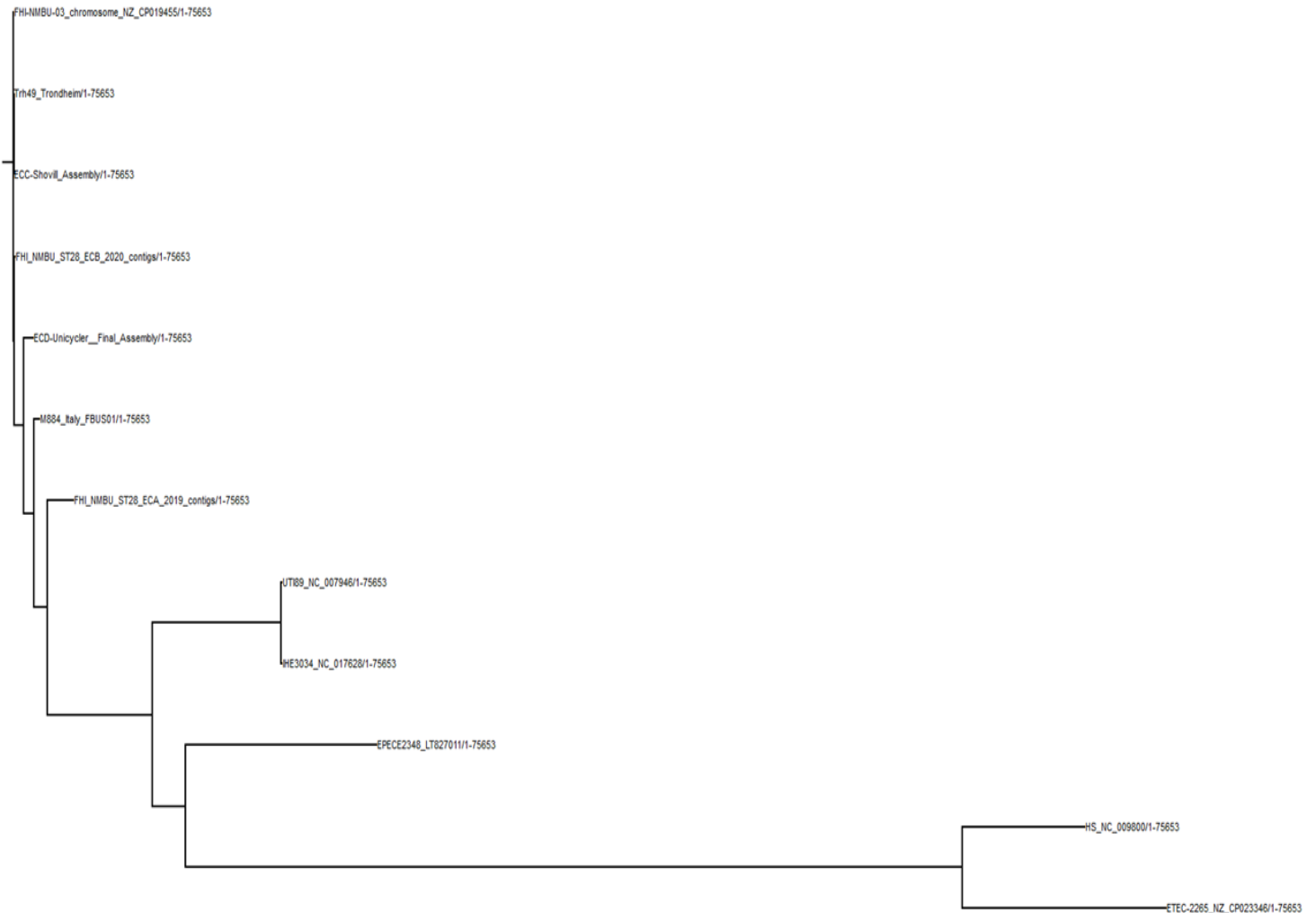


Figur 13. Fylogenetiske trær produsert ved å bruke verktøyet ClustalOmega sequence alignment, hvor virulensgenene eae, tir og vat ble analysert på aminosyrenivå for de ulike stammene.

ClustalOmega analysene viser at Eae, Tir og Vat sekvensene på aminosyrenivå var svært like hverandre mellom de ulike stammene. Likt som for analysene på nukleinsyrenivå, viser denne analysen at ECB for proteinet intimin har en 100 aminosyre kortere sekvens fra starten. Andre ulikheter var for M884Italy med hensyn til proteinet tir hvor denne sekvensen hadde ett tilfelle av aminosyren serin istedenfor prolin (aminosyrer med ulike egenskaper angitt av blankt felt). Med hensyn til Vat-proteinet var det ett tilfelle av aminosyren aspargin istedenfor serin for ECB (aminosyrer med svakt lignende egenskaper, angitt av et punktum), ett tilfelle av arginin istedenfor histidin for ECA (aminosyrer med sterkt lignende egenskaper, angitt av et kolon), ett tilfelle av aminosyren serin istedenfor aspargin for ECA (aminosyrer med svakt lignende egenskaper, angitt av et punktum).

3.11 SNP-analyse av de ulike stammene

Det ble utført en SNP-analyse av de ulike stammene ved å bruke CSIPhylogeny med *E. coli* K-12 som referansestamme for å få en oversikt over slektskap mellom stammene.



Figur 14. SNP-tre som viser slektskapet mellom de ulike stammene, hvor *E. coli* K-12 MG1655 er benyttet som referansestamme. Treet er generert ved å benytte verktøyet CSIPhylogeny.

SNP-analysen viser at stammene FHI_NMBU_03, Trh49, ECC og ECB er nærmest beslektet, også med referansestammen. Deretter er disse stammene nærmere beslektet med følgende stammer i rekkefølgen: ECD > M884Italy > ECA.

4.0 Diskusjon

Vacuolating autotransporter toxin (Vat) tilhører en undergruppe av autotransportere innenfor *Enterobacteriaceae* som blir omtalt som SPATEs (serine protease autotransporters of *Enterobacteriaceae*) (Pokharel et al., 2019). SPATEs er en viktig virulens- og overlevelsesmekanisme hos *E. coli*, hvor disse proteasene transporteres via type Va sekresjonsystem og har ulike biologiske virkninger slik som adheranse og cytotoxisk virkning. Vat spiller en rolle blant annet i ekstraintestinal infeksjon med UPEC, hvor UPEC er den vanligste årsaken til urinveisinfeksjoner (Nichols et al., 2016). Urinveisinfeksjoner er en av de mest vanlige infeksjonstypene, hvor særlig kvinner blir rammet (Foxman, 2010). Urinveisinfeksjoner kan involvere anatomiske systemer slik som blære og nyrer, men kan også videreutvikle seg til alvorlig urosepsis. Interessant er det at enkelte av de ulike SPATEs har en høy grad av aminosyresekvens homologi i form av konserverte områder og likheter i aminosyresekvensen, selv om de funksjonelle egenskapene kan være nokså ulike (Nichols et al., 2016; Pokharel et al., 2019). Vat som er assosiert med UPEC og APEC (avian pathogenic *Escherichia coli*), har en 78 % sekvenslikhet med SPATEs Tsh og Hbp, hvor Tsh og Hbp hovedsakelig er assosiert med APEC (*E. coli* infeksjoner hos fugler) (Pokharel et al., 2019). Tsh og Hbp har derimot en sekvenslikhet med hverandre som omtrent er identisk (>99 %). Den høye homologien mellom disse SPATEene gjør at disse er utsatt for å bli feilannotert og annotert ulikt avhengig av hvilket annoteringsverktøy man benytter seg av. Alle stammene analysert i denne masteroppgaven ble ulikt annotert med annoteringsverktøyene VirulenceFinder og Prokka, hvor VirulenceFinder annoterte sekvensene som Vat mens Prokka som Hbp. BLAST av aminosyresekvensene viste derimot at sekvensene kodet for Vat med 100 % match for alle stammene. Flere tilfeller i litteraturen har derimot vist at Vat, Tsh og Hbp har blitt forvirret, noe som belyses i oversiktsartikkelen av Pokharel et al (Pokharel et al., 2019). I denne artikkelen nevnes det at Vat feilaktig har blitt annotert som Tsh og Hbp i studier av UPEC stammen CFT073 (Heimer et al., 2004; Parham et al., 2005; Welch et al., 2002). I oversiktsartikkelen beskrives det videre at det er problematisk at ukarakteriserte SPATEs blir annotert som spesifikke karakteriserte SPATEs selv om sekvenslikheten ikke er tilstrekkelig og eksperimentelle forsøk mangler (Pokharel et al., 2019).

Helgenomsekvensering er i dag mulig å utføre effektivt med fremtredende NGS teknologi og tredje generasjons nanopore-sekvensering til stadig lavere kostnader og tidsbruk. Mulighetene som oppstår med høyteknologiske sekvenseringsmetoder sammen med lett tilgjengelige bioinformatiske verktøy vil forandre vårt syn på *E. coli* og virulens i fremtiden. *E. coli* sitt genom er i kontinuerlig forandring som gjør at typing av denne bakterien krever presise metoder som fanger opp de genetiske forandringene. Patotyping er en vanlig typingsmetode hos kliniske mikrobiologiske og referanselaboratorier hvor avføringsprøver fra pasienter eller mistenkte næringsmidler kan spores og testes med genetiske metoder. Metodene går i hovedsak ut på å identifisere spesifikke mål VAGs tilstede hos *E. coli* gjennom PCR amplifisering ved bruk av konstruerte primere rettet mot disse målene. For eksempel ved PCR patotyping av kjente IPEC patovarer vil man for EPEC påvise *eae*, EIEC *ipaH*, ETEC *elt*, *est*, EHEC *stx1* og *stx2*, EAEC *aggR*, *aatA* og *aaiC*, DAEC Afa/Dr adhesins (Lindstedt et al., 2018; Robins-Browne et al., 2016). For ExPEC patotyper er det derimot ikke et felles akseptert klassifiseringssystem, noe som også viser seg ved at det er langt mer vanlig i rutinesøk hos referanse- og diagnostiske mikrobiologiske laboratorier å kun søke etter IPEC assosierte VAGs (Lindstedt et al., 2018). Likevel viser ny forskning som undersøkte 168 *E. coli* stammer isolert fra avføringsprøver sendt til folkehelseinstituttet fra kliniske medisinske laboratorier rundt om i Norge, at 64.3 % av stammene hadde en kombinasjon av IPEC og ExPEC assosierte VAGs (Lindstedt et al., 2018). Funn i denne masteroppgaven viser at alle de 7 helgenomsekvenserte *E. coli* stammene hadde VAGs assosiert både med IPEC og ExPEC, med utgangspunkt i samme IPEC og ExPEC målgener som Lindstedt et al. En utfordring med dagens patotyping er at nye hybride stammer som ikke passer inn i etablerte kategorier ikke vil fanges opp i rutinesøk (Robins-Browne et al., 2016). For eksempel kan dette bety at mer aggressive hybride stammer klassifiseres innenfor en etablert patotype, men stammens fulle virulenspotensiale ikke vil tilfredsstillende dekkes av denne patotypen. Eksempler på slike hybride stammer som faller utenfor de etablerte patotypene som har blitt beskrevet i litteraturen er LT-produserende EPEC i Duetta et al., og shigatoksinproduserende EAEC som var ansvarlig for de store utbruddene i Europa 2011 (Dutta et al., 2015).

Fordelingen av VAGs assosiert med de ulike patotypene hos de ulike stammene, presentert i tabell 13, viser at alle stammene kan assosieres med flere enn én patotype. Ettersom alle stammene var isolert fra pasienter med diarè og tilhørte sekvenstype 28 assosiert med EPEC, er det ikke overraskende at det ble funnet flere EPEC karakteristiske VAGs hos stammene. For alle stammene bortsett fra M884Italy ble det identifisert CFA K88 fimbrie gener som koder for en type adhesin assosiert med ETEC (Kaper et al., 2004). K88 fimbrie er derimot ikke assosiert med humane ETEC stammer, men hos ETEC stammer som gir infeksjon hos dyr. En studie av Knutton et al., viste at K88 positive stammer adhererte dårlig til humane enterocytter i duodenum sammenliknet med human spesifikke ETEC stammer (Knutton et al., 1985). En annen studie av Denek et al. viste derimot at K88 positive stammer adhererte like godt som human spesifikke stammer, men til human enterocytter i ileum (Deneke et al., 1983). Adherensen til humane enterocytter ex vivo viser derimot ikke at stammene effektivt kan kolonisere tarmen in vivo, hvor andre faktorer enn effektiv adherering spiller en viktig rolle, slik som overlevelse og vekst i tarmmiljøet. Generelt regnes ETEC som å ikke være zoonotisk på grunn av forskjeller i fimbrie spesifikk binding mellom dyr og menneske (Granum, 2015; Wasteson, 2002). For å klassifisere en stamme som ETEC er det derimot nødvendig å påvise ett av enterotoksinene ST eller LT i tillegg til CAFs, noe som ikke ble påvist hos stammene i denne masteroppgaven (Granum, 2015; Robins-Browne et al., 2016). Påvisningen av K88 gener viser likevel at disse stammene kan ta opp ETEC assosierte CAFs i sitt genom, og en kan hypotesere om det er mulig at ST28 også kan ta opp human spesifikke CAFs og eventuelt ETEC spesifikke enterotoksiner. En slik hybrid EPEC/ETEC stamme (kalt for «strain 639») har tidligere blitt isolert fra et barn med diarè og beskrevet i litteraturen, hvor denne stammen ble bestemt til å være en ST328 (Dutta et al., 2015).

Urinveisinfectionsjoner (UTI) forårsakes hovedsakelig av tarmbakterier, hvor 75 % skyldes *E. coli* (NHI, 2018). Globalt er UTI en av de vanligste bakterielle infectionsjonene, hvor de fleste er ukompliserte infectionsjoner som rammer kvinner oftere enn menn (Geerlings, 2017). Den årlige selvrapporterte insidensen av UTI er 12 %, hvor en kvinne som er fylt 32 år vil ha en 50 % sjans for å minst hatt UTI en gang. Selv om kompliserte UTI er sjeldnere, har spesifikke grupper som menn, gravide kvinner, immunsvekkede og de med anatomiske eller funksjonelle unormalheter i urinveissystemet en høyere risiko. I Norge er helsetjenesteassosierte urinveisinfectionsjoner (HAIU) er et problem både i sykehus og sykehjem (Thale C. Berg et al., 2019). I 2018 var forebygging og behandling av urinveisinfectionsjoner den hyppigste årsaken til

antibiotika bruk i sykehjem. Prevalensen for HAIU i sykehus var i 2018 0.7 % og utgjorde 16.2 % av totalt antall HAI i sykehus. For sykehjem hadde HAIU en prevalens på 2.5 % og utgjorde 51 % av totalt antall HAI i sykehjem. Det følger derfor at UPEC er en utfordring både globalt og i Norge, som også er med å øke bruken av antibiotika som er ugunstig. En økt utbredelse av UPEC stammer, også mer aggressive hybridstammer vil særlig kunne utgjøre en trussel for populasjonsgrupper som er utsatt for kompliserte UTI. Slike pasientgrupper finner vi blant annet i sykehjem, hvor UTI utgjør en høy andel HAIU og kan være en trussel for immunsvekkede eldre personer. Generelt vil også overforbruk av antibiotika gi økt risiko for resistente bakterier som blir vanskeligere å behandle. For fem av stammene i denne masteroppgaven ble 5 UPEC assosierte VAGs identifisert. Disse var knyttet til jernopptak (*iron*) som er viktig for UPEC overlevelse i urinsystemet hvor jerntilgangen er begrenset, adheranse (*papC* og *papD*) som deler av *pap* operonet som koder for Pap-pili, *yfcV* som koder for en viktig del av chaperone-usher fimbriae, og *usp* som koder for et protein som oftere er assosiert med UPEC stammer enn andre *E. coli* isolater, hvor funksjonen er foreslått til å være et bakteriosin med genotoksisk virkning (Nipič et al., 2013; Spurbeck et al., 2012; Terlizzi et al., 2017; Zaw et al., 2013). Fire UPEC assosierte VAGs har blitt foreslått å være sterkt assosiert med UPEC, disse er: *vat*, *fyuA*, *chuA*, *yfcV* (Spurbeck et al., 2012). Alle disse genene ble identifisert hos alle stammene i denne masteroppgaven (*chuA*, *fyuA* er ikke presentert resultatkapittelet, men finnes i tilleggsfil 1 «Eget_virulenssøk_ST28» slik som beskrevet i vedlegg). Det ble også funnet ulike VAGs assosiert med antibiotikaresistens for alle stammene, slik som vist i tabell 11.

Et interessant funn, som tidligere har blitt beskrevet i Grande et al., er M884Italy stammen som er en ST28 med EPEC VAGs som i tillegg har tatt opp shigatoksiner i genomet, og kan derfor klassifiseres som en EHEC. Denne stammen kan sammenliknes med mye fryktede *E. coli* O157:H7 til en viss grad, som har forårsaket alvorlig sykdom på verdensbasis (Lim et al., 2010). Denne bakterien har hovedresorvar i storfe, og har blitt omtalt som «hamburgerbakterien», da den ved flere tilfeller har smittet gjennom utilstrekkelig varmebehandlede hamburgere ved hamburgerkjeder i USA. Bakterien kan også smitte gjennom kontaminerte vanningskilder, kontaktsmitte mellom person til person og mellom dyr og mennesket. Vanligvis vil sykdomsforløpet med O157:H7 utarte seg som andre *E. coli* enteritter med ikke-blodig diarè som går over av seg selv. Men i enkelt tilfeller vil denne bakterien kunne føre til hemoragisk kolitt (HC), og hos 5-10 % av pasienter med HC videreutvikle seg til livstruende tilstander som

HUS og trombotisk trombocytopenisk purpura (TTP). Pasientgruppene eldre og barn har en økt risiko for utvikling av alvorlige komplikasjoner med O157:H7. O157:H7 en rekke virulensfaktorer som gjør den til en svært effektiv patogen slik som blant annet økt syretoleranse, overlevelse i ulike miljøer og et høyt konservert plasmid kalt for pO157 som bidrar med en rekke virulensfaktorer. Virulensfaktoren som koder for syretoleranse *gad* ble funnet hos M884Italy (tabell 8), men det mangler fortsatt flere virulensfaktorer for å sidestille denne med O157:H7. Likevel er det interessant at en ST28 kan ta opp shigatoksiner, da 9 aEPEC stammer (Trh2, Trh3, Trh4, Trh17, Trh18, Trh27, Trh49, Trh50 og Trh52) av denne sekvenstypen ble identifisert fra fekale prøver av 37 barn med diarè og 19 friske barn i Norge med en alder på <5 år (Afset et al., 2008).

Noe som ikke ble undersøkt i denne masteroppgaven var forekomsten av genomiske øyer og integrerte profager i genomet til stammene, som ville kunne gi en oversikt over grad av horisontal genoverføring. Ulike bioinformatiske verktøy finnes for å undersøke dette. Horisontal genoverføring direkte mellom bakterier, fra miljøet eller via profager er en viktig årsak til utvikling av patogene stammer av ulike bakterier, inkludert *E. coli* (Busby et al., 2013).

Analyse av nukleotid- og aminosyresekvens til *eae*, *tir* og *vat* med ClustalOmega sequence alignment viste at disse var svært like hverandre. Dette viser at disse virulensfaktorene er godt bevart mellom stammene, noe som indikerer at disse faktorene er effektive med hensyn til virulens. Det er mulig å se fra resultatene at enkelte punktmutasjoner forekommer, og for ECB og *eae* kan den kortere nukleotid/aminosyresekvensen skyldes for eksempel innsetning av et transposon.

5.0 Konklusjon

Funnene viser at kombinasjoner av VAGs assosiert med IPEC og ExPEC finnes i ST28 stammer isolert fra norske pasienter med diarè. Dette kan utgjøre en trussel for den norske befolkningen, særlig for VAGs som kan bidra til å forårsake systemisk i utsatte pasientgrupper. Den høye forekomsten av høyst prediktive UPEC VAGs hos flere av stammene er bekymringsfullt, særlig med hensyn til hvor vanlig denne typen infeksjon er både globalt og nasjonalt i Norge.

6.0 Referanser

Adams, M. R. & Moss, M. O. (2008). *Food microbiology*. 3rd ed. utg. Cambridge, UK: RSC Pub.

Afset, J. E., Bevanger, L., Romundstad, P. & Bergh, K. (2004). Association of atypical enteropathogenic *Escherichia coli* (EPEC) with prolonged diarrhoea. *Journal of medical microbiology*, 53 (11): 1137-1144.

Afset, J. E., Anderssen, E., Bruant, G., Harel, J., Wieler, L. & Bergh, K. (2008). Phylogenetic backgrounds and virulence profiles of atypical enteropathogenic *Escherichia coli* strains from a case-control study using multilocus sequence typing and DNA microarray analysis. *Journal of clinical microbiology*, 46 (7): 2280-2290.

Bielaszewska, M., Prager, R., Köck, R., Mellmann, A., Zhang, W., Tschäpe, H., Tarr, P. I. & Karch, H. (2007). Shiga toxin gene loss and transfer in vitro and in vivo during enterohemorrhagic *Escherichia coli* O26 infection in humans. *Appl. Environ. Microbiol.*, 73 (10): 3144-3150.

Bolger, A. M., Lohse, M. & Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, 30 (15): 2114-2120. doi: 10.1093/bioinformatics/btu170.

Busby, B., Kristensen, D. M. & Koonin, E. V. (2013). Contribution of phage-derived genomic islands to the virulence of facultative bacterial pathogens. *Environmental microbiology*, 15 (2): 307-312.

Carrico, J., Rossi, M., Moran-Gilad, J., Van Domselaar, G. & Ramirez, M. (2018). A primer on microbial bioinformatics for nonbioinformaticians. *Clinical Microbiology and Infection*, 24 (4): 342-349.

Claverie, J.-M. & Notredame, C. (2006). *Bioinformatics for dummies*: John Wiley & Sons.

Deamer, D., Akeson, M. & Branton, D. (2016). Three decades of nanopore sequencing. *Nature biotechnology*, 34 (5): 518.

Deneke, C., McGowan, K., Thorne, G. & Gorbach, S. (1983). Attachment of enterotoxigenic *Escherichia coli* to human intestinal cells. *Infection and immunity*, 39 (3): 1102-1106.

Donnenberg, M. (2013). *Escherichia coli: pathotypes and principles of pathogenesis*: Academic Press.

Dubreuil, J. D. (2019). EAST1 toxin: An enigmatic molecule associated with sporadic episodes of diarrhea in humans and animals. *Journal of Microbiology*, 57 (7): 541-549.

Dutta, S., Pazhani, G. P., Nataro, J. P. & Ramamurthy, T. (2015). Heterogenic virulence in a diarrheagenic *Escherichia coli*: evidence for an EPEC expressing heat-labile toxin of ETEC. *International Journal of Medical Microbiology*, 305 (1): 47-54.

Eichhorn, I., Heidemanns, K., Ulrich, R. G., Schmidt, H., Semmler, T., Fruth, A., Bethe, A., Goulding, D., Pickard, D. & Karch, H. (2018). Lysogenic conversion of atypical enteropathogenic *Escherichia coli* (aEPEC) from human, murine, and bovine origin with bacteriophage Φ 3538 Δ stx2:: cat proves their enterohemorrhagic *E. coli* (EHEC) progeny. *International Journal of Medical Microbiology*, 308 (7): 890-898.

Folkehelseinstituttet. (2019). *E. coli-enteritt (inkludert EHEC-infeksjon og HUS) - veileder for helsepersonell*. Tilgjengelig fra: <https://www.fhi.no/nettpub/smittevernveilederen/sykdommer-a-a/e.-coli-enteritt-inkludert-ehc-inf/> (lest: 22.05.2020).

Foxman, B. (2010). The epidemiology of urinary tract infection. *Nature Reviews Urology*, 7 (12): 653.

Geerlings, S. E. (2017). Clinical presentations and epidemiology of urinary tract infections. *Urinary Tract Infections: Molecular Pathogenesis and Clinical Management*: 27-40.

Grad, Y. H., Lipsitch, M., Feldgarden, M., Arachchi, H. M., Cerqueira, G. C., FitzGerald, M., Godfrey, P., Haas, B. J., Murphy, C. I. & Russ, C. (2012). Genomic epidemiology of the *Escherichia coli* O104: H4 outbreaks in Europe, 2011. *Proceedings of the national academy of sciences*, 109 (8): 3065-3070.

Grande, L., Michelacci, V., Bondi, R., Gigliucci, F., Franz, E., Badouei, M. A., Schlager, S., Minelli, F., Tozzoli, R. & Caprioli, A. (2016). Whole-genome characterization and strain comparison of VT2f-producing *Escherichia coli* causing hemolytic uremic syndrome. *Emerging infectious diseases*, 22 (12): 2078.

Granum, P. E. (2015). *Matforgiftning: smitte gjennom mat og vann*. 4. utg. utg. Oslo: Cappelen Damm akademisk.

Harrington, S. M., Dudley, E. G. & Nataro, J. P. (2006). Pathogenesis of enteroaggregative *Escherichia coli* infection. *FEMS microbiology letters*, 254 (1): 12-18.

Heimer, S. R., Rasko, D. A., Lockett, C. V., Johnson, D. E. & Mobley, H. L. (2004). Autotransporter genes *pic* and *tsh* are associated with *Escherichia coli* strains that cause acute pyelonephritis and are expressed during urinary tract infection. *Infection and immunity*, 72 (1): 593-597.

Helmer Fjellvåg, Terje Berg & Knut Arnold Rosvold. (2019). *nanoteknologi*. Tilgjengelig fra: <https://snl.no/nanoteknologi> (lest: 25.02.2020).

Illumina. (2017). *An introduction to Next-Generation Sequencing Technology*. Tilgjengelig fra: https://www.illumina.com/documents/products/illumina_sequencing_introduction.pdf (lest: 18.02.2020).

Institute, R. K. (2011). *Report: Final Presentation and Evaluation of Epidemiological Findings in the EHEC O104: H4 Outbreak, Germany 2011*: Robert Koch Institute Berlin.

Kaper, J. B., Nataro, J. P. & Mobley, H. L. (2004). Pathogenic *Escherichia coli*. *Nature reviews microbiology*, 2 (2): 123-140.

Knutton, S., Lloyd, D. R., Candy, D. & McNEISH, A. S. (1985). Adhesion of enterotoxigenic *Escherichia coli* to human small intestinal enterocytes. *Infection and immunity*, 48 (3): 824-831.

Kono, N. & Arakawa, K. (2019). Nanopore sequencing: Review of potential applications in functional genomics. *Development, growth & differentiation*, 61 (5): 316-326.

Köhler, C.-D. & Dobrindt, U. (2011). What defines extraintestinal pathogenic *Escherichia coli*? *International Journal of Medical Microbiology*, 301 (8): 642-647.

Lim, J. Y., Yoon, J. W. & Hovde, C. J. (2010). A brief overview of *Escherichia coli* O157: H7 and its plasmid O157. *Journal of microbiology and biotechnology*, 20 (1): 5.

Lindstedt, B.-A., Finton, M. D., Porcellato, D. & Brandal, L. T. (2018). High frequency of hybrid *Escherichia coli* strains with combined Intestinal Pathogenic *Escherichia coli* (IPEC) and Extraintestinal Pathogenic *Escherichia coli* (ExPEC) virulence factors isolated from human faecal samples. *BMC infectious diseases*, 18 (1): 544.

Maiden, M. C., Van Rensburg, M. J. J., Bray, J. E., Earle, S. G., Ford, S. A., Jolley, K. A. & McCarthy, N. D. (2013). MLST revisited: the gene-by-gene approach to bacterial genomics. *Nature Reviews Microbiology*, 11 (10): 728-736.

Momand, J., McCurdy, A., Heubach, S. & Warter-Perez, N. (2016). *Concepts in Bioinformatics and Genomics*: Oxford University Press, Inc.

Nakazato, T., Ohta, T. & Bono, H. (2013). Experimental design-based functional mining and characterization of high-throughput sequencing data in the sequence read archive. *PLoS One*, 8 (10).

Nataro, J. P., Yikang, D., Cookson, S., Cravioto, A., Savarino, S. J., Guers, L. D., Levine, M. M. & Tacket, C. O. (1995). Heterogeneity of Enteropathogenic *Escherichia coli* Virulence Demonstrated. *Journal of Infectious Diseases*, 171 (2): 465-468.

NHI. (2018). *Urinveisinfeksjon hos kvinner*. Tilgjengelig fra: <https://nhi.no/sykdommer/nyrerurinveier/urinveisinfeksjoner/urinveisinfeksjon-hos-kvinner/?page=all> (lest: 29.05.2020).

Nichols, K. B., Totsika, M., Moriel, D. G., Lo, A. W., Yang, J., Wurpel, D. J., Rossiter, A. E., Strugnell, R. A., Henderson, I. R. & Ulett, G. C. (2016). Molecular characterization of the vacuolating autotransporter toxin in uropathogenic *Escherichia coli*. *Journal of bacteriology*, 198 (10): 1487-1498.

- Nipič, D., Podlesek, Z., Budič, M., Črnigoj, M. & Žgur-Bertok, D. (2013). Escherichia coli uropathogenic-specific protein, Usp, is a bacteriocin-like genotoxin. *The Journal of infectious diseases*, 208 (10): 1545-1552.
- Nyholm, O., Heinikainen, S., Pelkonen, S., Hallanvuo, S., Haukka, K. & Siitonen, A. (2015). Hybrids of Shigatoxigenic and enterotoxigenic Escherichia coli (STEC/ETEC) among human and animal isolates in Finland. *Zoonoses and public health*, 62 (7): 518-524.
- Parham, N. J., Pollard, S. J., Desvaux, M., Scott-Tucker, A., Liu, C., Fivian, A. & Henderson, I. R. (2005). Distribution of the serine protease autotransporters of the Enterobacteriaceae among extraintestinal clinical isolates of Escherichia coli. *Journal of clinical microbiology*, 43 (8): 4076-4082.
- Pokharel, P., Habouria, H., Bessaiah, H. & Dozois, C. M. (2019). Serine Protease Autotransporters of the Enterobacteriaceae (SPATEs): Out and About and Chopping It Up. *Microorganisms*, 7 (12): 594.
- Rasko, D. A., Webster, D. R., Sahl, J. W., Bashir, A., Boisen, N., Scheutz, F., Paxinos, E. E., Sebra, R., Chin, C.-S. & Iliopoulos, D. (2011). Origins of the E. coli strain causing an outbreak of hemolytic–uremic syndrome in Germany. *New England Journal of Medicine*, 365 (8): 709-717.
- Robins-Browne, R. M., Holt, K. E., Ingle, D. J., Hocking, D. M., Yang, J. & Tauschek, M. (2016). Are Escherichia coli pathotypes still relevant in the era of whole-genome sequencing? *Frontiers in cellular and infection microbiology*, 6: 141.
- Rodrigues, J., Thomazini, C. M., Lopes, C. A. & Dantas, L. O. (2004). Concurrent infection in a dog and colonization in a child with a human enteropathogenic Escherichia coli clone. *Journal of clinical microbiology*, 42 (3): 1388-1389.
- Seemann, T. (2020). *Shovill*. Tilgjengelig fra: <https://github.com/tseemann/shovill> (lest: 27.02.2020).
- Shin, S., Castanie-Cornet, M. P., Foster, J. W., Crawford, J. A., Brinkley, C. & Kaper, J. B. (2001). An activator of glutamate decarboxylase genes regulates the expression of

enteropathogenic *Escherichia coli* virulence genes through control of the plasmid-encoded regulator, *Per*. *Molecular microbiology*, 41 (5): 1133-1150.

Spurbeck, R. R., Dinh, P. C., Walk, S. T., Stapleton, A. E., Hooton, T. M., Nolan, L. K., Kim, K. S., Johnson, J. R. & Mobley, H. L. (2012). *Escherichia coli* isolates that carry *vat*, *fyuA*, *chuA*, and *yfcV* efficiently colonize the urinary tract. *Infection and immunity*, 80 (12): 4115-4122.

Starčič Erjavec, M. & Žgur-Bertok, D. (2015). Virulence potential for extraintestinal infections among commensal *Escherichia coli* isolated from healthy humans—the Trojan horse within our gut. *FEMS microbiology letters*, 362 (5): fnu061.

Terlizzi, M. E., Gribaudo, G. & Maffei, M. E. (2017). UroPathogenic *Escherichia coli* (UPEC) infections: virulence factors, bladder responses, antibiotic, and non-antibiotic antimicrobial strategies. *Frontiers in microbiology*, 8: 1566.

Thale C. Berg, Løwer, H. L., Alberg, T. & Eriksen, H. M. (2019). *Årsrapport 2018: Helsetjenesteassosierte infeksjoner, antibiotikabruk (NOIS), antibiotikaresistens (MSIS) og Verdens håndhygienedag*. Folkehelseinstituttet.

Tronsmo, A. (2016). *Innføring i mikrobiologi*. Oslo: Universitetsforl.

Vallance, B. & Finlay, B. (2000). Exploitation of host cells by enteropathogenic *Escherichia coli*. *Proceedings of the National Academy of Sciences*, 97 (16): 8799-8806.

Wasteson, Y. (2002). Zoonotic *Escherichia coli*. *Acta Veterinaria Scandinavica*, 43 (1): S79.

Watson, J. D. (2014). *Molecular biology of the gene*. 7th ed. utg. Boston: Pearson.

Welch, R. A., Burland, V., Plunkett, G., Redford, P., Roesch, P., Rasko, D., Buckles, E., Liou, S.-R., Boutin, A. & Hackett, J. (2002). Extensive mosaic structure revealed by the complete genome sequence of uropathogenic *Escherichia coli*. *Proceedings of the National Academy of Sciences*, 99 (26): 17020-17024.

Wirth, T., Falush, D., Lan, R., Colles, F., Mensa, P., Wieler, L. H., Karch, H., Reeves, P. R., Maiden, M. C. & Ochman, H. (2006). Sex and virulence in *Escherichia coli*: an evolutionary perspective. *Molecular microbiology*, 60 (5): 1136-1151.

Yu, F., Chen, X., Zheng, S., Han, D., Wang, Y., Wang, R., Wang, B. & Chen, Y. (2018). Prevalence and genetic diversity of human diarrheagenic *Escherichia coli* isolates by multilocus sequence typing. *International Journal of Infectious Diseases*, 67: 7-13.

Zaw, M. T., Yamasaki, E., Yamamoto, S., Nair, G. B., Kawamoto, K. & Kurazono, H. (2013). Uropathogenic specific protein gene, highly distributed in extraintestinal uropathogenic *Escherichia coli*, encodes a new member of HNH nuclease superfamily. *Gut pathogens*, 5 (1): 13.

7.0 Vedlegg

7.1 Matrix benyttet i SNP-analysene

	ECC-Shovill	ECD-Unicycler	EPECE2348_LT827011	ETEC-2265	FHI-NMBU-03	FHI_NMBU_ST28_ECA	FHI_NMBU_ST28_ECB	HS_NC_009800	IHE3034_NC_017628	M884_Italy	Trh49_Trondheim	UTI89_NC_007946
ECC-Shovill_Assembly/1-75653	0	1890	23988	47020	103	4919	190	45081	19281	2475	33	19321
ECD-Unicycler_Final_Assembly/1-75653	1890	0	24165	46757	1941	5433	1908	44757	19251	2482	1899	19289
EPECE2348_LT827011/1-75653	23988	24165	0	47595	23981	23762	23997	45760	23817	23621	23993	23851
ETEC-2265_NZ_CP023346/1-75653	47020	46757	47595	0	47037	47658	47029	22016	47531	47715	47031	47520
FHI-NMBU-03_chromosome_NZ_CP019455/1-75653	103	1941	23981	47037	0	4996	245	45098	19286	2426	112	19326
FHI_NMBU_ST28_ECA_2019_contigs/1-75653	4919	5433	23762	47658	4996	0	4961	45764	18924	4162	4922	18968
FHI_NMBU_ST28_ECB_2020_contigs/1-75653	190	1908	23997	47029	245	4961	0	45083	19293	2493	199	19333
HS_NC_009800/1-75653	45081	44757	45760	22016	45098	45764	45083	0	45654	45838	45088	45644
IHE3034_NC_017628/1-75653	19281	19251	23817	47531	19286	18924	19293	45654	0	18597	19286	311
M884_Italy_FBUS01/1-75653	2475	2482	23621	47715	2426	4162	2493	45838	18597	0	2476	18637
Trh49_Trondheim/1-75653	33	1899	23993	47031	112	4922	199	45088	19286	2476	0	19326
UTI89_NC_007946/1-75653	19321	19289	23851	47520	19326	18968	19333	45644	311	18637	19326	0

Figur 15. Matrix benyttet for å generere SNP-treet i figur 12. Minimum SNP forskjell er 33 og maksimum 47715. Grønn indikerer nært beslektet med en SNP forskjell mellom 1 og 500, mens gult indikerer beslektet med en SNP forskjell mellom 501 og 5000.

7.2 Tilleggsfil 1: Egendefinert virulenssøk

Excel-fil med alle data fra det egendefinerte virulenssøket utført av hovedveileder finnes i tilleggsfilen 1: «Eget_virulenssøk_ST28». Spesifikke fargekoder benyttet i sammenlikning arkfanen: Blå → finnes hos >1 men ikke alle stammene, gul → unike for en enkelt stamme, grønn/oransje → ulike genvarianter.



Norges miljø- og biovitenskapelige universitet
Noregs miljø- og biovitenskapelige universitet
Norwegian University of Life Sciences

Postboks 5003
NO-1432 Ås
Norway