ISBN: 978-82-575-1686-4
ISSN: 1894-6402

Norwegian University
of Life Sciences

Postboks 5003
NO-1432 Ås, Norway
+47 67 23 00 00
www.nmbu.no

Philosophiae Doctor (PhD), Thesis 2020:23

Misti Dawn Finton

106747 / ANDVORDGRAFISK.NO

Norwegian University of Life Sciences
Faculty of Chemistry, Biotechnology and Food Sciences

Philosophiae Doctor (PhD)
Thesis 2020:23

Characterization of heteropathogenic and multidrugresistant (MDR) bacterial isolates from clinical andenvironmental sources using combined nanopore and next-generation sequencing technologies

Karakterisering av heteropatogene og multiresistente (MDR) bakterieisolater fra kliniske- og miljø-kilder ved bruk av kombinert nanopore og neste generasjonssekvenseringsteknologi

Misti Dawn Finton

Characterization of heteropathogenic and multidrug resistant (MDR) bacterial isolates from clinical and environmental sources using combined nanopore and next-generation sequencing technologies

Karakterisering av heteropatogene og multiresistente (MDR) bakterieisolater fra kliniske- og miljø- kilder ved bruk av kombinert nanopore og neste generasjons sekvenseringsteknologi

Philosophiae Doctor (PhD) Thesis

Misti Finton

Norwegian University of Life Sciences

Faculty of Chemistry, Biotechnology and Food Science

Ås (2020)

Thesis number 2020:23

ISSN 1894-6402

ISBN 978-82-575-1686-4

Table of Contents

Acknowledgements I

Summary III

Norsk Sammendrag V

Abbreviations VII

List of papers IX

1. Introduction 1

1.1 Enterobacteriaceae 1

1.2 *Klebsiella pneumoniae* 1

1.3 *Escherichia coli* 1

1.3.1 Intestinal Pathogenic *E. coli* (IPEC) 3

1.3.2 Extraintestinal pathogenic *E. coli* (ExPEC) 7

1.3.3 Heteropathogenic *E. coli* 9

1.4 Treatment of *E. coli* infections 9

1.5 Antimicrobial agents 10

1.6 β-lactam and extended-spectrum β-lactam antibiotics 11

1.7 What is antibiotic resistance? 12

1.8 Antibiotic susceptibility testing 13

1.9 Intrinsic vs. acquired antibiotic resistance 14

1.10 Mechanisms of β-lactam resistance 15

1.10.1 β-lactamases 15

1.10.2. Extended-spectrum β-lactamases 16

1.11 ESBL classification systems 17

1.12 Dissemination of ESBLs (and other resistance factors) 18

1.12.1 Vertical transfer (clonal spread) 18

1.12.2 Horizontal gene transfer (HGT) 19

1.12.13 Mobile genetic elements (MGEs) 20

1.12.14 Plasmids 20

1.13 Surveillance of ESBL-producing Enterobacteriaceae in Europe 21

1.14 DNA – The Molecule of Life 23

1.14.1 DNA sequencing technology 25

1.14.2 Whole genome sequencing (WGS) 26

2. Aims of the thesis 31

3. Main results and discussion 33

4. Concluding remarks 43

5. Future perspectives 43
6. References 45
Enclosed papers 59

-

Acknowledgements

The present work originated from a faculty funded PhD fellowship grant at the Norwegian University of Life Sciences. The work of this thesis was carried out between 2016 and 2020 as a member of the research group for Dairy Technology and Food Quality in the faculty of Chemistry, Biotechnology, and Food Science department at the Norwegian University of Life Sciences.

First, my sincere gratitude goes to my supervisor, Bjørn-Arne Lindstedt. Only with your continued support and guidance was this PhD made possible. You have been a true mentor, and I whole-heartedly appreciate that you gave this Wisconsin girl the chance to work on such an exciting project. Secondly, I would like to thank my co-supervisor Lin T. Brandal for your thoughtful contributions during manuscript preparations and for the numerous strains of *E. coli* that you contributed to my projects.

Thank you to May-Helene Aalberg, Kari Olsen, and Bjørg Holter for being so helpful in every way, and for making every day brighter with such contagious smiles. Thank you to Ahmed Abdelghani for always lending a helping hand when I needed it most (no matter how busy you were!). A huge thank you to Davide Porcellato for including several of our isolates in your Illumina runs. And, to both Davide and Roger Meisal for sharing your brilliance in bioinformatics -- your contributions are greatly appreciated! A huge thank you to my office mate Martine Andrea Olsen and to my 'office neighbors' Sara Muhamed Gaber and Anna Dysvik for always giving insightful advice. Not to mention, our fantastic pizza Fridays!

Finally, I would like to thank Erik for the constant encouragement every single step of the way – you believed in me even when I didn't believe in myself. To my loving parents and sister who gave moral support throughout the entirety of my PhD. To Ellen, Anne, and Maya for always lending a helping hand (and paw!). To my best friend Dawn, we may live on completely opposite sides of the world, but you were my backbone no matter the time of day or night. Last but not least, thank you to my grandma. Gone, but not forgotten, you are my inspiration for everything.

"We are the Music Makers and we are the Dreamers of Dreams"
Willy Wonka/Roald Dahl

Summary

Escherichia coli (*E. coli*) are usually harmless commensals that colonize the gastrointestinal tract of humans and other warm-blooded mammals. However, pathogenic counterparts have acquired virulence associated genes (VAGs) that allow for colonization and infection of several anatomical sites. Intestinal pathogenic *E. coli* (IPEC) strains are important foodborne pathogens that cause diarrheal disease in humans and are one of the major causes of mortality amongst children <5 years of age. On the other hand, extraintestinal pathogenic *E. coli* (ExPEC) strains are common causes of urinary tract infections (UTIs), bacteremia, neonatal meningitis, and wound infections. Traditionally, IPEC and ExPEC have been considered distinct from one another, and ExPEC VAGs were thought to be completely absent in IPEC strains. However, novel combinations of VAGs from multiple pathotypes (both IPEC and ExPEC) may exist in a single strain and are often referred to as a 'heteropathogen' or 'hybrid'. Heteropathogens are an emerging public health threat that may cause a more serious infection and have an increased spreading potential. Therefore, a large portion of this PhD work (paper I and paper III) focused on expanding our knowledge of heteropathogenic *E. coli* by examining isolates from fecal samples of patients in Norwegian clinics that exhibited signs of gastrointestinal infection. In paper I, we looked at a wider range of VAGs and were especially interested in detecting known ExPEC virulence factors (VFs), as there is little information regarding the frequency of ExPEC strains in the human gut. To gain even more knowledge on the heteropathogenic nature of the strains of paper I, we chose several for whole genome sequencing (WGS) through a combination of Oxford Nanopore's MinION and Illumina's Miseq.

Our research also focused on the detection of extended-spectrum β-lactamases (ESBLs) in environmental water sources. ESBLs encode resistance to the most widely used class of drugs in human and veterinary medicine, β-lactams, and constitute a serious risk of community-acquired infections with limited treatment options. Their presence in the environment was of special interest, as soil and aquatic ecosystems are rich reservoirs for the acquisition, evolution, and dissemination of ARB. This particular project was originally thought to be a separate initiative than the heteropathogenic *E.*

coli of papers I and III, however we found heteropathogenic strains of ESBL-producing *E. coli* in the environment, tying the two seemingly different research topics together.

Overall, the backbone of this PhD work was Oxford Nanopore's MinION and Illumina's MiSeq. The long reads of the MinION and the short reads of the MiSeq put together allowed us to assemble multiple whole genomes, and even discover a number of novel genes. Currently, both heteropathogenic *E. coli* and ESBL-producing bacteria are hot topics, and the use of WGS enabled us to generate more knowledge for the never-ending quest of scientists.

Norsk Sammendrag

Escherichia coli (*E. coli*) er vanligvis en ufarlig bakterie som koloniserer mage-tarmkanalen hos mennesker og andre varmblodige pattedyr. Imidlertid bærer noen patogene varianter virulensassosierte gener (VAG) som muliggjør kolonisering og infeksjon flere steder i kroppen. Intestinale patogene *E. coli* (IPEC) stammer er viktige matbårne patogener som forårsaker diarésykdom hos mennesker og er en av de viktigste årsakene til dødelighet blant barn <5 år. På den annen side er stammer av ekstraintestinale patogene *E. coli* (ExPEC) vanlige årsaker til urinveisinfeksjoner (UTI), bakteremi, neonatal meningitt og sårinfeksjoner. Tradisjonelt har IPEC og ExPEC blitt ansett som forskjellige fra hverandre, og ExPEC virulensassosierte gener (VAG) ble antatt å være helt fraværende i IPEC-stammer. Imidlertid kan nye kombinasjoner av VAG-er fra flere patotyper (både IPEC og ExPEC) eksistere sammen i en enkelt stamme, og denne blir ofte referert til som en 'heteropatogen' eller 'hybrid' stamme. Heteropatogener er en voksende trussel mot folkehelsen som kan forårsake en mer alvorlig infeksjon og ha økt spredningspotensiale. Derfor fokuserte en stor del av dette doktorgradsarbeidet (artikkel I og artikkel III) på å utvide kunnskapen vår om heteropatogene *E. coli* ved å undersøke isolater fra fekale prøver av norske pasienter innsendt til referanselaboratoriet ved Folkehelseinstituttet. Vi undersøkte et bredt spekter av VAG-er og var spesielt interessert i å oppdage kjente ExPEC-virulensfaktorer, siden det er lite informasjon angående hyppigheten av ExPEC-stammer i den menneskelige tarmen. For å få dypere kunnskap om den heteropatogene karakteren av utvalgte stammer, valgte vi ut flere for helgenomsekvensering (WGS) gjennom en kombinasjon av Oxford Nanopores MinION og Illuminas Miseq.

Forskningen vår fokuserte også på påvisning av utvidet spektrum ß-laktamaser (ESBL) i vannkilder fra ytre miljø. ESBL-er koder for resistens mot den mest brukte klassen av legemidler i human- og veterinærmedisin, ß-laktamer, og utgjør en alvorlig risiko for infeksjoner med begrensede behandlingsalternativer. Tilstedeværelse av ESBL i miljøet var av spesiell interesse, ettersom jord og akvatiske økosystemer utgjør rike reservoarer for utvikling og spredning av antibiotikaresistens. Dette prosjektet var opprinnelig adskilt fra arbeidet med heteropatogene *E. coli* (artikkel I og III), men vi oppdaget tidlig i studiet heteropatogene stammer av ESBL-produserende *E. coli* i miljøprøvene, og dette bandt studiene sammen. Det virker som om andelen av heteropatogene *E. coli* er mye høyere enn først antatt både isolert fra humane pasienter, og nå også fra ytre miljø. Det er urovekkende at de heteropatogene stammene også kan tilegne seg multiresistens.

Mye av arbeidet i dette doktorgradsstudiet var fokusert på å integrere data generert fra Oxford Nanopores MinION og Illuminas MiSeq sekvensatorer som arbeidverktøy for kartlegging av patogener og antibiotikaresistens. MinIONs «long-read» sekvenser og MiSeqs «short read» sekvenser ble kombinert for mange stammer, og tillot oss å sette sammen flere helgenomer og komplette store plasmider. Nye varianter av resistensgener og uvanlige genkombinasjoner ble oppdaget. Interessen for, og forskning på heteropatogene *E. coli* og ESBL-produserende bakterier er sterkt økende

internasjonalt, og resultatene fra denne oppgaven vil bidra til viktig kunnskap om heteropatogenisitet både fra humane- og miljø- kilder.

Abbreviations

APEC	avian pathogenic *E. coli*
bla	β-lactamase
CMY	cefamycinase
CTX-M	cefotaximase-Munich
DAEC	diffusely adherent *E. coli*
DNA	deoxyribose nucleic acid
EAEC	enteroaggregative *E. coli*
EHEC	enterohemorrhagic *E. coli*
EIEC	enteroinvasive *E. coli*
EPEC	enteropathogenic *E. coli*
ESBL	extended-spectrum β-lactamase
ETEC	enterotoxigenic *E. coli*
ExPEC	extraintestinal pathogenic *E. coli*
HGT	horizontal gene transfer
HUS	hemolytic uremic syndrome
IPEC	intestinal pathogenic *E. coli*
KPC	*K. pneumoniae* carbapenamse
LEE	locus of enterocyte effacement
MDR	multidrug resistant (i.e. resistant to ≥ three antibiotic classes)
MGE	mobile genetic element
MLST	multilocus sequence typing
NCBI	National Center for Biotechnology Information
NGS	next generation sequencing
NMEC	neonatal meningitis *E. coli*
ONT	Oxford Nanopore Technologies
PCR	polymerase chain reaction
SHV	sulfhydryl-variable (β-lactamase/ESBLA-type)
STEC	shiga toxin-producing *E. coli*

Stx	shiga toxin
TEM	temoneira (β-lactamase/ESBLA-type; named after a patient)
UPEC	uropathogenic *E. coli*
UTI	urinary tract infection
WHO	World Health Organization
WIMP	What's in My Pot
VAG	virulence associated gene
VF	virulence factor
WGS	whole genome sequencing

List of papers

I. **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**
Lindstedt BA, Finton MD, Porcellato D, Brandal LT. BMC Infect Dis. 2018: **18**(1): 544.

II. **Whole genome sequencing and characterization of Multidrug-resistant (MDR) isolates from Norwegian environmental sources**
Finton MD, Meisel R, Porcellato D, Brandal LT, Lindstedt BA**.** *Manuscript.*

III. **Characterization of Clinical Heteropathogenic *Escherichia coli* Strains in Norway by Whole Genome Sequencing**
Finton MD, Meisel R, Porcellato D, Brandal LT, Lindstedt BA. *Manuscript.*

1. Introduction

1.1 Enterobacteriaceae

The Enterobacteriaceae represent a large family of gram-negative non-spore forming, facultative anaerobic rods that consists of over 50 genera with the number of species continually rising (Farmer *et al.*, 2007). Biochemical characteristics include reduction of nitrate to nitrite, oxidase negative, and catalase positive. Enterobacteriaceae are ubiquitous, living in both terrestrial and aquatic environments, but are also an important component of the mammalian intestine. However, there are pathogenic counterparts that can cause infection at multiple anatomical sites of the body. The most clinically relevant members of Enterobacteriaceae include *Citrobacter* spp., *Enterobacter* spp., *Klebsiella* spp., *Shigella* spp., *Salmonella*, and *Yersinia* spp. For the purpose of this thesis, *Escherichia coli* (*E. coli*) will be the main focus and *Klebsiella pneumoniae* (*K. pneumoniae*) will be included on occasion.

1.2 *Klebsiella pneumoniae*

K. pneumoniae is a common component of the gastrointestinal tract of humans, domestic animals, and livestock (Davis *et al*., 2015). This species is also prevalent in nature and can be found in both soil and water environments (Davis *et al.*, 2015). *K. pneumoniaeae* was first described by the German microbiologist and pathologist Carl Friedländer in 1882, and soon became known as a causative agent of pneumonia (Friedländer, 1882). However, pathogenic strains can cause a range of other infections including UTIs, pyelonephritis, meningitis, bacteremia, septicemia, and wound infections (Paczosa and Mecsas, 2016). Not only is this species an important hospital-acquired pathogen, but has also been isolated from raw vegetables, powdered infant formula, meat, fish, and street foods as a food-borne pathogen (Puspanadan *et al.*, 2012) (Overdevest *et al.*, 2014) (Sun *et al.*, 2010) (Davis and Price, 2016) (Haryani *et al.*, 2007). In addition, *K. pneumoniae* express numerous VFs necessary to establish infection, including capsular polysaccharides, lipopolysaccharides, fimbrial adhesins, and siderophores (Broberg *et al*., 2014) (Li *et al*., 2014).

1.3 *Escherichia coli*

E. coli colonize the newborn gut after exposure to maternal fecal matter and is predominant in the newborn gut microbiota (Adlerberth and Wold, 2009). It was first

described by the German-Austrian pediatrician and University professor Theodor Escherich in 1885 after its isolation from the feces of a newborn. *E. coli* remain the predominant facultative anaerobe of the human colonic microflora throughout life and is a beneficial member that aids in both digestion and defense against opportunistic pathogens. It is shed in the feces of warm-blooded animals and humans and is the most frequently observed species of facultative anaerobes in the environment (Biswas *et al.*, 2010). However, pathogenic counterparts are a serious public health problem as they have acquired the VAGs necessary to cause disease in multiple anatomical sites. Pathogenic *E. coli* is one of the most important causes of diarrheal disease and are responsible for approximately 70-90% of all human UTIs (Campos *et al.*, 2018), a leading cause of bacteremia (Laupland and Church, 2014), and the second most common cause of neonatal meningitis (Poolman and Wacker, 2016).

Transmission is primarily through the consumption of contaminated foodstuffs, including raw or undercooked ground meat products. Raw vegetables are also a significant contributor to infection due to fecal contamination from wild or domestic animals. The ingestion of contaminated water or cross-contamination during food preparation can also lead to colonization and infection (WHO; https://www.who.int/foodsafety/areas_work/foodborne-diseases/ecoli/en/). In addition, person-to-person contact by the fecal-oral route has been implicated as a transmission route.

Pathogenic *E. coli* are divided into intestinal pathogenic *E. coli* (IPEC) and extraintestinal pathogenic *E. coli* (ExPEC) depending on the location of the infection they are causing. Pathogenic strains that cause diarrheal disease are broadly regarded as intestinal pathogenic *E. coli*, and are divided into 6 well-described pathotypes: diffusely adherent, enteroaggregative, shiga toxin-producing/enterohemorrhagic, enteroinvasive, enteropathogenic, and enterotoxigenic *E. coli* (DAEC, EAEC, STEC/EHEC, EIEC, EPEC, ETEC, respectively) (Kaper and Nataro, 2004) (Croxen *et al.*, 2013). ExPEC subtypes include avian-pathogenic *E. coli* (APEC), neononatal meningitis *E. coli* (EPEC), sepsis-associated *E. coli*, and uropathogenic *E. coli* (UPEC) (Robins-Browne, *et al.*, 2016) (Sarowska *et al.*, 2019).

1.3.1 Intestinal Pathogenic *E. coli* (IPEC)

Diarrheal disease is one of the major causes of mortality amongst children <5 years of age, and IPEC has a significant role as a causative agent of these infections worldwide (Rojas-Lopez *et al.*, 2018). Each pathotype has distinct molecular markers that allow for their classification (Table 1).

1.3.1.1 Diffusely adherent *E. coli* (DAEC)

DAEC strains are defined by their diffuse adherence pattern on HeLa and Hep-2a cells in which the AFA/Dr adhesion is thought to be responsible (Scaletsky *et al.*, 2002). In addition, the fimbrial adhesion designated F1845 has also been reported as a cause of the diffuse adherence of this subtype, and the *daaD* gene is believed to be the F1845 determinant (Servin, 2005). However, Afa/Dr DAEC has been isolated from asymptomatic carriers, which suggests that a healthy and mature intestinal epithelial barrier can tolerate pathogenic DAEC strains (Servin, 2014). Nevertheless, Afa/Dr still has a major role in intestinal infection by triggering cell membrane receptor clustering, activation of the linked cell signaling pathways, and promoting structural and functional cell lesions and injuries to the intestinal barrier (Servin, 2014). The main transmission route for DAEC is food or water contaminated with human or animal fecal matter, and primarily causes diarrhea in children over 24 months of age (Scaletsky *et al.*, 2002).

1.3.1.2 Enteroaggregative *E. coli* (EAEC)

EAEC are emerging as causative agents of persistent diarrhea, especially in children living in developing countries (Aslani *et al.*, 2011) (Wanke *et al.*, 1991), and is a frequent cause of traveler's diarrhea of those from developed countries visiting undeveloped countries (Cabada Bauche *et al.*, 2011). The molecular marker that is used to identify this pathotype is *aggR*, which modulates fimbrial adhesion and facilitates penetration through the intestinal mucus by binding to lipopolysaccharide and altering the electrostatic properties (Okhuysen and DuPont, 2010). EAEC has a characteristic stacked 'brick-like' arrangement on the surface of cells in cell cultures. Contaminated food appears to be a prominent infection route for EAEC and has been identified as a cause outbreaks including an outbreak in Japanese children that consumed contaminated school lunches (Tokuda *et al.*, 2010).

1.3.1.3 Enteroinvasive *E. coli* (EIEC)

EIEC is a common cause of diarrheal disease in developed countries (Schaechter and Lederberg, 2004), but is also an important pathogen in developing countries due to poor sanitation and hygiene practices (Schaechter, 2009). EIEC is closely related to *Shigella* spp. as both are capable of causing dysentery through the same mechanisms of invasion. (Lan *et al.*, 2004). Adhesins are used to enter, multiply, and move in between the adjacent epithelial cells of the colon (Pasqua *et al.*, 2017), seriously damaging the intestinal wall. Although EIEC is not typically identified as a foodborne pathogen, several outbreaks have been associated with water and cheese (Bhavnani *et al.*, 2016).

1.3.1.4 Enterohemorrhagic *E. coli* (EHEC)/Shiga toxin-producing *E. coli* (STEC)

EHEC strains are a subgroup of STEC that are connected with certain serotypes that are frequently associated with outbreaks and severe clinical illness (Karmali *et al.*, 2003). The most significant serotype associated with outbreaks is O157:H7, however many other serotypes are able to cause disease. The major virulence factor behind STEC and EHEC infection is the production of shiga toxins (*Stx*), which may lead to the development of hemolytic uremic syndrome (HUS). However, the presence of the toxin itself may not be sufficient to cause HUS, and other bacterial factors may also play a role, including the locus of enterocyte effacement (LEE) region, a novel catalase-peroxidase encoded by the large plasmid of EHEC (*katP*), and the metalloprotease *stcE* (Griffin and Karmali, 2017) (Kobayashi *et al*., 2013). After ingestion of EHEC, the bacteria that pass through the acidic barrier of the stomach and adhere to the epithelial cells of the large intestine. The production of *Stx* begins, which is taken up by intestinal cells and enters circulation to the target organs (typically the kidneys) (Noris *et al.*, 2005). The intestinal tract of cattle is the primary reservoir, and undercooked meat is the major cause of infection. However, contaminated vegetables, juice, and unpasteurized milk have also been implicated in outbreaks (Pommerville, 2011). As the infectious dose is quite low (50-200 organisms), person-to-person contact may be a feasible mode of transmission (Schaechter and Lederberg, 2004).

1.3.1.5 Enteropathogenic *E. coli* (EPEC)

EPEC is responsible for thousands of deaths worldwide and has been identified as one of the most important pathogens that cause childhood diarrhea (Ochoa *et al.*, 2008).

Infection occurs upon colonization and T3SS injection of effector proteins into the host epithelial cell, which subsequently induces lesions in the cytoskeleton (Mansan-Almeida *et al.*, 2013). Virulence markers include the attaching and effacing (A/E) lesion encoded by the *eae* gene that is located on the LEE pathogenicity island. The bundle-forming pili (*bfp*) is another virulence trait that is responsible for the localized adherence of EPEC to host epithelial cells, however, *bfp* is not necessary for EPEC to cause infection. Strains without *bfp* (only containing *eae*) are referred to as atypical EPEC (aEPEC), and those containing both *eae* and *bfp* are characterized as typical EPEC (tEPEC). Epidemiological studies have shown that aEPEC has a higher prevalence of infection in both developed and developing countries and that the frequency of infection with tEPEC drops with age (Croxen *et al.*, 2013) (Afset *et al.*, 2006) (Ochoa *et al.*, 2008). Although tEPEC has only been found in humans, aEPEC has been isolated from a variety of animal species such as cattle, chicken, goats, pigeons and sheep (Cortes *et al.*, 2005).

1.3.1.6 Enterotoxigenic *E. coli* (ETEC)

ETEC is a primary cause of diarrhea in the developing world, particularly in children, and is the leading cause of traveler's diarrhea (Qadri *et al.*, 2005). The ETEC pathotype is molecularly defined as possessing one or two of the plasmid-borne heat-labile (LT) and/or heat stable (ST) enterotoxins that cause disease by adhering to the epithelium by means of colonization factors, secreting one or both of the enterotoxins that in return deregulate ion changes, thus promoting the loss of salt and water in the lumen resulting in diarrhea (Fleckenstein *et al.*, 2010). The intake of contaminated food or water is a primary mode of infection. The result of infection is often watery diarrhea, which contains up to 10^9 ETEC bacteria per ml, therefore a shared contaminated toilet or water may also pose a risk of transmission (Gonzales-Siles and Sjöling, 2015).

Table 1. Characteristics and molecular markers of IPEC pathotypes.

IPEC Pathotype	Type of diarrhea	Molecular Markers	References
ETEC	Acute, self-limited secretory diarrhea in children. Primarily endemic in developing countries and a significant cause of traveler's diarrhea (travelers from developed to developing countries)	Thermo-labile enterotoxin (LT) Thermo-stable enterotoxin (ST)	Veira *et al.*, 2007
EAEC	Mainly causes persistent diarrhea in children in developing countries	Aggregative regulator (*aggR*)	Nataro *et al.*, 1998
EPEC	Primarily endemic in developing countries, largely affects infants.	tEPEC: *eae* and *bfpA* aEPEC: *eae*	Veira *et al.*, 2007
EIEC	Primarily endemic in developing countries and causes dysentery, Watery diarrhea, inflammatory colitis	*ipaH*	Veira *et al.*, 2007
EHEC/STEC	Diarrhea, Hemorrhagic colitis, Hemolytic uremic syndrome. Causes worldwide outbreaks, HUS, and kidney failure	*Stx1*, *Stx2*, *eae*	Rojas-Lopez *et al.*, 2018
DAEC	Diarrhea in immunologically immature or malnourished children, usually over the age of 24 months. Primarily effects adults and children in developing countries	Afa/Dr *daad*	Meraz *et al.*, 2007 Scaletsky *et al.*, 2002 Servin *et al.*, 2005

1.3.2 Extraintestinal pathogenic *E. coli* (ExPEC)

ExPEC possess VFs that allow for the survival and invasion of anatomical sites outside of the intestinal tract. ExPEC is responsible for high morbidity and mortality rates that surpass those of IPEC infections (Riley, 2014). Unlike most of the IPEC pathovars, there are not a set of ExPEC VAGs that are universally accepted as unique genetic markers to positively identify each pathotype. VAGs associated with extraintestinal survival and disease have been identified amongst all ExPEC, and include those that enable iron acquisition, autotransporters, adhesins, protectins, toxins, and invasins.

1.3.2.1 Uropathogenic *E. coli* (UPEC)

It is estimated that approximately 150 million people across the world develop a UTI each year, and UPEC is the primary culprit (Terlizzi *et al.*, 2017). The main reservoir is the human intestinal tract, and due to the proximity of the urinary tract to the rectum, bacteria are capable of ascending the urethra into the bladder. Occasionally, the bacteria may migrate to the kidneys and cause pyelonephritis. Upon entrance to the urinary tract, UPEC must bind to host epithelial cells by expressing pili or fimbria that aid in attachment and reduce the risk of removal by urine flow (Lewis *et al*, 2016). UPEC then invades the epithelial cells and are transported into compartments that provide nutrients and allow for the evasion of antibiotics (Lewis *et al.*, 2016). Additionally, this translocation may become a reservoir for bacterial populations that parlay into chronic UTIs (Blango and Mulvey, 2010). The bacteria may also break free from superficial epithelial cells and multiply rapidly, forming biofilms that establish infection in the host (Lewis *et al.*, 2016). The molecular markers for UPEC are any two of the genes *chuA* (encodes a heme binding protein), *yfcV* (encodes the major subunit of a putative chaperone-usher fimbria), or *vat* (encodes an autotransporter serine protease toxin), when detected along with the gene *fyuA* (encodes the yersiniabactin receptor) (Spurbeck *et al.*, 2012).

1.3.2.2 Avian pathogenic *E. coli* (APEC)

APEC is a major pathogen associated with extraintestinal infections in poultry that result in large economic losses to poultry farms (Kemmett *et al.*, 2014). Infection leads to colibacillosis which is a syndrome that includes air sacculitis, respiratory tract infections, yolk sac infection, and swollen head syndrome (among others) (Markland *et*

al., 2015). However, the zoonotic potential of APEC cannot be brushed aside, as this pathotype shares common VFs with UPEC (Vincent *et al.*, 2010) (Nordstrom *et al.*, 2013). Therefore, poultry that is infected with this pathogen may serve as a transmission route to humans (Mitchell *et al.*, 2015) (Ewers *et al.*, 2007). A potential scenario is the ingestion of contaminated poultry, allowing the pathogen to gain access to the urinary tract in the same manner as UPEC, as UTIs are typically caused by one's own fecal matter.

1.3.2.3 Sepsis-associated *E. coli* (SEPEC)

In developed countries, SEPEC is an important pathogen involved in bloodstream infections with a fatality rate between 5 to 30% (Russo *et al.*, 2003) (Owrangi *et al.*, 2018). SEPEC strains may survive the bloodstream through serum resistance genes (*iss*) and the K2 capsular antigen (*kpsM*II) (Micenková *et al.*, 2017) (Russo *et al.*, 1998). A pore-forming endotoxin, HlyA, was also found to contribute to virulence during bloodstream infections and sepsis (Sonnen and Henneke, 2013). Although the occurrence of SEPEC has increased, the exact mechanisms of pathogenicity have yet to be resolved. The primary cause of a SEPEC infection is due to complications in patients with UTIs followed by complications due to intestinal disease (Laupland *et al.*, 2013) (Micenková *et al.*, 2017).

1.3.2.4 Neonatal meningitis *E. coli* (NMEC)

Those with the greatest risk for an infection with NMEC are Neonates under 28 days of age (Basmaci *et al.*, 2015). This pathogen possesses the ability to survive in the bloodstream and invade the meninges (Wijetunge *et al.*, 2015). Unfortunately, there is great heterogeneity among NMEC strains, making it difficult to distinguish from commensals (Wijetunge *et al.*, 2015). However, important VFs to pathogenicity have been identified, and include iron acquisition systems (*iro fyuA*), adhesins (*sfa, foc,* pap*G*II/III), toxins (*hly, hra, hek*), serum resistance (K1 antigen, *iss*) and those that promote invasion into cells and tissues (*ibeA*, *cnf*1). Currently, the mechanisms behind the pathogenesis of NMEC and SEPEC pathotypes have not been fully elucidated (Sarowska *et al.*, 2019).

1.3.3 Heteropathogenic *E. coli*

The different pathovars of IPEC were traditionally thought to be distinct from one another and were identified by the molecular markers associated with their pathogenicity. In addition, ExPEC was thought to harbor VAGs absent in IPEC strains. However, it is now clear that novel combinations of VAGs from multiple pathotypes (both IPEC and ExPEC) may exist in a single strain, often referred to as a 'heteropathogen' or 'hybrid' (Kaper *et al.*, 2004). As VAGs are frequently carried on mobile genetic elements (MGEs), genetic traits can be transferred between bacteria. This exchange of can lead to divergent pathogroups from the traditional *E. coli* classification system, making the correct diagnostic classification difficult and may subsequently stall further course of action. These strains may result in a more severe disease that is capable of increased spreading potential, as was demonstrated during the 2011 STEC/EAEC heteropathogen outbreak in Germany that spread to other European countries and North America. Approximately 4000 people were affected, over 900 developed HUS, and there were 54 deaths (Karch *et al.*, 2012). The seriousness of heteropathogenic strains has been demonstrated in patients with a gastrointestinal infection that subsequently became a systemic infection (Daga *et al.*, 2019) (Vaishnavi, 2013). In addition, studies have shown that the mortality rate was higher among those whose sepsis originated from the digestive tract rather than infections that originated from UTIs (Lefort *et al.*, 2011).

Heteropathogens have been reported in the environment (Gati *et al.*, 2019), clinical settings (Lindstedt *et al.*, 2018) (Khairy *et al.*, 2019), livestock (Johura *et al.*, 2016), domestic animals (Johura *et al.*, 2016), and wildlife (Bai *et al.*, 2019), which highlights their presence in many niches that may give ample opportunity for further exchange of genetic traits.

1.4 Treatment of *E. coli* infections

Diarrheagenic *E. coli* do not typically require treatment with antibiotics, and it is commonly recommended that those at risk of dehydration simply opt for fluid replacement. Due to the production of *Stxs*, antibiotics aren't a recommended path of therapy, as they may increase production of the toxins and escalate the risk of HUS (Karch *et al.*, 2005). In contrast, ExPEC infections range from mild (bladder

inflammation) to life threatening (sepsis, neonatal meningitis) and rely on the use of antibiotics to fight infection. The common antibiotics prescribed in cases of *E. coli* infection include extended spectrum β-lactams, fluoroquinolones, trimethoprim-sulfamethoxazole, and nitrofurantoin (Daneman *et al.*, 2019). Although these antibiotics have been successful options for treating infection in the past, due to persistent exposure, resistance has become a prominent global issue. Not to mention, the emergence of STEC/ExPEC heteropathotypes presents a serious problem, as the use of antibiotics to combat an extraintestinal infection may increase the production of toxins of STEC and escalate the risk of HUS (Karch *et al.*, 2005).

1.5 Antimicrobial agents

The modern era of antibiotics began with Sir Alexander Fleming's accidental discovery of penicillin in 1928. Since then, antibiotics have revolutionized modern medicine and saved millions of lives. Additionally, medical care from surgery to cancer therapy rely greatly on the accessibility of effective antibiotics. Due to the discovery of antibiotics, there has also been a significantly improved life expectancy – the average person was only expected to live for approximately 47 years prior to the availability of antimicrobial drugs (Adedeji, 2016). Commonly referred to as the 'golden era of antibiotics', the years between 1940 and 1962 proved to be an ambitious period in the discovery of new classes of antibiotics. Aminoglycosides were brought to market in 1944, tetracyclines in 1950, macrolides in 1952, polymyxins in 1958, and fluoroquinolones in 1962 (Figure 1). Science discovered a continuous new armory of drugs that kept pace with emerging resistance to the available drugs. Eventually, the development of new antibiotics dwindled, and as a result our world's 'medicine cabinet' has taken a hit. Although there have been modifications to antibiotic classes already in existence, there have been no new classes of antibiotics brought to market since the discovery of trimethoprim in 1968. A concern is that the funding for antibiotic research is of little interest to investors, as there is simply not a high enough return on the investment to gain appropriate attention and funding. Not to mention, the development of an entirely new drug is a challenge; it must not only be able to kill the bacteria that is causing the infection but must also be low-cost to manufacture and safe for consumption.

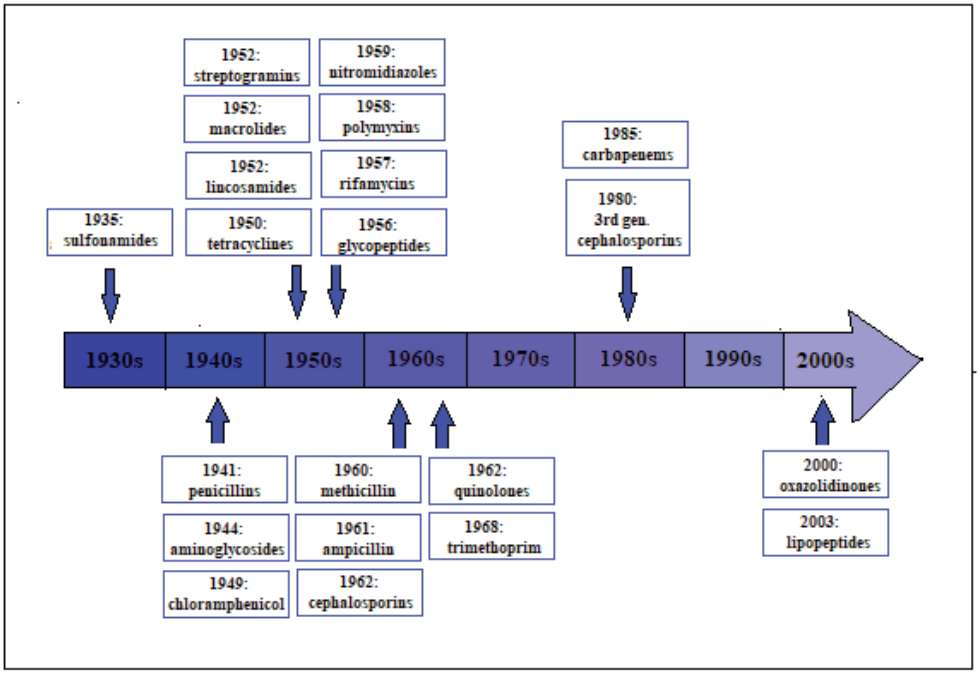

Figure 1. Milestones in the introduction of new antibiotic classes.

Antibiotics vary in their mechanisms of action, antimicrobial spectra, and toxicity. They are typically distributed amongst five groups according to their mode of action and antibacterial target: (i) inhibitors that target cell wall synthesis (i.e. β-lactams, bacitracin), (ii) inhibitors of cell membrane function (i.e. polymyxins), (iii) inhibitors of protein biosynthesis (i.e. aminoglycosides, macrolides, tetracyclines), (iv) inhibitors of DNA synthesis (i.e. quinolones), and (v) folic acid metabolism inhibitors (i.e. sulfonamides, trimethoprim) (Kapoor *et al.*, 2017).

1.6 β-lactam and extended-spectrum β-lactam antibiotics

β-lactams are the most important and widely used therapeutic class of drugs in human and veterinary medicine due to their broad spectrum of activity and low toxicity (Bush and Bradford, 2016). All β-lactam antibiotics have the same core 3-carbon and 1-nitrogen ring, called the β-lactam ring (Figure 2). Primary targets of the β-lactams are the PBPs, which are bacterial enzymes involved in peptidoglycan synthesis. Peptidoglycans are an essential component of the bacterial cell wall. The β-lactam ring

mimics the D-alanyl D-alanine region of the peptide chain and acts as a false substrate for d-alanyl-d-alanine transpeptidases (Kapoor *et al.*, 2017). This class of drugs acts by a covalent attachment to penicillin binding proteins (PBPs), thus impeding the growth of sensitive bacteria by inhibiting cell wall synthesis (Kong *et al.*, 2010). The spectrum and effects of the various β-lactams are determined by the PBPs to which these antibiotics bind (Vincent *et al.*, 2011). Based on side chain variations, the β-lactams are classified as either penicillins, cephalosporins, monobactams, or carbapenems (Yao *et al.*, 2011).

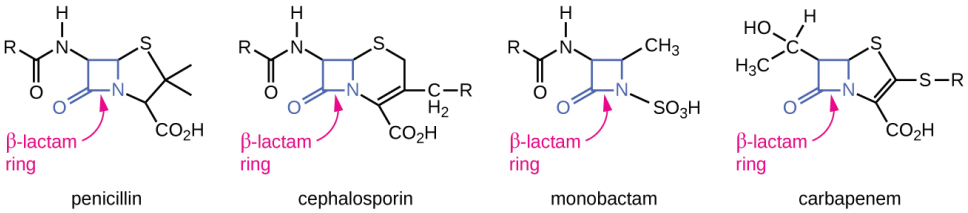

Figure 2. Chemical structure of β-lactam antibiotics. All β-lactam antibiotics have the same core 3-carbon and 1-nitrogen ring, called the β-lactam ring. Obtained from Parker *et al.,* 2016.

1.7 What is antibiotic resistance?

"It is not difficult to make microbes resistant to penicillin in the laboratory by exposing them to concentrations not sufficient to kill them, and the same thing has occasionally happened in the body." – Alexander Fleming in his Nobel Lecture "Penicillin", December 11, 1945

Antibiotic resistance is defined as the reduction in effectiveness of a drug, such as an antibiotic, in the treatment of a disease or infection. Antibiotic resistance is naturally occurring, as many of the antibiotics that we use today (i.e. penicillins, streptomycin, monobactams, tetracycline, cephalosporins, and carbapenems) are produced by soil dwelling fungi for their survival in a competitive environment (Clardy *et al.*, 2009) (Newton *et al.*, 1956) (Kahan *et al.*, 1979) (Sykes *et al.*, 1986). This may present a problem if resistance genes are transferred to pathogenic strains capable of causing

bacterial infections in humans or animals. An example of this is the CTX-M gene that encodes resistance to third generation cephalosporins, which has become one of the most problematic resistance genes to date. In addition, CTX-M is thought to have originated from environmental *Kluyvera*, a genus rarely associated with clinical disease (Bevan *et al.*, 2017) (Narchi, 2005). Although antibiotic resistance occurs naturally as a survival mechanism against the bacteria and fungi that produce them, the misuse and overuse of antibiotics by humans is one of the major selection pressures on the bacterial population. Natural selection will favor the resistant bacteria while the susceptible bacteria are killed, leaving resistant bacteria to multiply.

After penicillin brought antibiotics to the forefront of the scientific community, there were numerous companies invested in the production of new antibiotic classes. As newer and more powerful agents were continually developed, concerns about emerging resistance were alleviated. Eventually, the development of new antibiotics dwindled, and as a result our world's 'medicine cabinet' has taken a hit. As the same antibiotics are used, bacteria become multidrug resistance, and the available options for treatment diminish, becoming a substantial clinical problem. In fact, the average amount of time that it takes between the introduction of a new antibiotic into clinical practice and bacterial resistance to many of those antibiotics has been less than 10 years (CDC, 2013) (Humphries, 2015). As extended-spectrum β-lactam antibiotics are the most widely prescribed antibiotics in the world, the production of extended-spectrum enzymes (ESBLs) is a serious global public health threat, and it is essential that novel classes are discovered and placed onto the market. As with all antibiotics, resistance will be eventually be gained, but a novel class will reset the countdown towards a period where antibiotics can no longer help.

1.8 Antibiotic susceptibility testing

Antibiotic gradient strips are the most commonly used method for susceptibility testing of bacterial isolates in laboratories, and was the primary method used in this thesis. The strips contain an antimicrobial concentration gradient that allows the user to visualize the minimum inhibitory concentration (MIC) on agar plates. The MIC is the lowest antimicrobial concentration that inhibits visible growth of a microorganism after an overnight incubation and is used in microbiological laboratories to measure

the activity of an antibiotic against a specific microorganism. We used the MIC clinical breakpoints, which is a selected concentration of an antibiotic which defines the microorganisms as susceptible or resistant, from the European Committee on Antimicrobial Susceptibility Testing (EUCAST) to determine susceptibility (EUCAST, 2019). In addition, multidrug resistance was defined as phenotypic resistance to three or more classes of antibiotics.

1.9 Intrinsic vs. acquired antibiotic resistance

Antibiotic resistance can be intrinsic or acquired. Intrinsic resistance is naturally occurring and present in all bacterial species (Cox and Wright, 2013). As mentioned earlier, intrinsic resistance occurred even prior to clinical use. This may be due to a combinations of mechanisms, including an altered target site that would be necessary for the antibiotic to bind to, the existence of efflux pumps that actively transport the drugs out of the cell, or due to a low permeability to a certain antibiotic due differences in the bacterial membrane and the chemical nature of the drug (among other mechanisms featured in Figure 3). In addition, a fully susceptible microbe may also act resistant if present in a biofilm.

Acquired resistance occurs when a naturally susceptible bacterium develops or acquires resistance through mutations or horizontal gene transfer. For instance, mutations in chromosomal genes such as *gyrA, parC, parE* cause altered antibacterial or transcriptional changes and lead to fluoroquinolone resistance (Martinez *et al.*, 1998) (Huseby *et al*, 2017) (Moon *et al.*, 2010). The acquisition of resistance genes can also occur through horizontal gene transfer (HGT), for example, plasmid-mediated acquisition of β-lactamase encoding genes (Jacoby and Sutton, 1991). Another mechanism would be mutations in previously acquired genes, such as TEM-1, that result in the production of enzymes with a broader spectrum (Jacoby and Medeiros, 1991).

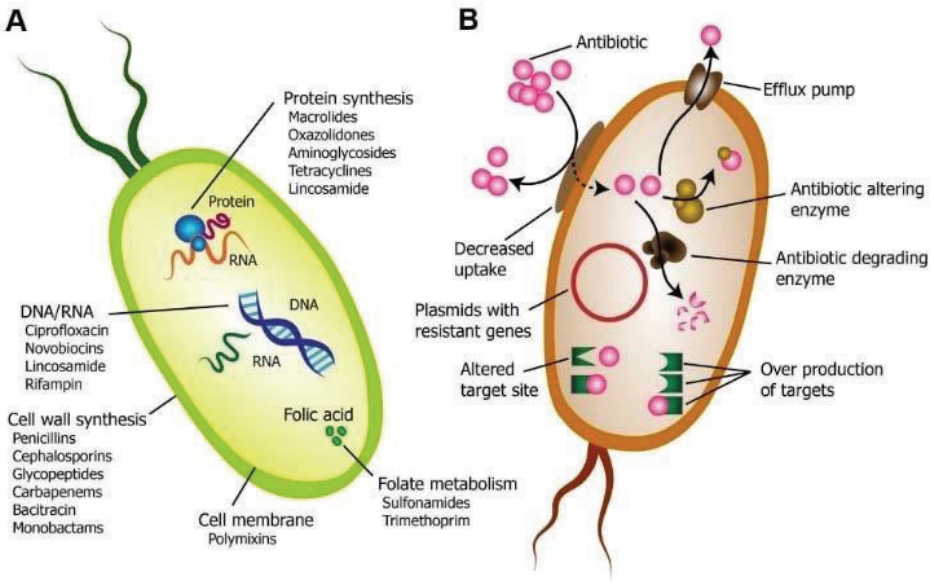

Figure 3. A. Common antibiotic target sites. **B.** Mechanisms of resistance. Rights to reprint from Elsevier (Ganewatta and Tang, 2015).

1.10 Mechanisms of β-lactam resistance

To evade β-lactam antibiotics, bacteria halve evolved several mechanisms of resistance, including: i) variation of target PBPs, ii) alteration of the peptidoglycan structure, and iii) expression of efflux pumps to reduce drug permeation across the outer membrane, and iv) the production of β-lactamases (Hakenbeck and Coyette, 1998). Additionally, the accumulation of point mutations in β-lactamases can cause resistance (Sun *et al.*, 2014).

1.10.1 β-lactamases

β-lactamases are ancient bacterial enzymes that existed before therapeutic antibiotics were brought to market, and plasmid-borne β-lactamases are estimated to have appeared millions of years ago (Barlow and Hall, 2002). These enzymes have been observed in several isolated South American populations that have rarely, if ever, been treated with β-lactam antibiotics (Bartoloni *et al.*, 2004) (Bartoloni *et al.*, 2009), as well as in ice cores of glaciers outside of Antarctica (Segawa *et al.*, 2013), and multiple pristine Antarctic soil samples (Van Goethem *et al.*, 2018). The first β-lactamase

(penicillinase) was detected in *E. coli* in 1940, even before penicillin was in clinical use (Abraham, 1940). However, by 1941 penicillin was used in clinical practice, which ultimately contributed to the emergence of plasmid-mediated penicillinases (Unemo and Shafer, 2014). To compete with emerging resistance, new molecules were added to the β-lactam ring. By the 1960s, the first plasmid-mediated β-lactamase was found in Greece and was designated TEM after the patient (Temoneira) who was infected with the pathogen (Paterson and Bonomo, 2005). The TEMs quickly spread around the globe, and soon SHV (sulphydryl variable) was discovered. Additionally, AmpC type β-lactamases were primarily encoded on the chromosome of various gram-negative bacteria, and through mutations in the promoter region, the bacteria gained the ablility to express high enough levels to confer clinical resistance to ampicillin and cephalosporins (Batchelor *et al.*, 2005).

β-lactamase production is the most common mechanism of β-lactam resistance in gram-negative bacteria. These bacterial enzymes inactivate β-lactam antibiotics by hydrolyzing the β-lactam ring (Hakenbeck and Coyette, 1998). β-lactamases can be found on the bacterial chromosome or a transmissible plasmid, with plasmid-encoded β-lactamases often expressed at high levels and chromosomally located β-lactamases express at low levels until the introduction of a substrate (Williams, 1999) (Jacoby, 1994). β-lactamases vary in their spectrum of activity depending on the structure of their side chains, including: i) narrow-spectrum β-lactamases, such as penicillinases, ii) broad-spectrum β-lactamases (ampicillinases), iii) extended-spectrum β-lactamases that hydrolyze 3rd and 4th generation cephalosporins and monobactams, and iv) carbapenamases, which hydrolyze all β-lactams including the last resort carbapenems.

1.10.2. Extended-spectrum β-lactamases

ESBLs and carbapenemases are currently considered one of the major public health threats worldwide. ESBLs commonly evolved from enzymes with a narrower spectrum, for instance, from the TEM-1 and SHV-1 enzymes (Paterson and Bonomo, 2005) (Bush and Jacoby, 2010). By the early 2000s, the CTX enzyme rapidly emerged in both hospital and community settings, quickly becoming the most wide-spread and clinically relevant ESBL worldwide. As stated previously, CTX-Ms are thought to have transferred from the environmental *Kluyvera* spp., and not by mutations (Cantón,

2008). However, it was only a matter of time until mutations of the CTX-M group emerged due to antibiotic selective pressure (Cantón *et al.*, 2012), leading to the development of new variants that ultimately decreased susceptibility to cephalosporins. Of the ESBLs, the CTX-M group is the predominant type worldwide (Adamski *et al.*, 2014). The CTX-M genes are also highly mobile, due to their presence on MGEs such as plasmids, transposons, or integrons, all of which enable their ability to effectively spread between bacteria. Although AmpC has been located on the chromosome, the genes that encode AmpC have become mobile and are increasingly found on plasmids. Thus, they have gained the ability to spread resistance to 3rd generation cephalosporins horizontally between different species of Enterobacteriaceae (Alvarez *et al.*, 2004), with CMY-2 having the broadest geographic dissemination. It is also common that co-resistance to fluoroquinolones, aminoglycosides, and trimethoprim is detected, all of which are choice antibiotics for the treatment of critically ill patients (Brolund *et al.*, 2014).

1.11 ESBL classification systems

β-lactamases are classified based on amino acid sequence according to Ambler (Ambler, 1991) or by functional characteristics according to Bush-Jacoby (Bush and Jacoby, 2010). The Ambler classification system divides β-lactamases into four major classes (A, B, C, and D). Classes A, C, and D function by the serine ester hydrolysis mechanism, and class B are metallo-β-lactamases that utilize zinc ions to facilitate β-lactam hydrolysis (Zhang and Hao, 2011). The Bush-Jacoby method is more applicable to the physician or microbiologist as it considers substrate and inhibitor profiles that group the enzymes in ways that can be correlated with their phenotype in clinical isolates. The three major groups in the Bush-Jacoby classification are: Group 1 cephalosporinases (class C), Group 2 that are broad spectrum, inhibitor resistant, and ESBLs and carbapenemases (Class A and D), and Group 3 metallo β-lactamases (Class B), all of which have corresponding sub-groups (Table 2). As of 2016, β-lactamases are numbered at close to 2,800 unique proteins (Bush, 2018).

β-lactamases also vary in the spectrum of their activity against antibiotics depending on the structure of their side chains. β-lactamases may be divided into: (i) narrow-spectrum β-lactamases (i.e. penicillinases), (ii) broad-spectrum β-lactamases (i.e.

ampicillinases), (iii) extended spectrum β-lactamases (ESBLs), which may hydrolyze 3rd and 4th generation cephalosporins and monobactams, and (iv) carbapenemases, which may hydrolyze all β-lactams including the last resort carbapenems.

AMBLER[A]	BUSH-JACOBY[B]		EXAMPLES	DEFINITIONS
Class A Serine	Group 2be L2 - group 2e	Class A ESBLs	CTX-M, TEM, SHV, and L2	Non-susceptibility to extended-spectrum cephalosporins and monobactams. Inhibited by clavulanic acid
Class A Serine	Group 2de	OXA-ESBLs	OXA-11, OXA-15	Inhibitor not known
Class A Serine	Group 2f	KPC and other class A carbapenemases	KPC and GES	Inhibited by boronic acid
Class B Metallo	Group 3a	MBLs	IMP, VIM, NDM, and L1	Inhibited by EDTA and dipicolinic acid
Class C AmpC	Group 1	Plasmid-mediated AmpC	CMY	Inhibited by cloxacillin and boronic acid
	Group 3b	Carbapenems	L1	Inhibited be EDTA
Class D Serine	Group 2df	OXA-carbapenemases	OXA-23 and OXA-48	Inhibitor not known

Table 2. The ESBL classification scheme of Ambler and Bush/Jacoby.
[a]From Ambler, 1980 [b]From Bush and Jacoby, 2010

1.12 Dissemination of ESBLs (and other resistance factors)

1.12.1 Vertical transfer (clonal spread)

Antibiotic resistance genes that are located on the chromosome are passed vertically from the mother to daughter bacterial cell during cell division (i.e. clonal spread). This is the mechanism that causes multiplication and dissemination into the community or hospital, and frequently causes outbreaks.

1.12.2 Horizontal gene transfer (HGT)

Horizontal gene transfer is the movement of genetic material between microorganisms and is the primary mechanism for the spread of antibiotic resistance. The mechanisms of HGT are (i) transformation, which is a form of genetic recombination where a competent recipient uptakes DNA from the environment or when MGEs move around to different positions within the genome of a single cell (ii) transduction, which involves the transfer of DNA from one bacterium to another through a bacteriophage and (iii) conjugation, which is the bacterial equivalent of sexual reproduction. One bacterium must carry the transferrable plasmid (referred to as F+) and the second bacteria must not (referred to as F-). Once the transfer of DNA (i.e. plasmids and transposons) by direct cell-to-cell contact is complete, then the second bacterial cell is now F+ (Figure 4) (Sun, 2018).

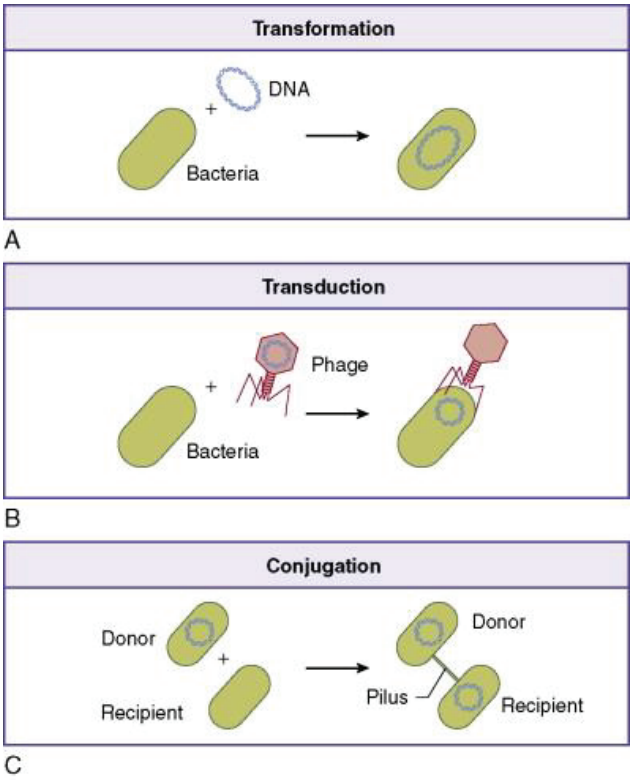

Figure 4. Mechanisms for horizontal transfer of genetic material in bacteria. **A.** Transformation **B.** Transduction **C.** Conjugation Rights to reprint via Elsevier (Actor, 2012).

1.12.13 Mobile genetic elements (MGEs)

Bacteria often evolve to resist antimicrobials through the acquisition of pre-existing resistance determinants. This is accomplished by MGEs that move within and between DNA and include insertion sequences, transposons, and gene cassettes/integrons, and those that have the ability to transfer between bacterial cells, such as bacteriophages and plasmids (Pfeifer *et al.*, 2010) (Mathers, 2016) (Brolund and Sandegren, 2016). MGEs may encode VFs and resistance determinants that are passed between bacteria and are considered significant contributors to the dissemination of resistance genes. Insertion sequences and transposons (or 'jumping genes') are stretches of DNA that are capable of moving themselves (and the determinants that they carry) within a genome or between genomes. Integrons use site-specific recombination to relocate resistance determinants between defined sites (Domingues *et al*, 2012). As figure 4B shows, intercellular mechanisms of genetic exchange include bacteriophage elements that integrate randomly into the genome by transduction and conjugation. Figure 4C illustrates conjugation, which is mediated by plasmids. When resistance genes are mobilized by phages or plasmids, their spread can be significantly accelerated (Peterson and Kaur, 2018). Conjugative plasmids are the most important contributor in the dissemination of antibiotic resistance genes (Milan, 2018).

1.12.14 Plasmids

Plasmids are small, circular, double-stranded DNA molecules within most bacterial cells that are an extrachromosomal genetic element and replicate independently. Plasmids can be transferred both vertically and horizontally and often contain VFs and resistance genes that may give the bacteria a selection advantage (Sherratt, 1982) (Hasegawa *et al*., 2018). Their spread is problematic, as they evolve and adapt to a variety of environments by sharing and rearranging genetic elements. Both narrow spectrum (i.e. penicillins) and ESBLs are frequently found on plasmids in Enterobacteriaceae. In addition, VFs can be spread on plasmids, which not only promotes the stability and maintenance of the bacterium, but also contributes to the observed heteropathogenicity of *E. coli* (Carattoli, 2013). ESBLs located on plasmids are often found with resistance genes to other antibiotics, such as aminoglycosides, trimethoprim, sulfonamides, fluoroquinolones, and tetracyclines other (Paterson, 2000). As a result, multidrug resistant organisms to some of the most important antibiotics are becoming more and more common and pose a major problem for the

effective treatment of bacterial infections. The IncF plasmids are among the most common types that carry ESBL-encoding genes, namely CTX-M ESBLs, and is also the most common incompatibility group found in Norway (Brolund, 2014) (Naseer *et al.*, 2009) (Brolund *et al.*, 2013). IncF plasmids are widely distributed in Enterobacteriaceae, and replicons can be divided into IncFIA, IncFIB, and IncFIC. The IncFII, IncA/C, IncL/M, IncN, and IncI plasmids are also associated with ESBLs carriage.

1.13 Surveillance of ESBL-producing Enterobacteriaceae in Europe

The incidence of bacterial infections caused by ESBL-producing Enterobacteriaceae have only been increasing around the world, even in low prevalence countries like Norway. The Norwegian Surveillance System for Antimicrobial Drug Resistance (NORM) is a national health registry that monitors antimicrobial drug resistance and its changes over time. Norway has taken the initiative to limit antimicrobial consumption as a preventative measure to reduce the occurrence of antimicrobial resistance. Although the overall presence of ESBL-producing Enterobacteriaceae in clinical isolates of Norway is low compared to other European countries, their presence has still increased since 2003 (Figure 5).

Since 2004, NORM has been the main driver behind Norway's participation in the European Antimicrobial Resistance Surveillance Network (EARS-NET), which is a publicly funded system for the surveillance of antimicrobial resistance throughout Europe. In addition, the European Antimicrobial Resistance Genes Surveillance Network (EURGen-Net) is a network for the surveillance of clinically important multidrug resistance bacteria and is coordinated by the European Center of Disease Prevention and Control (ECDC). Currently, a total of 37 European countries participate in the EURGen-Net Europe-wide survey of carbapenem and/or colistin resistant Enterobacteriaceae. Over the course of 6 months, hospitals and associated microbiology laboratories in the participating countries collect carbapenem and/or colistin resistant *K. pneumoniae* and *E. coli* isolates and forward the isolates to the national reference laboratory for confirmatory testing, and then to a central strain collection for whole for whole genome sequencing. Figure 6 illustrates the increasing prevalence between (2013 and 2018) of resistance in participating European countries

to the last resort drug given for serious bacterial infections (carbapenems and colistin). The geographical differences may be a reflection of less strict regulations on antibiotic use in European regions outside of Scandinavia.

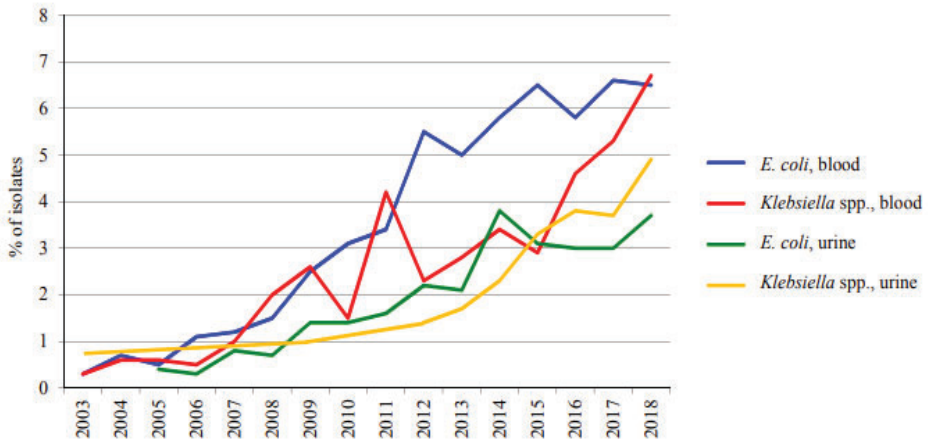

Figure 5. Prevalence of ESBL-producing *E. coli* and *K. pneumoniae* spp. in blood and urine from Norwegian patients between 2003 and 2018. (Source/obtained from: NORM/NORM-VET).

Antibiotic resistance constitutes a global health threat, with international travel contributing greatly to worldwide spread. ESBL producing *Enterobacteriaceae* has become the most common type of MDR resistant bacteria and is highly prevalent in developing regions of the world, with South Asia and Southeast Asia identified as one of the highest risk factors (Tängden *et al.*, 2010). The vast majority of visitors to these high-risk locations may remain asymptomatic with no infection. However, those asymptomatic travelers may then carry these ESBL-producing bacteria back to their respective countries, further disseminating resistance genes (Woerther *et al.*, 2017).

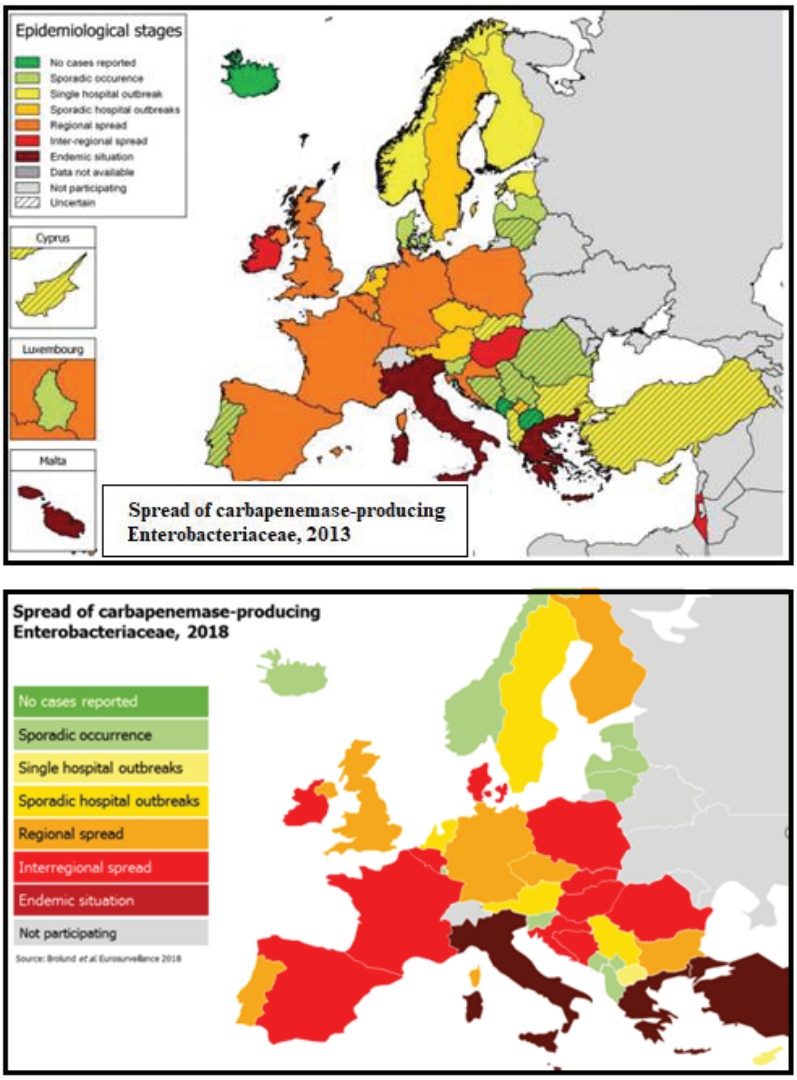

Figure 6. Proportion of carbapenemase-producing Enterobacteriaceae isolates in European countries in 2013 and 2018. (Source: ECDC/EARS-Net Maps, http://www.ecdc.europa.eu).

1.14 DNA – The Molecule of Life

"DNA is like Midas's gold, everyone who touches it goes mad." – Maurice Wilkins

The twentieth century was full of curiosity, intrigue, and breakthroughs regarding the structure of DNA (Figure 7), which is found in all prokaryotic and eukaryotic cells. In

1949, Ewin Chargaff discovered that the 'alphabet of the DNA' – four nitrogenous bases named adenine (A), thymine (T), cytosine (C), and guanine (G), that selectively bind to one other through complementary base pairing (adenine binds to thymine, and cytosine to guanine). In 1951, Rosalind Franklin used X-ray diffraction on DNA fibers to form critical evidence in the identification of the physical structure of DNA. In 1953, James Watson and Francis Crick utilized the information from Franklin's DNA image to determine the chemical structure of DNA – 4 nitrogenous bases connected to a sugar-phosphate backbone in the formation of a double helix. The discovery of the structure of DNA opened the door to the development of efficient and promising sequencing technology.

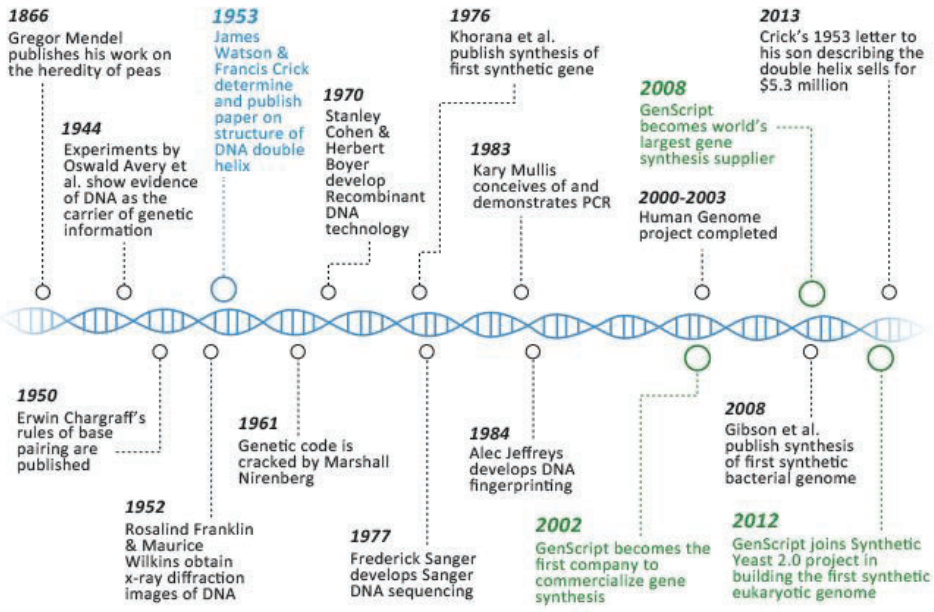

Figure 7. Timeline of significant contributions to knowledge of DNA.

1.14.1 DNA sequencing technology

The first sequencing technology was invented by Frederick Sanger and colleagues in 1977 and was termed Sanger sequencing. This First generation sequencing approach was the most widely used sequencing method for approximately 40 years, and is based on the selective termination of chain-elongation by using chemically modified nucleotides called dideoxynucleotides (ddNTPs), which is a nucleotide analogue that lacks a 3'-hydroxyl group necessary for the incorporation of the next nucleotide to be attached to. The chain terminating nucleotide is identified by its fluorescent dye, giving 4 different colors to each of the 4 different bases (Sanger, 1977). The samples are then separated according to their size using a gel, where an X-ray or UV light allows for the visualization of bands that correspond to the size of the DNA fragments. Although Sanger sequencing offers a high accuracy (99.9%), the speed and cost of analysis outweighs the benefits of this method in comparison to newer technologies.

Next generation sequencing (NGS) technology can be divided into second generation sequencing, which is distinguished based on the need to prepare amplified products prior to sequencing, and third generation sequencing technology, which does not require amplification (Kchouk *et al.*, 2017). The instruments used for second generation include pyrosequencing (Roche 454), Illumina sequencing technology (MiniSeq, MiSeq, NextSeq, HiSeq), ion torrent technologies, and SOLiD sequencing technologies. Third generation sequencing instrumentation are PacBio technologies and the Oxford Nanopore sequencers (MinION, PromethION, and GridION). NGS platforms can sequence millions to billions of reads in a single run and the time required to generate a GigaBase in sized reads in only a few days or hours, making it much more efficient with more information than Sanger sequencing. For example, sequencing the human genome with Sanger took multiple laboratories almost 15 years, not to mention the cost was approximately 800 million Norwegian Kroners (approximately 100 million US dollars) (Wheeler *et al.*, 2008). On the other hand, NGS of the human genome with the 454 Genome Sequencer FLX took approximately 2 months for about one hundredth of the cost (Wheeler *et al.*, 2008).

1.14.2 Whole genome sequencing (WGS)

'DNA sequencing' is the process of determining the sequence of the 4 nitrogenous bases in a strand of DNA. WGS is the process of sequencing an organism's entire genome. In 2003, the completion of the Human Genome Project marked the beginning of an era that introduced rapidly evolving technology that would make WGS more accessible, easier, and more affordable. Due to rapidly dropping sequencing costs, WGS is a future option for diagnostics and surveillance testing in clinics and hospitals, as this method gives the user the ability to look deep into the genome of a selected organisms to explore areas of the DNA that weren't previously accessible. The information obtained from WGS allows the user to bypass previously laborious and time-consuming methods and use the resulting data to determine species identification, genotype determination, present resistance genes, and VAGs. In this thesis, we submitted samples for Illumina MiSeq sequencing and conducted Oxford Nanopore MinION sequencing in-house. Through the compilation of data from both instruments together, we obtained several whole genomes.

1.14.2.1 Illumina MiSeq

The Illumina MiSeq process is divided into 4 different steps: i) library preparation ii) cluster generation iii) sequencing and iv) data analysis. During sample preparation, the DNA is fragmented and adapter-ligated fragments (containing the sequence binding site, indexes, and regions that are complementary to the oligos on the Illumina chip), are added to the end of the DNA fragments. For cluster generation, the library is loaded into a flow cell that contains a 'lawn' of two types of oligos that are complementary to the library adapters. The adapters attached to the DNA will then hybridize with the complementary oligo on the flow cell and each fragment will be amplified into distinct clonal clusters through bridge amplification. In this process, the strand folds over and the adapter region hybridizes to the second type of oligo on the flow cell. Polymerases generate the complementary strand, forming a double-stranded bridge. This bridge is denatured, resulting in two single-stranded copies that are secured to the flow cell. This process is repeated and occurs simultaneously for millions of clusters. When cluster generation is complete, the reverse strands are cleaved and washed off and sequencing begins. With each cycle, fluorescently tagged nucleotides compete for the addition to the growing chain which minimizes

incorporation bias and reduces the error rate. After the addition of each nucleotide, the clusters are excited by a light source, and the characteristic fluorescent signal is emitted (sequencing by synthesis) and emission wavelength and signal intensity determines the basecall. During data analysis, fragments are quality filtered, reads are merged, and similar fragments are locally clustered creating contigs or aligned to a reference genome (Illumina, 2015).

1.14.2.2 Oxford Nanopore MinION Sequencer

The Oxford Nanopore MinION has emerged as a competitive, portable technology (weighing under 100 grams) and was released in 2013. The device measures 4 inches in length and plugs into a PC or laptop using a high-speed USB 3.0 cable. With the MinION, it is only required to have an internet-connected laptop with the necessary software installed to run the experiment, making it possible to establish sequencing-on-the-field where access to lab equipment is limited. Oxford Nanopore Technologies has continually evolved, and the instruments GridION and PromethION have emerged, which are scaled-up versions of the MinION that are able to run 5 and 48 flow cells, respectively, at the same time. The MinION has garnered a lot of attention in the scientific community in large part due to the generation of long reads (exceeding 150 kbp) that will ensure a higher resolution to repeat content, low cost, and small size. There are three components needed to run the MinION sequencing device: the MinION sequencing device, the consumable flow cells, and the MinKNOW sequencing software. In addition, the appropriate kit for your sequencing objectives must be purchased. Each flow cells contains up to 2048 individual nanopores. Each flow cell can generate up to 30 Gb of DNA sequence data, or 7-12 million reads with RNA. Once the nanopores in the flow cell has 'died', it must be returned to Oxford Nanopore and a new flow cell much be purchased for future sequencing experiments.

With this sequencing technology, the first strand of DNA is linked by a hairpin to its complementary strand of DNA. The strands of DNA pass through a protein nanopore that is embedded into a membrane by action of a motor protein attached to the pore. An ionic current goes through each pore, and by measuring the variation in the current, single bases can be identified based on their distinct disruption (Figure 8). 1D (or one directional) reads are when the sequencing is made on the template strand,

then the hairpin structure is read, followed by the inverse strand generating the complement read. 2D (or two directional) reads are also an option in which the template and complement reads are combined. The nanopore will continue to sequence DNA strands until pore life is exhausted or until the user determines that enough real time data or sequence coverage was generated and terminates the run. The ability to analyze data in real time is a major advantage of the ONT system in clinical scenarios, as real time detection can provide quick epidemiological information, such as the relatedness of outbreak strains, antibiotic resistance genes, or virulence genes that directly concern the course of action during hospital outbreaks. Although this instrument is a major asset to the scientific community, the basecalls have a rather high error rate compared to second and third generation sequencers. In 2018, the error rates ranged from 5% to 20%, dependent on what material is sequenced and library preparation methods (Rang *et al.*, 2018). Although the error rate has greatly improved since its introduction in 2013, researchers often opt to produce hybrid assemblies, which is the combination of the error-prone long reads of the MinION with highly accurate short read sequence data.

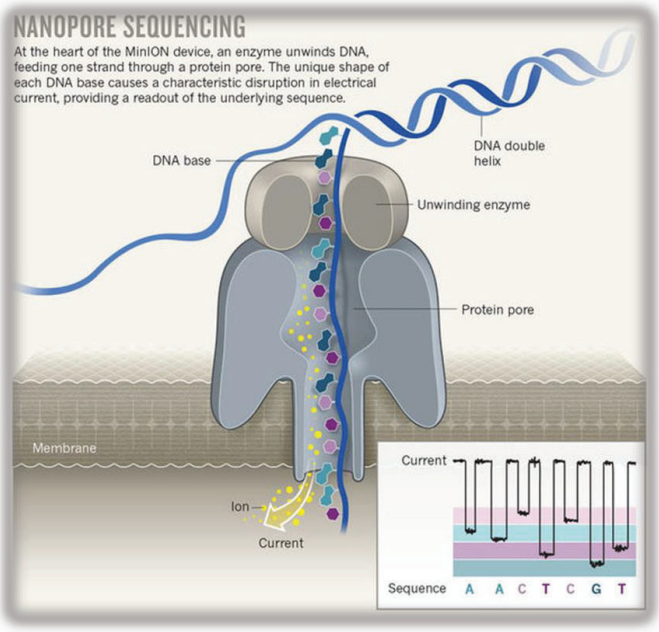

Figure 8. Theory behind nanopore sequencing. Obtained from Perkel, 2017.

1.14.2.3 Bioinformatics tools

Bioinformatics tools are used to analyze the sequenced data. The initial limiting factor in utilizing WGS data was knowledge about bioinformatics process and the manipulation of the data into a useable form. The first runs required the help of a skilled bioinformation, however, sequencing data from runs within the last 2 years were able to be processed solely through the Galaxy online platform.
All FastQ reads within each MinION barcode was concatenated using the Cygwin cat command. The online Galaxy platform was utilized to for all of the next steps, starting with the removal of adapters from the MiSeq and MinION data with Trimmomatic and Porechop, respectively. MinION data was then filtered by sequence length with a minimum threshold set at 800bp and above. If only working with MinION data, the resulting files can be placed into CANU for an assembly. If only processing MiSeq data, the resulting files can be placed into Shovill for an assembly. However, to process both MiSeq and MinION data togther, the resulting files can be placed into Unicycler to assemble the sequences into contigs.

Additionally, the sequence data was annotated using the NCBI Prokaryotic Genome Annotation Pipeline (Angiuoli *et al.*, 2008). To assess the genomes for acquired antibiotic resistance genes and VFs, ResFinder v 3.2, and VirulenceFinder v 2.0 servers were used (Center for Genomic Epidemiology, Technical University of Denmark, Lyngby, Denmark) with the following settings: selected ID threshold 95.0%, selected minimum length 60%. The ABRicate Mass screening of contigs for antimicrobial and virulence genes (Galaxy Version 0.9.8) tool combined with the NCBI National Database of Antibiotic Resistant Organisms were also used. Additionally, the Comprehensive Antibiotic Resistance Database (CARD) was used to search the genome for acquired antibiotic resistance genes. MLST v 2.0.4, PlasmidFinder v 2.1, and SerotypeFinder 1.1 (Center for Genomic Epidemiology) were used with default settings to determine MLST type, plasmid types, and serotypes of the isolates, with the exception of strain R2 in which the online database PubMLST (www.pubmlst.org) (Jolley *et al.*, 2018) was used for MLST results. Phylogenetic groups were determined using the publicly available ClermonTyper (http://clermontyping.iame-research.center/). In addition, a custom-made virulence database of over 760 genes was employed to assess for VFs with the CGE website. All data analysis tools used in our studies are found in Table 3.

Table 3. WGS data analysis tools used in our studies.

Software/database	Action	Reference or Link
NCBI Prokaryotic Genome Annotation Pipeline	Genome Annotation	Angiuoli *et al.*, 2008
BASys Bacterial Annotation System	Genome Annotation	Van Domselaar *et al.*, 2005
Rast	Genome Annotation	Aziz *et al.*, 2008
Prokka	Genome Annotation	Seemann, 2014
VirulenceFinder	Determination of virulence genes	https://cge.cbs.dtu.dk/services/VirulenceFinder/
SerotypeFinder	Determination of O and H type coding gene	https://cge.cbs.dtu.dk/services/SerotypeFinder/
MLST	Determination of multilocus sequence type	https://cge.cbs.dtu.dk/services/MLST/
ResFinder	Determination of antibiotic resistance genes	https://cge.cbs.dtu.dk/services/ResFinder/
SpeciesFinder	Determination of species	https://cge.cbs.dtu.dk/services/SpeciesFinder/
PHAge Search Tool Enhanced Release (PHASTER)	Bacteriophage prediction	https://phaster.ca/
PlasmidFinder	Plasmid prediction	https://cge.cbs.dtu.dk/services/PlasmidFinder/
Comprehensive Antibiotic Resistance Database (CARD)	Determination of resistance genes	https://card.mcmaster.ca/
MyDB Finder (input of our own database against bacterial isolate genomes)	Determination of virulence genes	https://cge.cbs.dtu.dk/services/MyDbFinder/
ABRicate	Mass screening of contigs for antimicrobial and virulence genes (Galaxy Version 0.9.8)	https://usegalaxy.org.au/
PubMLST	Determination of MLST	www.pubmlst.org
ClermonTyper	Determination of Phylogenetic grouping	http://clermontyping.iame-research.center/

2. Aims of the thesis

The objective of this thesis was to investigate the heteropathogenic nature of clinical isolates of *E. coli* from Norwegian patients that exhibited intestinal disease. In addition, we aimed to explore the existence of ESBL-producing bacteria in various soil and aquatic locations at the Norwegian University of Life Sciences in Ås, Norway.

Specific objectives:

I. To investigate the genomes of clinical heteropathogenic *E. coli* isolates from patients that showed signs of gastrointestinal disease in Norway.
   a. Screen 168 fecal isolates submitted to the Norwegian Institute of Public Health (NIPH) from clinical microbiological laboratories throughout Norway for the presence of 33 IPEC or ExPEC VAGs through multiplex PCR.
   b. Determine the frequency and combination of VAGs including those used for IPEC pathovar classification and a selection of VAGs related to ExPEC pathovars.
   c. WGS one isolate through a combination of Oxford Nanopore's MinION and Illumina's MiSeq to obtain a deeper assessment of its heteropathogenic nature.

II. To explore the presence of diverse ESBL-producing bacteria in samples obtained from the campus soil and pond ("Andedammen") at the Norwegian University of Life Sciences in Ås, Norway.
   a. Isolate potential ESBL-producing bacteria through growth on selective media
   b. Determine the presence of ESBLs or carbapenemases through multiplex PCR.
   c. Identify the susceptibility profiles of ESBL-positive isolates against a panel of the most commonly used antibiotics in clinical settings, including β-lactamases and carbapenemases.

d. WGS all ESBL-producing isolates through a combination of Oxford Nanopore's MinION and Illumina's MiSeq.

III. Due to the overwhelming amount of heteropathogenic *E. coli* in Paper I, we aimed to WGS more strains from our first experiment, as well as two additional strains from St. Olav's hospital in Trondheim, Norway.

a. WGS a total of 13 strains of *E. coli* isolates from patients that exhibited gastrointestinal disease.
b. Identify both IPEC and ExPEC VAGs present throughout the isolates as well as present antibiotic resistance genes.
c. Describe the extent of the heteropathogenic genetic profiles and determine which pathovars of IPEC and ExPEC may be represented in each strain.

3. Main results and discussion

The manuscripts presented in this thesis were initially conducted as two distinct studies: (i) the investigation of heteropathogenic *E. coli* from patients in clinical settings of Norway and (ii) the exploration and identification of diverse ESBL-producing bacteria from environmental sources. However, the results of WGS revealed that multiple ESBL-producing environmental isolates were also classified as heteropathogens, forming an unexpected commonality between the two originally separate studies.

In **Paper I**, we aimed to examine the presence of ExPEC VAGs in 168 fecal *E. coli* isolates from patients throughout Norway that exhibited signs of gastrointestinal infection. In circumstances of an intestinal infection, clinical diagnostic testing will typically only investigate the presence of IPEC genetic markers, and the results of these assays will allow for the classification of the *E. coli* isolate into one of the IPEC pathovars and determine further course of action. In the case of no target amplification, the tested strain will be classified as a non-enteropathogenic or commensal strain. When testing fecal isolates suspected of causing a gastrointestinal infection, ExPEC VAGs are not typically examined. Therefore, we examined both IPEC and ExPEC associated VAGs through multiplex PCR. Our results showed a surprisingly high frequency (93.9%) of strains that microbiological laboratories had previously classified as IPEC also harbored ExPEC VFs, and a very interesting observation was that 29.7% of the isolates only carried ExPEC VAGs with no IPEC VAGs. The *ehaG* gene was detected in 64.3% of the isolates and was the most common VAG in our collection, and it was determined that the *ehaG* gene was present in 48% of the strains with one or more ExPEC VAGs and no IPEC VAGs. Since *ehaG* mediates specific adhesion to colorectal epithelial cells, the presence of this gene may have been a contributing factor in the ability of these strains to adhere to colorectal epithelial cells in the absence of typical IPEC genetic markers. We also selected one heteropathogenic strain of *E. coli* for WGS through a combination of Oxford Nanopore's MinION and Illumina's MiSeq, and assembled a complete closed circular genome and a complete circular virulence plasmid from the combined runs. The isolate was identified as aEPEC that also harbored VAGs related to ETEC. Additionally, the genome showed hallmarks of the ExPEC pathovars UPEC, APEC, and NMEC, thus constituting a truly heteropathogenic

strain that was heavily influenced by horizontal gene transfer. It was also determined that closely related strains may have been causing disease for some time in Norway, as 56 isolates classified as aEPEC of the same sequence type and phylogenetic grouping were previously detected in fecal specimens from children < 5 years old that exhibited signs of intestinal infection. Heteropathogenic *E. coli* have serious clinical implications that may result in a more severe disease that is capable of increased spreading potential, therefore it is suggested that a wider range of genetic targets be considered for clinical diagnostic testing.

The motivation behind **Paper II** was to obtain more information on the presence of ESBL-producing bacteria in soil and water environments in Ås, Norway. As β-lactam antibiotics are the most widely prescribed antibiotics worldwide, the presence of ESBLs that render these critical drugs ineffective are considered a serious public health problem. Although commonly considered a clinical issue, their presence in the environment is gaining attention as an exposure route to humans through food items, drinking water, and direct contact with water bodies. The environment, especially water bodies, is a rich reservoir for microorganisms from a variety of sources. Therefore, it is a prime location for the acquisition, evolution, and dissemination of antibiotic resistant bacteria. Furthermore, the added selective pressures of antimicrobials, biocides, heavy metals, and disinfectants promote the selection of bacterial defense mechanisms that uphold their survival and spread. In our study, we found that ESBL-producing Enterobacteriaceae was predominant, which is not surprising as Enterobacteriaceae are the main producers of ESBLs. In addition, we isolated an ESBL-producing *Paraburkholderia* spp. strain and a carbapenemase *Stenotrophomonas maltophilia* strain, however not as much information was obtained from these two isolates, likely due to low level DNA. Amongst the Enterobacteriaceae, CTX-M-15, TEM-1B, and CMY-42 enzymes were the most commonly detected ESBLs, with the majority of these enzymes located on plasmids. In addition, no identical proteins were found on BLAST searches of one isolate's CMY-variant and another isolate's KPC and CTX-M variants; therefore, this may constitute a novel plasmid-borne variants distinctly different from the other known β -lactamases, which also may be indicative of a high diversity of β-lactamase genes in the "Andedammen" pond. WGS revealed the presence of a myriad of resistance genes throughout the isolates,

including determinants against aminoglycosides, bacitracin, bicyclomycin, chloramphenicol, fluoroquinolones, fosfomycin, fosmidomycin, macrolides, sulfonamides, tetracyclines and trimethoprims. Efflux pumps that confer resistance to multiple classes of antibiotics were also observed and several *E. coli* isolates contained chromosomal mutations in *gyrA* (S83L or D87N), *parC* (S80I, E84K), and *parE* (S458A) that confer fluoroquinolone resistance. Resistance genes to metals including arsenic, cadmium/zinc/lead, copper, mercury, and tellurite was observed. In addition, resistance genes to quaternary ammonium were detected amongst the isolates, which is used as a disinfectant in medical and food environments. Silver resistance determinants was detected in a *K. pneumoniae* isolate, which is a worrisome finding due to silver being a common component in a variety of healthcare and consumer products, including silver-treated catheters and wound dressings. An alarming observation was that all of the isolates exhibited a multidrug resistance profile, including to some of the most commonly prescribed antibiotics in clinical settings in Norway (and worldwide). For example, all seven strains were determined to be phenotypically resistant against the β-lactams penicillins (ampicillin) and third and fourth generation cephalosporins (cefepime and cefotaxime). Cross resistance to other classes of antibiotics is quite common in ESBL-producers, and the isolates from our study are no exception; 86.0% were resistant to macrolides (erythromycin), 43.0% were resistant to trimethoprim, and 43.0% were resistant to fluoroquinolones (ciprofloxacin). Fortunately, no acquired resistance towards the last resort carbapenems or colistin were observed in our study. The majority of our Enterobacteriaceae isolates possessed a variety of virulence traits central to pathogenicity, namely those necessary to cause UTIs. This is significant, as swimming in freshwater has been labeled as a risk factor for community acquired UTIs by Enterboacteriaceae (Søraas *et al.*, 2013). Although the locations in this study are not regularly used for recreational purposes, swimming does occur on occasion. In addition, the spread of the ESBL-producing bacteria is as a valid concern, as water uptake by wild or companion animals can disseminate these pathogens to other locations that have a greater extent of human contact.

A very interesting finding was that several of the *E. coli* isolates were heteropathogens characterized as DAEC/UPEC or aEAEC/UPEC, and therefore have the potential to

cause both diarrheal disease and UTIs. Three of the *E. coli* sequence types (STs) discovered amongst the environmental water isolates (ST38, ST69 and ST405), belong to well-known clones that seem to be spreading throughout the world in a way reminiscent of the spread of the ExPEC *E. coli* ST131 clone, which is now a major clinical problem. ST69 is a well-known cause of bloodstream infections (BSI) and UTIs, and in the UK, ST69 is one of the most frequent recovered ST from BSIs and UTIs (Ciesielczuk *et al.*, 2016). ST69 was also found among 104 ExPEC blood-culture isolates from Turkey (Bozcal *et al.*, 2018) and among patients with clinically confirmed UTI in Saudi Arabia (Alghoribi *et al.,* 2015). The link with possible severe systemic infections seems clear regarding *E. coli* ST69. However, ST69 has also been recovered from other sources. Characterization of a multi-resistant ST69 isolated in 2016 from a diseased broiler chicken in Germany was recently published (Hornsey M *et al.,* 2019). A ST69 strain was also recently detected in raw meat-based diets for companion animals in Switzerland, which may indicate a zoonotic potential (Nüesch-Inderbinen *et al.*, 2019). *E. coli* ST405 is also associated with systemic disease, and was recovered from a study of 2,427 *E. coli* bloodstream isolates sampled from a Spanish hospital over a 12-Year Period (2000 to 2011) (Mamani *et al.*, 2019), as well as among clinical isolates collected from tertiary hospitals in Thailand (Bubpamala *et al.*, 2018), and ST405 strains are among the leading uropathogenic *E. coli* clones in Riyadh, Saudi Arabia (Alghoribi *et al.,* 2015). Four ST405 isolates were detected among 27 carbapenem resistant *E. coli* strains isolated from 27 hospitalized patients at the American University of Beirut Medical Center (AUBMC) in Lebanon between 2012 and 2016 (Dagher *et al.*, 2018). ST405 was additionally the ST of the reported first two clinical cases of NDM-5 producing *E. coli* in Northern Italy (Bitar *et al.*, 2017), showing its ability to pick up an extended repertoire of resistance. *E. coli* ST405 was recently reported isolated from diseased dogs in Beijing, China, in a period from 2012 to 2017 (Chen *et al.*, 2019), which also points to a zoonotic potential. *E. coli* ST38 was the ST most frequently found in the study of clinical isolates collected from tertiary hospitals in Thailand (Bubpamala *et al.,* 2018), and was frequently encountered in the Saudi Arabian study (Alghoribi *et al.,* 2015). In a German study of a possible transmission of $bla_{\text{CMY-2}}$ along the food production chain, ST38 was one of only two STs to be isolated from all tested reservoirs (human, animal, and food) (Pietsch *et al.*, 2018). E. coli ST1286 is a single locus MLST variant of the more common ST10 clone and have been

isolated from chicken and swine in China (Wu et al., 2018), chicken and egg surface in Finland (Oikarainen et al., 2019), chicken in Argentina (Dominguez et al., 2019), as well as from urban rats, chicken and hen in Germany (Guenther et al., 2012, Pietsch et al., 2018, https://publikationsserver.tu-braunschweig.de/servlets/MCRFileNodeServlet/dbbs_derivate_00045243/Diss_Pietsch_Michael.pdf), and broiler and broiler meat in Denmark (https://www.danmap.org/-/media/arkiv/projekt-sites/danmap/danmap-reports/danmap-2018/web-annex-danmap-2018-250919.pdf?la=en).

These results clearly show that our environmental *E. coli* isolates of STs 38, 69 and 405 are linked to several cases of severe human disease and may have many reservoirs including surface water, as in our case. All of our surface water isolates were multidrug resistant, indicating a treacherous combination of pathogenic potential and possible treatment failure if infection should occur. Furthermore, ST38, ST69, and ST405 types detected in this study were also observed in human clinical samples from Norway (Lin T. Brandal *pers. comm.* and Enterobase; https://enterobase.warwick.ac.uk/), further correlating the clinical relevance of some of our environmental isolates.

The purpose of **Paper III** was similar to paper I. Ten additional isolates (besides one isolate that was already subjected to WGS and included in the first publication) were selected from the strains included in paper I, and two isolates from St. Olav's Hospital (not included in paper I), were recovered from patients that exhibited signs of gastrointestinal infection. WGS through Oxford Nanopore's MinION and the Illumina MiSeq was conducted to uncover the heteropathogenicity profiles of the strains. The genetic targets of STEC, aEPEC, tEPEC, ETEC, and EIEC were identified in 11 of the isolates, with two isolates containing the molecular markers of both STEC and EIEC, creating STEC/EIEC heteropathogens. UPEC predictors *chuA, yfcV*, and *fyuA* were observed in 53.8% of the remaining isolates which classified them as STEC/UPEC, aEPEC/UPEC, and tEPEC/UPEC heteropathogens. All isolates harbored a wide range of ExPEC-associated virulence genes, including iron acquisition systems, protectins, invasins, adhesins, and toxins. And several strains contained genes associated with ETEC virulence (K88 fimbriae, CFA/I fimbriae). Additionally, two isolates did not contain any of the genetic targets used in the identification of enteropathogens, and solely harbored VAGs that classified the strain as UPEC. However, WGS data showed

the presence of several IPEC-associated VAGs that may have promoted intestinal infection, including the heat-stable enterotoxin 1 (EAST1), *eptA* which mediates attachment to mamamalian host cells, long polar fimrbia (*lfpA*), and a putative fimbrial-like adhesin protein gene (*stcD*). Still, these strains could present a clinical problem as diagnostic testing typically only targets the well-defined molecular markers of IPEC. Therefore, these strains could create difficulties in the diagnosis and determining the proper course of treatment for the patient. In addition, all strains exhibited a variety of antibiotic resistance genes, and one isolate harbored the β-lactamase gene TEM-1B. Unlike paper II, phenotypic susceptibility tests were not conducted in this study.

The high frequency of strain with combined IPEC/ExPEC VAGs found in both the clinical and environmental isolates in this thesis are displayed in a comparison chart (Table 4). Additionally, a phylogenetic tree showing the relatedness of our isolates is found in Figure 9.

Table 4. Various IPEC and ExPEC VFs harbored by the heteropathogenic *E. coli* isolates from all 3 studies. This is not an exhaustive list, but rather a synopsis of the more interesting VFs. The category 'other VFs' contain those that are difficult to classify as solely IPEC or ExPEC.

STRAIN	SERO.	ST	PHYLO.	fimH type	SOURCE	IPEC VFs	ExPEC VFs	Other VFs
NMBU_W10C18	O102:H6	405	D	fimH27	Environmental water	air, eilA, ehaB, ehaC, espL1, espY1, espY3, espY4, ShET2/EspL2	aec35-37, chuA, fyuA, irp1/irp2, kpsS, sitABCD, traT, yfcV	Ag-43, CRISPR, fecA, ibrA, pgaA, rafA, fmlA (F9-fimbria gene), hutX, sinH, sipB, tssH
NMBU_W05E18	O86:H18	38	D	fimH5	Environmental water	air, eilA, draP, ehaB, ehaC, espL1, espY1, espY3, espY4, iha, nfaA, ShET2/EspL2	chuA, eitA, fyuA, irp1/irp2, iss, iuc-Aerobactin, iutA, kpsS, sat, sitABCD, traT, yfcV	cjrC, CRISPR, ibrA, BREX1, fmlA (F9-fimbria gene), hutX, sipB, tssH
NMBU_W06E18 _Strain1	O102:H6	405	D	fimH27	Environmental water	air, eilA, ehaB, ehaC, espL1, espY1, espY3, espY4, ShET2/EspL2	aec35-37, chuA, fyuA, irp1/irp2, kpsS, sitABCD, yfcV	Ag-43, CRISPR, fecA, ibrA, fmlA (F9-fimbria gene), hutX, sinH, sipB, tssH
NMBU_W12E19	O10:H32	1286	A	fimH30	Environmental water	bfp, aap, aatA, aatP (dispersin), EAST1-toxin, ehaC, ehaG, espL1, espY1, iha, matB	fyuA, irp1/irp2, iuc-Aerobactin, iutA, traT	deoK, Ag-43, eaeH, ibrA, mchB, sigA, terA, terB, terY1, terY2, terY3, yeeJ
NMBU_W13E19	O25:H18	69	D	fimH27	Environmental water	air, eilA, espL1, espR2, espY3, espY4, lpfA, ShET2/EspL2	aufA, chuA, fyuA, irp1/irp2, iss, kpsS, papC, evfC, sitABCD, traT, yfcV	Ag-43, CRISPR, fecA, ibrA, fmlA (F9-fimbria gene), hutX, sinH, tagH, tssA
FHI_NMBU_04	O128ab:H2	4748	B1	fimH32	Human feces	Stx1c, Stx2b, subAB, cfaB, ehaA, ehaB, iha, lpfA	hlyF, hmuR, ireA, iro-Salmochelin, iss, iuc-Aerobactin, iutA, kpsS, malX, sitABCD, ssbL, traJ, traT	deoK, Ag-43, cba, celb, cma, CRISPR, Mig-14, herA, cvaC, cvi, mchB, mkaC, pic, sepA, fmlA

								(F9-fimbria gene), pduC, vpeC, yadN, yeeJ
FHI_NMBU_05	O161:H4	117	F	fimH97	Human feces	cfaB, ehaB, iha, lpfA	hlyF, hmuR, ireA, iro-Salmochelin, iss, iuc-Aerobactin, iutA, kpsS, malX, sitABCD, ssbL, traJ, traT	deoK, Ag-43, cba, celb, cma, CRISPR, Mig-14, herA, cvaC, cvi, mchB, mkaC, pic, sepA, fmlA (F9-fimbria gene), pduC, vpeC, yeeJ
FHI_NMBU_03	O81:H6	28	B2	fimH90	Human feces	eae, faeC	aatB, upaB, chuA, fyuA, hbp, iro-Salmochelin, irp1/irp2, iss, malX, papC, traT, usp, yfcV	herA, cyclic-1,2-glucan synthase, elfC, fmlA (F9-fimbria gene), hutX, sinH, xhlA, yfaL, yjaA
Trh52	O84:H7	28	B2	fimH90	Human feces	eae, cdt, faeC	aatB, upaB, chuA, fyuA, ibeA, iro-Salmochelin, irp1/irp2, iss, malX, papC, traT, usp, vat, yfcV	herA, cyclic-1,2-glucan synthase, elfC, ferritin-B, ibrA, fmlA (F9-fimbria gene), sinH, yjaA
Trh50	O81:H6	28	B2	fimH90	Human feces	eae, cdt, cif, ehaB, faeC	aatB, upaB, chuA, fyuA, ibeA, iro-Salmochelin, irp1/irp2, iss, malX, papC, traT, usp, vat , yfcV	Ag-43, herA, cyclic-1,2-glucan synthase, elfC, fmlA (F9-fimbria gene), hutX, sinH, xhlA, yfaL, yjaA
FHI_NMBU_08	O137:H6	2678	B2	fimH138	Human feces	eae, cdt, fedC (F18 fimbrial gene)	aatB, upaB, chuA, fyuA, ibeA, irp1/irp2, iss, malX, usp, vat, yfcV	herA, elfC, mkaC, fmlA (F9-fimbria gene), hutX, pduC, sinH, yfaL, yjaA
FHI_NMBU_06	O156:H1	941	B2	fimH450	Human feces	eae, bfp, perA, EAST1-toxin	chuA, fyuA, ibeA, irp1/irp2, usp, yfcV	herA, elfC, mkaC, fmlA (F9-fimbria gene), hutX, sinH, yfaL, yjaA

FHI_NMBU_07	O145:H28	32	D	fimH331	Human feces	eae, EAST1-toxin, ehaA, ehaB, espR2, espX6, espY1, espY3, iha, ipaH, paa, Z2098	chuA, iss, iuc-Aerobactin, iutA, malX	fbpB, fecA, ibrA, fmlA (F9-fimbria gene), shuX, terB
FHI_NMBU_09	O145:H28	137	D	fimH331	Human feces	eae, EAST1-toxin, ecf1, efa1/lifA, ehaA, ehaC, ehaG, espL1, espR2, espX6, espY1, espY3, iha, ipaH, paa, pagC, Z2098	chuA, hlyA, hlyB, iuc-Aerobactin, iutA, traT	Ag-43, celb, fbpB, fecA, shuX, terA, terB
FHI_NMBU_11	O128ab:H2	25	B1	fimH32	Human feces	Stx1a, subAB, cfaB, EAST1-toxin, ehaA, ehaC, espL1, iha, lpfA	fyuA, hlyF, hmuR, ireA, iro-Salmochelin, irp1/irp2, iss, iutA, ssbL, traJ, traT	deoK, Ag-43, cma, CRISPR, Mig-14, herA, cvaC, cvi, mchB, mkaC, pic, fmlA (F9-fimbria gene), vpeC, yadN
FHI_NMBU_12	O78:H10	173	B1	fimH32	Human feces	capU, eatA, ehaB, ehaC, ehaG, eltAB, espL1, lpfA, matB, tibA	fyuA, irp1/irp2, iss	deoK, Ag-43, CRISPR, eafC, fecA, fmlA (F9-fimbria gene), vpeC, yadN
FHI_NMBU_10	O146:H21	829	B1	fimH32	Human feces	Stx2d, subAB, cfaB, ehaA, ehaB, ehaC, ehaG, eibG, espL1, iha, lpfA, matB, hesH, pagC	iss, hmuR, ireA, traT	deoK, Ag-43, CRISPR, cka, celb, eafC, fbpB, pic, sepA, vpeC, yadN, yeeJ

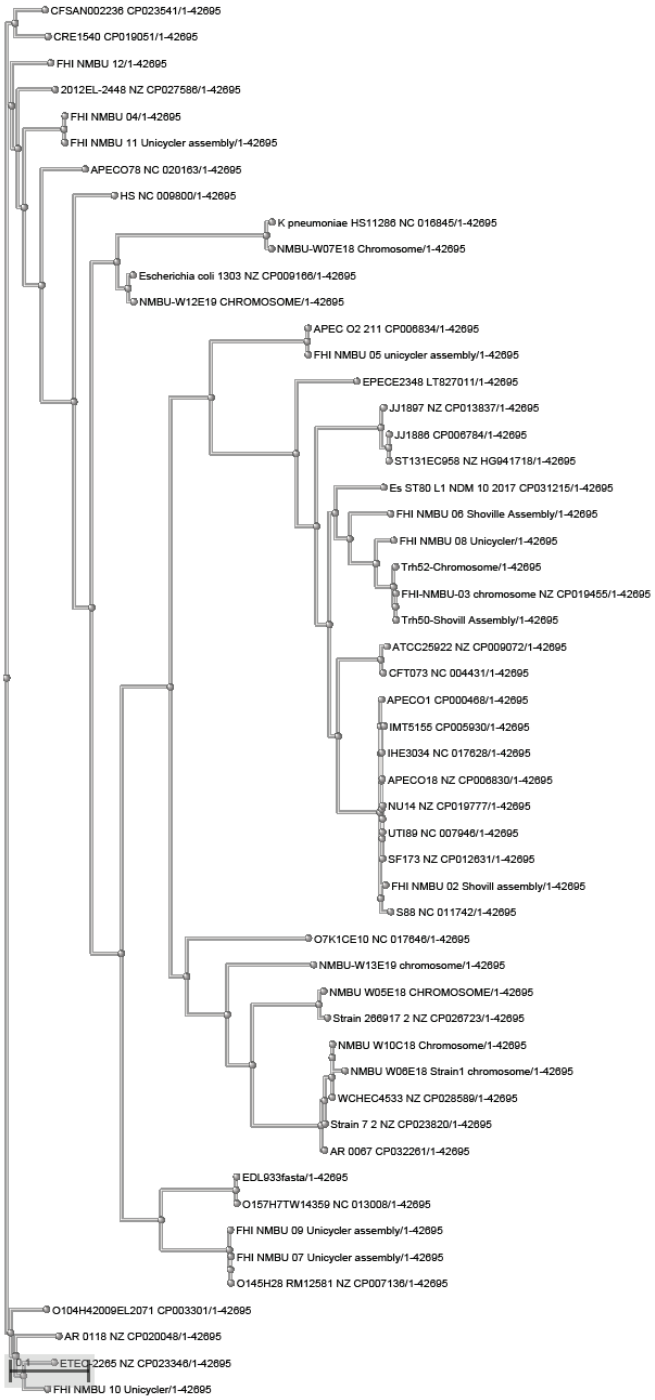

Figure 9. SNP based phylogenetic tree of our isolates compared to other strains.

4. Concluding remarks

The heteropathogenic strains may not only be capable of causing both intestinal- and extraintestinal disease, a more serious infection, and have increased spreading potential, but combined with the presence of β-lactamase genes may become very difficult to treat with currently available antibiotics.

5. Future perspectives

The clinical isolates were not assessed with phenotypic susceptibility testing, as was done in Paper II. Although we did determine the present resistance genes, assessing phenotypic susceptibility would be an interesting avenue to take further. Another future perspective could be a conjugation assay to assess the ability of the plasmids that harbor the β-lactamase genes to transfer into a non-plasmid bearing recipient strain. Additionally, we did not assess biofilm formation of our environmental isolates, although we did detect VAGs for biofilm formation. As biofilm formation is important when establishing colonization and infection, this would be an interesting future experiment. In addition, the results of WGS allowed us to determine that NMBU_W06E18 was a mixture of two *E. coli* isolates. We did not obtain much data on the contaminant strain, and there were ultimately too few sequences for a Genbank entry. However, we were able to locate TEM-1B in the sequences as well as silver resistance genes (*silES*). A future study could use silver-containing media to isolate the contaminant *E. coli*, as the first strain was not silver resistant. Although this would take considerably more time than the other perspectives, the screening of the healthy human population for the presence of heteropathogens and/or ESBL carriage would be interesting, as it may give insight into possible transmission routes via asymptomatic carriers.

6. References

Farmer JJ, Boatwright KD, Janda M. *Enterobacteriaceae*: introduction and identification. In: Murray PR, Baron EJ, Jorgensen JK, Landry ML, Pfaller MA., editors. Manual of clinical microbiology. 9th ed. Washington DC: American Society for Microbiology. 2007; 49-669.

Davis GS, Waits K, Nordstrom L, Weaver B, Aziz M, Price LB. Intermingled *Klebsiella pneumoniae* Populations Between Retail Meats and Human Urinary Tract Infections. Clin Infect Dis. 2015; **61**(6): 892-899.

Friedländer C. Ueber die Schizomyceten bei der acuten fibrösen Pneumonie. Archiv Patholog. Anat. Physiol. Klinische Med. 1882. **87**: 319–324.

Paczosa MK and Mecsas J. *Klebsiella pneumoniae*: Going on the Offense with a Strong Defense. Microbiol Mol Biol Rev. 2016; **80**(3): 629-661.

Puspanadan S, Afsahhejri L, Loo YY, Nillian E, Kuan CH, Goh SG, *et al.* Detection of *Klebsiella pneumoniae* in raw vegetables using most probable number-polymerase chain reaction (MPN-PCR). Int. Food Res. J. 2012; **19**: 1757–1762.

Overdevest IT, Heck M, van der Zwaluw K, Huijsdens X, van Santen M, Rijnsburger M, *et al.* Extended-spectrum β-lactamase producing *Klebsiella* spp. in chicken meat and humans: a comparison of typing methods. Clin. Microbiol. Infect. 2014; **20**: 251–255.

Sun F, Wu D, Qiu Z, Jin M, Wang X, Li J. Development of real time PCR systems based on SYBR Green fro specific detection and quantification of *Klebsiella pneumoniae* in infant formula. Food Control. 2010; **21**: 487–491.

Davis GS, Price LB. Recent research examining links among *Klebsiella pneumoniae* from food, food animals, and human extraintestinal infections. Curr. Environ. Health Rep. 2016; **3**: 1–8.

Haryani Y, Noorzaleha AS, Fatimah AB, Noorjahan BA, Patrick GB, Shamsinar AT, *et al.* Incidence of *Klebsiella pneumoniae* in street foods sold in Malaysia and their characterization by antibiotic resistance, plasmid profiling, and RAPD–PCR analysis. Food Control. 2007; **18**: 847–853.

Broberg CA, Palacios M, Miller VL. *Klebsiella*: a long way to go towards understanding this enigmatic jet-setter. F1000Prime Rep. 2014; **6**:64:

Li B, Zhao Y, Liu C, Zhou D. Molecular pathogenesis of *Klebsiella pneumonia*. Future Microbiol. 2014; **9**:1071–1081.

Adlerberth I and Wold AE. Establishment of the gut microbiota in Western infants. Acta Paediatr. 2009; **98**: 229–238.

Biswas S, Parves MAK, Shafiquzzaman M, Nahar S, Rahman MN. Isolation and Characterization of *Escherichia coli* in Ready-to-eat Foods Vended in Islamic University. J. Bio-sci. 2010; **18**: 99-103.

Campos AN, Andrade NL, Ferdous M, Chlebowicz MA, Santos CC, Rossen WA. Comprehensive Molecular Characterization of *Escherichia coli* Isolates from Urine Samples of Hospitalized Patients in Rio de Janeiro, Brazil. Front. Micriol. 2018; **9**(243).

Laupland KB, Church DL. Population-based epidemiology and microbiology of community-onset bloodstream infections. Clin Microbiol Rev. 2014; **27**(4):647-64.

Poolman JT and Wacker M. Extraintestinal Pathogenic *Escherichia coli*, a Common Human Pathogen: Challenges for Vaccine Development and Progress in the Field. J Infect Dis. 2016; **213**(1): 6-13.

WHO World Health Organization. Antimicrobial resistance: global report on surveillance (WHO, 2014).

Kaper JB, Nataro JP, Mobley HL. Pathogenic *Escherichia coli*. Nature Reviews Microbiology. 2004; **2**: 123–140.

Croxen MA, Law RJ, Scholz R, Keeney KM, Wlodarska M, Finlay BB. Recent advances in understanding enteric pathogenic *Escherichia coli*. Clin Microbiol Rev. 2013; **26**(4):822-80.

Robins-Browne RM, *et al.* Are *Escherichia coli* Pathotypes Still Relevant in the Era of Whole-Genome Sequencing? Front. Cell. Infect. Microbiol. 2016; **(6)**:141.

Sarowska J, Futoma-Koloch B, Jama-Kmiecik A, Frej-Madrzak M, Ksiazczyk M, Bugla-Ploskonska G, Choroszy-Krol I. Virulence factors, prevalence and potential transmission of extraintestinal pathogenic *Escherichia coli* isolated from different sources: recent reports. Gut Pathog. 2019; **11**(10).

Rojas-Lopez M, Monterio R, Pizza M, Desvaux M, Rosini R. Intestinal Pathogenic *Escherichia coli*: Insights for Vaccine Development. Front Microbiol. 2018; 9: 440.

Scaletsky IC, Fabbricotti SH, Carvalho RL, Nunes CR, Maranhao HS, Morais MB, *et al.* Diffusely adherent *Escherichia coli* as a cause of acute diarrhea in young children in Northeast Brazil: a case-control study. J Clin Microbiol. 2002; **40**(2):645–8.

Servin AL. Pathogenesis of Afa/Dr diffusely adhering *Escherichia coli*. Clin Microbiol Rev. 2005; **18**(2):264–92.

Servin AL. Pathogenesis of Human Diffusely Adhering *Escherichia coli* Expressing Afa/Dr Adhesins (Afa/Dr DAEC): Current Insights and Future Challenges. Clinical Microbiology Reviews. 2014; **27**(4): 823-869.

Aslani MM, Alikhani MY, Zavari A, Yousefi R, Zamani AR. Characterization of enteroaggregative *Escherichia coli* (EAEC) clinical isolates and their antibiotic resistance pattern. International Journal of Infectious Diseases. 2011; **15**: e136-e139.

Wanke CA, Schorling JB, Barrett LJ, Desouza MA, Guerrant RL. Potential role of adherence traits of *Escherichia coli* in persistent diarrhea in an urban Brazilian slum. Pediatr Infect Dis J. 1991; **10**: 746-751.

Cabada Bauche J, Herbert MD, DuPond HL. New Developments in Traveler's Diarrhea. Gastroenterol Hepatol. 2011. **7**(2): 88-95.

Okhuysen PC, DuPont HL. Enteroaggregative *Escherichia coli* (EAEC): A Cause of Acute and Persistent Diarrhea of Worldwide Importance. The Journal of Infectious Diseases. 2010; **202**(4):503–505.

Tokuda K, Nishi J, Imuta N, Fujiyama R, Kamenosono A, Kawano Y. EAEC japanese children school lunch. 2010; **54**(6): 320-329.

Schaechter M, Lederberg J. The Desk Encyclopedia of Microbiology. Elsevier Academic Press. London, UK. 2004.

Schaechter M. Encyclopedia of Microbiology. 3rd ed. Elsever Academic Press. San Diego, CA. 2009.

Lan R, Alles MC, Donohoe K, Martinez MB, Reeves PR. Molecular evolutionary relationships of enteroinvasive *Escherichia coli* and Shigella spp. Infect Immun. 2004; **72**(9):5080-8.

Pasqua M, *et al.* The Intriguing Evolutionary Journey of Enteroinvasive *E. coli* (EIEC) toward Pathogenicity. Front Microbiol. 2017; **8**: 2390.

Bhavnani D *et al.* Distribution of Enteroinvasive and Enterotoxigenic *Escherichia coli* across Space and Time in Northwestern Ecuador. Am J Trop Med Hyg. 2016; **94**(2): 276-284.

Karmali MA, Mascarenhas M, Shen S, Ziebell K, Johnson S, Kaper JB. Association of genomic O island 122 of *Escherichia coli* EDL 933 with verocytotoxin-producing *Escherichia coli* seropathotypes that are linked to epidemic and/or serious disease. J Clin Microbiol. 2003; **41**: 4930–4940 10.1128/JCM.41.11.4930-4940.2003.

Griffin PM and Karmali MA. Emerging Public Health Challenges of Shiga Toxin–Producing *Escherichia coli* Related to Changes in the Pathogen, the Population, and the Environment. Clinical Infectious Diseases. 2017; **64**: 371-376.

Kobayashi N, Lee K, Yamazaki A, Saito S, Furukawa I, Hara-Kudo Y. Virulence Gene Profiles and Population Genetic Analysis for Exploration of Pathogenic Serogroups of Shiga Toxin-Producing *Escherichia coli*. J Clin Microbiol. 2013; **51**(12): 4022-4028.

Noris M, Remuzzi G. Hemolytic uremic syndrome. J Am Soc Nephrol. 2005; **16**: 1035-50.

Pommerville JC. Alcamo's Fundamentals of Microbiology. 9th ed. Jones and Bartlett Publishers. Sudbury, MA. 2010.

Ochoa TJ, Barletta F, Contreras C, Mercado E. New insights into the epidemiology of enteropathogenic *Escherichia coli* infection. Transactions of the Royal Society of Tropical Medicine and Hygiene. 2008; **102**(9): 852-856.

Mansan-Almeida R, Pereria AL, Gimenes Giugliano L. Diffusely adherent *Escherichia coli* strains isolated from children and adults constitute two different populations. BMC Microbiol. 2013; **13**(22).

Afset JE, Bruant G, Brousseau R, Harel J, Anderssen E, Bevanger L, Bergh K. Identification of virulence genes linked with diarrhea due to atypical enteropathogenic *Escherichia coli* by DNA microarray analysis and PCR. J Clin Microbiol. 2006; **44**: 3703––3711

Cortes C, De la Fuente R, Blanco J, Blanco M, Blanco JE, Orden JA. Serotypes, virulence genes and intimin types of verotoxin-producing *Escherichia coli* and enteropathogenic *E. coli* isolated from healthy dairy goats in Spain. Vet. Microbiol. 2005; **110**:67-76.

Quadri F, Svennerholm AM, Sack B. Enterotoxigenic *Escherichia coli* in Developing Countries: Epidemiology, Microbiology, Clinical Features, Treatment, and Prevention. Clin Microbiol Rev. 2005; **18**(3): 465-483.

Fleckenstein JM, Hardwidge PR, Munson GP, Rasko DA, Sommerfelt, H, Steinsland H. Molecular mechanisms of enterotoxigenic *Escherichia coli* infection. Microbes Infect. 2010; **12**:89 –98.

Gonzales-Siles L and Sjöling Å. The different ecological niches of enterotoxigenic *Escherichia coli*. Environmental Microbiology. 2016; **18**(3): 741-751.

Vieira M, Gomes T, Ferreira A, Knöbl T, Servin A, Moal V. Two atypical enteropathogenic *Escherichia coli* strains induce the production of secreted and membrane-bound mucins to benefit their own growth at the apical surface of human. Mucin- Secreting Intestinal HT29-MTX cells. *Infect. Immun.* 2010; **78**: 927–938.

Nataro JP, Kaper JB. Diarrheagenic *Escherichia coli*. Clin Microbiol Rev. 1998; **11**(1): 143-201.

Meraz IM, Arikawa K, Nakamura H, Ogasawara J, Hase A, Nishikawa Y. Association of IL-8-inducing strains of diffusely adherent *Escherichia coli* with sporadic diarrheal patients with less than 5 years of age. Braz J Infect Dis. 2007; **11**(1).

Riley LW. Pandemic lineages of extraintestinal pathogenic *Escherichia coli*. Clin Microbiol Infect. 2014; **20**(5):380–90.

Terlizzi ME, Gribaudo G, Maffei ME. UroPathogenic *Escherichia coli* (UPEC) Infections: Virulence Factors, Bladder Responses, Antibiotic, and Non-antibiotic Antimicrobial Strategies. Front Microbiol. 2017; **8**: 1566.

Lewis AJ, Richards AC, Mulvey MA. Invasion of Host Cells and Tissues by Uropathogenic Bacteria. Microbiol Spectr. 2016; **4**(6).

Blango MG and Mulvey MA. Persistence of Uropathogenic *Escherichia coli* in the Face of Multiple Antibiotics. Antimicrob Agents Chemother. 2010; **54**(5): 1855-1863.

Spurbeck RR, Dinh PC, Walk ST, Stapleon AE, Hooton TM, *et al. Escherichia coli* Isolates That Carry vat, fyuA, chuA, and yfcV Efficiently Colonize the Urinary Tract. Infect Immun. 2012; **80**(12): 4115-4122.

Kemmett K, Williams NJ, Chaloner G, Humphrey S, Wigley P, Humphrey T. The contribution of systemic *Escherichia coli* infection to the early mortalities of commercial broiler chickens. Avian Pathol. 2014; **43**(1): 37-42.

Markland SM, LeStrange KJ, Sharma M, Kniel KE. Old Friends in New Places: Exploring the Role of Extraintestinal *E. coli* in Intestinal Disease and Foodborne Illness. Zoonoses Public Health. 2015; **62**(7):491-6.

Vincent JL, Abraham E, Moore FA, Kochanek PM, Fink MP. Textbook of Critical Care. Elsevier Saunders. Philadelphia, PA. 2011.

Nordstrom L, Liu CM, Price LB. Foodborne urinary tract infections: a new paradigm for antimicrobial-resistant foodborne illness. Front. Microbiol. 2013; **4**:29.

Mitchell NM, Johnson JR, Johnston B, Curtiss R, Mellata M. Zoonotic Potential of *Escherichia coli* Isolates from Retail Chicken Meat Products and Eggs. Applied and Environmental Microbiology. 2015; **81**(3): 1177-1187.

Ewers C, Li G, Wilking H, Kiessling S, Alt K, Wieler LH. Avian pathogenic, uropathogenic, and newborn meningitis-causing *Escherichia coli*: how closely related are they? Int J Med Microbiol. 2007; **297**(3): 163-176.

Russo TA, Johnson JR. Medical and economic impact of extraintestinal infections due to *Escherichia coli*: focus on an increasingly important endemic problem. Microbes Infect. 2003; **5**(5): 449-56.

Owrangi B, Masters N, Kuballa A, O'Dea C, Vollmerhausen TL, Katouli M. Invasion and translocation of uropathogenic *Escherichia coli* isolated from urosepsis and patients with community-acquired urinary tract infection. Eur. J. Clin. Microbiol. Infect. Dis. 2018; **37**: 833–839.

Micenková L, Benováa A, Frankovicováb L, Bosáka J, Vrbac M, Sevcíkovác A. Human *Escherichia coli* isolates from hemocultures: septicemia linked to urogenital tract infections is caused by isolates harboring more virulence genes than bacteraemia linked to other conditions. Int. J. Med. Microbiol. 2017; **307**: 182–189.

Russo TA, Wenderoth S, Carlino UB, Merrick JM, Lesse AJ. Identification, genomic organization, and analysis of the group III capsular polysaccharide genes kpsD, kpsM,

kpsT, and kpsE from an extraintestinal isolate of *Escherichia coli* (CP9, O4/K54/H5). J. Bacteriol. 1998; **180**, 338–49.

Sonnen,A, Henneke P. Role of pore-forming toxins in neonatal sepsis. Clin. Dev. Immunol. 2013; 608456.

Laupland KB. Incidence of bloodstream infection: a review of population-based studies. Clin. Microbiol. Infect. 2013; **19**: 492–500.

Lefort A, Panhard X, Clermont O, Woerther PL, Branger C, Mentré F, *et al.* Host factors and portal of entry outweigh bacterial determinants to predict the severity of *Escherichia coli* bacteremia. J. Clin. Microbiol. 2011; **49**: 777–783.

Daga AP, Koga VL, Vespero EC. *Escherichia coli* Bloodstream Infections in Patients at a University Hospital: Virulence Factors and Clinical Characteristics. Front Cell Infect Microbiol. 2019; **9**(191).

Basmaci R, Bonacorsi S, Bidet P, Biran V, Aujard Y, Levy C. *Escherichia coli* Meningitis Features in 325 Children From 2001 to 2013 in France. Clinical Infectious Disease. 2015; **61**(5): 779-786.

Wijetunge DS, Gongati S, DebRoy C, Kim KS, Couraud PO, Romero IA, Weksler B, Kariyawasam S. Characterizing the pathotype of neonatal meningitis causing *Escherichia coli*. BMC Microbiology. 2015; **15** :211

Karch H, Denamur E, Vicente M. The enemy within us: lessons from the 2011 European *Escherichia coli* O104:H4 outbreak. EMBO Mol Med. 2012; **4**(9): 841-848.

Vaishnavi C. Translocation of gut flora and its role in sepsis. Ind. J. Med. Microbiol. 2013; **31**: 334–342.

Gati NS, Middendorf-Bauchart B, Bletz S, Dobrindt U, Mellmann A. Origin and Evolution of Hybrid Shiga Toxin-Producing and Uropathogenic *Escherichia coli* Strains of Sequence Type 141. Journal of Clinical Microbiology. 2019; **58**(1): e01309-19.

Lindstedt BA, Finton MD, Porcellato D, Brandal LT. 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 Infect Dis. 2018; **18**(1):544.

Khairy RM, Mohamed ES, Ghany HM, Abdelrahim SS. Phylogenic classification and virulence genes profiles of uropathogenic *E. coli* and diarrhegenic *E. coli* strains isolated from community acquired infections. PLoS ONE. 2019; **14**(9): e0222441.

Johura FT, Parveen R, Alam M. Occurrence of Hybrid *Escherichia coli* Strains Carrying Shiga Toxin and Heat-Stable Toxin in Livestock of Bangladesh. Front Public Health. 2016; **4**(287).

Bai X, Zhang J, Ambikan A, Jernberg C, Ehricht R, Sheutz F, *et al.* Molecular Characterization and Comparative Genomics of Clinical Hybrid Shiga Toxin-Producing and Enterotoxigenic *Escherichia coli* (STEC/ETEC) Strains in Sweden. 2019; **9**(5619).

Karch H, Tarr PI, Bielaszewska M. 2005. Enterohaemorrhagic *Escherichia coli* in human medicine. Int. J. Med. Microbiol. **295**:405–418.

Daneman N, Chateau D, Dahl M, Zhang J, Fisher A, *et al.* Fluoroquinolone use for uncomplicated urinary tract infections in women: a retrospective cohort study. Clinical Microbiology and Infection. 2019.

Adedeji WA. The Treasure Called Antibiotics. Ann lb Postgrad Med. 2016. **14**(2): 56-57.

Kapoor G, Saigal S, Elongavan A. Action and resistance mechanisms of antibiotics: A guide for clinicians. J Anaesthesiol Clin Pharmacol. 2017; **33**(3): 300-305.

Bush K and Bradford PA. β-Lactams and β-Lactamase Inhibitors: An Overview. Cold Spring Harb Perspect Med. 2016; **6**(8).

Kong KF, Schneper L, Mathee K. Beta-lactam Antibiotics: From Antibiosis to Resistance and Bacteriology. APMIS. 2010; **118**(1): 1-36.

Vincent JL, Abraham E, Moore FA, Kochanek PM, Fink MP. Textbook of Critical Care. Elsevier Saunders. Philadelphia, PA. 2011.

Yao J and Moellering R. Antibacterial Agents. Manual of clinical microbiology. 10th ed. Washington DC, ASM Press. 2011: 1043-1081.

Parker N, Schneegurt M, Thi Tu AH, Lister P, Forster BM. Microbiology. 1st ed. XanEdu Publishing Inc. Ann Arbor, MI. 2016.

Clardy J, Fischbachbach M, Currie C. The natural history of antibiotics. 2009; 19(11): R437-R441.

Newton GGF, Abraham EP. 1956. Isolation of cephalosporin C, a penicillin-like antibiotic containing d-alpha aminoadipic acid. Biochem J. 62:651–658.

Kahan JS, Kahan FM, Goegelman R, Currie SA, Jackson M, Birnbaum J. Thienamycin, a new beta-lactam antibiotic. I. Discovery, taxonomy, isolation and physical properties. J Antibiot. 1979; **32**:1–12.

Sykes RB, Wells JS, Parker WL, Koster WH, Cimarusti CM. Aztreonam: discovery and development of the monobactams. N J Med Spec. 1986; :**8**–15.

Bevan ER, Jones AM, Hawkey PM. Global epidemiology of CTX-M β-lactamases: temporal and geographical shifts in genotype. Journal of Antimicrobial Chemotherapy. 2017; **72**(8): 2145-2155.

Narchi H. Kluyvera urinary tract infection: case report and review of the literature. Pediatr Infect Dis J. 2005; **24**:570–572.

CDC Center for Disease Control and Prevention. 2019. https://www.cdc.gov/ecoli/2019/o157h7-11-19/index.html

Humphries, R. M. *et al.* First Report of Ceftazidime-Avibactam Resistance in a KPC3-Expressing *Klebsiella pneumoniae* Isolate. Antimicrobial Agents and Chemotherapy. 2015; **59**: 6605-6607.

European Committee on Antimicrobial Susceptibility Testing. Clinical breakpoints; Breakpoint table for interpretation of MICs and zone diameters [Internet]. Available from: http://www.eucast.org/clinical_breakpoints/. Accessed September 20, 2019.

Cox G, Wright GD. Intrinsic antibiotic resistance: Mechanisms origins, challenges and solutions. International Journal of Medical Microbiology. 2013. **303** (6-7): 287-292.

Martinez JL, Alonso A, Gomez-Gomez JM, *et al.* Quinolone resistance by mutations in chromosomal gyrase genes. Just the tip of the iceberg? J Antimicrob Chemother. 1998; **42**: 683-688.

Huseby DL, Pietsch F, Brandis G, Garoff L, Tegehall A, Hughes D. Mutation Supply and Relative Fitness Shape the Genotypes of Ciprofloxacin-Resistant *Escherichia coli*. Mol. Biol. Evol. 2017; 34(5):1029-1039.

Moon DC, Seol SY, Gurung M, Jin JS, Choi CH, Kim J, *et al.* Emergence of a new mutation and its accumulation in the topoisomerase IV gene confers high levels of resistance to fluoroquinolones in *Escherichia coli* isolates. International Journal of Antimicrobial Agents. 2010; 35:76-79.

Jacoby GA, Sutton L. Properties of plasmids responsible for production of extended-spectrum beta-lactamases. Antimicrob Agents Chemother. 1991; **35**: 164-169.

Jacoby GA, Medeiros AA. More extended-spectrum beta-lactamases. Antimicrob Agents Chemother. 1991; **35**: 1697-1704.

Ganewatta MS and Tang C. Controlling macromolecular structures towards effective antimicrobial polymers. Polymer. 2015; **63**, A1-A29.

Hakenbeck R, Coyette J. Resistant penicillin-binding proteins. Cell Mol Life Sci. 1998;54:332–340.

Sun S, Selmer M, Andersson DI. Resistance to β-Lactam Antibiotics Conferred by Point Mutations in Penicillin-Binding Proteins PBP3, PBP4 and PBP6 in Salmonella enterica. PloS One. 2014; 9(5): e97202.

Barlow M, Hall BG. Phylogenetic analysis shows that the OXA beta-lactamase genes have been on plasmids for millions of years. J Mol Evol. 2002; **55**:314–321.

Bartoloni A, Bartalesi F, Mantella A, Dell'Amico E, Rossolini GM. High prevalence of acquired antimicrobial resistance unrelated to heavy antimicrobial consumption. J Infect Dis. 2004; **189**: 1291–1294.

Bartoloni A, Pallecchi L, Rodriguez H, Fernandez C, Rossolini GM. Antibiotic resistance in a very remote Amazonas community. Int J Antimicrob Agents. 2009; **33**: 125–129.

Segawa T, Takeuchi N, Rivera A, Yamada A, K. Distribution of antibiotic resistance genes in glacier environments. Environ Microbiol Rep. 2013; **5**:127–134.

Van Goethem MW, Pierneef R, Bezuidt OKI, Van De Peer Y, Cowan DA, Makhalanyane TP. A reservoir of 'historical' antibiotic resistance genes in remote pristine Antarctic soils. Microbiome. 2018; **6**:40.

Abraham EP, Chain E. An enzyme from bacteria able to destroy penicillin. Nature. 1940; **146**: 837.

Unemo M, Shafer WM. Antimicrobial Resistance in Neisseria gonorrhoeae in the 21st Century: Past, Evolution, and Future. Clin Microbiol Rev. 2014; **27**(3): 587-613.

Paterson DL, Bonomo RA. Extended-spectrum β-lactamases: a clinical update. Clin. Microbiol. Rev. 2005; 18: 657–686.

Batchelor M, Hopkins KL, Liebana E. Characterization of AmpC-Mediated Resistance in Clinical Salmonella Isolates Recovered from Humans during the Period 1992 to 2003 in England and Wales. J Clin Microbiol. 2005; **43**(5): 2261-2265.

Williams JD. Beta-lactamases and beta-lactamase inhibitors. Int J Antimicrob Agents. 1999; 12:S3–S7.

Jacoby GA. Genetics of extended-spectrum beta-lactamases. Eur J Clin Microbiol Infect Dis. 1994;13.

Bush K, Jacoby GA. Updated functional classification of β-lactamases. Antimicrob. Agents Chemother. 2010; 54, 969–976.

Cantón R. Epidemiology and evolution of β-lactamases. Evolutionary Biology of Bacterial and Fungal Pathogens. ASM Press. Washington. 2008; 249–270.

Cantón R, González-Alba JM, Galán JC. CTX-M Enzymes: Origin and Diffusion. Front Microbiol. 2012; **3**(110).

Adamski CJ, Cardenas AM, Brown NG, Horton LB, Sankaran B, Prasad BV, et al. Molecular basis for the catalytic specificity of the CTX-M extended-spectrum β-lactamases. Biochemistry. 2014; **54**(2): 447–457.

Alvarez M, Tran JK, Chow N, Jacoby GA. Epidemiology of conjugative plasmid-mediated AmpC beta-lactamases in the United States. Antimicrob Agents Chemother. 2004; **48**(2): 533-537.

Brolund A, Edquist PJ, Mäkitalo B, Olsson-Liljequist B, Söderblom T, Wisell KT, Giske CG. Epidemiology of extended-spectrum β-lactamase-producing *Escherichia coli* in Sweden 2007-2011. Clin Microbiol Infect. 2014; 20(6):O344-52.

Ambler RP, Coulson AF, Frere JM, Ghuysen JM, Joris B, Forsman M, *et al.* A standard numbering scheme for the class A beta-lactamases. Biochem J. 1991; 276:269–70.

Zhang HM, Hao Q. Crystal structure of NDM-1 reveals a common β-lactam hydrolysis mechanism. FASEB J. 2011; 25:2574–2582.

Bush K. Past and Present Perspectives on -Lactamases. Past and Present Perspectives on -Lactamases. Antimicrob Agents Chemother. 2018; 62(10): e01076-18.

Actor JK. Elsevier's Integrated Review Immunology and Microbiology (Second Edition). Elsevier Saunders. Philadelphia, PA. 2012.

Pfeifer Y, Cullik A, Witte W. Resistance to cephalosporins and carbapenems in Gram-negative bacterial pathogens. Int J Med Microbiol. 2010; **300**: 371–379.

Mathers A. Mobilization of carbapenemase-mediated resistance in Enterobacteriaceae. Microbiol Spectr. 2016. EI10-0010-2015.

Brolund A, Sandegren L. Characterization of ESBL disseminating plasmids. Infect Dis. 2016; **48**:18–25.

Domingues S, da Silva GJ, Nielsen KM. Integrons. Mob Genet Elements. 2012; **2**(5): 211-223.

Peterson E, Kaur P. Antibiotic Resistance Mechanisms in Bacteria: Relationships Between Resistance Determinants of Antibiotic Producers, Environmental Bacteria, and Clinical Pathogens. Front Microbiol. 2018; **9**: 2928.

Milan AS. Antibiotic Resistance in the Clinical Context. Trends in Microbiology. 2018; **26**(12).

Sherratt DJ. The Maintenance and Propagation of Plasmid Genes in Bacterial Populations. J Gen Microbiol. 1982; **128**: 655-661.

Hasegawa H, Suzuki E, Maeda S. Horizontal Plasmid Transfer by Transformation in Escherichia coli: Environmental Factors and Possible Mechanisms. Front Microbiol. 2018; **9**(2365).

Carattoli A. Plasmids and the spread of resistance. International Journal of Medical Microbiology. 2013; **303**(6-7): 298-304.

Paterson DL. Recommendation for treatment of severe infections caused by Enterobacteriaceae producing extended-spectrum β-lactamases (ESBLs) Clin Microbiol Infect. 2000;**6**:460–3.

Naseer U, Haldorsen B, Tofteland S, Hegstad K, Scheutz F, Simonsen GS, Sundsfjord A, Norwegian ESBL Study Group. Molecular characterization of CTX-M-15-producing clinical isolates of *Escherichia coli* reveals the spread of multidrug-resistant ST131 (O25:H4) and ST964 (O102:H6) strains in Norway. APMIS. 2009; **117**(7):526-36.

Brolund A, Franzén O, Melefors O, Tegmark-Wisell K, Sandegren L. Plasmidome-analysis of ESBL-producing *Escherichia coli* using conventional typing and high-throughput sequencing. One. 2013; 8(6):e65793.

Sun D. Pull in and Push Out: Mechanisms of Horizontal Gene Transfer in Bacteria. Front. Microbiol. 2018; **9**(2154).

NORM/NORM-VET. Usage of Antimicrobial Agents and Occurrence of Antimicrobial Resistance in Norway [Report]. NORM/NORM-VET. (2016).

Tängden T, Cars O, Melhus A, Lowdin E. Foreign travel is a major risk factor for colonization with *Escherichia coli* producing CTX-M-type extended-spectrum beta-lactamases: a prospective study with Swedish volunteers. Antimicrob Agents Chemother 2010; 54: 3564–8.

Woerther PL, Andremont A, Kantele A. Travel-acquired ESBL-producing Enterobacteriaceae: impact of colonization at individual and community level. J Travel Med. 2017; **24**: S29-S34.

Sanger F, Nicklen D. DNA sequencing with chain-terminating. Proc Natl Acad Sci U S A. 1977; **74**(12):5463–5467.

Kchouk M, Gibrat JF, Elloumi M. Generations of Sequencing Technologies: From First to Next Generation. Biol Med. 2017; **9**:3.

Wheeler DA, Srinivasan M, Egholm M, Shen Y, Chen L, *et al.* The complete genome of an individual by massively parallel DNA sequencing. Nature. 2008. **452**: 872-876.

Illumina. An introduction to Next-Generation Sequencing Technology. https://www.illumina.com/documents/products/illumina_sequencing_introduction.pdf. (accessed 10 January 2020).

Rang FJ, Kloosterma, WP, De Ridder J. From squiggle to basepair: Computational approaches for improving nanopore sequencing read accuracy. Genome Biology, 2018; **19**: 90.

Perkel J. TechBlog: The nanopore toolbox. NatureJobs. 2017. http://blogs.nature.com/naturejobs/2017/10/16/techblog-the-nanopore-toolbox/

Angiuoli SV, Gussman A, Klimke W, Cochrane G, Field D, Garrity G, *et al.* Toward an online repository of Standard Operating Procedures (SOPs) for (meta)genomic annotation. OMICS. 2008; **12**(2):137-141.

Jolley KA, Bray JE, Maiden MCJ. Open-access bacterial population genomics: BIGSdb software, the PubMLST.org website and their applications. Wellcome Open Res. 2018; **3**(124).

Van Domselaar GH, Stothard P, Shrivastava S, Cruz JA, Guo A, Dong X, Lu P, Szafron D, Greiner R, Wishart DS. BASys: a web server for automated bacterial genome annotation. Nucleic Acids Res. 2005; **33**:W455–9

Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Zagnitko O. The RAST Server: Rapid Annotations using Subsystems Technology. BMC Genomics. 2008; **9**(75).
Seemann T. Prokka: rapid prokaryotic genome annotation. Bioinformatics. 2014; **30**(14):2068–9

Seeman T. Prokka: rapid prokaryotic genome annotation. Bioinformatics. 2014; **30** Seemann, 2014 Seemann, 2014; (14): 2068-2069.

Søraas A, Sundsfjord A, Sandven I, Brunborg C, Jenum PA. Risk Factors for Community-Acquired Urinary Tract Infections Caused by ESBL-Producing Enterobacteriaceae – A Case-Control Study in a Low Prevalent Country. PLoS One. 2013; **8**(7).

Ciesielczuk H, Jenkins C, Chattaway M, Doumith M, Hope R, Wareham DW. Trends in ExPEC serogroups in the UK and their significance. Eur J Clin Microbiol Infect Dis. 2016; **35**: 1661-1666.

Bozcal E, Eldem V, Aydemir S, Skurnik M. The relationship between phylogenetic classification, virulence and antibiotic resistance of extraintestinal pathogenic *Escherichia coli* in İzmir province, Turkey. *PeerJ*. 2018; ***6***: 5470.

Alghoribi MF, Gibreel TM, Farnham G, Al Johani SM, Balkhy HH, Upton M. Antibiotic-resistant ST38, ST131 and ST405 strains are the leading uropathogenic *Escherichia coli clones* in Riyadh, Saudi Arabia. J Antimicrob Chemother. 2015; **70**: 2757-2762.

Hornsey M, Betts JW, Mehat JW, Wareham DW, van Vliet AHM, La Ragione RM. Characterization of a colistin-resistant Avian Pathogenic *Escherichia coli* ST69 isolate recovered from a broiler chicken in Germany. J Med Microbiol. 2019; **68**: 111-114.

Nüesch-Inderbinen M, Treier A, Zurfluh K, Stephan R. Raw meat-based diets for companion animals: a potential source of transmission of pathogenic and antimicrobial-resistant Enterobacteriaceae. R. Soc. open sci. 2019; **6**: 191170.

Mamani R, Flament-Simon SC, Garcia V, Mora A, Alonso MP, Blanco J. Sequence Types, Clonotypes, Serotypes, and Virotypes of Extended-Spectrum β-Lactamase-Producing *Escherichia coli* Causing Bacteraemia in a Spanish Hospital Over a 12-Year Period (2000 to 2011). Front. Microbiol. 2019; **10**(1530).

Bubpamala J, Khuntayaporn P, Thirapanmethee K, Montakantikul P, Satanirand P, Chomnawang MT. Phenotypic and genotypic characterizations of extended-spectrum

beta-lactamase-producing *Escherichia coli* in Thailand. Infect Drug Res. 2018; **11**: 2151-2157.

Dagher C, Salloum T, Alousi S, Arabaghian H, Araj GF, Tokajian S. Molecular characterization of Carbapenem resistant *Escherichia coli* recovered from a tertiary hospital in Lebanon. PLoS One **13**(9): e0203323.

Bitar I, Piazza A, Gaiarsa S, Villa L, Pedroni P, Migliavacca R. ST405 NDM-5 producing *Escherichia coli* in Northern Italy: the first two clinical cases. Clin Microbiol Infect. 2017; **23**: 489-490.

Chen Y, Liu Z, Zhang Y, Zhang Z, Lei L, Xia Z. Increasing Prevalence of ESBL-Producing Multidrug Resistance *Escherichia coli* From Diseased Pets in Beijing, China From 2012 to 2017. Front. Microbiol. 2019; **10**(2852).

Pietsch M, Irrgang A, Roschanski N, Brenner MG, Hamprecht A, Werner G. Whole genome analyses of CMY-2-producing *Escherichia coli* isolates from humans, animals and food in Germany. BMC Genomics. 2018; **19**(1).

Enclosed papers

Paper I

Lindstedt et al. BMC Infectious Diseases (2018) 18:544
https://doi.org/10.1186/s12879-018-3449-2

BMC Infectious Diseases

RESEARCH ARTICLE Open Access

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

Bjørn-Arne Lindstedt[1*], Misti D. Finton[1], Davide Porcellato[1] and Lin T. Brandal[2]

Abstract

Background: Classification of pathogenic *Escherichia coli* (*E. coli*) has traditionally relied on detecting specific virulence associated genes (VAGs) or combinations thereof. For *E. coli* isolated from faecal samples, the presence of specific genes associated with different intestinal pathogenic pathovars will determine their classification and further course of action. However, the *E. coli* genome is not a static entity, and hybrid strains are emerging that cross the pathovar definitions. Hybrid strains may show gene contents previously associated with several distinct pathovars making the correct diagnostic classification difficult. We extended the analysis of routinely submitted faecal isolates to include known virulence associated genes that are usually not examined in faecal isolates to detect the frequency of possible hybrid strains.

Methods: From September 2012 to February 2013, 168 faecal isolates of *E. coli* routinely submitted to the Norwegian Institute of Public Health (NIPH) from clinical microbiological laboratories throughout Norway were analysed for 33 VAGs using multiplex-PCR, including factors associated with extraintestinal pathogenic *E. coli* (ExPEC) strains. The strains were further typed by Multiple Locus Variable-Number Tandem-Repeat Analysis (MLVA), and the phylogenetic grouping was determined. One isolate from the study was selected for whole genome sequencing (WGS) with a combination of Oxford Nanopore's MinION and Illumina's MiSeq.

Results: The analysis showed a surprisingly high number of strains carrying ExPEC associated VAGs and strains carrying a combination of both intestinal pathogenic *E. coli* (IPEC) and ExPEC VAGs. In particular, 93.5% (101/108) of isolates classified as belonging to an IPEC pathovar additionally carried ExPEC VAGs. WGS analysis of a selected hybrid strain revealed that it could, with present classification criteria, be classified as belonging to all of the Enteropathogenic *Escherichia coli* (EPEC), Uropathogenic *Escherichia coli* (UPEC), Neonatal meningitis *Escherichia coli* (NMEC) and Avian pathogenic *Escherichia coli* (APEC) pathovars.

Conclusion: Hybrid ExPEC/IPEC *E. coli* strains were found at a very high frequency in faecal samples and were in fact the predominant species present. A sequenced hybrid isolate was confirmed to be a cross-pathovar strain possessing recognised hallmarks of several pathovars, and a genome heavily influenced by horizontal gene transfer.

Keywords: *Escherichia coli*, Pathogenic, ExPEC, IPEC, Hybrid strains, MinION

* Correspondence: bjorn-arne.lindstedt@nmbu.no
[1]Faculty of Chemistry, Biotechnology and Food Science, Norwegian University of Life Sciences, P.O. Box 5003, N-1432 Ås, Norway
Full list of author information is available at the end of the article

Background

Escherichia coli (*E. coli*) is a highly diverse and predominant species among facultative anaerobic bacteria of the human gastrointestinal tract [1]. *E. coli* comprises non-pathogenic commensals as well as strains causing a range of diseases. *E. coli* strains capable of causing extra-intestinal infections are designated as extraintestinal pathogenic *E. coli* (ExPEC) to distinguish them from strains causing intestinal disease, commonly designated as intestinal pathogenic *E. coli* (IPEC).

ExPEC can cause a wide variety of extraintestinal infections at multiple anatomical sites. ExPEC frequently cause urinary tract infection (UTI), septicaemia, meningitis, as well as causing soft tissue damage [2, 3]. ExPEC includes, among others, the pathovars uropathogenic *E. coli* (UPEC) associated with urinary tract infection in human and animals, neonatal meningitis-associated *E. coli* (NMEC), septicaemic *E. coli* (SePEC) causing systemic infection in human and animals, avian pathogenic *E. coli* (APEC) that cause avian colibacillosis, and a potentially emerging ExPEC lineage named endometrial pathogenic *E. coli* (EnPEC) [4, 5].

A wide range of VAGs have been associated with ExPEC and common virulence attributes among ExPEC strains are those enabling their extraintestinal lifestyle e.g. genes coding for the production of adhesins, toxins, protectins, siderophores, iron transport systems, and invasins [2, 6–9]. It is believed that ExPEC are facultative pathogens, which reside in the normal gut flora as commensals in some groups of the healthy population [8]. However, there are no universal accepted concrete genetic criteria for defining an *E. coli* strain as ExPEC nor for definite pathovar classification within the ExPEC group. Thus, the true pathovar classification can only be done on the basis of the isolation source for the majority of ExPECs.

There is limited information regarding the frequency of ExPEC strains in the human intestine, however a recent meta study of more than 500 published papers assessed a prevalence of ExPEC strains among faecal isolates of about 10% in healthy individuals [10]. Reference laboratories or diagnostic microbiological laboratories routinely search for only the established IPEC virulence factors in faecal samples from symptomatic patients. There exist little data on the frequency of ExPEC related virulence factors among these strains.

The aim of this study was to investigate the frequency and combination of virulence markers including VAGs used for IPEC pathovar classification and a selection of VAGs related to ExPEC pathovars among *E. coli* strains submitted from individuals showing signs of gastrointestinal infections. We assessed the frequency of ExPEC and IPEC strains, phylogenetic grouping and the MLVA-genotype.

In light of the large German O104:H4 outbreak in 2011 [11], which was caused by a hybrid Enteroaggregative *E. coli* (EAEC)/Shiga toxin producing *E. coli* (STEC) strain [12], the monitoring of isolates to detect new or altered combinations of VAGs is important as it may give a pre-warning of emerging strains harbouring novel VAG combinations, which should be studied in closer detail to assess whether they also have altered virulence capabilities.

Methods

Bacterial isolates

All 168 *E.coli* strains were obtained from the culture collection at the National Reference Laboratory for Enteropathogenic Bacteria at the Norwegian Institute of Public Health (NIPH).

VAG PCR

ExPEC VAGs, cnf1, cnf2, cnf3, ehaA and ehaG PCR

PCR-primers for amplication of the following VAGs were constructed using primer3 software (http://www.ncbi.nlm.nih.gov/tools/primer-blast/) and DNASTAR's Lasergene software module "Primer Select" (DNAstar, Inc., Madison, WI): cytotoxic necrotising factors 1–3 *cnf1, cnf2, cnf3*; autotransporters (ATs) *sat, tsh, vat, ehaA*, and *ehaG*; iron acquisition *iutA*, *sitA, iucD, iroC, fbpB,* and *fyuA;* adhesins *sfaS, papC,* and *tosA*; protectins *kpsS, traT* and *iss*; the invasin gene *ibeA,* and primers directed at orf5 in the gimB genetic island (sequence acc. no. AY170898). Primers directed at the *etsA* gene encoding the macrolide-specific efflux protein EtsA were also designed (see Additional file 1). PCR primers aimed at ExPEC VAGs and ehaA + ehaG were combined in four multiplex reaction mixes as follows: Multiplex 1 (*cnf1, cnf2, cnf3, iutA, ibeA and sitA*), Multiplex 2 (*iucD, iss, traT, iroC, sat, papC and ehaA*), Multiplex 3 (*tsh*, gimB-genetic island, *etsA, kpsS* and *sfaS*), and Multiplex 4 (*tosA, vat, fbpB, fyuA* and *ehaG*).

All primers had a final concentration of 5 μM. The PCR was run on a GeneAmp 9700 thermocycler (Applied-Biosystems, Foster City, CA, USA) with the following conditions: multiplexes 1, 2 and 4; 95 °C for 15 min, then 25 cycles of 94 °C for 30 s, 58 °C for 90 s and 72 °C for 90 s, followed by a hold on 72 °C for 10 min after temperature cycling has ended. Multiplex 3; 95 °C for 15 min, then 25 cycles of 94 °C for 30 s, 60 °C for 90 s and 72 °C for 90 s, followed by a hold on 72 °C for 10 min after temperature cycling has ended. The multiplexes were diluted 1:25 and run in separate capillaries on an ABI 3130 Genetic Analyzer (Applied-Biosystems, Foster City, CA, USA) with GS 600LIZ as internal size standard.

IPEC VAGs

PCR for detecting common IPEC VAGs was performed as previously published [13–15]. In all, primers for the following IPEC VAGs were included: *stx1, stx2, eaeA, ipaH, LTI, STIa, STIb, aggR, ehxA, bfp* with 16S control *rrs* (see Additional file 1).

Phylogenetic group PCR

The improved phylogenetic PCR-assay [16] of the original assay described by Clermont [17] was used to assign the *E. coli* isolates to major phylogenetic groups and subgroups.

MLVA

Multi-locus variable-number tandem repeats analysis was performed using a modified version of the 10-loci generic *E. coli* MLVA scheme previously published [18]. The PCR-amplicon of the published CCR001 locus contains two variable repeated elements, and the modified scheme allows typing of both these variable elements increasing the number of the generic *E. coli* MLVA to 11-loci. The modification consists of a change of dyes and an additional new reverse-primer at the CCR001 locus as follows: the 6FAM dye was removed from the published CCR001 forward primer [18] and the published unlabelled CCR001 reverse primer was labelled with 6FAM and renamed CCR001aR. A new second VIC-labelled reverse primer was added "CCR001bR: 5' - VIC-CGCATTTTATCTGTCTGTACGGC – 3'". The combination of both reverse primers made it possible to simultaneously separate both repeat containing regions at the CCR001 locus.

Stx subtyping

Subtyping of *stx1* and *stx2* was performed as described in Brandal et al. 2015 [15].

Oxford Nanopore MinION sequencing

The hybrid ExPEC/IPEC strain FHI_NMBU_03 identified by PCR, was chosen for sequencing by the MinION MK1 device. DNA was quantified using the Qubit fluorometer (Life Technologies, Paisley, UK) and 200 ng of DNA was used for library preparation. The strain was sequenced using the R9.4 SpotON flow cell and the SQK-RAD002 rapid sequencing kit. All runs were prepared according to the standard protocol of Oxford Nanopore Technologies (Oxford, UK). The flow cells were primed with a priming solution that consisted of a mixture of nuclease free water and Fuel Mix. The library was then loaded into the MinION SpotON port and the 48-h sequencing protocol was selected in the MinKNOW software. The basecalling was done through the Metrichor Desktop Agent using 1D Basecalling for the SQK-RAD002 protocol.

Illumina MiSeq sequencing

Illumina sequencing was performed on an Illumina MiSeq platform (Illumina Inc., San Diego, CA, USA). Library was prepared using the Nextera XT kit (Illumina Inc) according to manufacturer's instructions and was sequenced using a 300 bp paired-end sequencing kit (Illumina Inc).

Sequence analysis

Raw Illumina reads were paired and quality filtered using Trimmomatic [19] and bases with low quality (< q20) were discarded. MinION reads were extracted using poRe [20] and both read types were assembled using SPAdes [21] version 3.5.0 using the option "--nanopore".

Using combined MiSeq and MinION data, the sequences were assembled into a large contig constituting the genome and a contig containing a large virulence plasmid.

The sequence data was annotated using four different services, the NCBI Prokaryotic Genome Annotation Pipeline [22], the BASys Bacterial Annotation System [23], The RAST Annotation Server [24] and Prokka [25]. The sequences were further analysed using a variety of free and publicly available software. Integrated prophages and genomic islands (GIs) were searched using PHASTER [26] and Island Viewer 4 [27] respectively, and the final location of prophages and GIs was determined using a combination of the resulting data. Multilocus sequence typing (MLST)-type, Fim-type, antibiotic resistance genes, and virulence genes were searched using online services from the Center for Genomic Epidemiology (CGE) at the Danish Technical University (DTU), Lyngby, Denmark (http://www.genomicepidemiology.org/). Assembly and annotation of the isolate FHI_NMBU_03 and its plasmid are publicly available at NCBI (accession number CP019455 and CP019456, respectively).

Results

PCR

The pathovar distribution among the 168 *E. coli* faecal isolates were as follows: 53 non-IPEC (31.5%), (including 2 strains harbouring *ehaG* only and 1 strain negative for all tested VAGs). One hundred eight IPEC (64.3%), (including 49 atypical-EPEC (aEPEC) (29.2%), 31 STEC (18.5%), 21 enterotoxigenic *Escherichia coli* (ETEC) (12.5%), 7 necrotoxin producing *E. coli* (NTEC) (4.2%), 3 enteroinvasive *Escherichia coli* (EIEC) (1.8%), 2 EAEC (1.2%), 1 typical-EPEC (tEPEC) (0.6%), and 1 STEC/ETEC (*stx2d, LTI, iss, traT* and *ehaG*) hybrid strain (0.6%)). A total of 108 isolates (64.3%) contained both recognised IPEC and ExPEC VAGs, thus 93.9% (108/115) of the IPEC isolates also carried ExPEC VAGs. Fifty isolates (29.7%) carried only recognised ExPEC VAGs without any accompanying IPEC associated genes (Table 1). The frequency of the phylogenetic subgroups

Table 1 Distribution of pathotypes in *E. coli* faecal isolates

Pathotype	ExPEC[a]	All IPEC[b]	Other[c]	IPEC VF only[d]	IPEC/EXPEC[e]
Number	50	115	3	7	108
Percent	29.7%	68.5%	1.8%	4.2%	64.3%

[a]Number of isolates with ExPEC VAGs only
[b]Number of isolates containing an IPEC VAG
[c]Two isolates positive for the *ehaG* gene only, and one isolate negative for all 33 markers
[d]Number of isolates with IPEC VAGs exclusively
[e]Number of isolates positive for combinations of both IPEC and ExPEC VAGs

were: 15 A0 (8.9%), 35 A1 (20.8%), 64 B1 (38.1%), 7 B2_2 (4.2%), 21 B2_3 (12.5%), 20 D1 (11.9%) and 6 D2 (3.6%). The phylogenetic group distribution within each pathovar can be seen in Table 2. The highest frequency of combinatory IPEC/ExPEC strains was seen in phylogenetic subgroup B2_2 and group B1 (100 and 75%, respectively). The frequency of the tested ExPEC related VAGs among all isolates can be seen in Table 3. The *ehaG* gene was detected in 64.3% of the isolates and was the most common VAG in our collection. *eae, ehaA, ehxA* and the gimB genetic island marker were present in 44.6, 38.7, 15.5, and 1.2% of the isolates, respectively. When we looked at the average number of VAGs within all phylogenetic subgroups, we found that subgroup B2_2 carried most VAGs (7 VAGs) followed by B2_3 (6.9 VAGs), D2 (5.7 VAGs), D1 (5.4 VAGs), B1 (5.1 VAGs), A1 (4.1 VAGs), and A0 (3.3 VAGs).

The 168 isolates grouped into 131 different MLVA-profiles (1.23 isolates/MLVA-profile), where six clusters of identical MLVA-profiles containing three or more isolates were detected. Cluster 1 consisted of five ExPEC isolates of phylogenetic group A1, all from December 2012. Four of the isolates shared the same VAGs (*sitA, iss, traT, kpsS* and *ehaG*), while the fifth isolate had a deviating VAG composition (*iutA, cnf2, iucD, iss, traT, ehaA, fbpB* and *ehaG*) and was designated NTEC due to the presence of the gene for cytotoxic necrotising factor 2 (*cnf2*). Cluster 2 comprised of three aEPEC strains of phylogenetic group D1 isolated in October and November 2012, all of serogroup O55 with identical VAGs (*eae, iss, ehaA, fbpB* and *ehaG*). Cluster 3 contained six phylogenetic group B1 isolates from December 2012, where five isolates shared the same VAGs (*LTI, iss, fyuA* and *ehaG*) and was designated ETEC due to the presence of the *LTI* gene. Of these five isolates, four were serotyped into serogroup O78 while no serogroup could be assigned to the fifth isolate. The sixth isolate of MLVA-cluster 3 was also an O78 B1 isolate, but with different VAGs (*sitA, iss, traT, kpsS, fyuA* and *ehaG*). Cluster 4 consisted of four phylogenetic group B1 serogroup O103 STEC isolates from September to December 2012, all with identical VAGs (*stx1a, eae, ehxA, traT, ehaA* and *ehaG*). Cluster 5 consisted of six phylogenetic group B1 serotype O103:H2 STEC isolates from October and

Table 2 Phylogenetic group distribution within each pathovar

Phylogroup	A		B1	B2		D	
Subgroup	A0	A1	B1	B2_2	B2_3	D1	D2
aEPEC	5	7	17	7	8	4	1
tEPEC	0	0	0	0	1	0	0
STEC	0	1	24	0	0	6	0
ETEC	0	8	11	0	0	1	1
EAEC	0	0	1	0	0	1	0
EIEC	2	0	1	0	0	0	0
STEC/ETEC	0	0	0	0	0	0	1
NTEC	0	1	0	0	6	0	0
NON-IPEC	8	18	10	0	6	8	3
Σ Subgroups	15	35	64	7	21	20	6
Σ Phylogroups	50 (29.8%)		64 (30.1%)	28 (16.7%)		26 (15.5%)	

Table 3 Frequency of ExPEC associated virulence genes (PCR screening)

ExPEC associated VAG	Comment	Frequency
iss	Increased serum survival gene	48.8%
traT	Gene encoding complement resistance protein	45.2%
fyuA	Ferric yersiniabactin uptake receptor gene	42.3%
iucD	Aerobactin biosynthesis gene	24.4%
iutA	Ferric aerobactin receptor gene	23.8%
sitA	Iron/manganese transport system periplasmic binding protein gene	23.2%
kpsS	Capsule polysaccharide export protein gene	18.5%
tsh	Temperature-sensitive hemagglutinin autotransporter gene	15.5%
iroC	Salmochelin siderophore system gene	12.5%
vat	Vacuolating autotransporter toxin gene	11.3%
fbpB	Gene associated with urinary tract infections	10.1%
sat	Secreted autotransporter toxin gene	9.5%
ibeA	Invasion protein gene	8.3%
etsA	Macrolide-specific efflux protein gene	4.2%
cnf1	Gene encoding the cytotoxic necrotizing factor 1	3.6%
sfaS	S-fimbrial adhesin gene	1.8%
papC	P-fimbriae outer membrane usher protein gene	1.8%
tosA	Repeat-in-toxin gene	1.8%
cnf2	Gene encoding the cytotoxic necrotizing factor 2	0.6%
cnf3	Gene encoding the cytotoxic necrotizing factor 3	ND[a]

[a]Not detected

November 2012 submitted from the same Norwegian hospital with identical VAGs (*stx1a, eae ehxA, traT, ehaA* and *ehaG*). Cluster 6 contained four phylogenetic group B2_2 aEPEC isolates from September and October 2012 where three of the isolates showed the same VAGs (*eae, ibeA, iss, traT, iroC, tsh, vat* and *fyuA*), while the fourth isolate had the following VAGs (*eae, ibeA, tsh, vat* and *fyuA*).

Among the 49 *eae* containing aEPEC isolates, the following VAGs were additionally detected: *sitA, iss, ehaA, ehaG, papC, tsh, kpsS, vat, fyuA, iutA, iucD, fbpB, ehxA, sat, tsh, traT, ibeA, iroC, etsA, tosA*, as well as a marker in the gimB genetic island. Only 4 of 49 aEPEC isolates (8.2%) did not carry any VAGs previously associated with ExPEC strains. Thus, the majority (91.8%) of our aEPEC faecal isolates contained VAGs related to ExPEC strains. The most common ExPEC related VAGs among the aEPEC isolates were: *traT* (49%), *iss* (38.8%), *fyuA* (32.7%), *tsh* (26.5%) and *ibeA* (26.5%). When we divided the aEPEC isolates by phylogenetic group, we observed that the *ibeA* gene was present in 86.7% (13/15) of the aEPEC B2 strains, and the VAGs *ehaA* and *ehaG* were also frequently present, 49 and 51% respectively.

The 31 STEC isolates contained 18 *stx1* only positive strains and 9 *stx2* only positive strains. The remaining four strains contained both *stx1* and *stx2*. Among the STEC isolates, the following VAGs were additionally found: *eae, iutA, iucD, iss, traT, iroC, ehaA, ehaG, etsA, fyuA, kpsS, ehxA* and *fbpB*. The most common ExPEC related VAGs were: *traT* (58%), *iss* (35.5%), *iucD* (29%) and *iutA* (25.8%). Additional prevalent non-ExPEC factors present were: *ehaA* (96.8%), *ehaG* (90.3%), *ehxA* (74.2%) and *eae* (71%).

Among the 21 ETEC isolates, *ehaG* was detected in 12 strains (57%), but *ehaA* was not detected in any of the ETEC isolates.

When we looked at pair-clustering of the VAGs we found that the most common pairs (in more than 20% of isolates) of VAGs included: *ehaA* and *ehaG* in 60/168 (35.7%) of the isolates, *ehaG* and *traT* or *iss* both combinations in 49/168 (29.2%) of the isolates, *eae* and *ehaA* in 48/168 (28.6%) of the isolates, *eae* and *ehaG* in 46/168 (27.3%) of the isolates, *iss* and *fyuA* in 43/168 (25.6%) of the isolates, *traT* and *eae* or *ehaA* both combinations in 40/168 (23.8%) of the isolates, *iucD* and *iutA* in 40/168 (23.8%) of the isolates and *traT* and *iss* in 38/168 (22.6%) of the isolates.

Sequencing

One strain from this study designated FHI_NMBU_03 from MLVA-cluster 6 was selected for whole genome sequencing using a combination of long- and short- read technologies, Oxford Nanopore MinION (91,865 reads) and Illumina MiSeq (361,031 reads), respectively. We were able to assemble a complete closed circular genome (4,685,056 bp acc. nr. CP019455) and a complete circular virulence plasmid (159,821 bp acc. nr. CP019456) pFHI_NMBU_03–1 from the combined runs. The genome sequence (coverage 21.6x) contained 4954 genes (gene density 1.057 genes/Kbp) and 200 pseudogenes, with a GC content of 51%. The chromosome contains five integrated prophages according to PHASTER analysis [26], and 19 genomic islands (phages excluded) according to the Island Viewer 4 software [27]. FHI_NMBU_03 showed a surprising collection of both IPEC and ExPEC related VAGs as indicated by the PCR-analysis. It contained the locus of enterocyte effacement (LEE)-region of EPEC/EHEC as well as recognized markers for ExPEC subtypes of UPEC/APEC and NMEC. The LEE region of FHI_NMBU_03 contains 36 recognized genes, four open reading frames (ORFs) of unknown function as well as two pseudogenes, and is inserted in the *selC* tRNA gene. The *eae*-intimin subtype of FHI_NMBU_03 is *β2*. The LEE-encoded Tir protein of FHI_NMBU_03 is, by BLAST search, identical to three Tir proteins from EPEC strains and one protein from a human strain designated as UPEC (upec-202, SAMN02802023), as well as eight animal strains. Additionally the genome encodes the intimin-like proteins FdeC and a SinH-variant. FHI_NMBU_03 was also positive for a cluster of the non–LEE-encoded effectors *nleB, nleC, nleG, nleH* and a frameshifted *nleA* pseudogene, located within a phage-region identified by PHASTER. Using CGE the MLST type was predicted to be ST28 and the *fimH* subtype was predicted to fimH90. A selection of chromosomal genes found by sequencing associated with virulence can be seen in Table 4. On the large virulence plasmid, ExPEC pathogenicity associated genes include: *bor* (an *iss* homologue), *traT* (serum resistance associated), the pyelonephritis-associated pilus *pap* operon; *papABCDEFHJK*, a putative *pixG* adhesin related gene encoding a protein 99% identical to a protein (EQZ28352.1) from the *E. coli* human UTI strain UMEA-3585-1 (PRJNA186355), a putative autotransporter gene encoding an uncharacterized protein identical to protein EQZ28355.1 from UMEA- 3585-1, *iroN* (catecholate siderophore receptor), an AppA (HlyII) hemolysin protein and the leukotoxin genes *lktBCD*.

The *alkB* gene coding for the alkylated DNA repair protein AlkB has an internal frameshift, and is probably inactive in FHI_NMBU_03. Several loci pertaining to fimbrial structures were found and noteworthy are genes related to K88-fimbria, 987P-fimbria and colonization factor antigen I fimbriae (CFA/I), which are all associated with ETEC strains. FHI_NMBU_03 is also positive for the YghJ protein gene, also known as SslE (Secreted and surface associated lipoprotein), which is a cell surface associated and secreted lipoprotein harbouring M60 metalloprotease domain [28].

Table 4 Selected virulence associated genes found on the FHI-NMBU-03 chromosome by nBLAST

Gene name (FHI-NMBU-03 chromosome)	Comment	Associated pathovar	% identity	Cover.	BLAST sequence
aatB	Autotransporter adhesin and virulence factor of avian pathogenic *Escherichia coli*.	APEC	98.43	1017 / 1017	JX402062
herA	Archaeal bi-polar DNA helicase	Unknown	99.47	1686 / 1686	NZ_NLRN01000019
aslA	Arylsulfatase gene	ExPEC (Invasive K1 strains)	98.55	1656 / 1656	CU928163
aufC	Fimbrial usher protein gene	UPEC	99.50	2595 / 2595	KE702411
cesAB	Enteropathogenic *Escherichia coli* chaperone for the type-III translocator proteins	EPEC/STEC	100.00	324 / 324	FM986651
cesD2	A second chaperone for the type III secretion translocator protein EspD	EPEC/STEC	98.28	407 / 408	NC013364
cesT	A bivalent enteropathogenic Escherichia coli chaperone required for translocation of both Tir and Map	EPEC/STEC	100.00	471 / 471	LT903847
chuA	*E.coli* hemeutilization protein A gene	ExPEC	99.65	1983 / 1983	LT827011
cif	Type III secreted effector	EPEC/STEC	100.00	849 / 849	AF497476
csgA	Major curlin subunit	Several	99.56	459 / 459	CP023388
csgB	Minor curlin subunit	Several	99.56	456 / 456	CP027060
csgE	Curli production assembly/transport component	Several	98.70	386 / 390	NC_011750
csgF	Curli production assembly/transport component	Several	98.08	417 / 417	NC_011750
csgG	Curli production assembly/transport component	Several	97.72	834 / 834	CP003034
cvaA	Colicin V secretion protein gene	Several	100.00	1242 / 1242	GG773553
Death on curing RelE/ParE family toxin gene	Component of Toxin-antitoxin (TA) system	Several	100.00	272 / 276	CP023388
eae (subtype Beta2)	Intimin - Necessary for the production of attaching and effacing lesions on tissue culture cells	EPEC/STEC	100.00	2820 / 2820	AB647493
ecpA	Common pilus major fimbrillin subunit	Several	98.47	588 / 588	BA000007
ecpD	Fimbria adhesin of the *E.coli* common pilus	Several	99.64	1644 / 1644	CP019777
elfC	Putative fimbrial usher protein	Several	99.73	2595 / 2595	CP021288
entA	Enterobactin biosynthesis gene	Several	96.12	747 / 747	CP027060
entE	Enterobactin biosynthesis gene	Several	95.65	1611 / 1611	CP027060
entH	Enterobactin biosynthesis gene	Several	94.93	414 / 414	CP027060
escC	Outer membrane secretin	EPEC/STEC	91.68	1539 / 1539	AP010958
escD	Type III secretion system inner membrane ring protein	EPEC/STEC	99.10	1221 / 1221	BA000007
escF	Type III secretion system needle major subunit	EPEC/STEC	100.00	222 / 222	NC_002695
escJ	Required for the formation of the type III Secretion Apparatus	EPEC/STEC	91.62	573 / 573	AP010958
escN	Type III secretion ATPase	EPEC/STEC	100.00	1341 / 1341	BA000007
escR	Type III secretion system export apparatus protein gene	EPEC/STEC	99.69	654 / 654	BA000007
escS	Type III secretion system export	EPEC/STEC	100.00	270 / 270	BA000007

Table 4 Selected virulence associated genes found on the FHI-NMBU-03 chromosome by nBLAST *(Continued)*

Gene name (FHI-NMBU-03 chromosome)	Comment	Associated pathovar	% identity	Cover.	BLAST sequence
	apparatus protein gene				
escT	Type III secretion system export apparatus protein gene	EPEC/STEC	99.61	777 / 777	BA000007
escU	Type III secretion system LEE export apparatus switch protein gene	EPEC/STEC	96.15	1038 / 1038	AP010958
escV	Translocase of the type III secretion system	EPEC/STEC	99.70	2028 / 2028	BA000007
espA	Type III secretions system gene	EPEC/STEC	100.00	573 / 573	AJ225016
espG	Type III secretion system effector, which localize to the Golgi apparatus and disrupt its architecture	EPEC/STEC	98.41	1197 / 1197	BA000007
etgA	Lytic transglycosylase	EPEC/STEC	100.00	459 / 459	FM986650
fdeC	Mediates *E. coli* adhesion to mammalian cells and extracellular matrix	ExPEC/STEC	97.86	4251 / 4251	CP019777
sfaH	S-fimbrial protein subunit gene	ExPEC	98.56	903 / 903	KT444704
flgD	Flagellar basal body rod modification protein gene	Several	96.55	696 / 696	CP027060
flgM	Negative regulator of flagellin synthesis	Several	98.97	290 / 294	CP028192
fmlA	Major F9-fimbrial subunit	ExPEC/IPEC	96.81	564 / 564	BA000007
fyuA	Ferric yersiniabactin uptake receptor	ExPEC	99.51	2022 / 2022	CP016828
gad (1)	Glutamate decarboxylase gene	Several	99.64	1401 / 1401	CP001671
gad (2)	Glutamate decarboxylase gene	Several	99.79	1401 / 1401	FM180568
grlA	Global regulator of LEE activator	EPEC/STEC	97.56	409 / 414	AP010958
gtrA	Type IV O-antigen modification gene (*Shigella flexneri*)	Unknown	90.08	363 / 363	AF288197
hbp	Hemoglobin-binding protease hbp autotransporter gene	ExPEC	99.95	4131 / 4131	CP009072
hlyIII	Gene encoding inner membrane protein, hemolysin III family	ExPEC/IPEC	98.41	690 / 690	CP003034
ibeA	Invasion protein gene	NMEC/APEC/AIEC	98.61	1371 / 1371	CP001855
ibeB	Invasion protein gene	Several	98.55	1383 / 1383	AF094824
ibeC/yijP/cptA	Invasion protein gene	Several	99.77	1734 / 1734	CP019777
irp1	HMWP1 nonribosomal peptide/ polyketide synthase	ExPEC	99.65	9492 / 9492	CU928163
irp2	HMWP2 Yersiniabactin biosynthetic protein	ExPEC	98.85	6106 / 6108	CP006834
ler	Negative autoregulator of the LEE1 operon	EPEC/STEC	99.49	390 / 390	BA000007
malX	*Escherichia coli* pathogenicity island-marker	ExPEC	98.61	1581 / 1581	AF003742
MAP	LEE effector protein gene	EPEC/STEC	97.06	612 / 612	LC053401
MBL-fold metallo hydrolase gene	Putative phylogroup B2 specific marker	ExPEC	99.52	1044 / 1044	CP023388
mdtH	Multidrug resistance protein gene	Several	99.83	1209 / 1209	CP019777
mpc	Type III secretion system regulator gene	EPEC/STEC	92.09	354 / 354	AP010953
mviM	Putative virulence factor	Several	98.70	924 / 924	CU928164
nleA[a]	Non-LEE encoded effector A	EPEC/STEC	99.84	1239 / 1239	AB303062
nleB	Non-LEE encoded effector B	EPEC/STEC	100.00	981 / 981	AB303062

Table 4 Selected virulence associated genes found on the FHI-NMBU-03 chromosome by nBLAST *(Continued)*

Gene name (FHI-NMBU-03 chromosome)	Comment	Associated pathovar	% identity	Cover.	BLAST sequence
nleC-like gene	T3SS secreted effector NleC-like protein gene	EPEC/STEC	100.00	264 / 264	CYEL01000033
nleG	Non-LEE encoded effector G	EPEC/STEC	100.00	576 / 576	AB303062
nleH	Non-LEE encoded effector H	EPEC/STEC	99.75	812 / 812	AP010958
usp/putative colicin	Uropathogenic specific protein gene	UPEC	97.14	1782 / 1782	CU651637
sepL	Secretion switching protein gene	EPEC/STEC	94.93	1046 / 1056	BA000007
sepQ	T3SS structure protein	EPEC/STEC	95.53	918 / 918	CP003109
sinH	Intimin-like inverse autotransporter	ExPEC	100.00	2178 / 2178	NZ_NMHI01000013
stcD	Putative fimbrial-like adhesin protein gene	IPEC	99.71	1035 / 1035	NC_018658
stfD	Fimbrial protein gene	Unknown	100.00	753 / 753	LOFW01000008
tir	Translocated intimin receptor protein gene	EPEC/STEC	99.88	1650 / 1650	DQ206455
xhlA	*Xenorhabdus nematophila* haemolysin	Unknown	99.73	372 / 372	LDCR01000046
ybtA	Yersiniabactin transcriptional regulator	ExPEC	99.79	960 / 960	CP028714
ydeR	Fimbrial-like protein gene	Several	98.41	504 / 504	CU928163
yfcV	Major subunit of a putative chaperone-usher fimbria	ExPEC	97.18	567 / 567	NC_011750

[a]Frameshifted

A previously reported insertion of unknown origin with a base composition suggestive of horizontal gene transfer in a genetic region between *mutS* and *rpoS*, associated with phylogroup B2 and uropathogens [29] is additionally present. This region has later been named the *o454-nlpD* region [30].

Discussion

Clinical microbiological laboratories and reference laboratories rely increasingly on genetic testing of faeces to identify possible pathogenic microbes. For enteric bacteria, a widely used practice is to perform PCR or real-time PCR assays, or other amplification methodology, to detect specific genes used for pathogen identification. For *E. coli*, PCR on faecal isolates [13] is used to detect the well-recognized IPEC pathovars EPEC, STEC, ETEC, EAEC and EIEC [31]. These pathovars all have genetic targets used for identification and classification. The most common genetic targets are the *eae* and *bfp* genes for EPEC, *stx1* and *stx2* genes for STEC, genes encoding the thermostable (ST) and thermolabile (LT) toxins for ETEC, the *aggR* gene for EAEC, and the *ipaH* gene for EIEC. These targets are also candidate targets for automatic pathogen identification systems, especially in a culture-independent diagnostic tests (CIDTs) workflow. The results from these assays will be a classification of the *E. coli* isolates into one of the recognized pathovars or, in case of no target amplification, a classification as a non-enteropathogenic or commensal strain.

In the present study, we looked at a wider range of virulence factors in faecal *E. coli* isolates submitted to the Reference Laboratory for Enteropathogenic Bacteria at the Norwegian Institute of Public Health (NIPH). We especially searched for known ExPEC VAGs as in recent years a heighten interest in the frequency of ExPEC strains in the human gut has emerged, however there are few studies examining the selection of VAGs used in the present study.

One surprising finding in our study was the high frequency of *E. coli* strains (64.3%) with a combination of recognized IPEC and ExPEC VAGs. There are limited data on how common these IPEC/ExPEC hybrid strains are. In a study of 265 *E. coli* isolates from hospital inpatients and outpatients with UTIs, 10.6% of isolates harboured at least one IPEC virulence factor [32]. In previous studies of human faecal isolates, the *E. coli* strains are separately designated as IPEC or as commensal strains harbouring ExPEC VAGs, thus it is unclear how high of a percentage may be IPEC/ExPEC combinatory strains. The IPEC/ExPEC combination was especially high among the aEPEC strains (91.8%).

One notable finding was that 13 out of 14 (92.9%) *ibeA* positive isolates was an EPEC strains of phylogenetic group B2. Thus, *ibeA* carriage in faeces seems to be associated with a distinct group of IPEC strains in our material. The *ibeA* gene is a known virulence factor of *E. coli* strains responsible for neonatal meningitis in humans (NMEC) by contributing to the invasion of

brain microvascular endothelial cells (BMEC) [33]. It has also been described that *ibeA* plays an important role in the invasion of intestinal epithelial cells, as the absence of *ibeA* accounted for a reduction in invasion of ca. 67% compared to wild type in experiments with the adherent-invasive *E. coli* (AIEC) strain NRG857c and an *ibeA* deletion mutant strain (NRG857cΔ*ibeA*) [34]. Furthermore, *ibeA* was present in the genome of 26% of pathogenic isolates from chicken (APEC), but absent from the genome of non-pathogenic isolates of avian origin [35]. The *ibeA* gene was positively linked to the pathogenicity of the APEC strains, and it was additionally shown that *ibeA* was involved in the invasion of human BMEC by the APEC strain BEN 2908 [35].

An interesting observation was the high number of strains harbouring genes coding for the trimeric autotransporter proteins (TAAs) EhaA and EhaG. Especially finding the *ehaG* gene in 48% of the strains with one or more ExPEC VAGs and no IPEC VAGs, since EhaG mediates specific adhesion to colorectal epithelial cells [36]. This indicates that 48% of our isolates carrying solely ExPEC VAGs may have the capacity to adhere to colorectal epithelial cells in humans. Both *ehaA* and *ehaG* are most prevalent in the phylogenetic groups B1 and D, while a difference between *ehaA* and *ehaG* was observed in phylogenetic group A where *ehaA* was not detected but *ehaG* was present in 34% of the isolates. The distribution pattern of *ehaA* and *ehaG* was in the same range as results from a study by Zude et al. 2014 [37], with the exception of phylogenetic group B2 where Zude et al. 2014 report that 21.9% of the strains carry the *ehaG* gene, while in the present study 7.1% of the B2 strains were positive for *ehaG*. EhaG is localized at the bacterial cell surface and, in addition to colorectal epithelial cell adhesion, promotes cell aggregation, biofilm formation, and adherence to a range of extracellular matrix (ECM) proteins [36]. TAAs are regarded as important virulence factors of many Gram-negative bacterial pathogens. We are aware that our PCR-based phylogrouping results may show minor differences from the 2013 Clermont method [38]. Non-IPEC strains are not stored at NIPH thus a re-typing of all strains using the 2013 Clermont method on all strains in this study is not possible, however the findings and conclusions are valid, and in future our phylogrouping will be sequenced-based e.g. by using online tools [39].

The fully sequenced FHI_NMBU_03 phylogroup B2 strain (with plasmid) from this study shows hallmarks of ExPEC pathovars UPEC, APEC, NMEC and the IPEC pathovar aEPEC with some VAGs related to ETEC (K88-, 987P- and CFA/I- fimbrial genes), thus it constitutes a truly pathovar-hybrid strain (Additional file 3). The *eae* gene alone will classify it as an aEPEC by most molecular diagnostics tests.

It was previously reported that YghJ caused extensive haemorrhage in mouse ileum in a dose dependent manner and it was suggested that YghJ could be a virulence factor of enteric pathogens associated with haemorrhagic diarrhoea [28]. A recent study additionally showed that the YghJ protein from a neonatal septicaemic *E. coli* altered cellular morphology of various cell lines and triggered the induction of several proinflammatory cytokines, which are attributed as one of the key mediators in the pathogenesis of sepsis [40].

Several factors classify this strain as UPEC (e.g. *usp*, *fyuA*, *sfaS*, the *pap* fimbrial operon, *chuA* and *yfcV*). It has previously been reported that any two of *yfcV*, *vat*, or *chuA* along with *fyuA* could be used to differentiate UPEC from diarrheagenic *E. coli* (DEC), human commensal, or animal commensal isolates. However, to differentiate UPEC from APEC, *vat*, *fyuA*, and *yfcV* together are necessary, where the presence of the putative fimbrial subunit gene *yfcV* is highly predictive of UPEC, increasing the odds of a strain being UPEC by 99.5-fold [41].

The fimH90 subtype was also an interesting finding as it appears to be rare among *E. coli* strains and was not found among 243 draft genomes of *E. coli* isolates in a study using the CGE FimTyper Web tool [42]. However, BLAST searches found an identical *fimH* gene in a sequence scaffold from a human aEPEC strain (702898_aEPEC) isolated in Pakistan (GenBank: CYBW01000017.1). The CGE FimTyper confirmed this *fimH* gene to also be of subtype fimH90.

The comparison of sequence data with PCR typing revealed PCR positive results for *tsh* and *vat* while sequencing showed the presence of the highly related *hbp* gene on the chromosome and a putative related autotransporter on the virulence plasmid (locus tag: BXO92_24355). The PCR results can be explained by the similarity of the intended target genes, and the considerable confusion in GenBank submitted sequences on the correct nomenclature. The Tsh and Hbp proteins differ by only two amino acid residues. In addition, Vat and Tsh/Hbp are 77.5% identical in amino acids.

The plasmid located putative autotransporter protein (protein id: PRJNA362852:BXO92_24355) show 43.7% AA identity and 56.6% AA similarity to Tsh. RAST annotates this protein as EspC, while BASys annotates it as Hbp.

The number of GIs and integrated prophages indicate that FHI_NMBU_03 has obtained a high number of virulence factors by horizontal gene transfer and this may have been facilitated by a defect in the DNA-repair system with a frameshifted *alkB* gene. It is known that AlkB relevant lesions appear to represent strong blocks to replication, but these blocks can be bypassed by error-prone translesion DNA polymerases as a part of the SOS-system, leading to mutagenesis [43].

The *o454-nlpD* region was shown to consist of several genetic patterns, where pattern III (the FHI_NMBU_03 sequence contains pattern III) had significant associations with phylogenetic group B2 strains, representing the most virulent members of the ExPEC group. This *o454-nlpD* region pattern was proposed as a tool to identify highly extraintestinal virulent strains among a mixed population of *E. coli* [30].

Strains closely related to FHI_NMBU_03 may have caused disease in Norway for an extended period of time as nine aEPEC intimin *eae*-β2 carrying B2 strains of sequence type ST28 was previously detected among 56 aEPEC isolates from faecal specimens from children < 5 years old in Norway (five strains were from community-acquired diarrhoea samples) [44]. All nine strains where shown by microarray analysis to contain the *ibeA*, *malX* and *usp* genes as FHI_NMBU_03.

The high frequency of strains with combined IPEC/ExPEC VAGs found in this study is worrisome as they might be capable of causing both intestinal- and extraintestinal disease. One scenario could be a general weakening of the immune system caused by ongoing intestinal disease, thereby creating an opportunity for spread of bacteria with ExPEC VAGs to other anatomical sites where the ExPEC VAGs may contribute to severe extraintestinal disease.

Conclusion

We report that a high frequency (> 93%) of routinely submitted faecal *E. coli* strains from Norwegian hospitals, previously characterized as IPEC, also harbour ExPEC virulence factors. Traditionally IPEC is regarded as a diarrhoeagenic pathogen with a set of virulence genes that is absent in ExPEC strains e.g. UPEC. This very high frequency of combined IPEC/ExPEC was an unexpected finding warranting further studies, as they may provide a rich source of opportunistic extraintestinal infections. WGS of one selected strain confirmed the pathovar-hybrid nature and revealed a genome heavily influenced by horizontal gene transfer (HGT). Sequence complex ST28 has previously been assigned to a hybrid group that was named "phylogroup ABD" [45], which supports our finding of the hybrid nature for strain FHI_NMBU_03.

Additional files

Additional file 1: PCR primers used in study. Sequences of all PCR-primers used in this study, with references. (DOCX 18 kb)

Additional file 2: The Excel sheet contains VAGs PCR, Phylogenetic PCR and MLVA results for all *E. coli* strains included in this study. - PCR positive amplicons are listed as well as the MLVA profile and the results from the phylogenetic group PCR. (XLSX 21 kb)

Additional file 3: FHI-NMBU-03 SNPtree03 slanted. The image shows results from comparing the genome of FHI-NMBU-03 with a selection of *E. coli* whole genomes with *E. coli* K-12 MG1655 as reference. The SNP based phylogenetic tree was constructed using CSI Phylogeny 1.4 (https://cge.cbs.dtu.dk/services/CSIPhylogeny/). (PDF 10 kb)

Abbreviations

aEPEC: Atypical enteropathogenic *Escherichia coli*; AIEC: Adherent-invasive *Escherichia coli*; APEC: Avian pathogenic *Escherichia coli*; BMEC: Brain microvascular endothelial cells; CGE: Center for Genomic Epidemiology; cnf: Cytotoxic necrotising factor; DEC: Diarrheagenic *Escherichia coli*; *E. coli*: *Escherichia coli*; EAEC: Enteroaggregative *Escherichia coli*; EIEC: Enteroinvasive *Escherichia coli*; EnPEC: Endometrial pathogenic *Escherichia coli*; EPEC: Enteropathogenic *Escherichia coli*; ETEC: Enterotoxigenic *Escherichia coli*; ExPEC: Extraintestinal pathogenic *Escherichia coli*; GIs: Genomic islands; HGT: Horizontal gene transfer; IPEC: Intestinal pathogenic *Escherichia coli*; LEE: The locus of enterocyte effacement; MLST: Multilocus sequence typing; MLVA: Multiple Locus Variable-Number Tandem-Repeat Analysis; NIPH: Norwegian Institute of Public Health; NMEC: Neonatal meningitis *Escherichia coli*; NTEC: Necrotoxin producing *Escherichia coli*; ORFs: Open reading frames; SePEC: Septicaemic *Escherichia coli*; STEC: Shiga toxin producing *Escherichia coli*; TAAs: Trimeric autotransporter proteins; tEPEC: Typical enteropathogenic *Escherichia coli*; UPEC: Uropathogenic *Escherichia coli*; UTI: Urinary tract infection; VAGs: Virulence factors; WGS: Whole genome sequencing

Acknowledgements

We would like to thank Inger Løbersli and Marit Hindrum for technical help and support with strain characterization, MLVA, phylogeny and VAGs PCR-analysis.

Funding

No funding was obtained for this study.

Availability of data and materials

All results from PCR analyses (VAGs and Phylogeny) and MLVA genotyping is included in "Additional file 2" as an Excel spreadsheet. The datasets used and/or analysed during the current study are additionally available from the authors on reasonable request. All bacterial strains containing IPEC VAGs are available from the strain collection at the Norwegian Institute of Public Health, Oslo, Norway (https://www.fhi.no/en/more/access-to-data/), a fee might be applied. The nucleotide sequence of the hybrid strain FHI_NMBU_03 and its large virulence plasmid can be downloaded from NCBI's nucleotide database (https://www.ncbi.nlm.nih.gov/nucleotide/) with accession numbers NZ_CP019455.1 and NZ_CP019456.1 respectively.

Authors' contributions

BAL designed the project and participated in the experimental design, implementation and data analysis, wrote the first draft of the manuscript and was responsible for the final submission of the manuscript and submission of sequencing data to GenBank. MDF and DP contributed to WG-sequencing with MinION and MiSeq, data analysis pipeline (bioinformatics), annotations and strain comparisons. LTB participated in the experimental design, implementation and analysis of PCR results, strain selection and data analysis. All authors contributed to manuscript revisions. All authors have read and approved the final manuscript.

Ethics approval and consent to participate

Not applicable. The study did not involve human subjects, human material, nor human data.

Consent for publication

Not applicable. The study does not include details, images, or videos relating to any individual person.

Competing interests

The authors declare that they have no competing interests.

Publisher's Note

Author details
[1]Faculty of Chemistry, Biotechnology and Food Science, Norwegian University of Life Sciences, P.O. Box 5003, N-1432 Ås, Norway. [2]Department of Zoonotic, Food- and Waterborne Infections, Norwegian Institute of Public Health, Oslo, Norway.

Received: 14 June 2018 Accepted: 16 October 2018
Published online: 01 November 2018

References

1. Chaudhuri RR, Henderson IR. The evolution of the Escherichia coli phylogeny. Infect Genet Evol. 2012;12(2):214–26.
2. Dale AP, Woodford N. Extra-intestinal pathogenic Escherichia coli (ExPEC): disease, carriage and clones. J Infect. 2015;71(6):615–26.
3. Johnson JR, Russo TA. Extraintestinal pathogenic Escherichia coli: "the other bad E coli". J Lab Clin Med. 2002;139(3):155–62.
4. Goldstone RJ, Popat R, Schuberth HJ, Sandra O, Sheldon IM, Smith DG. Genomic characterisation of an endometrial pathogenic Escherichia coli strain reveals the acquisition of genetic elements associated with extra-intestinal pathogenicity. BMC Genomics. 2014;15:1075.
5. Sheldon IM, Rycroft AN, Dogan B, Craven M, Bromfield JJ, Chandler A, Roberts MH, Price SB, Gilbert RO, Simpson KW. Specific strains of Escherichia coli are pathogenic for the endometrium of cattle and cause pelvic inflammatory disease in cattle and mice. PLoS One. 2010;5(2):e9192.
6. Di Lorenzo M, Stork M. Plasmid-encoded Iron uptake systems. Microbiol Spectr. 2014;2(6). https://doi.org/10.1128/microbiolspec.PLAS-0030-2014.
7. Kaper JB, Nataro JP, Mobley HL. Pathogenic Escherichia coli. Nat Rev Microbiol. 2004;2(2):123–40.
8. Kohler CD, Dobrindt U. What defines extraintestinal pathogenic Escherichia Scoli? Int J Med Microbiol. 2011;301(8):642–7.
9. Luthje P, Brauner A. Virulence factors of uropathogenic E. coli and their interaction with the host. Adv Microb Physiol. 2014;65:337–72.
10. Starcic Erjavec M, Zgur-Bertok D. Virulence potential for extraintestinal infections among commensal Escherichia coli isolated from healthy humans--the Trojan horse within our gut. FEMS Microbiol Lett. 2015; 362(5). https://doi.org/10.1093/femsle/fnu061.
11. Karch H, Denamur E, Dobrindt U, Finlay BB, Hengge R, Johannes L, Ron EZ, Tonjum T, Sansonetti PJ, Vicente M. The enemy within us: lessons from the 2011 European Escherichia coli O104:H4 outbreak. EMBO Mol Med. 2012; 4(9):841–8.
12. Jandhyala DM, Vanguri V, Boll EJ, Lai Y, McCormick BA, Leong JM. Shiga toxin-producing Escherichia coli O104:H4: an emerging pathogen with enhanced virulence. Infect Dis Clin N Am. 2013;27(3):631–49.
13. Brandal LT, Lindstedt BA, Aas L, Stavnes TL, Lassen J, Kapperud G. Octaplex PCR and fluorescence-based capillary electrophoresis for identification of human diarrheagenic Escherichia coli and Shigella spp. J Microbiol Methods. 2007;68(2):331–41.
14. Brandal LT, Sekse C, Lindstedt BA, Sunde M, Lobersli I, Urdahl AM, Kapperud G. Norwegian sheep are an important reservoir for human-pathogenic Escherichia coli O26:H11. Appl Environ Microbiol. 2012;78(12):4083–91.
15. Brandal LT, Wester AL, Lange H, Lobersli I, Lindstedt BA, Vold L, Kapperud G. Shiga toxin-producing escherichia coli infections in Norway, 1992-2012: characterization of isolates and identification of risk factors for haemolytic uremic syndrome. BMC Infect Dis. 2015;15:324.
16. Doumith M, Day MJ, Hope R, Wain J, Woodford N. Improved multiplex PCR strategy for rapid assignment of the four major Escherichia coli phylogenetic groups. J Clin Microbiol. 2012;50(9):3108–10.
17. Clermont O, Bonacorsi S, Bingen E. Rapid and simple determination of the Escherichia coli phylogenetic group. Appl Environ Microbiol. 2000;66(10):4555–8.
18. Lobersli I, Haugum K, Lindstedt BA. Rapid and high resolution genotyping of all Escherichia coli serotypes using 10 genomic repeat-containing loci. J Microbiol Methods. 2012;88(1):134–9.
19. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 2014;30(15):2114–20.
20. Watson M, Thomson M, Risse J, Talbot R, Santoyo-Lopez J, Gharbi K, Blaxter M. poRe: an R package for the visualization and analysis of nanopore sequencing data. Bioinformatics. 2015;31(1):114–5.
21. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol. 2012;19(5):455–77.
22. Angiuoli SV, Gussman A, Klimke W, Cochrane G, Field D, Garrity G, Kodira CD, Kyrpides N, Madupu R, Markowitz V, et al. Toward an online repository of standard operating procedures (SOPs) for (meta)genomic annotation. OMICS. 2008;12(2):137–41.
23. Van Domselaar GH, Stothard P, Shrivastava S, Cruz JA, Guo A, Dong X, Lu P, Szafron D, Greiner R, Wishart DS. BASys: a web server for automated bacterial genome annotation. Nucleic Acids Res. 2005;33(Web Server issue):W455–9.
24. Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA, Formsma K, Gerdes S, Glass EM, Kubal M, et al. The RAST server: rapid annotations using subsystems technology. BMC Genomics. 2008;9:75.
25. Seemann T. Prokka: rapid prokaryotic genome annotation. Bioinformatics. 2014;30(14):2068–9.
26. Arndt D, Grant JR, Marcu A, Sajed T, Pon A, Liang Y, Wishart DS. PHASTER: a better, faster version of the PHAST phage search tool. Nucleic Acids Res. 2016;44(W1):W16–21.
27. Bertelli C, Laird MR, Williams KP, Simon Fraser University Research Computing G, Lau BY, Hoad G, Winsor GL, Brinkman FS. IslandViewer 4: expanded prediction of genomic islands for larger-scale datasets. Nucleic Acids Res. 2017;45(W1):W30–5.
28. Tapader R, Bose D, Pal A. YghJ, the secreted metalloprotease of pathogenic E. coli induces hemorrhagic fluid accumulation in mouse ileal loop. Microb Pathog. 2017;105:96–9.
29. Culham DE, Wood JM. An Escherichia coli reference collection group B2- and uropathogen-associated polymorphism in the rpoS-mutS region of the E. coli chromosome. J Bacteriol. 2000;182(21):6272–6.
30. Ewers C, Dematheis F, Singamaneni HD, Nandanwar N, Fruth A, Diehl I, Semmler T, Wieler LH. Correlation between the genomic o454-nlpD region polymorphisms, virulence gene equipment and phylogenetic group of extraintestinal Escherichia coli (ExPEC) enables pathotyping irrespective of host, disease and source of isolation. Gut Pathog. 2014;6:37.
31. Croxen MA, Law RJ, Scholz R, Keeney KM, Wlodarska M, Finlay BB. Recent advances in understanding enteric pathogenic Escherichia coli. Clin Microbiol Rev. 2013;26(4):822–80.
32. Toval F, Kohler CD, Vogel U, Wagenlehner F, Mellmann A, Fruth A, Schmidt MA, Karch H, Bielaszewska M, Dobrindt U. Characterization of Escherichia coli isolates from hospital inpatients or outpatients with urinary tract infection. J Clin Microbiol. 2014;52(2):407–18.
33. Huang SH, Wan ZS, Chen YH, Jong AY, Kim KS. Further characterization of Escherichia coli brain microvascular endothelial cell invasion gene ibeA by deletion, complementation, and protein expression. J Infect Dis. 2001;183(7): 1071–8.
34. Cieza RJ, Hu J, Ross BN, Sbrana E, Torres AG. The IbeA invasin of adherent-invasive Escherichia coli mediates interaction with intestinal epithelia and macrophages. Infect Immun. 2015;83(5):1904–18.
35. Germon P, Chen YH, He L, Blanco JE, Bree A, Schouler C, Huang SH, Moulin-Schouleur M. ibeA, a virulence factor of avian pathogenic Escherichia coli. Microbiology. 2005;151(Pt 4):1179–86.
36. Totsika M, Wells TJ, Beloin C, Valle J, Allsopp LP, King NP, Ghigo JM, Schembri MA. Molecular characterization of the EhaG and UpaG trimeric autotransporter proteins from pathogenic Escherichia coli. Appl Environ Microbiol. 2012;78(7):2179–89.
37. Zude I, Leimbach A, Dobrindt U. Prevalence of autotransporters in Escherichia coli: what is the impact of phylogeny and pathotype? Int J Med Microbiol. 2014;304(3–4):243–56.
38. Clermont O, Christenson JK, Denamur E, Gordon DM. The Clermont Escherichia coli phylo-typing method revisited: improvement of specificity and detection of new phylo-groups. Environ Microbiol Rep. 2013;5(1):58–65.
39. Beghain J, Bridier-Nahmias A, Le Nagard H, Denamur E, Clermont O. ClermonTyping: an easy-to-use and accurate in silico method for Escherichia genus strain phylotyping. Microb Genom. 2018;4(7). https://doi.org/10.1099/mgen.0.000192.
40. Tapader R, Bose D, Basu P, Mondal M, Mondal A, Chatterjee NS, Dutta P, Basu S, Bhadra RK, Pal A. Role in proinflammatory response of YghJ, a secreted metalloprotease from neonatal septicemic Escherichia coli. Int J Med Microbiol. 2016;306(7):554–65.
41. Spurbeck RR, Dinh PC Jr, Walk ST, Stapleton AE, Hooton TM, Nolan LK, Kim KS, Johnson JR, Mobley HL. Escherichia coli isolates that carry vat, fyuA, chuA, and yfcV efficiently colonize the urinary tract. Infect Immun. 2012;80(12):4115–22.
42. Roer L, Tchesnokova V, Allesoe R, Muradova M, Chattopadhyay S, Ahrenfeldt J, Thomsen MCF, Lund O, Hansen F, Hammerum AM, et al. Development of a web tool for Escherichia coli subtyping based on fimH alleles. J Clin Microbiol. 2017;55(8):2538–43.

43. Falnes PO, Klungland A, Alseth I. Repair of methyl lesions in DNA and RNA by oxidative demethylation. Neuroscience. 2007;145(4):1222–32.
44. Afset JE, Anderssen E, Bruant G, Harel J, Wieler L, Bergh K. Phylogenetic backgrounds and virulence profiles of atypical enteropathogenic Escherichia coli strains from a case-control study using multilocus sequence typing and DNA microarray analysis. J Clin Microbiol. 2008;46(7):2280–90.
45. Wirth T, Falush D, Lan RT, Colles F, Mensa P, Wieler LH, Karch H, Reeves PR, Maiden MCJ, Ochman H, et al. Sex and virulence in Escherichia coli: an evolutionary perspective. Mol Microbiol. 2006;60(5):1136–51.

Paper II

Whole genome sequencing and characterization of Multidrug-resistant (MDR) isolates from Norwegian environmental sources

Misti D. Finton[a]*, Roger Meisal[a,c], Davide Porcellato[a], Lin T. Brandal[b], and Bjørn-Arne Lindstedt[a]

[a] Faculty of Chemistry, Biotechnology and Food Science, Norwegian University of Life Sciences, Ås, Norway

[b] Department of Zoonotic, Food- and Waterborne Infections, Norwegian Institute of Public Health, Oslo, Norway

[c] Present address: Møreforsking Ålesund AS, Ålesund, Norway

E-mail addresses:

Misti D. Finton: misti.dawn.finton@nmbu.no

Bjørn-Arne Lindstedt: bjorn-arne.lindstedt@nmbu.no

Roger Meisal: roger.meisal@moreforsk.no

Davide Porcellato: davide.porcellato@nmbu.no

Lin T. Brandal: LinCathrineT.Brandal@fhi.no

* Corresponding author:

Misti D. Finton

Address: Norwegian University of Life Sciences

P.O. Box 5003

N-1432 Ås

Norway

Abstract

The presence of extended-spectrum β-lactamase (ESBL)-producing bacteria in environmental sources has been reported worldwide and constitutes a serious risk of community-acquired infections with limited treatment options. The current study aimed to explore the presence of ESBL-producing bacteria obtained from aquatic locations at the Norwegian University of Life Sciences in Ås, Norway. A total of 98 bacterial isolates survived growth on selective chromogenic media and were identified by 16S rRNA Sanger sequencing. All strains were evaluated for the presence of β-lactamases (*bla*$_{TEM}$, *bla*$_{CMY}$, *bla*$_{CTX-M}$, and *bla*$_{SHV}$) and carbapenemases (*bla*$_{OXA-48}$, *bla*$_{IMP}$, *bla*$_{VIM}$, *bla*$_{NDM}$, *bla*$_{SFC1}$, *bla*$_{KPC}$) through multiplex PCR. Overall, 8 ESBL-producing bacterial strains were detected. Phenotypic resistance to 18 antimicrobial agents was assessed and isolates were subjected to whole genome sequencing through a combination of Oxford Nanopore's MinION and Illumina's MiSeq. Results revealed the presence of ESBL-producing *Escherichia coli*, *Klebsiella pneumoniae*, *Stenotrophomonas maltophilia* and one *Paraburkholderia* spp. Detected β-lactamases included *bla*$_{CTX-M}$, *bla*$_{TEM}$, *bla*$_{CMY}$, *bla*$_{SHV}$ and a possible *bla*$_{KPC}$-like gene, with both documented and novel sequences established. The metallo-β-lactamases *bla*$_{L1}$ and *bla*$_{L2}$ were also detected, which confer resistance to the crucial last resort antibiotics, the carbapenems. All strains were determined to be multidrug resistant and numerous virulence factors and resistance genes to non-β-lactams were observed. In conclusion, this study demonstrates that environmental sources are a potential reservoir of clinically relevant ESBL-producing bacteria that may pose a health risk to humans upon exposure.

Keywords: Extended-spectrum β-lactamase (ESBL), carbapenemase, multidrug resistant, heteropathogenic *Escherichia coli*, whole genome sequencing, MinION

1. Introduction

Antibiotic resistant bacteria is a global public health threat that jeopardizes the successful treatment of infectious disease (WHO, 2014). Currently, the most widely used class of drugs in human and veterinary medicine are the β-lactams, including 3rd generation cephalosporins and carbapenems (Bush and Bradford, 2016). Due to persistent exposure of bacterial strains to these antibiotics, extended-spectrum β-lactamases (ESBLs) have evolved to be one of the most clinically significant resistance mechanisms associated with limited therapeutic options (Reeba *et al.*, 2019). ESBLs inactivate β-lactams by hydrolyzing their β-lactam ring, rendering these vital antibiotics ineffective (Pitout *et al.*, 2005). The predominant ESBLs detected in clinical settings are derived from *bla*$_{\text{CTX-M}}$, *bla*$_{\text{TEM}}$, and *bla*$_{\text{SHV}}$ (Bora *et al.*, 2014.) and are most frequently identified in *Enterobacteriaceae* (Pitout *et al.*, 2005). ESBL-producing bacteria also commonly exhibit co-resistance to many other classes of antibiotics, further impeding the successful treatment of bacterial infections (Chaudhary and Aggarwal, 2004).

Although ESBL-producing bacteria have traditionally been considered a clinical problem, their presence in the environment has gained attention as an exposure route to humans through food items, drinking water, and direct contact with water bodies. Previous studies have identified aquatic ecosystems as “hot spots” for microorganisms from a variety of sources as well as the acquisition, evolution, and dissemination of ARB (Baquero, Martínez, & Cantón, 2008), (Machado & Bordalo, 2014), (Tokajian *et al.*, 2018). In addition, the added selective pressures of antimicrobials, biocides, heavy metals, and disinfectants promote the selection of bacterial defense mechanisms that uphold their survival and spread (Baquero, Martínez, & Cantón., 2008), (Tokajian *et al.*, 2018), (Marti, Variatza, & Balcazar, 2014).

Previous studies have typically focused on the detection of specific types of ESBL-producers in the environment, namely *Enterobacteriaceae* (Zurfluh *et al.*, 2013) (Zarfel *et al.*, 2017), (Kittinger *et al.*, 2016), (Jang *et al.*, 2013). However, ESBLs are commonly located on plasmids that are readily transferrable between a variety of bacterial species (Jacoby, 1997). Therefore, the current study aimed to explore the presence of diverse types of β-lactamase and carbapenemase-producing bacteria in water sources of the Norwegian environment. Furthermore, we assessed resistance determinants, virulence factors (VFs), and phylogenetic traits characteristic of pathogenicity through whole genome sequencing (WGS).

2. MATERIALS AND METHODS

2.1 Sampling sites and collection. Between August 2017 and January 2019, a total of 29 water samples were obtained from the campus pond ("Andedammen") at the Norwegian University of Life Sciences in Ås, Norway (Figure 1). Samples were collected in sterile Pyrex® wide mouth storage bottles. All samples were stored at 4°C until analysis. Water quality testing was performed using an EPA-approved commercial kit, Colilert, which simultaneously detects and quantifies both total coliforms and *E. coli* (IDEXX Laboratories, Inc., Maine 04092 USA).

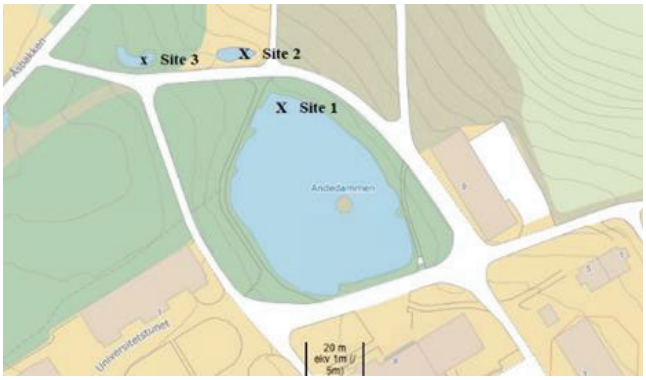

Figure 1. Map showing soil and water sampling sites selected on the NMBU campus where 98 water samples were collected for screening for ESBL and carbapenemase-producing bacterial isolates. Sample site 1 (59°40'02.6"N 10°46'09.0"E), sample site 2 (59°40'03.2"N 10°46'08.2"E), and sample site 3 (59°40'03.2"N 10°46'06.4"E).

2.2 Bacterial isolation. Samples were coarsely filtered through Whatman® Quantitative filter paper (589/1, black ribbon, 15 cm diameter). The filtered water from each site was divided into 100 ml portions in separate sterile glass bottles and vacuum-filtrated through EZ-Pak® filters with a pore size of 12-25 µm (Merck, Darmstadt, Germany) on a Millipore Microfil Support Frit (Merck). The filter membranes were then transferred onto Brilliance^{TM} ESBL agar and Brilliance CRE^{TM} agar (Oxoid) with a sterile tweezer and incubated for 24 hours at 37°C. Per the manufacturer's recommendations, plates that displayed weak growth were incubated for an additional 24 hours and re-assessed. Single colonies with phenotypic differences were sub-cultured onto fresh plates and incubated at 37°C for approximately 24 hours. To verify survival on the selective media, isolates were sub-cultured once more in the same manner.

2.3 DNA extraction and identification. Genomic DNA (gDNA) was extracted with the GenElute Bacterial Genomic DNA Kit (Sigma-Aldrich, St. Louis, USA). The kit's elution buffer contains EDTA and was therefore substituted with the SequalPrep Normalization Elution Buffer

(Oxoid). All other manufacturer's instructions were followed. DNA was assessed for purity and quantity with Nanodrop (Oxoid) and Qubit® 2.0 (Life Technologies, Grand Island, NY), respectively.

Species identification was determined through amplification of the 16S rRNA gene using the primers listed in Table 1. The reaction was conducted with 0.02 U/µl, iProof High-Fidelity DNA Polymerase (Bio-rad Laboratories Inc., Hercules, CA, USA), 1X iProof Buffer, 0.25 µM of each primer, 200 nM dNTPs (Oxoid), and sterile H_2O. All PCR reactions were performed for 35 cycles in a 40 µl final reaction mixture. The amplification conditions were as follows: 98°C for 30 seconds, followed by 35 cycles of 98°C for 15 seconds, 53°C for 30 seconds, 72°C for 20 seconds, followed by 72°C for 10 minutes.

2.4 Detection of β-lactamase genes. All isolates were screened for the most common genes that confer resistance to β-lactams (*bla*$_{TEM}$, *bla*$_{CTX-M}$ groups 1, 2 and 9, *bla*$_{SHV}$ and *bla*$_{OXA}$) and carbapenems (*bla*$_{CMY}$, *bla*$_{OXA-48}$, *bla*$_{IMP}$, *bla*$_{VIM}$, *bla*$_{NDM}$, *bla*$_{SFC1}$, *bla*$_{KPC}$) via multiplex PCR (Tables 1 and 2) using DNA purified as described above. DNA amplification was performed in an Applied Biosystems^{TM} SimpliAmp Thermal cycler (Applied Biosystems, MA, USA) in a 25 µL final reaction mixture and amplified products were analyzed on a 1% agarose gel at 80 V-cm for 80 minutes. For all positive results, the process was repeated with individual primer pairs in separate reactions. The amplification conditions were as follows: 95°C for 15 minutes, followed by 28 cycles (ESBL reaction) or 30 cycles (carbapenemase reaction) of 94°C for 30 seconds, 60°C (carbapenemase) or 62°C (ESBL) for 90 seconds, 72°C for 90 seconds and 72°C for 10 minutes.

ESBL positive and 16S rRNA PCR products were purified with the GenElute^{TM} PCR Clean-up Kit (Sigma-Aldrich) and quantified with Qubit 2.0 (Life Technologies) following manufacturer's instructions. Sanger sequencing was performed by GATC Biotech (GATC, Konstanz, Germany), and the nucleotide sequences were assessed with online similarity searches performed with the Basic Local Alignment Search Tool (BLAST). Isolates were selected for further testing based on the presence of ESBLs or an interest in the bacterial species.

2.5 Antimicrobial Susceptibility Testing. M.I.C.Evaluator (Oxoid), ETEST® (bioMérieux, Marcy l'Étoile, France), or MIC (Liofilchem, Roseto degli Abruzzi, Italy) gradient strips were used to evaluate susceptibility. The following panel of 18 antibiotics spanning 13 classes were used: ampicillin and amoxicillin/clavulanic acid (penicillins); cefotaxime and cefepime (cephalosporins); ciprofloxacin (fluoroquinolones); amikacin, gentamicin, and streptomycin (aminoglycosides); trimethoprim (trimethoprims); trimethoprim/sulfamethoxazole (trimethoprim/sulfonamides); erythromycin (macrolides); tetracycline (tetracyclines); nitrofurantoin (nitrofurans); bacitracin (bacitracins); fosfomycin (fosfomycins); and imipenem and meropenem (carbapenems). Susceptibility profiles to colistin (polymyxins) was determined through the broth micro-dilution method using ComASPTM Colistin (Liofilchem). All testing was performed in accordance to the manufacturer's instructions. MIC clinical breakpoints from the European Committee on Antimicrobial Susceptibility Testing (EUCAST) were used to determine susceptibility (EUCAST, 2019). Multidrug resistance was defined as phenotypic resistance to three or more classes of antibiotics.

153 **Table 1.** Primers used for the screening of main ESBL genes.

Target genes	Primer Sequence (5'-3')	Amplicon Size (bp)	Reference
Multiplex 1			
bla$_{OXA-48}$	F- GCTTGATCGCCCTCGATT R- GATTTGCTCCGTGGCCGAAA	281	(Dallenne *et al.*, 2010)
bla$_{CTX-M}$ (gr. 2)	F- CGTTAACGGCACGATGAC R- CGATATCGTTGGTGGTTCCAT	404	(Dallenne *et al.*, 2010)
bla$_{OXA}$	F- GGCACCAGATTCAACTTTCAAG R- GACCCCAAGTTTCCTGTAAGTG	564	(Dallenne *et al.*, 2010)
bla$_{SHV}$	F- AGCCGCTTGAGCAAATTAAAC R- ATCCCGCAGATAAATCACCAC	713	(Dallenne *et al.*, 2010)
Multiplex 2			
bla$_{CTX-M}$ (gr. 9)	F- TCAAGCCTGCCGATCTGGT R- TGATTCTCGCCGCTGAAG	561	(Dallenne *et al.*, 2010)
bla$_{CTX-M}$ (gr. 1)	F- TTAGGAARTGTGCCGCTGYA R- CGATATCGTTGGTGGTRCCAT	688	(Dallenne *et al.*, 2010)
bla$_{TEM}$	F- CATTTCCGTGTCGCCCTTATTC R- CGTTCATCCATAGTTGCCTGAC	800	(Dallenne *et al.*, 2010)
Multiplex 3			
bla$_{NDM}$	F- TGGCCCGCTCAAGGTATTTT R- GTAGTGCTCAGTGTCGGCAT	157	This study
bla$_{VIM}$	F- ATAGAGCTCAGTGTGTCGGCAT R- TTATTGGTCTATTTGACCGCGT	564	This study
bla$_{KPC}$	F- TCCGTTACGGCAAAAATGCG R- GCATAGTCATTTGCCGTGCC	460	This study
Multiplex 4			
rpoB	F- CAGGTCGTCACACGGTAACAAG R- GTGGTTCAGTTTCAGCATGTAC	512	Universal primers
16S rRNA	F- AGAGTTTGATCMTGGCTCAG R- GYTACCTTGTTACGACTT	1505	Universal primers

154

155 **Table 2.** Primers used for screening of main carbapenemase genes.

Target genes	Primer Sequence (5'-3')	Amplicon Size (bp)	Reference
Multiplex 1			
bla$_{CMY}$	F- GCATCTCCCAGCCTAATCCC R- TTCTCCGGGACAACTTGACG	188	This study
bla$_{OXA-48}$	F- GCTTGATCGCCCTCGATT R- GATTTGCTCCGTGGCCGAAA	281	(Dallenne *et al.*, 2010)
bla$_{IMP}$	F- ACAGGGGGAATAGAGTGGCT R- AGCCTGTTCCCATGTACGTT	393	This study
bla$_{VIM}$	F- ATAGAGCACACTCGCAGACG R- TTATTGGTCTATTTGACCGCGT	564	This study
Multiplex 2			
bla$_{NMD}$	F- TGGCCCGCTCAAGGTATTTT R- GTAGTGCTCAGTGTCGGCAT	157	This study
bla$_{SFC}$	F- GGAGGGCGAATTGGGGTTTA R- CACTGTACTGCAGAGTGGCA	268	This study
bla$_{KPC}$	F- TCCGTTACGGCAAAAATGCG R- GCATAGTCATTTGCCGTGCC	460	This study

156

2.6 Oxford Nanopore MinION sequencing. DNA library was prepared using the SQK-RBK004 Rapid Barcoding Kit and loaded into the MinION SpotON R9.4.1 flow cell in a MinION MK1 sequencer. The 48-h (MinKNOW v.19.06.8 and older versions) or 72-h (MinKNOW 19.10.1) live basecalling sequencing protocol was selected in the MinKNOW software and allowed to complete basecalling after the completion of the sequencing run. With MinKNOW v.19.06.8 and older, EPI2ME was used to demultiplex the barcodes. With MinKNOW v.19.10.1 and after, barcode demultiplexing was conducted during the run while basecalling. Individual barcodes were examined through the WIMP-ARMA applications to identify resistance genes and confirm the species and of each isolate. All runs were according to the standard protocol of Oxford Nanopore Technologies (Oxford, UK).

2.7 Illumina MiSeq sequencing and assembly. Paired-end libraries (2 x 300 bp) were prepared with Nextera™ DNA Flex Tagmentation (Illumina Inc., San Diego, CA, USA) and sequenced on an Illumina MiSeq platform using a 2 x 300 paired-end approach with v3 chemistry by the Norwegian Sequencing Center (Oslo, Norway).

2.8 Sequence Analysis. All FastQ reads within each MinION barcode was concatenated using the Cygwin cat command. The online Galaxy platform was utilized to remove adapters from the MiSeq and MinION data with Trimmomatic and Porechop, respectively. MinION data was filtered by sequence length with a minimum threshold set at 800bp and above. MiSeq and MinION data were combined in Unicycler to assemble sequences into contigs. Additionally, the sequence data was annotated using the NCBI Prokaryotic Genome Annotation Pipeline (Angiuoli *et al.*, 2008). To assess the genomes for acquired antibiotic resistance genes and VFs,

ResFinder v 3.2 and VirulenceFinder v 2.0 (Center for Genomic Epidemiology, Technical University of Denmark, Lyngby, Denmark) servers were used with the following settings: selected ID threshold 95.0%, selected minimum length 60%. The ABRicate Mass screening of contigs for antimicrobial and virulence genes (Galaxy Version 0.9.8) tool combined with the NCBI National Database of Antibiotic Resistant Organisms were also used. Additionally, the Comprehensive Antibiotic Resistance Database (CARD) was used to search the genome for acquired antibiotic resistance genes. MLST v 2.0.4, PlasmidFinder v 2.1, and SerotypeFinder 1.1 (Center for Genomic Epidemiology) were used with default settings to determine MLST type, plasmid types, and serotypes of the isolates, with the exception of strain NMBU_R2 in which the online database PubMLST (www.pubmlst.org) (Jolley *et al.*, 2018) was used for MLST results. Phylogenetic groups were determined using the publicly available ClermonTyper (http://clermontyping.iame-research.center/). In addition, a custom-made virulence database of over 760 genes was employed to assess for VFs with the CGE website. Upon analysis of WGS results, it was determined that NMBU_W06E18 was two separate strains rather than a single strain. In this case, MetaSPades (Nurk *et al.*, 2017), MaxBin2 (Wu *et al.*, 2016), Unicycler (all from the Galaxy platform), and NCBI Blast were used to separate the strains.

3. RESULTS

3.1 Isolation and characterization of bacteria. Among the 98 environmental isolates that survived growth on Brilliance^{TM} ESBL or Brilliance^{TM} CRE agar, ESBLs were detected in 8 (8.2%) isolates. Species identification revealed six isolates within *Enterobacteriaceae* including five *Escherichia coli* (*E. coli*) strains and one *Klebsiella pneumoniae* (*K. pneumoniae*) strain.

Two additional strains were identified within *Xanthomonadaceae* (*Stenotrophomonas maltophilia*) and *Burkholderiaceae* (*Paraburkholderia* spp.).

The phylogenetic analysis of the five *E. coli* isolates revealed that the commensal-associated group A and the pathogenic-associated group D were represented. One strain was classified as phylogroup A and ST1286 (Table 3). Four strains were classified as phylogroup D and were of three different ST types: ST405 (n = 2), ST69 and ST38. The *K. pneumoniae* isolate was classified as ST659 and *S. maltophilia* as ST31. The ST of the *Paraburkholderia* spp. could not be determined.

Table 3. Genotypic characteristics of ESBL-producing isolates, isolation date, site, species, serotype and ST
.

	Isolation month/year	**Site**	**Species**	**Serotype**	**Phylo.**	**ST**	**ESBLs**	**Plasmid incompatibility groups**
NMBU_W05E18	February 2018	2	*E. coli*	O86:H18	D	38	CTX-M-15, TEM-1B, CMY-42-like *bla*AMPH *bla*EC-8	Col156, IncFIB(AP001918), IncFII
NMBU_W06E18_Strain1	February 2018	3	*E. coli*	O102:H6	D	405	TEM-84 CMY-42 *bla*EC	IncFII, IncI, IncX4
NMBU-W07E18	February 2018	2	*K. pneumoniae*	O2:K107	-	659	CTX-M-15, TEM-1B, SHV-11 *bla*AMPH	IncFIB(AP001918), IncFII
NMBU_W010C18	February 2018	2	*E. coli*	O102:H6	D	405	TEM-1B, CMY-42 *bla*EC *bla*AMPH	IncFIA, IncI, IncX4
NMBU_W12E19	January 2019	1	*E. coli*	O10:H32	A	1286	CTX-M-15 *bla*EC-8	IncFII, Incl1-Iγ, Incl2
NMBU_W13E19	January 2019	1	*E. coli*	O25:H18	D	69	CTX-M-15 *bla*EC-8	Col(BS512), IncFIA, IncFIB(AP001918), IncFIC, IncFII, IncI, IncX4, IncX5, p0111
NMBU_R2	December 2018	2	*S. maltophilia*	ND	-	31	L1, L2	ND
NMBU_R16	August 2018	2	*Paraburkholderia spp.*	ND	-	ND	CTX-M-like KPC-like	ND

3.2 Antibiotic Susceptibility Testing. The *Enterobacteriaceae* isolates showed a range of susceptibility profiles towards 18 antibiotics (supplementary Table 1). All strains were determined to be phenotypically resistant against ampicillin, bacitracin, cefotaxime, cefepime and erythromycin. Resistance to amoxicillin with clavulanic acid (3, 50.0%), ciprofloxacin (3, 50.0%), tetracycline (2, 33.0%), trimethoprim (2, 33.0%), fosfomycin (1, 16.7%), and nitrofurantoin (1, 16.7%) was also observed. No *Enterobacteriaceae* strains exhibited phenotypic resistance to amikacin, gentamicin, streptomycin, imipenem, meropenem, colistin, or trimethoprim/sulfamethoxazole. It is notable that all of the ESBL-producing *Enterobacteriaceae* strains were sensitive to carbapenems and colistin, as these are the last resort antibiotics for ESBL-producing bacterial infections in humans. Nevertheless, all *Enterobacteriaceae* isolates were classified as multidrug-resistant. It is important to note that the WGS results of NMBU_W06E18 indicated that this was a mixed sample of two strains of *E. coli* during susceptibility testing.

Although *Stenotrophomonas maltophilia* (*S. maltophilia*) had high MIC values against many of the tested antibiotics, results for agents other than trimethoprim-sulfamethoxazole should be treated with caution, as EUCAST has not determined clinical breakpoint values for this particular type of bacteria. Therefore, the relationship between various susceptibility testing results and the clinical outcome for *S. maltophilia* infection cannot be confidently determined (Gülmez, 2010). However, the MIC values for a variety of antibiotics (streptomycin, imipenem, meropenem, cefotaxime, cefepime, ampicillin, amoxicillin with clavulanic acid, bacitracin, nitrofurantoin, and trimethoprim) exceeded the testable limits, and treatment levels would likely not approach those MIC values. Therefore, this strain was characterized as multidrug resistant. In addition, the

Paraburkholderia spp. had difficulties growing, thus phenotypic antibiotic susceptibility could not be determined, and additional results that correspond to this strain are limited.

3.3 Detection of β-lactamases and co-resistance genes.

Enterobacteriaceae. Data on β-lactamases and co-resistance genes for all isolates is presented in Tables 3 and 4. All *Enterobacteriaceae* harbored at least one β-lactamase gene, with $bla_{\text{CTX-M-15}}$ (4, 66.7%) as the most prevalent variant, followed by $bla_{\text{TEM-1B}}$ (3, 50.0%). Other detected types were $bla_{\text{CMY-42}}$ (2, 33.3%) and $bla_{\text{TEM-84}}$ (1, 16.7%). All isolates harbored 2 or more β-lactamase genes. A novel $bla_{\text{CMY-42-like}}$ gene was documented, which may be indicative of a high diversity of β-lactamase genes in the pond. A chromosomally located $bla_{\text{CTX-M-15}}$ was identified in strain NMBU_W05E18. Chromosomal integration of usually plasmid encoded ESBL genes in *E. coli* is a rather new development (Rodríguez *et al.*, 2014) (Hirai I *et al.*, 2013). All of the *E. coli* strains harbored a genetic region containing an AmpC/CMY-like gene, *blc* and *sugE* as previously described (Verdet *et al.*, 2009) (Singh *et al.*, 2018). In strain NMBU-W05E18 this region is situated both on the chromosome and on a plasmid where the NCBI annotation pipeline reports a frameshifted CMY gene, however on closer inspection a full ORF can be reconstructed by selecting an upstream initiator methionine. This translated ORF shows 99% amino acid identity with CMY-42 with two AA substitutions at codon 303 (Ser303ile) and at codon 318 (Thr318ala). No identical proteins were found on BLAST searches; therefore, this may constitute a novel plasmid-borne CMY-variant (supplementary file 1). Additionally, no identical proteins were found on BLAST searches for the KPC-like and CTX-M-like genes from NMBU_R16 and are therefore novel and distinctly different from the other known β -lactamases.

A wide variety of resistance mechanisms to non-β lactam drugs were identified amongst the *Enterobacteriaceae* isolates, including determinants against aminoglycosides [*AAC3-lb*]; bacitracin (*bacA*); bicyclomycin resistance (*bcr*); chloramphenicol (*catb4*); fluoroquinolones (*oqxA, oqxB, qnrS1*); fosfomycin (*fosA*); fosmidomycin (*fsr*); macrolides (*macA*, *mphA*, Mrx); sulfonamides (*sul1*); tetracyclines [*tet*(*A*), *tet(B)*] and trimethoprims (*dfrA*). Efflux pumps that confer resistance to multiple classes of antibiotics were also observed (*acrAB*, *emrAB, mdtH, mdfA*). In addition, several *E. coli* isolates contained chromosomal mutations in *gyrA* (S83L or D87N), *parC* (S80I, E84K), and *parE* (S458A) that confer fluoroquinolone resistance (Huseby *et al.*, 2017) (Moon *et al.*, 2010). Mutations in *folP* (F4Y, A247T) that confer resistance to sulfonamides were also observed (Buwembo *et al.*, 2013). It is notable that no acquired genes encoding carbapenemases or resistance to colistin were found.

Resistance genes against metals were identified, including arsenic (*ArsHR*), cadmium/zinc/lead (*zntA*), copper (*cus*, *copA*), mercury (*merA*), and tellurite (*tehA*). In addition, quaternary ammonium is used as a disinfectant in medical and food environments, and the gene that confers that resistance (*sugE*) was observed throughout the isolates (Sundheim G *et al.*, 1998). The qacEΔ1 resistance gene against biocides was also detected in one isolate. Silver resistance determinants (*silES*) were also observed in the *K. pneumoniae* isolate, which is an emerging public health issue, as silver is commonly used as a disinfectant and preservative in a variety of healthcare and consumer products, for instance silver-treated catheters and wound dressings (Silver *et al.*, 2006). Three of the strains were positive for the CRISPR-Cas bacteriophage adaptive immunity system (NMBU_W05E18, NMBU_W13E19 and NMBU_W10C18) and one strain (NMBU_W05E18) additionally contained the novel bacteriophage exclusion (BREX)

resistance system, which has an important role in innate defense against phages (Goldfarb *et al.*, 2015).

***S. maltophilia*.** The metallo-β-lactamases bla_{L1} and bla_{L2} were identified. Resistance genes towards aminoglycosides [*aph*(*3*'')- *IIc*], bicyclomycin (*bcr*), fluoroquinolones (*oqxB*), and polymyxins (*arnA*) were additionally detected. Resistance genes to metals including arsenic (*acr3*), copper (*copAC*), cobalt/zinc/cadmium (*czcA*), mercury (*merR1*), nickel/cobalt (*cnrA*), and quaternary ammonium (*sugE*) were observed. In addition, efflux pumps associated with resistance to aminoglycosides, β-lactams, quinolones (SmeABC), chloramphenicol, tetracycline, and quinolones (SmeDEF) were detected. The multidrug efflux pumps *emrKY, mdtH*, *mexAB*, *norM* and *bmrA* were also identified. A resistance mechanism against a plant-produced organic hydroperoxide (*ohr*) was located, which allows for an advantage in the interaction of bacterial pathogens over symbionts in plant hosts (Caswell *et al.*, 2012). The persistence and stress resistance toxin, *pasT*, was identified and is documented to have a role in persistence after antibiotic exposure and survival of nitrosative stress.

***Paraburkholderia* spp**. This isolate harbored novel $bla_{CTX\text{-}M\text{-}like}$ and $bla_{KPC\text{-}like}$ genes. As there were no identical proteins found on BLAST searches, these may constitute novel plasmid-borne variants (supplementary file 1). In addition, this strain contained resistance genes towards polymyxins (*pmrK*), arsenic (*arsHR*, *acr3*), chromate resistance (*chrAB*) cobalt/zinc/cadmium (*czcAD*), copper (*copCD*), tellurium resistance (*terC*) and quaternary ammonium (*sugE*). The multidrug efflux pumps *ermAB* and *mdtABC* were also identified. In addition, the 'Bacterial abortive infection' system was observed, which activates cell death upon phage infection thereby limiting viral replication and protecting the bacterial population (Dy *et al.*, 2014).

Table 4. Co-resistance genes identified throughout the isolates. Present resistance determinants are represented by a grey box.

		NMBU_W05E18	NMBU_W06E18_Strain1	NMBU_W010C18	NMBU-W12E19	NMBU-W13E19	NMBU-W07E18	NMBU_R2	NMBU_R16
Aminoglycosides	*AAC3-lb*					■			
	aph(3')- IIc							■	
Fluoroquinolones	*oqxA*						■		
	oqxB						■	■	
	gyrA (p.S83L)	■	■	■		■			
	gyrA (p.D87Y)	■	■	■					
	parC (p.S80I)	■	■	■					
	parC (p.E84K)				■	■			
	parE (S458A)	■	■	■					
	QnrS1				■				
Macrolides	*macA*				■		■		
	mphA		■		■	■			
	Mrx	■			■	■			
Polymyxins	*arnA*							■	
	pmrK								■
Trimethoprim	*dfrA7*	■							
	dfrA14					■			
Sulfonamides	*sul1*	■							
	folP (p. F4Y)			■					
	folP (A247T)								
Tetracyclines	*tet(A)*	■		■	■	■			
	tet(B)				■	■			
Fosfomycin	*fosA*						■		
Fosmidomycin	*fsr*	■	■	■	■	■			
Phenicol	*catb4*				■	■			
Antiseptics	qacEΔ1	■							
Arsenic	*arc3*						■	■	■
	arsHR						■		■
Bacitracin	*bacA*				■	■	■		
Chromate resistance	*chrAB*								■
Cobalt/zinc/cadmium	*czcA*						■	■	■
Cadmium/Zinc/Lead	*zntA*	■	■	■	■	■	■	■	■
Copper	*copA*	■	■	■	■	■			
	copC						■	■	
	copCD								■
	cus						■		
Nickel/cobalt	*cnrA*							■	
Mercury	*merR1*	■				■		■	
	merA	■				■			
Quaternary ammonium	*sugE*	■	■	■	■	■		■	■
Silver	*silES*						■		

Tellurium	*terC*								
Tellurite	*tehA*								
BREX	BREX								
Sulfonamide and bicyclomycin	*bcr*								
Multidrug efflux pumps	*emrKY*								
	mdf(A)								
	bmrA								
	MexAB								
	mdtA								
	mdtH								
	acrAB								
	emrAB								
	norM								
	smeABCDEF								

310

311 **3.4 Detection of Virulence Factors.**

312 ***Enterobacteriaceae.***

313 All isolates showed a wide array of UPEC-associated virulence genes, including iron acquisition
314 systems (*chuA*, *fyuA, iroN*, *irp1*, *irp2*, *iuc*, *iutA, sitA*), protectins (*kpsS, kpsM*), invasins (*cva, iss,*
315 *traT, kpsM*), adhesins (*csg, eilA, dr/afa, fdeC, fimH, iha, pap*), and toxins (*astA, hlyD, sat, vat,*
316 *hbp*) (Supplementary table 2).
317 In a previous study, we found that fecal samples from Norwegian patients with signs of intestinal
318 infection were colonized with a surprisingly high frequency of intestinal pathogenic *E. coli*
319 (IPEC)/extraintestinal pathogenic *E. coli* (ExPEC) heteropathogenic strains (Lindstedt *et al.*,
320 2018). Now, we have isolated multidrug resistant heteropathogens from environmental samples.
321 The *E. coli* surface water strain NMBU_W05E18 [O86:H18 (ST38 – phylogroup D)] contains
322 the *afa* gene cluster that encodes afimbrial adhesins, expressed by both uropathogenic and
323 diarrhea-associated *E. coli* strains and is also associated with mortality due to in bacteraemia
324 infections (Daga *et al.*, 2019). Interestingly, this strain has an *afa* gene cluster both
325 chromosomally and plasmid located. The genomes of *E. coli* strains NMBU_W05E18 and
326 NMBU_W12E19 [O10:H32 (ST1286 – phylogroup A)] contained the *pap* (pyelonephritis-

associated pili) operon associated with UPEC strains. All *E. coli* strains except NMBU_W12E19 harbored *eilA* and all *E. coli* strains except NMBU_W10C18 [O102:H6 (ST405 – phylogroup D)] harbored *ygeH*, which are both *hilA*-like regulator genes, with the *HilA* protein as the master regulator of the *Salmonella* pathogenicity island 1. The *eilA* gene appears to be restricted to *E. coli* 042 and other enteroaggregative strains, while *ygeH* is present in a variety of *E. coli* strains including K12, enterohemorrhagic O157:H7, enteroaggregative hemorrhagic O104:H4, among others (Hüttener *et al.*, 2014).

In all phylogroup D strains of this study, the *eilA* gene is located in close proximity with genes encoding Air (an enteroaggregative immunoglobulin repeat protein), an IpaD/SipD/SspD family (type III secretion system needle tip protein), and SipB (type III cell invasion protein). However, the translated Air protein is different between the strains, where the protein in NMBU_W06E18_Strain1 and NMBU_W10C18 has a length of 3806 AA and a length of 3418 AA and 4485 AA in NMBU_W13E19 and NMBU_W05E18, respectively. The difference in protein length seems to be attributed to different number of bacterial Ig-like domain 1 (BIG-1 domain) units in the Air proteins. The genomes of the phylogroup D strains also contain genes identified as effector proteins of a type III secretion system (T3SS) in enterohemorrhagic *E. coli* (EHEC) and enteropathogenic *E. coli* (EPEC): espX1, espX4, espX5, espY1, espY3 and espY4. The EspY1 and EspY4 proteins have been confirmed as genuine translocated T3SS effectors of EHEC. EspY1 was shown to participate in apoptosis and regulation of cell cycle. It has further been demonstrated that EspY3 is an effector protein translocated by the T3SS of both EHEC O157:H7 and EPEC O127:H6, and that EspY3 localizes in the pedestal region in EPEC (Larzábal *et al.*, 2018). The phylogroup A strain (NMBU_ W12E19) harbored espX1, espX4 and espX5 but espY1, espY3 and espY4 were absent.

The T3SS effector leucine-rich repeat protein EspR1 was present in strains NMBU_ W12E19 and NMBU_ W13E19, while only strain NMBU_W13E19 was positive for the EspR2 protein. Strain NMBU_W13E19 displayed the gene combination *eilA*, Air and *lpfA* (encoding for long polar fimbriae) that was recently reported from a $bla_{\text{CTX-M}}$-producing *E. coli* strain isolated from a wastewater treatment plant in China (Jiang *et al.*, 2019), and in eight food-producing animal isolates (seven turkey and one laying hen) in Poland (Zajac *et al.*, 2019). This particular gene combination was also seen in the first clinical isolate of an *E. coli* harboring the mcr-1 gene in Mexico, isolated in 2017 (Merida-Vieyra *et al.*, 2019), as well as in six ESBL-producing *E. coli* strains isolated from veterinary hospital staff and students in three UK veterinary hospitals (Royden *et al.*, 2019) .

The NMBU_W12E19 strain has an unusual array of virulence genes when compared to typical and atypical EAEC strains. It has the dispersin locus associated with typical EAEC but is also negative for AggR, which is often used to define a typical EAEC strain. NMBU_W12E19 further harbors the enteroaggregative *E. coli* heat-stable enterotoxin 1 (EAST1), which was originally discovered in EAEC but has also been associated with enterotoxigenic *E. coli* (ETEC). The NMBU_W12E19 *astA* gene has 100% nucleotide identity to the previously described *astA*-allele-2 that was proven to express enterotoxic activity (access nr. AF143819) and has the same genetic location on an insertion sequence (IS) element lying entirely within a transposase-like gene (McVeigh *et al.*, 2000). Interestingly, NMBU_W12E19 is also positive for the *bfpA* and *bfpB* genes of the bundle forming pilus (BFP) associated with typical EPEC (tEPEC), which are often used as genetic markers in intestinal pathogenic *E. coli* clinical diagnostic procedures. The genomic sequences of two O157:H7 strains, EDL933 and Sakai, contain a gene cluster predicted to encode an additional T3SS named ETT2, which is involved in virulence (Zhou *et al.*, 2014).

Strain NMBU_W05E18 and NMBU_ W13E19 carry an almost intact ETT2 where only the *eivH* gene is missing. Strain NMBU_ W12E19 carries a deleted ETT2 version where the genes between *epaO* and ECs3736/*pkgA* are deleted, *eivH* is however present. In strains NMBU_W10C18 and NMBU_W06E18_Strain1 almost the entire ETT2 locus is deleted between ECs3737 and *yqeH*.

Klebsiella pneumonia harbored several VFs including those that aid aerobactin transport (*entB*, *iutA*) as well as the enterobactin/salmochelin importer (*fepB*) which is required to establish an infection in iron poor areas of the body, as in the urinary tract or lungs (Palacios *et al.*, 2017). Major adhesive structures type 1 fimbriae (*fimA*) were present and are expressed in the bladder and have been shown to contribute to uropathogenicity (Paczosa and Mecsas, 2016). Type 3 (*mrk*) fimbrial adhesins were also detected and play an important role in adhesion to medical devices, such as in catheters (Stahlhut *et al.*, 2012). An AcrAB efflux pump was also identified and is recognized as a required virulence factor to resist immune defense mechanisms of the lung, thus facilitating the onset of pneumonia (Padilla *et al.*, 2010). In addition, *kvgAS* was detected which has only been found in virulent *K. pneumoniae* CG43 (Lai YC, *et al.*, 2003).

***S. maltophilia*.** This isolate harbored an aerobactin receptor (*fep*), iron receptor (*iroN*), and a gene that aids in the release of iron from other siderophores (*viuB*), which are all imperative to establishing infection extraintestinally. The type II secretion system (*gsp*) was detected that likely plays an important role in pathogenesis of the lungs (Karaba *et al.*, 2013.). In addition, a twitching motility protein (*pilG*) was identified that is known to promote attachment and translocation across host cells (Mattick *et al.*, 1996).

***Paraburkholderia* spp**. No known VFs were detected in this isolate.

3.5 Water quality testing.

The water quality of the "Andedammen" site was tested on October 21st, 2019 using the Colilert test. It showed a most probable number (MPN) of >24196.6 coliforms and 236 *E. coli* cells per 100 mL of water.

3.6 Plasmid identification. The IncFII plasmid type was observed to be the most prevalent (n = 5). Four plasmids of the IncF family were identified: IncFIA, IncFIB, IncFIC, IncFII, and all strains characterized as phylogroup D carried IncFIB and IncFII. IncI1-Iγ and Incl2 plasmids were only found in the Phylogroup A strain (NMBU-W12E19). No plasmids could be identified in the *Paraburkholderia* spp. or *S. maltophilia* strains.

3.7 Nucleotide sequence accession numbers. The sequence data of NMBU_W05E18, NMBU_W07E18, NMBU_W12E19 have been deposited in the GenBank database under accession numbers CP042878 through CP042892 and CP045712 through CP045717. NMBU_W06E18 has been deposited in the Genbank database under accession numbers CP047609 through CP047613. NMBU_R2 and NMBU_R16 was deposited at DDBJ/ENA/GenBank under the accession numbers JAAAYF000000000 and JAAAYE000000000, respectively. The versions described in this paper are version JAAAYF01000000 and JAAAYE010000000.

DISCUSSION

Our study explored the presence of ESBL-producing bacteria in soil and freshwater of Norway, a country with a low prevalence of antibiotic resistance (Brolund, 2014). An array of ESBL-

419 producing microorganisms were found, with *Enterobacteriaceae* as the most predominant, and
420 included strains of *E. coli* and *K. pneumoniae*. ESBL-producing *Enterobacteriaceae* are
421 identified as a serious public health risk and recognized by WHO as pathogens of critical priority
422 amongst MDR bacteria (WHO, 2014), with community-acquired UTIs as the most common
423 infection (Pitout and Laupland, 2008). In addition to *Enterobacteriaceae*, ESBL-producing *S.*
424 *maltophilia* and *B. multivorans* were observed. Both strains have largely been distinguished as
425 nosocomial pathogens that mainly affect immunocompromised individuals (Spencer, 1995),
426 however community-acquired infections have been reported. For instance, *S. maltophilia* has
427 been implicated in lower respiratory tract infections, perceived to be due to recreational activity
428 in aquatic environments (Gajdács and Urbán, 2019). Although there are some pathogenic
429 counterparts of the *Paraburkholderia* lineage, most seem unlikely to cause infection (Eberl and
430 Vandamme, 2016). However, the ESBLs harbored by this strain can transfer to pathogenic
431 counterparts.
432 It is concerning that all tested isolates in our study exhibited a multidrug resistance profile,
433 including to some of the most prescribed antibiotics in clinical settings worldwide. A very high
434 rate of resistance towards ampicillin (100%), bacitracin (100%), erythromycin (100%), and 3rd
435 and 4th generation cephalosporins (100%) was found. Fortunately, no acquired resistance to the
436 last-resort treatment options of carbapenems or colistin was observed. However, a recent study
437 has confirmed the presence of plasmid-mediated colistin-resistant ESBL-producing *E. coli* in a
438 Norwegian public beach (Jørgensen et al, 2017), indicating a risk of dispersal. Another alarming
439 finding was that a majority of our isolates possessed a multitude of virulence traits central to
440 pathogenicity, namely those necessary to cause UTIs. This is important, as previous research has
441 established that swimming in freshwater is a risk factor for community acquired UTIs with

Enterobacteriaceae (Søraas *et al*, 2013). Although the locations in this study are not regularly used for recreational purposes, swimming does occur on occasion. In addition, water uptake by wild or companion animals can disseminate these pathogens to other locations that have a greater extent of human contact. Another interesting finding was the heteropathogenic nature of various *E. coli* strains. The emergence of heteropathogens may lead to serious consequences to public health due to their enhanced virulence from different pathotypes. In our study, the isolates were characterized as DAEC/UPEC, aEAEC/UPEC and therefore have the potential to cause both diarrheal disease and UTIs. However, the majority of the *E. coli* isolates were characterized solely as UPEC, which is in agreement with a previous study that determined ExPEC as the main pathotype of *E. coli* isolates from water sources (Hamelin, 2006). An interesting finding was that 3 *E. coli* strains contained the *fec* locus, making them potential mastitis-causing strains in bovines, a disease that has a significant economic impact on global dairy production (Blum *et al.*, 2018). A worrisome finding was that a serious multidrug resistant strain of similar ST as our multidrug resistant *S. maltophilia* of ST31 was detected from a human blood sample in China (Zhao *et al.*, 2015). Additionally, the ST38, ST69, and ST405 types detected in this study were also observed in human clinical samples throughout Norway (Lin T Brandal *pers. comm*., Enterobase database search), which further connected the clinical relevance of some of our environmental isolates.

The $bla_{\text{CTX-M-15}}$ enzyme was the predominant variant detected in this study, and is currently the most prevalent in clinical isolates (Livermore and Hawkey, 2005), human and animal feces (Guenther *et al.*, 2011), and aquatic environments across the world (Zarfel *et al.*, 2017) (Zurfluh *et al.*, 2013). As in our study, other research has determined that $bla_{\text{CTX-M-15}}$ is strongly associated with human ExPEC strains, particularly UPEC (Mellata, 2013). An exciting result

was the detection of the novel variants *bla*$_{\text{CMY-42}}$-like, *bla*$_{\text{CTX-M-like}}$, and *bla*$_{\text{KPC-like}}$ which highlights the importance of monitoring the evolution of ESBL-producing bacteria. In addition, one isolate contained a chromosomally located *bla*$_{\text{CTX-M-15}}$ (ST38). Previous studies have identified a strong relationship between ST38 and a chromosomally located *bla*$_{\text{CTX-M-15}}$ (Dimou *et al.*, 2012) (Rodríguez *et al.*, 2014) not only in clinical settings (Coque *et al.*, 2008), but in Mongolian wild birds as well (Guenther *et al.*, 2017). As there is abundant wildlife around the sample locations, the wild birds may be a potential explanation for the source of the ESBL-producing isolates in this study. The investigated site "Andedammen" means "duck pond" and is a common gathering locale for wild birds that may have influenced the bacterial community with bird droppings. To supplement that speculation, previous studies have determined that ESBL-producing bacteria from both wild birds and humans are genetically similar (Atterby, 2017) (Bonnedahl, 2010). It is worrisome that the resistance determinants are commonly located on IncF-type plasmids, as they represent a serious risk for the further dissemination of resistance genes. The presence of IncF-type plasmids in a rich reservoir of freshwater increases the likelihood of acquiring additional resistance determinants as well as the evolution of novel resistance determinants (Amos *et al.*, 2014).

4. Conclusion

The present study confirms the existence of ESBL-producing and multidrug resistant bacteria in aquatic environments of Norway, a country with a low incidence of antibiotic resistance. It is also worrisome that many of the isolates contained virulence determinants of major pathogens and STs seen in human patient isolates, indicating the potential to cause serious infections. As humans can be exposed through the food chain or recreational activities, the presence of ESBLs

in the environment pose a critical threat to global public health, and our results provide additional support to this epidemiological association.

Acknowledgements

This work was supported with a grant from the Norwegian Surveillance System for Antimicrobial Drug Resistance (NORM). We are grateful to Senior Engineer Ahmed Abdelghani for performing the water quality test.

Conflict of interest

The authors declare that they have no conflicts of interest.

REFERENCES

1. WHO World Health Organization. Antimicrobial resistance: global report on surveillance (WHO, 2014).
2. Bush K, Bradford, PA. β-Lactams and β-lactamase inhibitors: an overview. Cold Spring Harbor Pers. Med. 2016; **6**(8). doi: 10.1101/cshperspect.a025247
3. Reeba J, Colley P, Nguyen HL, Berhe M. Outcomes analysis in patients with extended-spectrum beta-lactamase bacteremia empirically treated with piperacillin/tazobactam versus carbapenems. Proc (Bayl Univ Med Cent). 2019; **32**(2): 187-191. doi: 10.1080/08998280.2019.1582466
4. Pitout JD, Nordmann P, Laupland KB, Poirel L. Emergence of *Enterobacteriaceae* producing extended-spectrum beta-lactamases (ESBLs) in the community. J Antimicrob Chemother 2005; **56**(1): 52–59. doi:10.1093/jac/dki166
5. Bora A, Hazarika NK, Shukla SK, Prasad KN, Sarma JB, Ahmed G. Prevalence of *bla*$_{TEM}$, *bla*$_{SHV}$ and *bla*$_{CTX-M}$ genes in clinical isolates of *Escherichia coli* and *Klebsiella pneumoniae* from Northeast India. Indian J Pathol Microbiol. 2014; **57**(2):249-254. doi: 10.4103/0377-4929.134698
6. Chaudhary U, Aggarwal R. Extended spectrum-lactamases (ESBL) - an emerging threat to clinical therapeutics. Indian J Med Microbiol [serial online]. 2004 [cited 2019 Dec 05]; **22**(2):75-80. Available from: http://www.ijmm.org/text.asp?2004/22/2/75/8076
7. Baquero F, Martínez JL, Cantón, R. Antibiotics and antibiotic resistance in water environments. Curr Opin in Biotechnol. 2008; **19**(3), 260-265. doi: 10.1016/j.copbio.2008.05.006
8. Machado A, Bordalo AA. Prevalence of antibiotic resistance in bacteria isolated from drinking well water available in Guinea-Bissau (West Africa). Ecotoxicol Environ Saf. 2014; 106:188–194. doi: 10.1016/j.ecoenv.2014.04.037
9. Tokajian S, Moghnieh R, Salloum T, Arabaghian H, Alousi S, Moussa J, Husni R, *et al.* Extended-spectrum β-lactamase-producing *Escherichia coli* in wastewaters and refugee camp in Lebanon. Future microbiology. 2018; **13**(1): 81-95. doi:10.2217/fmb-2017-0093
10. Marti E, Variatza E, Balcazar JL. The Role of aquatic ecosystems as reservoirs of antibiotic resistance. Trends in Microbiology. 2014; **22**(1): P36-41. DOI: https://doi.org/10.1016/j.tim.2013.11.001
11. Zurfluh K, Hächler H, Nüesch-Inderbinen M, Stephan R. Characteristics of extended-spectrum β-lactamase- and carbapenemase-producing Enterobacteriaceae isolates from rivers and lakes in Switzerland. Appl Environ Microbiol. 2013; **79**: 3021-3026. doi: 10.1128/AEM.00054-13
12. Zarfel G, Lipp M, Gürtl E, Folli B, Baumert R, Kittinger C. Troubled water under the bridge: Screening of River Mur water reveals dominance of CTX-M harboring *Escherichia coli* and for the first time an environmental VIM-1 producer in Austria. Science of the Total Environment. 2017; **593-594**: 399-405. doi: 10.1016/j.scitotenv.2017.03.138
13. Kittinger C, Lipp M, Folli B, Kirschner A, Baumert R, Zarfel G, *et al.* Enterobacteriaceae Isolated from the River Danube: Antibiotic Resistances, with a Focus on the Presence of

ESBL and Carbapenemases. PLOS ONE. 2016; **11**(11): e0165820. doi: 10.1371/journal.pone.0165820
14. Jang J, Suh YS, Di DYW, Unnos T, Sadowsky MJ, Hur HG. Pathogenic *Escherichia coli* strains producing extended-spectrum β-lactamases in the Yeongsan River basin of South Korea. Environ Sci Technol. 2013; **47**(2): 1128-36. doi: 10.1021/es303577u
15. Jacoby GA. Extended-spectrum beta-lactamases and other enzymes providing resistance to oxyimino-beta-lactams. Infect Dis Clin North Am. 1997, **11**(4):875-87. https://doi.org/10.1016/S0891-5520(05)70395-0
16. European Committee on Antimicrobial Susceptibility Testing. Clinical breakpoints; Breakpoint table for interpretation of MICs and zone diameters [Internet]. Available from: http://www.eucast.org/clinical_breakpoints/. Accessed September 20, 2019.
17. Dallenne C, Da Costa A, Decré D, Favier C, Arlet G. Development of a set of multiplex PCR assays for the detection of genes encoding important beta-lactamases in Enterobacteriaceae. J Antimicrob Chemother. 2010; **65**:490-495. doi: 10.1093/jac/dkp498
18. Angiuoli SV, Gussman A, Klimke W, Cochrane G, Field D, White O, *et al.* Toward an online repository of Standard Operating Procedures (SOPs) for (meta)genomic annotation. OMICS. 2008; **12**(2):137-141. doi: 10.1089/omi.2008.0017
19. Jolley KA, Bray JE, Maiden MCJ. Open-access bacterial population genomics: BIGSdb software, the PubMLST.org website and their applications. Wellcome Open Res. 2018; **3**(124). doi: 10.12688/wellcomeopenres.14826.1. eCollection 2018
20. Nurk S, Meleshko D, Korobeynikov A, Pevzner PA. metaSPAdes: a new versatile metagenomic assembler. Genome Research 2017; **27**: 824-834. doi: 10.1101/gr.213959
21. Wu YW, Simmons BA, Singer SW. MaxBin 2.0: an automated binning algorithm to recover genomes from multiple metagenomic datasets. Bioinformatics. 2016; **32**(4): 605–607. doi: 10.1093/bioinformatics/btv638
22. Gülmez D, Cakar A, Sener B, Hascelik G. Comparison of different antimicrobial susceptibility testing methods for *Stenotrophomonas maltophilia* and results of synergy testing. J Infect Chemother. 2010; **16**(5): 322-328. doi: 10.1007/s10156-010-0068-2
23. Rodríguez I, Thomas K, Van Essen A, Schink AK, Day M, Guerra B. Chromosomal location of blaCTX-M genes in clinical isolates of *Escherichia coli* from Germany, The Netherlands and the UK. Int J Antimicrob Agents. 2014; 43(6):553-7. doi: 10.1016/j.ijantimicag.2014.02.019
24. Hiari I, Fukui N, Taguchi M, Yamauchi K, Nakamura T, Yamamoto Y, *et al.* Detection of chromosomal blaCTX-M-15 in *Escherichia coli* O25b-B2-ST131 isolates from the Kinki region of Japan. Inter J Antimicrob Agents. 2013; **42**(6): 500-506.
25. Verdet C, Gautier V, Chachaty E, Ronco E, Hidri N, Decré D, Arlet G. Genetic Context of Plasmid-Carried *bla*$_{CMY-2}$-Like Genes in *Enterobacteriaceae*. Antimicrobial Agents and Chemotherapy. 2009; **53**(9): 4002-4006. doi:10.1128/AAC.00753-08
26. Singh NS, Singhal N, Virdi JS. Genetic Environment of *bla*$_{TEM-1}$, *bla*$_{CTX-M-15}$, *bla*$_{CMY-42}$ and Characterization of Integrons of *Escherichia coli* Isolated From an Indian Urban Aquatic Environment. Front. Microbiol. 2018; **9**(382): https://doi.org/10.3389/fmicb.2018.00382

27. Huseby DL, Pietsch F, Brandis G, Garoff L, Tegehall A, Hughes D. Mutation Supply and Relative Fitness Shape the Genotypes of Ciprofloxacin-Resistant *Escherichia coli*. Mol. Biol. Evol. 2017; **34**(5):1029-1039. doi: 10.1093/molbev/msx052
28. Moon DC, Seol SY, Gurung M, Jin JS, Choi CH, Lee JC, *et al.* Emergence of a new mutation and its accumulation in the topoisomerase IV gene confers high levels of resistance to fluoroquinolones in *Escherichia coli* isolates. International Journal of Antimicrobial Agents. 2010; **35**(1): 76-79. doi:10.1016/j.ijantimicag.2009.08.003
29. Buwembo W, Aery S, Rwenyonyi CM, Swedberg G, Kironde F. Point Mutations in the folP Gene Partly Explain Sulfonamide Resistance of *Streptococcus mutans*. International Journal of Microbiology. 2013. http://dx.doi.org/10.1155/2013/367021
30. Sundheim G, Langsrud S, Heir E, Holck AL. Bacterial resistance to disinfectants containing quaternary ammonium compounds. International Biodeterioration & Biodegradation. 1998; **41**(3-4): 235-39. https://doi.org/10.1016/S0964-8305(98)00027-4
31. Silver S, Phung LT, Silver G. Silver as biocides in burn and wound dressings and bacterial resistance to silver compounds. J. Ind. Microbiol. Biotechnol. 2006; **33**(7): 627–634. http://dx.doi.org/10.1007/s10295-006-0139-7
32. Goldfarb T, Sberro H, Weinstock E, Cohen O, Doron S, Sorek R, *et al.* BREX is a novel phage resistance system widespread in microbial genomes. EMBO J. 2015; **34**:169-183. doi: 10.15252/embj.201489455
33. Caswell CC, Baumgartner JE, Martin DW, Roop MR. Characterization of the Organic Hydroperoxide Resistance System of *Brucella abortus* 2308. 2012; **194**(18): 5065-72. doi:10.1128/JB.00873-12
34. Dy RL, Przybilski R, Semeijn K, Salmond GPC, Fineran PC. A widespread bacteriophage abortive infection system functions through a Type IV toxin–antitoxin mechanism. Nucleic Acids Res. 2014; **42**(7):4590-05. doi: 10.1093/nar/gkt1419
35. Lindstedt BA, Finton MD, Porcellato D, Brandal LT. 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 Infect Dis. 2018; **18**(544). doi: 10.1186/s12879-018-3449-2
36. Daga AP, Koga VL, Soncini JGM, de Matos CM, Perugini MRE, Vespero EC, *et al.* *Escherichia coli* Bloodstream Infections in Patients at a University Hospital: Virulence Factors and Clinical Characteristics. Front. Cell. Infect. Microbiol. 2019; **9**:191. doi: 10.3389/fcimb.2019.00191
37. Hüttener M, Dietrich M, Paytubi S, and Juárez A. HilA-like regulators in *Escherichia coli* pathotypes: the YgeH protein from the enteroaggregative strain 042. BMC Microbiol. 2014; **14**(268). doi: 10.1186/s12866-014-0268-5
38. Larzábal M, Da Silva M, Riviere NA, Cataldi AA. Novel Effector Protein EspY3 of Type III Secretion System from Enterohemorrhagic *Escherichia coli* Is Localized in Actin Pedestals. Microorganisms. 2018; **6**(4): E112. doi: 10.3390/microorganisms6040112
39. Jiang X, Cui X, Xu H, Liu W, Tao F, Zheng B, *et al.* Whole Genome Sequencing of Extended-Spectrum Beta-Lactamase (ESBL)-Producing *Escherichia coli* Isolated From a Wastewater Treatment Plant in China. Front. Microbiol. 2019; **10**: 1797. doi: 10.3389/fmicb.2019.01797

40. Zajac M, Sztromwasser P, Bortolaia V, Leekitcharoenphon P, Cavaco LM, Wasyl D, *et al.* Occurrence and Characterization of mcr-1-Positive *Escherichia coli* Isolated From Food-Producing Animals in Poland, 2011–2016. Front Microbiol. 2019; **10**: 1753. doi: 10.3389/fmicb.2019.01753
41. Merida-Vieyra J, De Colsa-Ranero A, Arzate-Barbosa P, Arias-dela Garza E, Méndez-Tenorio A, Aquino-Andrade A, *et al.* First clinical isolate of *Escherichia coli* harboring mcr-1 gene in Mexico. PLoS One. 2019; **14**(4): e0214648. doi:10.1371/journal.pone.0214648
42. Royden A, Ormandy E, Pinchbeck G, Pascoe B, Hitchings MD, Williams NJ, *et al.* Prevalence of faecal carriage of extended-spectrum β-lactamase (ESBL)-producing *Escherichia coli* in veterinary hospital staff and students. Vet Rec Open. 2019; **6**(1): e000307. doi: 10.1136/vetreco-2018-000307
43. McVeigh A, Fasano A, Scott DA, Jelacic S, Moseley SL, Savarino SJ, *et al.* IS1414, an *Escherichia coli* Insertion Sequence with a Heat-Stable Enterotoxin Gene Embedded in a Transposase-Like Gene. Infect Immun. 2000; **68**(10): 5710-5715. doi: 10.1128/iai.68.10.5710-5715.2000
44. Zhou M, Guo Z, Duan Q, Hardwidge PR, Zhu G. *Escherichia coli* type III secretion system 2: a new kind of T3SS? Veterinary Research. 2014; **45**(32). doi: 10.1186/1297-9716-45-32
45. Palacios M, Broberg C, Walker KA, Miller VL. A Serendipitous Mutation Reveals the Severe Virulence Defect of a *Klebsiella pneumoniae* fepB Mutant. mSphere. 2017; **2**(4). doi: 10.1128/mSphere.00341-17
46. Paczosa MK, Mecsas J. *Klebsiella pneumoniae*: Going on the Offense with a Strong Defense. Microbiol Mol Rev. 2016; **80**(3):629-61. doi: 10.1128/MMBR.00078-15
47. Stahlhut SG, Struve C, Krogfelt KA, Reisner A. Biofilm formation of *Klebsiella pneumoniae* on urethral catheters requires either type 1 or type 3 fimbriae. FEMS Immunol Med Microbiol. 2012; **65**(2): 350-359. doi: 10.1111/j.1574-695X.2012.00965.x
48. Padilla E, Llobet E, Doménech-Sánchez A, Martínez-Martínez L, Bengoechea JA, Albertí S. *Klebsiella pneumoniae* AcrAB Efflux Pump Contributes to Antimicrobial Resistance and Virulence. Antimicrob Agents Chemother. 2010; **54**(1): 177-183. doi: 10.1128/AAC.00715-09
49. Lai YC, Lin GT, Yang SL, Chang HY, Peng HL. Identification and characterization of KvgAS, a two-component system in *Klebsiella pneumoniae* CG43. FEMS Immunol Med Microbiol. 2003; **218**(1): 121-26. https://doi.org/10.1111/j.1574-6968.2003.tb11507.x
50. Karaba SM, White RC, Cianciotto NP. *Stenotrophomonas maltophilia* Encodes a Type II Protein Secretion System That Promotes Detrimental Effects on Lung Epithelial Cells. Infection and Immunity. 2013; **81**(9): 3210-3219. doi:10.1128/IAI.00546-13
51. Mattick JS, Whitchurch CB, Alm RA. The molecular genetics of type-4 fimbriae in *Pseudomonas aeruginosa* - a review. 1996; **179**(1): 147-155. https://doi.org/10.1016/S0378-1119(96)00441-6
52. Brolund A. Overview of ESBL-producing *Enterbacteriaceae* from a Nordic perspective. Infection Ecology & Epidemiology. 2014; **4**(1). http://dx.doi.org/10.3402/iee.v4.24555

53. Pitout JD, Laupland KB. Extended-spectrum beta-lactamase-producing Enterobacteriaceae: an emerging public-health concern. Lancet Infect Dis. 2008; **8**(3): 159–166. doi: 10.1016/S1473-3099(08)70041-0
54. Spencer RC. The emergence of epidemic, multiple-antibiotic-resistant *Stenotrophomonas* (Xanthomonas) *maltophilia* and *Burkholderia* (*Pseudomonas*) *cepacia*. Journal of Hospital Infection. 1995; **30**: 453-464. https://doi.org/10.1016/0195-6701(95)90049-7
55. Gajdács M and Urbán E. Prevalence and Antibiotic Resistance of *Stenotrophomonas maltophilia* in Respiratory Tract Samples: A 10-Year Epidemiological Snapshot. Health Serv Res Manag Epidemiol. 2019; **6**: 1-9. doi: 10.1177/2333392819870774
56. Eberl L, Vandamme P. Members of the genus *Burkholderia*: good and bad guys [version 1; referees: 3 approved] F1000Research 2016, 5(F1000 Faculty Rev): 1007. doi: 10.12688/f1000research.8221.1
57. Jørgensen SB, Søraas A, Arnesen L, Leegaard T, Sundsfjord A, Jenum PA. First environmental sample containing plasmid-mediated colistin-resistant ESBL-producing *Escherichia coli* detected in Norway. APMIS. 2017; **125**(9):822-825. doi: 10.1111/apm.12720
58. Søraas A, Sundsfjord A, Sandven I, Brunborg C, Jenum PA. Risk Factors for Community-Acquired Urinary Tract Infections Caused by ESBL-Producing Enterobacteriaceae – A Case-Control Study in a Low Prevalent Country. PLoS One. 2013; **8**(7). doi: 10.1371/journal.pone.0069581
59. Hamelin K, Bruant G, El-Shaarawi A, Hill S, Edge TA, Brousseau R, *et al.* A virulence and antimicrobial resistance DNA microarray detects a high frequency of virulence genes in *Escherichia coli* isolates from Great Lakes recreational waters. Appl Environ Microb. 2006; 72(6):4200–4206. doi: 10.1128/AEM.00137-06.
60. Blum SE, Goldstone RJ, Connolly JPR, . Postgenomics Characterization of an Essential Genetic Determinant of Mammary Pathogenic *Escherichia coli*. mBio. 2018; **9**(2): e0042318 v
61. Zhao Y, Niu W, Sun Y, Hao H, Yu Dong, Li Y. Identification and Characterization of a Serious Multidrug Resistant *Stenotrophomonas maltophilia* Strain in China. BioMed Research International. 2015. http://dx.doi.org/10.1155/2015/580240.
62. Livermore DM, Hawkey PM. CTX-M: changing the face of ESBLs in the UK. Journal of Antimicrobial Chemotherapy. 2005; **56**: 451-54. doi:10.1093/jac/dki239
63. Guenther S, Ewers C, Wieler LH. Extended-Spectrum Beta-Lactamases Producing *E. coli* in Wildlife, yet Another Form of Environmental Pollution? Front Microbiol. 2011; 2: 246. doi: 10.3389/fmicb.2011.00246
64. Mellata M. Zoonotic Risks, and Antibiotic Resistance Trends. Foodborne Pathog Dis. 2013; **10**(11): 916-32. doi: 10.1089/fpd.2013.1533
65. Dimou V, Dhanji H, Pike R, Livermore DM, Woodford N. Characterization of Enterobacteriaceae producing OXA-48-like carbapenemases in the UK. J Antimicrob Chemother. 2012; **67**(7): 1660-65. https://doi.org/10.1093/jac/dks124
66. Coque TM, Novais A, Carattoli A, Poirel L, Pitout J, Cantón R, *et al.* Dissemination of clonally related *Escherichia coli* strains expressing extended-spectrum-lactamase CTX-M-15. Emerg Infect Dis 2008; **14**(2): 195–200. doi: 10.3201/eid1402.07035

67. Guenther S, Semmler T, Stubbe A, Stubbe M, Stubbe A, Wieler LH, Schaufler K. Chromosomally encoded ESBL genes in *Escherichia coli* of ST38 from Mongolian wild birds. J Antimicrob Chemother. 2017; **72**(5): 1310-1313. doi:10.1093/jac/dkx006
68. Atterby C, Börjesson S, Ny S, Järhult JD, Byfors S, Bonnedahl J. ESBL-producing *Escherichia coli* in Swedish gulls-A case of environmental pollution from humans? 2017; **12**(12): e0190380. doi: 10.1371/journal.pone.0190380
69. Bonnedahl J, Hernandez J, Stedt J, Waldenstrom J, Olsen B, Drobni M. Extended-spectrum beta-lactamases in *Escherichia coli* and *Klebsiella pneumoniae* in Gulls, Alaska, USA. Emerg Infect Dis. 2014; **20**(5):897–899. pmid:24750592. doi: 10.3201/eid2005.130325
70. Amos GCA, Hawkey PM, Gaze WH, Wellington EM. Waste water effluent contributes to the dissemination of CTX-M-15 in the natural environment. J Antimicrob Chemother. 2014; **69**(7): 1785-1791. doi: 10.1093/jac/dku079

769

Supplementary Table 1: MIC data for Environmental isolates.

	Aminoglycosides			Carbapenems		Cephalosporins		Fluoroquinolones
	AK-Amikacin	**CN-Gentamicin**	**S-Streptomycin**	**IMI-Imipenem**	**MRP-Meropenem**	**CTX-Cefotaxime**	**FEP-Cefepime**	**CIP-Ciprofloxacin**
NMBU_W05E18	1	0.5	3	0.38	0.047	≥ 256	32	≥32
NMBU_W06E18_Str1	0.5	0.25	3	0.38	0.032	≥ 256	8	≥32
NMBU_W10C18	0.5	0.25	2	0.21	0.064	≥ 256	8	≥32
NMBU_W12E19	1	0.37	3	0.185	0.023	32	1	0.25
NMBU_W13E19	1.5	0.5	3	0.12	0.016	32	1.25	0.12
NMBU_W07E18	0.75	2	16	0.12	0.032	≥ 256	14	2
NMBU_R2	8	12	≥64	≥32	≥32	≥ 256	≥32	0.5

The MIC values labeled with gray background indicate the resistance to the antimicrobial agents based on EUCAST clinical breakpoint values (v 9.0).
R2 (*Stenotrophomonas maltophilia*): Trimethoprim-sulfamethoxazole is the only agent for which EUCAST breakpoints are currently available.

770
771
772
773
774
775

					Miscellaneous		
Macrolides	**Penicillins**		**Tetracyclines**	**Bacitracins**	**Polymyxins**	**Fosfomycin**	**Nitrofurans**
E- Erythromycin	**AM- Ampicillin**	**AMC- Amoxicillin w/ clavulanic acid**	**Tet- Tetracycline**	**B - Bacitracin**	**CS- Colistin**	**FOF- Fosfomycin**	**F- Nitrofurantoin**
≥ 256	≥ 256	5	5	≥256	1.5	5	7
≥ 256	≥ 256	≥ 256	3	≥256	1.5	0.5	5
32	≥ 256	≥ 256	3.5	≥256	1.5	0.565	7
≥ 256	≥ 256	5	48	≥256	1.5	8	8
≥ 256	≥ 256	4	128	≥256	1.5	0.12	12
24	≥ 256	12	7	≥256	1.5	≥1024	≥512
12	≥ 256	≥ 256	14	≥256	8	96	≥512

Miscellaneous	
Trimethoprim	Trimethoprim/ sulfonamide
TM- Trimethoprim	SXT- Trimethoprim/ sulfamethoxazole
≥32	0.625
0.25	0.275
0.19	0.275
0.22	0.47
≥32	1.5
0.25	0.75
≥32	0.125

Supplemental file 1.

Novel CMY-42 variant gene.

NMBU_W05E18 CMY-42-variant gene: ORF nucleotide sequence (ACCESSION number CP042878).

ATGATGAAAAAATCGTTATGCTGCGCTCTGCTGCTGACAGCCTCTTTCTCCACATTTGCTGCCGCAAAAA

CAGAACAACAGATTGCCGATATCGTTAATCGCACCATCACCCCGTTGATGCAGGAGCAGGCTATTCCGGG

TATGGCCGTTGCCGTTATCTACCAGGGAAAACCCTATTATTTCACCTGGGGTAAAGCCGATATCGCCAAT

AACCACCCAGTCACGCAGCAAACGCTGTTTGAGCTAGGATCGGTTAGTAAGACGTTTAACGGCGTGTTGG

GCGGCGATGCTATCGCCCGCGGCGAAATTAAGCTCAGCGATCCGGTCACGAAATACTGGCCAGAACTGAC

AGGCAAACAGTGGCAGGGTATCCGCCTGCTGCACTTAGCCACCTATACGGCAGGCGGCCTACCGCTGCAG

ATCCCCGATGACGTTAGGGATAAAGCCGCATTACTGCATTTTTATCAAAACTGGCAGCCGCAATGGACTC

CGGGCGCTAAGCGACTTTACGCTAACTCCAGCATTGGTCTGTTTGGCGCGCTGGCGGTGAAACCCTCAGG

AATGAGTTACGAAGAGGCAATGACCAGACGCGTCCTGCAACCATTAAAACTGGCGCATACCTGGATTACG

GTTCCGCAGAACGAACAAAAAGATTATGCCTGGGGCTATCGCGAAGGGAAGCCCGTACACAGTTCTCCGG

GACAACTTGACGCCGAAGCCTATGGCGTGAAATCCAGCGTTATTGATATGGCCCGCTGGGTTCAGGCCAA

CATGGATGCCAGCCACGTTCAGGAGAAAACGCTCCAGCAGGGCATTGCGCTTGCGCAGTCTCGCTACTGG

CGTATTGGCGATATGTACCAGGGATTAGGCTGGGAGATGCTGAACTGGCCGCTGAAAGCTGATTCTAGCA

TCAACGGCAGCGACAGCAAAGTGGCATTGGCAGCGCTTCCCACCGTTGAGGTAAACCCGCCCGCCCCCGC

AGTGAAAGCCTCATGGGTGCATAAAACGGGCTCCACTGGTGGATTTGGCAGCTACGTAGCCTTCGTTCCA

GAAAAAAACCTTGGCATCGTGATGCTGGCAAACAAAAGCTATCCTAACCCTGTCCGTGTCGAGGCGGCCT

GGCGCATTCTTGAAAAGCTGCAATAA

NMBU_W05E18 CMY-42-variant gene: translated amino acid sequence.

SEQUENCE 381 AA.

MMKKSLCCAL LLTASFSTFA AAKTEQQIAD IVNRTITPLM QEQAIPGMAV AVIYQGKPYY
FTWGKADIAN NHPVTQQTLF ELGSVSKTFN GVLGGDAIAR GEIKLSDPVT KYWPELTGKQ
WQGIRLLHLA TYTAGGLPLQ IPDDVRDKAA LLHFYQNWQP QWTPGAKRLY ANSSIGLFGA
LAVKPSGMSY EEAMTRRVLQ PLKLAHTWIT VPQNEQKDYA WGYREGKPVH SSPGQLDAEA
YGVKSSVIDM ARWVQANMDA SHVQEKTLQQ GIALAQSRYW RIGDMYQGLG WEMLNWPLKA
DSSINGSDSK VALAALPTVE VNPPAPAVKA SWVHKTGSTG GFGSYVAFVP EKNLGIVMLA
NKSYPNPVRV EAAWRILEKL Q

Alignment of beta-lactamases from environmental strains and references (CLUSTAL Omega v.1.2.4).

```
CMY-variant_NMBU_W05E18          MMKKSLCCALLLTASFSTFAAAKTEQQIADIVNRTITPLMQEQAIPGMAVAVIYQGKPYY   60
F7057_00310                      MMKKSLCCALLLTASFSTFAAAKTEQQIADIVNRTITPLMQEQAIPGMAVAVIYQGKPYY   60
CMY_NMBU_W06E18                  MMKKSLCCALLLTASFSTFAAAKTEQQIADIVNRTITPLMQEQAIPGMAVAVIYQGKPYY   60
AID23886.1                       MMKKSLCCALLLTASFSTFAAAKTEQQIADIVNRTITPLMQEQAIPGMAVAVIYQGKPYY   60
AOG62294.1                       MMKKSLCCALLLTASFSTFAAAKTEQQIADIVNRTITPLMQEQAIPGMAVAVIYQGKPYY   60
WP_063610930.1                   -MFKTTLCALLITASCSTFAA---PQQINDIVHRTITPLIEQQKIPGMAVAVIYQGKPYY   56
FVP48_14090                      -MFKTTLCALLITASCSTFAA---PQKINDIVHRTITPLIEQQKIPGMAVAVIYQGKPYY   56
NP_313158.1                      -MFKTTLCALLITASCSTFAA---PQQINDIVHRTITPLIEQQKIPGMAVAVIYQGKPYY   56
FVA71_23515                      -MFKMTLCALLITASCSTFAA---PQQINDIVHRTITPLIEQQKIPGMAVAVIYQGKPYY   56
FZN32_14220                      -MFKMTLCALLITASCSTFAA---PQQINDIVHRTITPLIEQQKIPGMAVAVIYQGKPYY   56
F7057_02980                      -MFKMTLCALLITASCSTFAA---PQQINDIVHRTITPLIEQQKIPGMAVAVIYQGKPYY   56
beta_lactamase_C_NMBU_W06E18     -MFKMTLCALLITASCSTFAA---PQQINDIVHRTITPLIEQQKIPGMAVAVIYQGKPYY   56
WP_063860454.1                   -MFKTTLCALLITASCSTFAA---PQQINDIVHRTITPLIEQQKIPGMAVAVIYQGKPYY   56
NP_418574.1                      -MFKTTLCALLITASCSTFAA---PQQINDIVHRTITPLIEQQKIPGMAVAVIYQGKPYY   56
                                  * *   ****:*** *****    *:* ***:******:::* ****************

CMY-variant_NMBU_W05E18          FTWGKADIANNHPVTQQTLFELGSVSKTFNGVLGGDAIARGEIKLSDPVTKYWPELTGKQ  120
F7057_00310                      FTWGKADIANNHPVTQQTLFELGSVSKTFNGVLGGDAIARGEIKLSDPVTKYWPELTGKQ  120
CMY_NMBU_W06E18                  FTWGKADIANNHPVTQQTLFELGSVSKTFNGVLGGDAIARGEIKLSDPVTKYWPELTGKQ  120
AID23886.1                       FTWGKADIANNHPVTQQTLFELGSVSKTFNGVLGGDAIARGEIKLSDPVTKYWPELTGKQ  120
AOG62294.1                       FTWGKADIANNHPVTQQTLFELGSVSKTFNGVLGGDAIARGEIKLSDPVTKYWPELTGKQ  120
WP_063610930.1                   FTWGYADIAKKQPVTQQTLFELGSVSKTFTGVLGGDAIARGEIKLSDPATKYWPELTAKQ  116
FVP48_14090                      FTWGYADIAKKQPVTQQTLFELGSVSKTFTGVLGGDAIARGEIKLSDPATKYWPELTAKQ  116
NP_313158.1                      FTWGYADIAKKQPVTQQTLFELGSVSKTFTGVLGGDAIARGEIKLSDPATKYWPELTAKQ  116
FVA71_23515                      FTWGYADIAKKQPVTQQTLFELGSVSKTFTGVLGGDAIARGEIKLSDPATKYWPELTAKQ  116
FZN32_14220                      FTWGYADIAKKQPVTQQTLFELGSVSKTFTGVLGGDAIARGEIKLSDPATKYWPELTAKQ  116
F7057_02980                      FTWGYADIAKKQPVTQQTLFELGSVSKTFTGVLGGDAIARGEIKLIDPATKYWPELTAKQ  116
beta_lactamase_C_NMBU_W06E18     FTWGYADIAKKQPVTQQTLFELGSVSKTFTGVLGGDAIARGEIKLIDPATKYWPELTAKQ  116
WP_063860454.1                   FTWGYADIAKKQPVTQQTLFELGSVSKTFTGVLGGDAIARGEIKLIDPATKYWPELTAKQ  116
NP_418574.1                      FTWGYADIAKKQPVTQQTLFELGSVSKTFTGVLGGDAIARGEIKLSDPTTKYWPELTAKQ  116
                                 **** ****:::*****************.*************** **.********.**

CMY-variant_NMBU_W05E18          WQGIRLLHLATYTAGGLPLQIPDDVRDKAALLHFYQNWQPQWTPGAKRLYANSSIGLFGA  180
F7057_00310                      WQGIRLLHLATYTAGGLPLQIPDDVRDKAALLHFYQNWQPQWTPGAKRLYANSSIGLFGA  180
CMY_NMBU_W06E18                  WQGIRLLHLATYTAGGLPLQIPDDVRDKAALLHFYQNWQPQWTPGAKRLYANSSIGLFGA  180
AID23886.1                       WQGIRLLHLATYTAGGLPLQIPDDVRDKAALLHFYQNWQPQWTPGAKRLYANSSIGLFGA  180
AOG62294.1                       WQGIRLLHLATYTAGGLPLQIPDDVRDKAALLHFYQNWQPQWTPGAKRLYANSSIGLFGA  180
WP_063610930.1                   WNGITLLHLATYTAGGLPLQVPDEVKSSSDLLRFYQNWQPAWAPGTQRLYANSSIGLFGA  176
FVP48_14090                      WNGITLLHLATYTAGGLPLQVPDEVKSSSDLLRFYQNWQPAWAPGTQRLYANSSIGLFGA  176
NP_313158.1                      WNGITLLHLATYTAGGLPLQVPDEVKSSSDLLRFYQNWQPAWAPGTQRLYANSSIGLFGA  176
FVA71_23515                      WNGITLLHLATYTAGGLPLQVPDEVKSSSDLLRFYQNWQPAWAPGTQRLYANSSIGLFGA  176
FZN32_14220                      WNGITLLHLATYTAGGLPLQVPDDVKSSSDLLRFYQNWQPAWAPGTQRLYANSSIGLFGA  176
F7057_02980                      WNGITLLHLATYTAGGLPLQVPDDVKSSSDLLRFYQNWQPAWAPGTQRLYANSSIGLFGA  176
beta_lactamase_C_NMBU_W06E18     WNGITLLHLATYTAGGLPLQVPDDVKSSSDLLRFYQNWQPAWAPGTQRLYANSSIGLFGA  176
WP_063860454.1                   WNGITLLHLATYTAGGLPLQVPDDVKSSSDLLRFYQNWQPAWAPGTQRLYANSSIGLFGA  176
NP_418574.1                      WNGITLLHLATYTAGGLPLQVPDEVKSSSDLLRFYQNWQPAWAPGTQRLYANSSIGLFGA  176
                                 *:** ***************:**:*:..: **:******* *:**::************

CMY-variant_NMBU_W05E18          LAVKPSGMSYEEAMTRRVLQPLKLAHTWITVPQNEQKDYAWGYREGKPVHSSPGQLDAEA  240
F7057_00310                      LAVKPSGMSYEEAMTRRVLQPLKLAHTWITVPQNEQKDYAWGYREGKPVHSSPGQLDAEA  240
CMY_NMBU_W06E18                  LAVKPSGMSYEEAMTRRVLQPLKLAHTWITVPQNEQKDYAWGYREGKPVHSSPGQLDAEA  240
AID23886.1                       LAVKPSGMSYEEAMTRRVLQPLKLAHTWITVPQNEQKDYAWGYREGKPVHSSPGQLDAEA  240
AOG62294.1                       LAVKPSGMSYEEAMTRRVLQPLKLAHTWITVPQNEQKDYAWGYREGKPVHVSPGQLDAEA  240
WP_063610930.1                   LAVKPSGLSFEQAMQTRVFQPLKLNHTWINVPPPEEKNYAWGYREGKAVHVSPGALDAEA  236
FVP48_14090                      LAVKPSGLSFEQAMQTRVFQPLKLNHTWINVPPAEEKNYAWGYREGKAVHVSPGALDAET  236
NP_313158.1                      LAVKPSGLSFEQAMQTRVFQPLKLNHTWINVPPAEEKNYAWGYREGKAVHVSPGALDAET  236
FVA71_23515                      LAVKPSGLSFEQAMQTRVFQPLKLTHTWINVPSAEEKNYAWGYREGKAVHVSPGALDAEA  236
FZN32_14220                      LAVKPSGLSFEQAMQTRVFQPLKLTHTWINVPSAEEKNYAWGYREGKAVHVSPGALDAEA  236
F7057_02980                      LAVKPSGLSFEQAMQTRVFHPLKLTHTWINVPSAEEKNYAWGYREGKAVHVSPGALDAEA  236
beta_lactamase_C_NMBU_W06E18     LAVKPSGLSFEQAMQTRVFHPLKLTHTWINVPSAEEKNYAWGYREGKAVHVSPGALDAEA  236
WP_063860454.1                   LAVKPSGLSFEQAMQTRVFQPLKLNHTWINVPPAEEKNYAWGYREGKAVHVSPGALDAEA  236
NP_418574.1                      LAVKPSGLSFEQAMQTRVFQPLKLNHTWINVPPAEEKNYAWGYREGKAVHVSPGALDAEA  236
                                 *******:*:*:**  **::**** ****.**  *:*:********* ** *** ****:

CMY-variant_NMBU_W05E18          YGVKSSVIDMARWVQANMDASHVQEKTLQQGIALAQSRYWRIGDMYQGLGWEMLNWPLKA  300
F7057_00310                      YGVKSSVIDMARWVQANMDASHVQEKTLQQGIALAQSRYWRIGDMYQGLGWEMLNWPLKA  300
CMY_NMBU_W06E18                  YGVKSSVIDMARWVQANMDASHVQEKTLQQGIALAQSRYWRIGDMYQGLGWEMLNWPLKA  300
AID23886.1                       YGVKSSVIDMARWVQANMDASHVQEKTLQQGIALAQSRYWRIGDMYQGLGWEMLNWPLKA  300
AOG62294.1                       YGVKSSVIDMARWVQANMDASHVQEKTLQQGIALAQSRYWRIGDMYQGLGWEMLNWPLKA  300
WP_063610930.1                   YGVKSTIEDMARWVRSNMNPRDINDKTLQQGIQLAQSRYWQTGDMYQGLGWEMLDWPVNP  296
FVP48_14090                      YGVKSTIEDMACWVRSNMNPRDINDKTLQQGIQLAQSRYWQTGDMYQGLGWEMLDWPVNP  296
NP_313158.1                      YGVKSTIEDMACWVRSNMNPRDINDKTLQQGIQLAQSRYWQTGDMYQGLGWEMLDWPVNP  296
FVA71_23515                      YGVKSTIEDMARWVQSNLKPLDITEKTLQQGIQLAQSRYWQTGDMYQGLGWEMLDWPVNP  296
FZN32_14220                      YGVKSTIEDMARWVQSNLKPLDITEKTLQQGIQLAQSRYWQTGDMYQGLGWEMLDWPVNP  296
F7057_02980                      YGVKSTIEDMARWVQSNLKPLDITEKTLQQGIQLAQSRYWQTGDMYQGLGWEMLDWPVNP  296
beta_lactamase_C_NMBU_W06E18     YGVKSTIEDMARWVQSNLKPLDITEKTLQQGIQLAQSRYWQTGDMYQGLGWEMLDWPVNP  296
WP_063860454.1                   YGVKSTIEDMARWVQSNLKPLDITEKTLQQGIQLAQSRYWQTGDMYQGLGWEMLDWPVNP  296
NP_418574.1                      YGVKSTIEDMARWVQSNLKPLDINEKTLQQGIQLAQSRYWQTGDMYQGLGWEMLDWPVNP  296
                                 ******:: *** **::*:.  .: :******* *******: ************:**::

CMY-variant_NMBU_W05E18          DSSINGSDSKVALAALPTVEVNPPAPAVKASWVHKTGSTGGFGSYVAFVPEKNLGIVMLA  360
F7057_00310                      DSIINGSDSKVALAALPAVEVNPPAPAVKASWVHKTGSTGGFGSYVAFVPEKNLGIVMLA  360
CMY_NMBU_W06E18                  DSIINGSDSKVALAALPAVEVNPPAPAVKASWVHKTGSTGGFGSYVAFVPEKNLGIVMLA  360
AID23886.1                       DSIINGSDSKVALAALPAVEVNPPAPAVKASWVHKTGSTGGFGSYVAFVPEKNLGIVMLA  360
AOG62294.1                       DSIINGSDSKVALAALPAVEVNPPAPAVKASWVHKTGSTGGFGSYVAFVPEKNLGIVMLA  360
WP_063610930.1                   DSIINGSGNKIALAAPPVKAITPPTPAVRASWVHKTGATGGFGSYVAFIPEKELGIVMLA  356
FVP48_14090                      DSIINGSGNKIALAARPVKAITPPTPAVRASWVHKTGATGGFGSYVAFIPEKELGIVMLA  356
NP_313158.1                      DIIVNGSDNKIALAARPVKAITPPTPAVRASWVHKTGATGGFGSYVAFIPEKELGIVMLA  356
```

```
FVA71_23515                     DIIINGSDNKIALAARPVKAITPPAPAVRASWVHKTGATGGFGSYVAFIPEKELGIVMLA       356
FZN32_14220                     DIIINGSDNKIALAARPVKAITPPTPAVCASWVHKTGATGGFGSYVAFIPEKELGIVMLA       356
F7O57_02980                     DIIINGSDNKIALAARPVKAITPPTPAVRASWVHKTGATGGFGSYVAFIPEKELGIVMLA       356
beta_lactamase_C_NMBU_W06E18    DIIINGSDNKIALAARPVKAITPPTPAVRASWVHKTGATGGFGSYVAFIPEKELGIVMLA       356
WP_063860454.1                  DSIINGSDNKIALAARPVKAITPPTPAVRASWVHKTGATGGFGSYVAFIPEKELGIVMLA       356
NP_418574.1                     DSIINGSDNKIALAARPVKAITPPTPAVRASWVHKTGATGGFGSYVAFIPEKELGIVMLA       356
                                *  :***..*:**** *.  :.**:*** ********:**********:***:*******


CMY-variant_NMBU_W05E18         NKSYPNPVRVEAAWRILEKLQ        381
F7O57_00310                     NKSYPNPVRVEAAWRILEKLQ        381
CMY_NMBU_W06E18                 NKSYPNPVRVEAAWRILEKLQ        381
AID23886.1                      NKSYPNPVRVEAAWRILEKLQ        381
AOG62294.1                      NKSYPNPVRVEAAWRILEKLQ        381
WP_063610930.1                  NKNYPNPARVAAAWQILNALQ        377
FVP48_14090                     NKNYPNPARVTAAWQILNALQ        377
NP_313158.1                     NKNYPNPARVAAAWQILNALQ        377
FVA71_23515                     NKNYPNPARVAAAWQILNALQ        377
FZN32_14220                     NKNYPNPARVTAAWQILNALQ        377
F7O57_02980                     NKNYPNPARVAAAWQILNALQ        377
beta_lactamase_C_NMBU_W06E18    NKNYPNPARVAAAWQILNALQ        377
WP_063860454.1                  NKNYPNPARVTAAWQILNALQ        377
NP_418574.1                     NKNYPNPARVDAAWQILNALQ        377
```

Legend:

WP_063860454.1: cephalosporin-hydrolyzing class C beta-lactamase EC-8 [Escherichia coli]

WP_063610930.1: class C extended-spectrum beta-lactamase EC-15 [Escherichia coli]

AID23886.1: beta-lactamase CMY-42 [Escherichia coli]

AOG62294.1: CMY-2 [Escherichia coli]

NP_313158.1: beta-lactamase [Escherichia coli O157:H7 str. Sakai]

NP_418574.1 beta-lactamase [Escherichia coli str. K-12 substr. MG1655]

FVP48_14090: class C beta-lactamase from NMBU-W12E19

FVA71_23515: class C beta-lactamase from NMBU_W05E18 (chromosome)

FZN32_14220: class C beta-lactamase from NMBU-W13E19

F7O57_00310: class C beta-lactamase from NMBU-W10C18 (plasmid)

F7O57_02980: class C beta-lactamase from NMBU-W10C18 (chromosome)

beta_lactamase_C_NMBU_W06E18: class C beta-lactamase from NMBU_W06E18

CMY-variant_NMBU_W05E18: New CMY-42 variant from NMBU_W05E18 (plasmid)

Alignment of two novel class-A β-lactamases from strain NMBU_R16 (bla_1 and bla_2) with closest matches.

bla_1) Query = NMBU_R16_locus_tag="GWC77_08380"
Sbjct = class A beta-lactamase [Burkholderia sp. AD24], Sequence ID: WP_134964008.1

Score	Expect	Method	Identities	Positives	Gaps
354 bits(908)	5e-119	Compositional matrix adjust.	189/314(60%)	228/314(72%)	10/314(3%)

```
Query  1    MQRRQFIGTVASGLIAGLVTVPGAARAADAKSTARAGDDSPN-ELARS--IEARLAAIET  57
            + RR+F G +     IAG+       +  A  K+ ++AG+ +P  ELA++  I  RLA IE
Sbjct  2    ITRRKFTGAMLGVSIAGVAV---GSVGAFGKTASKAGNAAPGLELAKAAAIRQRLAQIEA  58

Query  58   RVAGRLGVSILDTDSGFAAGHRENERFPMCSTFKALAAAAVLERVDRQLDDLSRRIVFSR  117
               GRLGVSI+DT SG  AG R +ERFPMCSTFK LAA AVL RVDR  +DL+RR+VFS+
Sbjct  59   ESGGRLGVSIVDTSSGLHAGLRTDERFPMCSTFKVLAAGAVLTRVDRAQEDLARRVVFSQ  118

Query  118  EDLVPYSPETGKHA----GGTGMTLSELCEAAVTLSDNTAGNLLLASIGGPAGLTAFARH  173
             DLVP SP T +H     G  GM+++ELC+AA+TLSDNTA NLLLA+ GGPA LTAFAR
Sbjct  119  RDLVPNSPATSRHTRERTGDAGMSIAELCKAAITLSDNTAANLLLATFGGPAALTAFARD  178

Query  174  LGDTVTRLDRNEPTLNEALPGDPRDTTSPAAMRATLRTLLLGNRLSPNSRDRWLAWLEAN  233
            LGD +TRLDR EPTLNEA+PGDPRDTT+P AM   LR L+LG  LS +SR + LAWL AN
Sbjct  179  LGDGITRLDRIEPTLNEAIPGDPRDTTTPNAMLGNLRELVLGEHLSSSSRAQLLAWLVAN  238

Query  234  QTGGERIRAGLPTGWRVGDKTGTGERGTANDIAIVWPPGRGPIIVTAYLTATTAAAAQRN  293
            +TGG R+RA LP  W VGDKTGTG+ GTANDIAI+WPPGRGP++V  YLT T   AA+ N
Sbjct  239  ETGGARLRAKLPQDWGVGDKTGTGDHGTANDIAILWPPGRGPVLVAVYLTGTNGDAARCN  298

Query  294  AALAEVGRLAASIV  307
            AA+AEVG L    V
Sbjct  299  AAIAEVGALVVQSV  312
```

bla_2) Query = NMBU_R16_locus_tag="GWC77_24470"
Sbjct = class A beta-lactamase [Paraburkholderia sp. DHOM06], Sequence ID: WP_115533798.1

Score	Expect	Method	Identities	Positives	Gaps
394 bits(1011)	3e-135	Compositional matrix adjust.	189/258(73%)	220/258(85%)	0/258(0%)

```
Query  44   LANIEAESGGRLGFFAVDTGSGRTLAYRADERFLMCSTFKGILAAQIFARVDRGEERLDR  103
            LA+IE   GGRLG FAVDTGSGRTL++RADERFLMCSTFKGILAAQI AR D G+ERLDR
Sbjct  33   LADIERRHGGRLGVFAVDTGSGRTLSHRADERFLMCSTFKGILAAQILARTDSGQERLDR  92

Query  104  QIAYTQKDLIFTSPVTQANVARGAMSIDDLCRAILEESDNTAAVLLMRSAGGPEALTAFV  163
             + YT+ DLIFTSPVT+AN++RGAMS++ LC+AILEESDNTAA+LLMRSAGGP ALT F+
Sbjct  93   LVHYTKNDLIFTSPVTKANLSRGAMSVEALCQAILEESDNTAAILLMRSAGGPAALTRFI  152

Query  164  RRLGDTVTRSDRYEPQTNSYSGLLDTTSPRAIVTLAKTLLLGNALTAESRARLERGMINC  223
            R LGDTVTRSDRYEP +N YSG+LDTTSPRAIVT+A++LLLG+ L+ +SR RLERGMI C
Sbjct  153  RGLGDTVTRSDRYEPDSNRYSGVLDTTSPRAIVTVARSLLLGDVLSPKSRTRLERGMIAC  212

Query  224  KPGRSRIRAVLPQSWSCADRPGTSVGGETNDYALVRPIGRHPLIVAAYYDAPSLAMAERE  283
            +PG +RIRAVLP  W   DRPGTSV  ETNDYALVRP GR PL+VA Y DAP ++M +RE
Sbjct  213  RPGLNRIRAVLPAGWQAGDRPGTSVESETNDYALVRPPGRAPLLVAVYCDAPGVSMDDRE  272

Query  284  SVLRKAGEAFTQWAMEVA  301
            +VLR+AG+ F QWA  VA
Sbjct  273  AVLREAGKVFVQWAASVA  290
```

Supplementary table 2. Virulence gene profiles of *E. coli* isolates. Gray blocks represent the presence of a virulence gene.

VIRULENCE GENE	ACCESSION NUMBER	DESCRIPTION	NMBU_05E18	NMBU_W06E18_Strain 1	NMBU_W10C18	NMBU_W12E19	NMBU_W13E19
aapA	Z32523	Dispersin (anti-aggregation protein) gene				■	
aatA	FN554767	Dispersin transporter protein				■	
aatB	JX402062	APEC autotransporter adhesin	■	■	■		
aatD	HE603111	Dispersin (anti-aggregation protein) gene				■	
aec35	AY857617	Part of a selC-Associated Genomic Island of APEC strain BEN2908	■	■	■		
aec36	AY857617	Putative MFS superfamily hexuronate transporter	■	■	■		
aec37	AY857617	Putative glucosidase-family 31 of glycosyl hydrolases	■	■	■		
aec77	CYDF01000002	Similarity to YeeW, *E. coli*RW1374	■	■	■	■	■
afa/Dr	AY030355	DNA sequence specific to the C1845 Afa/Dr DAEC strain	■				
afaE-2	X85782	afa E-2 protein, adhesin of tDAEC	■				
AIDA-I	BA000007	Potent bacterial adhesin that mediates bacterial attachment to a broad variety of human and other mammalian cells			■		
Air	CP003034	Putative adhesin – (enteroaggregative immunoglobulin repeat protein)	■	■	■		■
antigen-43	U24429	AIDA-I–type AT protein	■	■	■	■	■
antigen-43b	AE014075	AIDA-I–type AT protein	■				
aslA	CU928163	Contributes to Invasion of Brain Microvascular Endothelial Cells In Vitro and In Vivo	■	■	■	■	■
bfpA	NZ_AIFX01000009	Bundle-forming pilus associated with typical EPEC				■	
bfpB	CABEEG010000008	Bundle-forming pilus associated with typical EPEC				■	
cas1	BA000007	CRISPR-associated endonuclease	■	■	■		■
cas2	BA000007	CRISPR-associated protein	■	■	■		■
cas3	BA000007	CRISPR-associated protein	■	■	■		■
casA	BA000007	CRISPR system Cascade subunit	■		■		■
casE	BA000007	CRISPR-associated protein Cas6/Cse3/CasE	■		■		■
cfaB	CYCQ01000001	Colonization factor antigen 1					
chuA	LT827011	Outer membrane hemin receptor	■	■	■		■
cirA	CP030337	Ferric iron-catecholate outer membrane transporter	■	■	■	■	■
csgA	CP023388	Major curlin subunit, from *E. coli*strain 1105	■	■	■	■	■
csgA	BA000007	Major curlin subunit, from *E. coli*strain O157:H7 Sakai		■			■
csgB	CP027060	Minor curlin subunit	■	■	■	■	
csgE	NC_011750	Curli production assembly/transport component	■	■	■	■	
csgF	NC_011750	Curli production assembly/transport component	■	■	■	■	■
csgG	CP003034	Curli production assembly/transport component	■	■	■	■	■
csgG	LT903847	Curli production assembly/transport component					■
cvaA	CU928146	Colicin-V immunity protein	■	■	■	■	■
eaeH	CP012635	Putative attaching and effacing protein				■	
EAST1/astA	AF143819	Heat-stable enterotoxin 1 (allele-2)				■	
ecpA	BA000007	Common pilus major fimbrillin subunit EcpA	■	■	■		■
ecpB	NZ_QOON01000045	Probably fimbrial chaperone	■	■	■	■	■
ecpD	CP019777	Common pilus fimbria adhesin EcpD	■	■	■	■	■
ECs3737	NC_002695	ETT2 gene	■	■	■	■	■
ehaA	BA000007	Autotransporter protein of EHEC O157:H7	■				■
ehaB	BA000007	The *E. coli* O157:H7 EhaB autotransporter protein binds to laminin and collagen I and induces a serum IgA response in O157:H7 challenged cattle		■	■	■	

ehaC	BA000007	Autotransporter that may play a role in aEPEC infection					
eilA	CP023364	transcriptional regulator HilA, locus					
eitA	DQ381420	putative iron transport system, periplasmic binding protein					
eitB	DQ381420	putative iron transport system, permease component					
eitC	DQ381420	putative iron transport system, ATP-binding component					
eivC	KU684470	ETT2 gene					
eivF	NC_002695	ETT2 gene					
eivG	NC_002695	ETT2 gene					
eivH	DQ077151	ETT2 gene					
eivI	KU684470	ETT2 gene					
eivJ	KU684470	ETT2 gene					
entA	CP027060	Involved in the biosynthesis of the siderophore enterobactin					
entE	CP027060	Involved in the biosynthesis of the siderophore enterobactin					
entH	CP027060	Involved in the biosynthesis of the siderophore enterobactin					
epaO	KU684470	ETT2 gene					
epaP	NC_002695	ETT2 gene					
epaQ	NC_002695	ETT2 gene					
epaR	KU684470	ETT2 gene					
epaS1	KU684470	ETT2 gene					
epaS2	KU684470	ETT2 gene					
eprH	NC_011750	Putative Type III secretion protein					
eprI	NC_002695	Type III secretion protein					
eprJ	KU684470	Type III secretion apparatus protein					
eprK	NC_011750	Type III secretion					
espR1	BA000007	Non-LEE-encoded type III secreted effector					
espR2	BA000007	Type III secretion effector					
espX1	CP003034	Putative type III secreted effector					
espX4	NC_017646	Type III secretion system (T3SS) in EHEC and EPEC					
espX5	NC_017646	Type III secretion system (T3SS) in EHEC and EPEC					
espY1	NC_002695	Type III secretion system (T3SS) in EHEC and EPEC					
espY3	NC_002695	Type III secretion system (T3SS) in EHEC and EPEC					
espY4	NC_002695	Type III secretion system (T3SS) in EHEC and EPEC					
etrA	KU684467	Type III secretion regulator					
fdeC	CP019777	Intimin-like protein that was shown to contribute to kidney colonization in a mouse urinary tract infection model					
fecA	NC_011751	Outer membrane ferric-siderophore receptor					
feoB	CP019777	Ferrous iron uptake system which is active under anaerobic growth conditions					
fepA	CP000468	Outer membrane (OM) protein that binds and transports ferric enterobactin (ferric enterochelin)					
fepB	AE014075	Binds ferrienterobactin; part of the binding-protein-dependent transport system for uptake of ferrienterobactin					
fepC	AE014075	Encodes the predicted ATP-binding subunit of a ferrric enterobactin ABC transporter complex					
fepD	AE014075	Subunit of a ferrric enterobactin ABC transporter complex					
fepE	AE014075	Ferric enterobactin transport protein					
fepG	AE014075	Subunit of a ferrric enterobactin ABC transporter comple					
fes	CP027060	Enterochelin esterase - upon internalization, ferric enterobactin is processed via an exquisitely specific pathway that is dependent on FES activity, making iron available for metabolic use					
fiu	CP027060	Involved in the active transport across the outer membrane of iron complexed with catecholate siderophores such as dihydroxybenzoylserine and dihydroxybenzoate					
flgD	CP027060	Deletion of flgD attenuated ExPEC strain PCN033 invasion and colonization in vivo, probably by affecting bacterial adhesion and invasion					
flgM	CP028192	Negative regulator of flagellin synthesis (Anti-sigma factor)					
fmlA	BA000007	F9/Yde/Fml pilus is involved in UPEC persistence in the inflamed urothelium					

fyuA	CP016828	The ferric yersiniabactin uptake receptor fyuA is required for efficient biofilm formation in UPEC					
gad	FN554766	Glutamate decarboxylase					
hlyD		Hemolysin secretion protein					
hlyE	BA000007	Toxin, which has some hemolytic activity towards mammalian cells					
hlyIII	CP003034	inner membrane protein, hemolysin III family					
ibeB	AF094824	IbeB is an important determinant contributing to *E. coli*K1 crossing of the blood–brain barrier					
ibeC	CP019777	Invasin of brain endothelial cells					
ibrA	CP003034	Immunoglobulin-binding regulator					
iha	AF399919	Virulence Factor in Murine Urinary Tract Infection					
invA	CU928163	May be a part of ETT2					
invE	NC_011751	May be a part of ETT2					
ipaH-like	CU928164	Invasion plasmid antigen of EIEC and Shigella					
irp1	CU928163	HMWP1 nonribosomal peptide/polyketide synthase, part of HPI					
irp2	CP006834	HMWP2 Yersiniabactin biosynthetic protein, part of HPI					
IS26 transposase	AP018456	Plays a major role in the acquisition and dissemination of antibiotic resistance					
iss	CP030791	Increased serum survival					
iucA	CU928163	Part of the aerobactin gene cluster					
iucB	AE014075	Part of the aerobactin gene cluster					
iucC	AE014075	Part of the aerobactin gene cluster					
iucD	CP001232	Part of the aerobactin gene cluster					
iutA	CP011134	Ferric aerobactin receptor					
kpsM	CU928163	The deletion of gene kpsM weakens the virulence of porcine ExPEC PCN033					
kpsMII	X53819	Associated with persistence or relapse in recurrent urinary tract infections caused by *E. coli*					
kpsM							
kpsS	CU928163	Capsule polysaccharide export protein					
lpfA	CP006834	Long polar fimbriae (LPF) are related to type I fimbriae					
lpfA-O113	AY057066	Long polar fimbriae (LPF) are related to type I fimbriae					
matD	HM102365	Meningitis-associated and temperature-regulated (Mat) fimbriae gene					
nfaA/dafaA	AF325672	Diffuse adherence fibrillar adhesin gene locus. Non-fimbrial adhesin 1. The *nfaA* gene encoding the antigen adhesive factor of enterotoxigenic *E. coli*.					
nfaB/dafaB	AF325672	*E. coli* diffuse adherence fibrillar adhesin gene locus. Diffuse adherence fibrillar adhesin *DafaB*					
ompA	CP027060	Required for the action of colicins K and L and for the stabilization of mating aggregates in conjugation.					
ompW	CP027060	Receptor for Colicin S4					
pap operon	AP018784/ CP003034	Pap (pyelonephritis-associated pili) operon associated with UPEC strains.					
pgaA	NZ_UGAE01000003	Biofilm related gene					
ppk	CP025268	In some mutants lacking *ppk*, the phenotypes included features indicative of decreased virulence such as: (i) growth defects, (ii) defective responses to stress and starvation, (iii) loss of viability, (iv) polymyxin sensitivity, (v) intolerance to acid and heat, and (vi) diminished invasiveness in epithelial cells					
rafA	NC_010558	Peripheral raffinose metabolic pathway					
rafB	NC_010558	Peripheral raffinose metabolic pathway					
rafD	NC_010558	Peripheral raffinose metabolic pathway					
sat	CU928163	The secreted autotransporter toxin of UPEC					
senB	AP018458	May have some role in enterotoxicity of EIEC					
sfmC	NSBV01000011	Could contribute to adhesion to various surfaces in specific environmental niches. Increases adhesion to eukaryotic T24 bladder epithelial cells in the absence of fim genes					
shET2	NZ_AYOG01000052	Enterotoxin					

sinH	CP023644	Recent work suggests that at least two other virulence-associated bacterial outer membrane proteins share a structural and evolutionary history with intimin and invasin					
sitA	FQ482074	Homolog of the iron transport system SitABCD encoded on SPI1, which is required for full virulence of Salmonella typhimurium					
sitB	NC_017659	Homolog of the iron transport system SitABCD encoded on SPI1, which is required for full virulence of Salmonella typhimurium					
sitC	CP000836	Homolog of the iron transport system SitABCD encoded on SPI1, which is required for full virulence of Salmonella typhimurium					
sitD	NC_017659	Homolog of the iron transport system SitABCD encoded on SPI1, which is required for full virulence of Salmonella typhimurium					
stcD	NC_018658	Fimbrial-like adhesin protein					
tia	DQ095216	Invasion determinant					
traJ	DQ381420	Contributes to *E. coli*K1 virulence in the neonatal rat					
traT	AY214164	May inhibit the formation of C5b6 complex or causes structural alteration of the complex to a nonfunctional form					
upaG-like		Autotransporter protein of UPEC					
ybtA	CP028714	Yersiniabactin transcriptional regulator					
ybtQ	CP003034	Yersiniabactin-iron ABC transporter permease					
ybtX	CP003034	Yersiniabactin-iron transporter permease					
ycgV	NC_011751	Putative outer membrane autotransporter/Predicted adhesin. The YcgV-encoding gene was found in the majority of commensal and DEC genomes (it was truncated in both EHEC strains); however, the entire gene was missing from every single UPEC genome					
ydeR	CU928163	Fimbrial protein that is significantly phylogroup B1-associated					
yfhA							
ygeF	NC_002695	ETT2 gene					
ygeG	NC_002695	ETT2 gene					
ygeH	NC_002695	EilA and YgeH proteins show a moderate similarity to HilA and are encoded in pathogenicity islands from several *E. coli*strains, both pathogenic and non-pathogenic					
ygeI	NC_011750	ETT2 gene					
yghJ	KX245009	Secreted metalloprotease of pathogenic *E. coli*induces hemorrhagic fluid accumulation in mouse ileal loop					
ypjA	NC_011751	Identified as a protein with similarity to antigen 43 (Ag43), a self-recognizing surface adhesin					
yqfA	CP027060	Hemolysin III family protein					
yqgB	CP023258	Acid stress response protein					

1017

Paper III

Characterization of Clinical Heteropathogenic *Escherichia coli* Strains in Norway by Whole Genome Sequencing

Misti D. Finton[a]*, Roger Meisal[a,c], Davide Porcellato[a], Lin T. Brandal[b], and Bjørn-Arne Lindstedt[a]

[a] Faculty of Chemistry, Biotechnology and Food Science, Norwegian University of Life Sciences, Ås, Norway

[b] Department of Zoonotic, Food- and Waterborne Infections, Norwegian Institute of Public Health, Oslo, Norway

[c] Present address: Møreforsking Ålesund AS, Ålesund, Norway

E-mail addresses:

Misti D. Finton: misti.dawn.finton@nmbu.no

Bjørn-Arne Lindstedt: bjorn-arne.lindstedt@nmbu.no

Roger Meisal: roger.meisal@moreforsk.no

Davide Porcellato: davide.porcellato@nmbu.no

Lin T. Brandal: LinCathrineT.Brandal@fhi.no

* Corresponding author:

Misti D. Finton

Address: Norwegian University of Life Sciences

P.O. Box 5003

N-1432 Ås

Norway

ABSTRACT

Background

Heteropathogenic *Escherichia coli* (*E. coli*) are emerging public health threats that may cause a more severe infection with increased spreading potential. Intestinal pathogenic *E. coli* (IPEC) has traditionally been grouped into subtypes based on the presence or absence of specific molecular markers thought to be absent in extraintestinal pathogenic *E. coli* (ExPEC). However, the gene content of heteropathogens is a combination of multiple pathovars, ultimately blurring the classical distinction between intestinal and extraintestinal *E. coli*.

Methods

We examined a total of 13 fecal *E. coli* isolates collected from patients that exhibited signs of gastrointestinal disease in Norwegian hospitals and clinics. Eleven of the isolates were submitted to, and collected from, the Norwegian Institute of Public Health (NIPH) from clinical microbiological laboratories of Norway. The remaining isolates were a kind gift from Dr. Afset at the Department of Medical Microbiology at St. Olav's University Hospital in Trondheim, Norway. Samples were subjected to whole genome sequencing (WGS) through a combination of Oxford Nanopore's MinION and the Illumina MiSEQ. All data was analyzed for virulence associated genes (VAGs) associated with IPEC and ExPEC pathogenicity.

Results

The genetic targets STEC, aEPEC, tEPEC, ETEC, and EIEC were identified in 11 of the isolates, with one isolate containing the molecular markers of both STEC and EIEC. UPEC predictors *chuA*, *yfcV,* and *fyuA* were observed in 53.8% of the remaining isolates which classified them as STEC/UPEC, aEPEC/UPEC, and tEPEC/UPEC heteropathogens. All isolates harbored a wide range of ExPEC-associated virulence genes, including iron acquisition systems, protectins,

invasins, adhesins, and toxins. Additionally, one isolate did not contain any of the genetic targets used in the identification of enteropathogens, and solely harbored VAGs that classified the strain as UPEC.

Conclusion

Current diagnostic testing only searches for the established enteropathogenic VAGs in fecal samples from patients that exhibit gastrointestinal infection. Heteropathogenic strains may contain gene content that diverge from the commonly targeted enteric molecular markers, thus making the correct diagnostic classification difficult. It is critical that microbiological laboratories take heteropathogenic profiles into consideration during clinical diagnostic testing.

Keywords: *Escherichia coli*, ExPEC, IPEC, heteropathogenic, hybrid, Oxford Nanopore MinION

INTRODUCTION

Escherichia coli (*E. coli*) is a harmless member of human gut microbiota as well as a human pathogen that is estimated to cause more than two million deaths per year (Ochoa *et al.*, 2008) (Maltby *et al.*, 2013). Those that cause diarrheal disease are divided into 6 well-described variants of intestinal pathogenic *E. coli*, including diffusely adherent, enteroaggregative, shiga toxin-producing/enterohemorrhagic, enteroinvasive, enteropathogenic, and enterotoxigenic *E. coli* (DAEC, EAEC, STEC/EHEC, EIEC, EPEC, ETEC, respectively) (Nataro and Kaper, 1998) (Croxen *et al.*, 2013). Extraintestinal pathogenic *E. coli* (ExPEC) strains are the leading cause of community-acquired urinary tract infections (UTIs), the most common cause of gram negative bacterium-associated bacteremia, and a frequent cause of neonatal meningitis (Riley, 2014) (Poolman and Wacker, 2016) (Laupland and Church, 2014) (Korczak *et al.*, 2005). ExPEC pathovars include avian-pathogenic *E. coli* (APEC), neononatal meningitis *E. coli* (EPEC), sepsis-associated *E. coli*, and uropathogenic *E. coli* (UPEC) (Sarowska *et al.*, 2019). Molecular targets commonly used for IPEC include the Afa/Dr adhesins for DAEC, the *aggR* gene for EAEC, genes encoding the thermostable (ST) and/or thermolabile (LT) toxins for ETEC, *eae* and *bfp* genes for tEPEC (aEPEC solely carries *eae*), *stx1* and *stx2* genes for STEC/EHEC, and the *ipaH* gene for EIEC. Unlike most of the IPEC pathovars, there are not a set of ExPEC VAGs that are universally accepted as unique genetic markers to positively identify each pathotype. VAGs associated with extraintestinal survival and disease have been identified amongst all ExPEC, and include those that enable iron acquisition, autotransporters, adhesins, protectins, toxins, and invasins. However, the possession of any two of the genes *chuA, yfcV*, or *vat,* along with *fyuA* has been reported as a predictor of UPEC strains (Spurbeck *et al.*, 2012).

The typical VAGs that are used to classify these pathogens are frequently located on mobile genetic elements (MGEs) that are capable of transferring between bacteria, thus facilitating the development of novel combinations of known virulence factors into a single strain of *E. coli*, often referred to as a "heteropathogen" or "hybrid" (Kaper et al, 2004). The emergence of heteropathogens have serious clinical implications that may result in a more severe disease that is capable of increased spreading potential, as was demonstrated during the 2011 STEC/EAEC heteropathogen outbreak in Germany that spread to other European countries and North America. Roughly 4000 people were affected, over 900 developed HUS, and 54 individuals died (Karch et al, 2012).

The high prevalence of clinical heteropathogenic *E. coli* in our previous study (Lindstedt *et al.*, 2018) prompted us to further investigate 13 clinical isolates submitted from patients that showed signs of gastrointestinal infection. Whole genome sequencing (WGS) was conducted through a combination of Oxford Nanopore's MinION and the Illumina MiSEQ and data was analyzed for virulence associated genes (VAGs) associated with IPEC and ExPEC pathogenicity.

MATERIALS AND METHODS

Bacterial isolates.

Eleven *E. coli* isolates were obtained from the culture collection at the National Reference Laboratory for Enteropathogenic Bacteria at the Norwegian Institute of Public Health (NIPH) and two isolates were a kind gift from Dr Jan E. Afset at St. Olav's Hospital, Trondheim, Norway.

Oxford Nanopore MinION sequencing. FHI_NMBU_02, FHI_NMBU_03, FHI_NMBU_04 and FHI_NMBU_05 were sequenced as previously published (Lindstedt *et al.*, 2018). DNA library for the remaining isolates was prepared using the SQK-RBK004 Rapid Barcoding Kit and loaded into the MinION SpotON R9.4.1 flow cell in a MinION MK1 sequencer. The 72-h live basecalling sequencing protocol and barcode demultiplexing was selected in the MinKNOW software and allowed to continue after the conclusion of the sequencing run. Individual barcodes were further processed through the Antimicrobial Resistance Mapping Application (ARMA) to identify the resistance genes of each sample. All runs were according to the standard protocol of Oxford Nanopore Technologies (Oxford, UK).

Illumina MiSeq sequencing and assembly. Paired-end libraries (2 x 300 bp) were prepared with Nextera™ DNA Flex Tagmentation (Illumina Inc., San Diego, CA, USA) and sequenced on an Illumina MiSeq platform using a 2 x 300 paired-end approach with v3 chemistry by the Norwegian Sequencing Center (Oslo, Norway).

Sequence Analysis. FHI_NMBU_02, FHI_NMBU_03, FHI_NMBU_04 and FHI_NMBU_05 were placed into contigs as previously described (Lindstedt *et al.*, 2018). The remaining 9 isolate's FastQ reads within each MinION barcode was concatenated using the Cygwin cat command. The online Galaxy platform was used to remove adapters from the MiSeq and MinION data with Trimmomatic and Porechop, respectively. MinION data was filtered by sequence length with a minimum threshold set at 800bp and above. MiSeq and MinION data were combined in Unicycler to assemble sequences into contigs. Additionally, the sequence data was annotated using the NCBI Prokaryotic Genome Annotation Pipeline (Angiuoli *et al.*, 2008)

and The RAST Annotation Server (Aziz, *et al.*, 2008). To assess the genomes for antibiotic resistance genes and virulence factors (VFs), ResFinder v 3.2, and VirulenceFinder v 2.0 servers were used (Center for Genomic Epidemiology, Technical University of Denmark, Lyngby, Denmark) with the following settings: selected ID threshold 95.0%, selected minimum length 60%. The ABRicate Mass screening of contigs for antimicrobial and virulence genes (Galaxy Version 0.9.8) tool combined with the NCBI National Database of Antibiotic Resistant Organisms were also used. Additionally, the Comprehensive Antibiotic Resistance Database (CARD) was used to search the genome for acquired antibiotic resistance genes. Whole genome sequencing data was examined for MLST-type, virulence factors (VFs), and plasmids (nucleotide identity threshold of 95% and minimum length of 60%) with the Center for Genomic Epidemiology (CGE), Technical University of Denmark (http://www.genomicepidemiology.org). In addition, a custom-made virulence database of over 760 genes was utilized to assess for VFs with the CGE website. Phylogrouping was performed using the Clermon Typing webpage (http://clermontyping.iame-research.center/index.php). SNP-typing was performed using CSI Phylogeny 1.4 (https://cge.cbs.dtu.dk/services/CSIPhylogeny/).

RESULTS

Characterization of MLST and Serotype

IPEC strains are generally classified as phylogenetic groups A and B1, while ExPEC strains are typically classified in phylogroups groups B2 and D. In the present study, 61.5% of the isolates belonged to phylogroups B2 and D, while only 30.8% belonged to phylogroups A and B1 (Table 1). FHI_NMBU_05 belonged to G, which is a group intermediate between the F and B2 phylogroups. Those that belonged to phylogroup D harbored the highest number of VAGs,

followed by group B2, B1, and A. Additionally, all aEPEC/UPEC and tEPEC/UPEC heteropathogens were solely categorized as phylogroup B2.

Isolate	Origin	Serotype	MLST	Phylo.	Pathotype	Plasmid incompatibility groups detected
FHI_NMBU_02	NIPH	O1:K1:H7	95	B2	UPEC	ND
FHI_NMBU_03	NIPH	O81:H6	28	B2	aEPEC/UPEC/ NMEC/APEC	IncFIB(AP001918)
FHI_NMBU_04	NIPH	O128ab:H2	4748	B1	STEC/ExPEC	IncFIB(AP001918), IncFII
FHI_NMBU_05	NIPH	O161:H4	117	G	UPEC	IncFIB(AP001918), IncFII
FHI_NMBU_06	NIPH	O156:H1	941	B2	tEPEC/UPEC	IncFIB(AP001918), IncY
FHI_NMBU_07	NIPH	O145:H28	32	D	STEC/EIEC	P0111
FHI_NMBU_08	NIPH	O137:H6	2678	B2	aEPEC/UPEC	IncY
FHI_NMBU_09	NIPH	O145:H28	137	D	STEC/EIEC	Col156, IncFIB, Incl2, IncY
FHI_NMBU_10	NIPH	O146:H21	829	B1	STEC/ExPEC	Col(MP18), IncFII
FHI_NMBU_11	NIPH	O128ab:H2	25	B1	STEC/ExPEC	IncFIB, IncFII
FHI_NMBU_12	NIPH	O78:H10	173	A	ETEC/ExPEC	IncFII
Trh50	St. Olav's	O81:H6	28	B2	aEPEC/UPEC	IncFIB(AP001918)
Trh52	St. Olav's	O84:H7	28	B2	aEPEC/UPEC	IncFII

Table 1. Genotypic characteristics of isolates, origin, pathotype and plasmids.

Amongst the 13 isolates, 10 different STs and 9 different serotypes were identified. ST28 was the most frequently observed and was classified as aEPEC/UPEC heteropathotype of phylogroup B2. In a previous study it was found that ST28 was detected in 56 aEPEC isolates from fecal samples of children < 5 years old that showed signs of gastrointestinal infection in Norway (Afset *et al.*, 2008). In the present study, FHI_NMBU_03, Trh50, Trh52 were of ST28, however only Trh50 and Trh52 were from children <5, while FHI_NMBU_03 was collected from an adult fecal sample. Another previous study found an association between *E. coli* O1:K1:H7 of ST95 and phylogroup B2 with APEC pathogenicity, which matches the attributes of FHI_NMBU_02 [O1:K1:H7 (ST95 – phylogroup B2)], and may serve as an indication that host specificity does not apply to isolates grouped into the same classifications (Mora *et al.*, 2009). Additionally, *E.*

coli O1:H7 has also been frequently implicated in neonatal meningitis, UTIs, and septicemia in humans (Mora *et al.*, 2009). It was also reported that APEC isolates are strongly associated with serogroup O78 and ST117. Furthermore, a study by Ronco *et al.* determined that those isolates also carried numerous VAGs associated with human UPEC, similar to FHI_NMBU_05 [O78:H4 (ST117 –phylogroup G) (Ewers *et al.*, 2007) (Olsen *et al.*, 2011) (Ronco, *et al.*, 2017). Additionally, FHI_NMBU_08 [O137:H6 (ST2678 – phylogroup B2)] was classified as an aEPEC/STEC, and similar heteropathogenic EPEC/STEC strains of ST2678 and serotype O137:H6 were isolated from small-sized psittacine birds raised as pets (Gioia-Di Chiacchio *et al.*, 2018).

Identification of virulence and resistance genes

In this study, the established genetic targets of the 6 IPEC pathovars were used as a reference to classify the 13 isolates. In total, ten strains were identified as either STEC, aEPEC, tEPEC, or ETEC (38.5%, 30,8%, 7.7%, 7.7%, respectively). Additionally, two isolates (15.4%) contained the hallmarks of both STEC and EIEC, thus being classified as a heteropathogen of two IPEC pathovars. An interesting finding was that, regardless of causing symptoms of gastrointestinal disease, FHI_NMBU_02 and FHI_NMBU_05 did not harbor any of the typical IPEC molecular targets that would be used in diagnostic testing. Both FHI_NMBU_02 and FHI_NMBU_05 carried numerous ExPEC VAGs, including those that qualified them as UPEC. FHI_NMBU_05 contained the IPEC associated genes *cfaB, ehaB, iha,* and *lpfA*, which may have contributed to the symptoms of gastrointestinal infection in the patient (supplementary table 1).

As the isolates were recovered from patients that exhibited signs of gastrointestinal infection, the presence of ExPEC VAGs were of interest. WGS revealed the presence of a wide variety of

VAGs associated with extraintestinal survival and disease throughout all of the isolates, including autotransporters (*vat, ehaA, ehaG*), adhesins (EAST1/*astA, papC*), protectins (*iss, traT)* toxins (*hlyA*), iron acquisition (*chuA, iutA, irp1, irp2, sitA, iucD, iroC, fbpB, fyuA*) and invasins (*ibeA*). In total, 53.8% of the strains contained the genes necessary to identify as UPEC and three (42.9%) of those isolates contained the pyelonephritis-associated pilus *pap* operon, which has been reported in more than 80% of all strains of *E. coli* that cause pyelonephritis and is recognized as a key determinant in promoting the virulence of UTIs (Tewari *et al.*, 1994). FHI_NMBU_04 contained the aerobactin (*iutA*, *iucABCD*), sit (*sitABCD*) and salmochelin (*iroBCDEN*) operons, which are strongly associated with ExPEC virulence as they have been reported in high frequency on the virulence plasmids of UPEC, APEC and NMEC (Johnson *et al.*, 2008). An interesting finding was that the *iro* locus encoding salmochelins is extremely rare in intestinal *E. coli*, however it was found in FHI_NMBU_03, FHI_NMBU_04, FHI_NMBU_05, Trh50, and Trh52 (Okeke *et al.*, 2004). Additionally, FHI_NMBU_03, FHI_NMBU_06, FHI_NMBU_08, Trh50 and Trh52 contained all three of the genes *malX, usp,* and *fyuA* which is positively associated with mortality and may underscore the severity of an infection with these strains (Johnson *et al.*, 2006). The genomic sequences of FHI_NMBU_04, FHI_NMBU_10, FHI_NMBU_11, FHI_NMBU_12, contain a gene cluster predicted to encode a T3SS named $ETT2_{sepsis}$, which is thought to be involved in attachment and effacement (Ideses *et al*, 2005). The three strains carry a deleted $ETT2_{sepsis}$ version where *kduE*, *epaP*, *yqeF*, and *pkgA* are missing and the genes between *eivA-G* are deleted, though *eivH* is present. Although, this region is degenerate, studies have shown that it still contributes to pathogenesis (Ideses *et al.*, 2005). This may be due to the possibility that some areas of the type III secretion system partake in other roles that add to virulence, even though the complex as a whole is non-functional.

From SNP analysis and virulence factor content, it was determined that FHI_NMBU_03, FHI_NMBU_06, FHI_NMBU_08, Trh52 and Trh50 belong to the same cluster of strains. These isolates were classified as phylogroup B2 and contained the *eae* gene, but not genes for *bfp*, thus categorizing these strains as aEPEC. Four (80.0%) isolates contained the *eae*-β2, while one (20.0%) harbored *eae*- ζ3. All isolates contained intestinal pathogenic VAGs including the EHEC autotransporter B gene, *ehaB*, and the non-LEE-encoded type III translocated virulence factors, *nleABCEGH*, as well as the *yjaA* gene involved in acid stress. FHI_NMBU_03, Trh50, and Trh52 also contained K88 reported to be associated with ETEC strains. The five aEPEC isolates harbored *chuA, yfcV,* and *fyuA*, making them true UPEC/aEPEC heteropathogens. In addition, the uropathogenic specific protein (*usp*), heme utilization carrier protein (*hutX*), Hemolysin III, and the yersiniabactin HPI operon (*ybtAEQSTUI*) were detected, further highlighting potential ExPEC pathogenicity. The five isolates also contained the *ibeABC*genes, which is a critical determinant of *E. coli* K1 to facilitate bacterial penetration through the blood-brain barrier in NMEC strains. Iron uptake is extremely important to survival when located in extraintestinal areas, therefore the presence of iron-regulated genes (*fyuA, irp1, irp2*, and *ybtAEQSTU*) and iron chelator genes (*entABCDEF* and *fepABCDE*) was further evidence of extraintestinal pathogenic potential. Additionally, FHI_NMBU_03, FHI_NMBU_08, Trh50, and Trh52 carried the autotransporter adhesin and virulence factor (*aatB*) associated with APEC pathogenicity, the *iss* protein that protects against phagocytosis and the bactericidal action of serum frequently found in APEC, NMEC, and SEPEC strains (Sarowska *et al.*, 2019), and the autotransporter *upaB* which contributes to UPEC colonization of the urinary tract (Allsopp *et al.*, 2011). Another interesting finding was the presence of bipolar helicase HerA which seems to be very rare among *E. coli*, however all five isolates in this cluster contained this gene as well as

FHI_NMBU04 and FHI_NMBU_11. The virulence plasmids of FHI_NMBU_03, Trh50 and Trh52 carried a large ORF encoding a Cyclic β-1,2-glucan synthase, which is present in a restricted number of symbiotic or pathogenic bacteria. When a nBLAST was performed only one *E. coli* was reported carrying a similar ORF (chromosomally located) at 90.37% identity (*Escherichia coli* strain NCTC11104 genome assembly, chromosome: 1).

Antibiotic resistance genes

One β-lactamase gene, $bla_{\text{TEM-1B}}$, was identified in FHI_NMBU_05. All of the *E. coli* strains harbored a genetic region containing an AmpC/CMY-like gene, *blc* and *sugE* as previously described (Verdet *et al.*, 2009) (Singh *et al.*, 2018). Resistance mechanisms to non-β lactam drugs were also identified, including determinants against aminoglycosides [*aadA1* and *aph(6)-Id*], bicyclomycin resistance (*bcr*), fosmidomycin (*fsr*); macrolides (*macA* and *mphA*), sulfonamides (sul1), tetracyclines [*tet*(*A*)] and trimethoprims (*dfrA*). Efflux pumps that confer resistance to multiple classes of antibiotics were also observed (*acrAB, emrAB, mdtH, mdfA*).

		FHI_NMBU_02	FHI_NMBU_03	FHI_NMBU_04	FHI_NMBU_05	FHI_NMBU_06	FHI_NMBU_07	FHI_NMBU_08	FHI_NMBU_09	FHI_NMBU_10	FHI_NMBU_11	FHI_NMBU_12	Trh50	Trh52
β-lactamases	TEM-1B				■									
Aminoglycosides	*aadA1*											■		
	aph(6)-Id			■										
Macrolides	*macA*	■	■	■	■	■	■	■	■	■	■	■	■	■
Polymyxins	*arnA*	■	■	■	■	■	■	■	■	■	■	■	■	■
	pmrK											■		
Trimethoprim	*dfrA1*											■		
Sulfonamides	*sul1*											■		
Tetracyclines	*tet(A)*				■									
Fosmidomycin	*fsr*	■	■	■	■	■	■	■	■	■	■	■	■	■
Cadmium/Zinc/Lead	*zntA*	■	■	■	■	■	■	■	■	■	■	■	■	■
Copper	*copA*	■	■	■	■	■	■	■	■	■	■	■	■	■
	cusA	■	■	■	■	■	■	■	■	■	■	■	■	■
Quaternary ammonium	*sugE*	■	■	■	■	■	■	■	■	■	■	■	■	■
Tellurite	*tehA*	■	■	■	■	■	■	■	■	■	■	■	■	■
Bicyclomycin	*bcr*	■	■	■	■	■	■	■	■	■	■	■	■	■
Multidrug efflux pumps	*emrABKY*	■	■	■	■	■	■	■	■	■	■	■	■	■
	mdf(A)	■	■	■	■	■	■	■	■	■	■	■	■	■
	MexAB	■				■	■	■	■					■
	mdtA	■	■	■	■	■	■	■	■	■	■	■	■	■
	mdtH	■	■	■	■	■	■	■	■	■	■	■	■	■
	acrAB	■	■	■	■	■	■	■	■	■	■	■	■	■

Table 2. Resistant genes identified throughout the isolates. Present resistance determinants are represented by a grey box.

Nucleotide sequence accession numbers.

Isolate	Accession Numbers
FHI_NMBU_02	JAAAGC000000000
FHI_NMBU_03	CP019455 and CP019456
FHI_NMBU_04	WXYX00000000
FHI_NMBU_05	WXYY00000000
FHI_NMBU_06	WXYW00000000
FHI_NMBU_07	WXYZ00000000
FHI_NMBU_08	JAAALJ000000000
FHI_NMBU_09	WXYV00000000
FHI_NMBU_10	CP046539 through CP046545
FHI_NMBU_11	CP045712 - CP045717CP
FHI_NMBU_12	WXZA00000000
Trh50	JAAALI000000000
Trh52	CP047600 and CP047601

DISCUSSION

Previous studies have reported heteropathogenic *E. coli* in the environment (Gati *et al.*, 2019), clinical settings (Lindstedt, *et al.*, 2018) (Khairy *et al.*, 2019), livestock (Johura *et al.*, 2017), domestic animals (Johura *et al.*, 2017), and wildlife (Bai *et al.*, 2019). In the present study, we observed aEPEC/UPEC, tEPEC/UPEC, ETEC/ExPEC, STEC/UPEC, and STEC/EIEC heteropathogens. EPEC are among the most important pathogens causing persistent diarrhea in children and adults worldwide (Ochoa *et al*, 2008). There are two distinct groups of EPEC, with atypical EPEC (aEPEC) as more prevalent than typical EPEC (tEPEC), and aEPEC being increasingly associated with outbreaks (Ochoa *et al.*, 2011). ETEC is a primary cause of diarrhea in the developing world, particularly in children, and is the leading cause of traveler's diarrhea. EIEC is closely related to *Shigella* spp. through virulence, phenotypic properties, and that the

clinical illness caused is similar between both types (Lan *et al.*, 2004). STEC infection may cause more severe gastrointestinal illness associated with a greater health impact that may lead to lethal complications, including hemolytic uremic syndrome (HUS). Infections with heteropathogenic IPEC strains can then migrate to other anatomical sites of the body to cause a bacterial infection. For instance, the human intestinal tract is the main reservoir of UPEC, and due to the proximity of the urinary tract to the rectum, these heteropathogenic strains may also be capable of causing an opportunistic infection by ascending the urethra into the bladder. Occasionally, the bacteria may even migrate to the kidneys and cause pyelonephritis.

In our previous study, we isolated several strains of extended-spectrum β-lactamase-producing heteropathogenic *E. coli* from environmental soil and aquatic locations in Norway (manuscript in prep – accession numbers found in supplementary table 2). The group of strains comprising FHI_NMBU_03, Trh50, Trh52, FHI_NMBU_08 and FHI_NMBU_06 was of particular interest as they have such a diverse VAG repertoire being truly at the crossroads of the IPEC and ExPEC designations. The closest grouped strains in this group FHI_NMBU_03, Trh50 and Trh52 are isolated over several years in Norway at different locations, and share some rather unusual genes e.g. genes encoding Cyclic-1,2-glucan synthase, the rare Xhla hemolysin and *herA* bipolar helicase, where the possible contribution to virulence has to be further investigated. One notable difference in the group of FHI_NMBU_03 like strains is that Trh50, Trh52 and FHI_NMBU_08 harbor the gene cluster for Cytolethal distending toxin (cdt), while FHI_NMBU_03 and FHI_NMBU_06 are negative. How cdt may contribute to virulence in this group is at present uncertain. In the present study, only one isolate (FHI_NMBU_05) was found to harbor the β-lactamase TEM-1, which has the ability to hydrolyze penicillins, rendering them ineffective. When considering the resistance profiles of our isolates, the emergence of STEC/ExPEC

heteropathogens raises questions on the ability to pursue antibiotic therapy. Due to the production of *Stxs*, antibiotics aren't a recommended path of therapy, as they may increase production of toxins and escalate the risk of HUS (Smith *et al.,* 2012). In contrast, ExPEC infections range from mild (bladder inflammation) to life threatening (sepsis, neonatal meningitis), and antibiotics are frequently prescribed. Therefore, a bacterial infection with a heteropathogenic STEC/ExPEC strain may cause serious clinical complications due to a heightened risk in the use of an antibiotic to treat an ExPEC infection.

Clinical microbiological laboratories commonly use classic PCR or real-time PCR assays to identify enteric pathogens by their genetic targets. However, heteropathogens may possess unique virulence profiles that don't fit the stereotypical genetic targets used in microbiological laboratory testing, making the correct diagnostic classification difficult. In the present study, FHI_NMBU_02 was initially classified as aEPEC by microbiological laboratories, however, WGS results indicated that it was a mix of two *E. coli* strains. Upon the isolation of a fresh sample, it was revealed that the original *eae* gene (which is necessary to classify this isolate as aEPEC) was from a contaminant strain. Therefore, FHI_NMBU_02 did not harbor any of the molecular targets used in clinical diagnostics and was categorized solely as UPEC. Similar isolates that solely carry ExPEC VAGs may evade proper identification and be classified as a non-enteropathogenic or a commensal strain, complicating further course of action. In addition, FHI_NMBU_12 had been classified as NTEC, but it was determined that this strain did not possess any of the necessary VAGs to be classified as such, and was instead classified as ETEC, further elucidating how difficult clinical diagnostic testing may be.

CONCLUSION

It is critical to take notice of hybrid strains, as they appear to be more common than previously thought and are able to cause infection in multiple anatomical sites of the body. Due to the severity of disease that may be caused by heteropathogenic *E. coli,* these strains should be of increased concern in the development of pathogen detection methods. Given the public health significance of heteropathogenic *E. coli*, it is suggested that a wider range of genetic targets be considered for routine clinical diagnostic testing, or that WGS approaches are combined with methodology to rapidly detect a wide array of VAGs (including both IPEC and ExPEC associated genes).

REFERENCES

Ochoa TJ, Barletta F, Contreras C, Mercado E. New insights into the epidemiology of enteropathogenic *Escherichia coli* infection. Trans R Soc Trop Med Hyg. 2008; **102**(9): 852-856.

Maltby R, Leatham-Jensen MP, Gibson T, Cohen PS, Conway T. Nutritional Basis for Colonization Resistance by Human Commensal *Escherichia coli* Strains HS and Nissle 1917 against *E. coli* O157:H7 in the Mouse Intestine. PLoS One. 2013; **8**(1): e53957.

Nataro JP and Kaper JB. Diarrheagenic *Escherichia coli*. Clin Microbiol Rev. 1998; **11**(1): 142-201.

Croxen MA, Law RJ, Scholz R, Keeney KM, Wlodarska M, Finlay BB. Recent advances in understanding enteric pathogenic *Escherichia coli*. Clin Microbiol Rev. 2013; **26**(4): 822–80.

Riley LW. Pandemic lineages of extraintestinal pathogenic *Escherichia coli*. Clin Microbiol Infect. 2014; **20**(5): 380-390.

Poolman JT and Wacker M. Extraintestinal Pathogenic *Escherichia coli*, a Common Human Pathogen: Challenges for Vaccine Development and Progress in the Field. J Infect Dis. 2016; **213**(1): 6-13.

Laupland KB and Church DL. Population-based epidemiology and microbiology of community-onset bloodstream infections. Clin Microbiol Rev. 2014; **27**(4): 647-664.

Korczak B, Frey J, Schrenzel J, Pluschke G, Pfister R, Kuhnert P, *et al.* Use of Diagnostic Microarrays for Determination of Virulence Gene Patterns of *Escherichia coli* K1, a Major Cause of Neonatal Meningitis. Journal of Clinical Microbiology. 2005; **43**(3): 1024-1031.

Sarowska J, Futoma-Koloch B, Jama-Kmiecik A, Frej-Madrzak M, Ksiazczyk M, Choroszt-Krol I. Virulence factors, prevalence and potential transmission of extraintestinal pathogenic *Escherichia coli* isolated from different sources: recent reports. Gut Pathog. 2019. **11**(10).

Spurbeck RR, Dinh PC, Walk ST, Stapleton AE, Hooton TM, Mobley HL. *Escherichia coli* isolates that carry vat, fyuA, chuA, and yfcV efficiently colonize the urinary tract. Infect Immun. 2012; **80**(12): 4115-4122.

Kaper JB, Nataro JP, Mobley HL. Pathogenic *Escherichia coli*. Nature Reviews Microbiology. 2004; **2**: 123–140.

Karch H, Denamur E, Dobrindt U, Finlay BB, Hengge R, Vincente M. The enemy within us: lessons from the 2011 European *Escherichia coli* O104:H4 outbreak. EMBO Mol Med. 2012; **4**(9): 841-848.

Lindstedt BA, Finton MD, Porcellato D, Brandal LT. 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 Infect Dis. 2018; **18**(1):544.

Angiuoli SV, Gussman A, Klimke W, Cochrane G, Field D, White O, *et al.* Toward an online repository of Standard Operating Procedures (SOPs) for (meta)genomic annotation. OMICS. 2008; **12**(2): 137-141.

Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Zagnitko O. The RAST Server: Rapid Annotations using Subsystems Technology. BMC Genomics. 2008; **9**(75).

Afset JE, Anderssen E, Bruant G, Harel J, Wieler L, Bergh K. Phylogenetic Backgrounds and Virulence Profiles of Atypical Enteropathogenic *Escherichia coli* Strains from a Case-Control Study Using Multilocus Sequence Typing and DNA Microarray Analysis. J Clin Microbiol. 2008; **46**(7): 2280-2290.

Mora A, López C, Dabhi G, Blanco M, Blanco JE, Blanco J. Extraintestinal pathogenic *Escherichia coli* O1:K1:H7/NM from human and avian origin: detection of clonal groups B2 ST95 and D ST59 with different host distribution. BMC Microbiol. 2009; 9(132).

Ewers C, Li G, Wilking H, Kiessling S, Alt K, Wieler LH, *et al.* Avian pathogenic, uropathogenic, and newborn meningitis-causing *Escherichia coli*: How closely related are they? Int J Med Microbiol. 2007; 297:163–76.

Olsen RH, Stockholm NM, Permin A, Christensen JP, Christensen H, Bisgaard M. Multi-locus sequence typing and plasmid profile characterization of avian pathogenic *Escherichia coli* associated with increased mortality in free-range layer flocks. Avian Pathol. 2011; 40(5): 437–444.

Ronco T, Stegger M, Heidemann Olsen R, Sekse C, Nordstoga AB, Pederson K. Spread of avian pathogenic *Escherichia coli* ST117 O78:H4 in Nordic broiler production. BMC Genomics. 2017; **18**(13).

Gioia-Di Chiacchio RM, Cunha MPV, Marques de Sá LR, Davies YM, Pacheco Pereira CB, Knöbl T. Novel Hybrid of Typical Enteropathogenic *Escherichia coli* and Shiga-Toxin-Producing *E. coli* (tEPEC/STEC) Emerging From Pet Birds. Font Microbiol. 2018; **6**(9): 2975.

Tewari R, Ikeda T, Malaviya R, MacGregor JI, Little JR, Abraham SN: The PapG tip adhesin of P fimbriae protects *Escherichia coli* from neutrophil bactericidal activity. Infect Immun 1994; 62(12): 5296-5304.

Johnson TJ, Wannemuehler YM, Nolan LK. Evolution of the iss gene in *Escherichia coli*. Appl Env Microbiol. 2008;74: 2360–2369.

Okeke IN, Scaletsky IC, Soars EH, Macfarlane LR, Torres AG. Molecular epidemiology of the iron utilization genes of enteroaggregative *Escherichia coli*. 2004; 42(1): 36-44.

Johnson JR, Clermont O, Menard M, Kuskowski MA, Picard B, Denamur E. Experimental Mouse Lethality of *Escherichia coli* Isolates, in Relation to Accessory Traits, Phylogenetic Group, and Ecological Source. The Journal of Infectious Diseases. 2006; 194(8): 1141-1150

Ideses D, Gophna U, Paitan Y, Chaudhuri RR, Pallen MJ, Ron EZ. A Degenerate Type III Secretion System from Septicemic *Escherichia coli* Contributes to Pathogenesis. J Bacteriol. 2005; **187**(23): 8164-8171.

Allsopp LP, Beloin C, Ulett G, Valle J, Totsika M, Schembri MA. Molecular Characterization of UpaB and UpaC, Two New Autotransporter Proteins of Uropathogenic *Escherichia coli* CFT073. Infect. Immun. 2012; **80**(1): 321-332.

Verdet C, Gautier V, Chachaty E, Ronco E, Hidri N, Decré D, Arlet G. Genetic Context of Plasmid-Carried blaCMY-2-Like Genes in Enterobacteriaceae. Antimicrob Agents Chemother. 2009; **53**(9): 4002-4006.

Singh NS, Singhal N, Virdi JS. Genetic Environment of $bla_{\text{TEM-1}}$, $bla_{\text{CTX-M-15}}$, $bla_{\text{CMY-42}}$ and Characterization of Integrons of *Escherichia coli* Isolated From an Indian Urban Aquatic Environment. Frontiers in Microbiology. 2018; **9**(382).

Gati NS, Middendorf-Bauchart, Bletz S, Dobrindt U, Mellmann A. Origin and Evolution of Hybrid Shiga Toxin-Producing and Uropathogenic *Escherichia coli* Strains of Sequence Type 141. J Clin Microbiol. 2019; **58**(1): e01309-19.

Khairy RM, Mohamed ES, Abdel Ghany HM, Abdelrahim SS. Phylogenic classification and virulence genes profiles of uropathogenic *E. coli* and diarrhegenic *E. coli* strains isolated from community acquired infections. PLoS One. 2019; **14**(9): e0222441.

Johura FT, Parveen R, Islam A, Sadique A, Rahim MN, Alam M. Occurrence of Hybrid *Escherichia coli* Strains of Carrying Shiga Toxin and Heat-Stable Toxin in Livestock of Bangladesh. Front Public Health. 2017; **4**(287).

Bai X, Zhang J, Ambikan A, Jernberg C, Ehricht R, Matussek A, *et al.* Molecular Characterization and Comparative Genomics of Clinical Hybrid Shiga Toxin-Producing and Enterotoxigenic *Escherichia coli* (STEC/ETEC) Strains in Sweden. 2019; **9**(5619).

Ochoa TJ, Contreras CA. Enteropathogenic *Escherichia coli* infection in children. Curr. Opin. Infect. Dis. 2011; **24**:478–483.

Lan R, Alles MC, Donohoe K, Martinez MB, Reeves PR. Molecular evolutionary relationships of enteroinvasive *Escherichia coli* and Shigella spp. Infect Immun. 2004; **72**(9): 5080-5088.

Smith KE, Wilker PR, Reiter PL, Hedican EB, Bender JB, Hedberg CW. Antibiotic treatment of *Escherichia coli* O157 infection and the risk of hemolytic uremic syndrome, Minnesota. Pediatr Infect Dis J. 2012; **31**(1): 37-41.

499 **Supplemental table 1**. Virulence gene profiles. Gray blocks represent the presence of a
500 virulence gene.

Virulence Genes	Description	Pathotype	FHI_NMBU_02	FHI_NMBU_03	FHI_NMBU_04	FHI_NMBU_05	FHI_NMBU_06	FHI_NMBU_07	FHI_NMBU_08	FHI_NMBU_09	FHI_NMBU_10	FHI_NMBU_11	FHI_NMBU_12	Trh50	Trh52
aatB	Autotransporter adhesin	APEC	■	■		■			■					■	■
Aec77	Gene on a selC-Associated Genomic Island	APEC	■		■	■		■		■	■	■	■		
Antigen 43	Self-recognizing surface adhesin	IPEC and ExPEC	■	■		■				■	■	■	■	■	
Antigen 43-A (Variant A)	Promotes long-term persistence in the urinary bladder	IPEC and ExPEC			■										
aslA	Fimbrial protein gene	ExPEC	■	■	■	■	■	■	■	■	■	■	■	■	■
aufC	'Another UPEC fimbriae' that may play a role in UTIs	UPEC	■	■	■										
Bfp	Bundle-forming pili	EPEC					■								
bor	*iss* homologue	ExPEC		■										■	■
cadA	Involved in acid stress resistance	IPEC	■	■		■	■	■	■	■	■	■	■	■	■
cas1	CRISPR-associated endonuclease	IPEC/ExPEC	■		■	■	■	■		■	■	■	■		
cas2	CRISPR-associated endonuclease	IPEC/ExPEC				■	■	■		■	■	■	■		
casA	CRISPR system cascade subunit	IPEC/ExPEC			■	■	■	■		■	■	■	■		
cesAB	Chaperone for the type-III translocator proteins EspA and EspB	EPEC/STEC		■			■	■	■	■				■	■
cesD2	The second chaperone for the type III secretion translocator protein EspD	EPEC/STEC		■			■	■	■	■				■	■
cesF	T3SS chaperone	EPEC/STEC		■			■	■	■	■				■	■
cesT	Chaperone for the type III secretion of Tir	EPEC/STEC		■			■	■	■	■				■	■
cfaB	Colonization factor antigen	ETEC		■	■	■	■		■		■	■	■	■	
cfaE	CFA/I fimbriae tip unit	ETEC		■	■				■		■		■		
chuA	Heme uptake	*E. coli* Isolates That Carry *vat, fyuA, chuA,* and *yfcV* efficiently colonize the urinary tract	■	■		■	■	■	■	■				■	■
cif	Cell cycle inhibiting factor	EPEC/STEC		■				■		■				■	■
cirA	Ferric iron-catecholate out membrane transporter	ExPEC	■	■	■	■	■	■	■	■	■	■	■	■	■
Cnf2	Gene encoding the cytotoxic necrotizing factor 2	ExPEC											■		
Colicin 10	Polypeptide toxins produced by and active	IPEC and ExPEC				■	■	■			■	■			

	against *E. coli* and closely related bacteria														
csgA	Curli major subunit protein	IPEC and ExPEC													
csgB	Curli nucleator protein	IPEC and ExPEC													
csgC	Periplasmic chaperone	IPEC and ExPEC													
csgD	A regulator that controls the transcription of curli genes	IPEC and ExPEC													
csgE	Curli accessory factor	IPEC and ExPEC													
csgF	Curli accessory factor	IPEC and ExPEC													
csgG	An outer membrane secretion pore essential for curli biogenesis	IPEC and ExPEC													
cvaA	Colicin V secretion protein	IPEC and ExPEC													
cvi	Colicin V immunity protein	IPEC and ExPEC													
eae	Intimin	EPEC/STEC		**β 2**			ζ 3	γ	**β 2**	γ				**β 2**	**β 2**
EAST1 (astA)	Heat-stable enterotoxin 1	IPEC, but doesn't play a substantial role in ETEC													
eatA	Serine protease activity	ETEC													
ecf1	A plasmid gene highly associated with *E. coli* strains that are positive for Stx, *eae*, and *ehxA*	STEC/EHEC													
efa1	Adherence factor	IPEC													
ehaA	Autotransporter protein that contributes to adhesion and biofilm formation	EHEC/STEC													
ehaB	Adhesin	EHEC/STEC													
ehaG	Mediates specific adhesion to colorectal epithelial cells	EHEC/STEC													
ehxA	Plasmid-encoded enterohemolysin	IPEC													
eivA	ETT2	IPEC and ExPEC													
eivC	ETT2	IPEC and ExPEC													
eivE	ETT2	IPEC and ExPEC													
eivF	ETT2	IPEC and ExPEC													
eivG	ETT2	IPEC and ExPEC													
eivH	ETT2	IPEC and ExPEC													
eivI	ETT2	IPEC and ExPEC													
eivJ	ETT2	IPEC and ExPEC													

elaD	Present in all IPEC strains but absent from ExPEC strains	IPEC													
eltAB	Enterotoxin LT-AB	ETEC													
entA	Enterobactin biosynthesis gene, upregulated in vivo during UTI	UPEC													
entE	Involved in the biosynthesis of the siderophore enterobactin (enterochelin)	UPEC													
entH	Required for optimal enterobactin synthesis	UPEC													
epaO	ETT2	IPEC and ExPEC													
epaP	ETT2	IPEC and ExPEC													
epaQ	ETT2	IPEC and ExPEC													
epaR	ETT2	IPEC and ExPEC													
epaS	ETT2	IPEC and ExPEC													
eprH	ETT2 Type III secretion protein	IPEC and ExPEC													
eprI	Type III secretion apparatus needle protein	IPEC and ExPEC													
eprJ	Type III secretion	IPEC and ExPEC													
eprK	Part of ETT2sepsis	IPEC and ExPEC													
escC	A predicted outer membrane secretin of the Type III Secretion System (LEE)	EPEC/STEC													
escD	Proteins *escD* and *escJ* are crucial for the oligomerization of *escC* (LEE)	EPEC/STEC													
escE	Type III secretion system LEE co-chaperone	EPEC/STEC													
escF	The Type III Needle Protein *escF* of Enteropathogenic *Escherichia coli* (LEE)	EPEC/STEC													
escI	Type III secretion system LEE inner rod protein	EPEC/STEC													
escJ	Proteins *escD* and *escJ* are crucial for the oligomerization of *escC* (LEE)	EPEC/STEC													
escN	Translocator (LEE)	EPEC/STEC													
escR	Type III secretion system protein, YscR/HrcR family (LEE)	EPEC/STEC													
escS	Type III secretion system protein (LEE)	EPEC/STEC													
escT	T3SS structure protein (LEE)	EPEC/STEC													
escU	Auto-cleavage promotes type III effector translocation into host cells (LEE)	EPEC/STEC													
escV	Translocator (LEE)	EPEC/STEC													

espA	Encodes an EPEC secreted protein that is necessary for activating epithelial signal transduction, intimate contact, and formation of attaching and effacing lesions	EPEC/STEC														
espB	The type III secretion protein *espB* is translocated into the host cells and plays an important role in adherence, pore formation and effector translocation during infection	EPEC/STEC														
espD	Protein required for the formation of bacterial surface appendages and is incorporated in the cytoplasmic membranes of target cell	EPEC/STEC														
espF	Type III secretion system LEE needle major subunit	EPEC/STEC														
espG	Binds the Golgi matrix protein GM130 and disrupts the Golgi structure and function	EPEC/STEC														
espH	T3SS effector	EPEC/STEC														
espJ	Effector *espJ* blocks Src kinase activity via amidation and ADP ribosylation	EPEC/STEC														
espK	Type III secreted effector protein of	EPEC/STEC														
espN	T3SS effector	EPEC/STEC														
espL1	Pseudogene	EPEC/STEC														
espL2	Induces actin microfilament aggregation through annexin 2 activation	EPEC/STEC														
espL4	T3SS effector-like	EPEC/STEC														
espR1	T3SS effector-like	EPEC/STEC														
espR2	T3SS effector-like	EPEC/STEC														
espX1	Putative type III secreted effector	EPEC/STEC														
espX4	Putative type III secreted effector	EPEC/STEC														
espX5	Putative type III secreted effector	EPEC/STEC														
espX6	T3SS effector-like	EPEC/STEC														
espY1	T3SS effector-like	EPEC/STEC														
espY3	T3SS effector-like	EPEC/STEC														
espY4	T3SS effector-like	EPEC/STEC														
espM	Type III secretion system effectors	EPEC/STEC														
espN	T3SS effector	EPEC/STEC														
espP	Proteolytic toxin	EPEC/STEC														
espZ	Secreted protein that inhibits host cell apoptosis	EPEC/STEC														
etgA	Muramidase is required for efficient type III secretion	EPEC/STEC														
etsC	Fimbria K88	ETEC														

eptA	Located on the tip of ETEC flagella mediates attachment to mammalian host cells	ETEC														
etpD	Type II secretion system	EHEC-associated														
etrA	Type III secretion system 2 regulator that promotes virulence	APEC														
evfA	Type VI secretion system (T6SS)	IPEC/ExPEC														
evfC	Type VI Secretion System (T6SS) gene and macrophage toxin	IPEC/ExPEC														
evfQ	Type VI Secretion System (T6SS) gene	IPEC/ExPEC														
evfW	Type VI secretion system (T6SS) gene	IPEC/ExPEC														
faeC	K88 fimbrial protein precursor	IPEC														
faeD	K88 fimbria	IPEC														
faeH	K88 minor fimbrial subunit	IPEC														
faeI	K88 minor fimbrial subunit	IPEC														
fbpb	Gene associated with urinary tract infections	Associated with UPEC														
fdeC	Intimin-like adhesin	ExPEC/STEC														
fecA	Ferric dicitrate uptake	Mammary pathogenic *E. coli* (bovines)														
fedC	Fimbrial usher	IPEC														
fepA	Ferric enterobactin transport	ExPEC														
fepB	Ferric enterobactin transport	ExPEC														
fepC	Ferric enterobactin transport	ExPEC														
fepD	Ferric enterobactin transport	ExPEC														
fepE	Ferric enterobactin transport	ExPEC														
fepG	Ferric enterobactin transport	ExPEC														
fes	Enterobactin/ferric enterobactin esterase	ExPEC														
fmlA	Major F9-fimbrial subunit	IPEC/ExPEC														
fyuA	Ferric yersiniabactin uptake receptor gene	ExPEC														
gad	Glutamate decarboxylase gene	Several														
gimB genetic island	A genetic island associated with human newborn meningitis	ExPEC														
grlA	LEE type III secretion system transcriptional regulator	EPEC/STEC														
grlR	Type III secretion system	EPEC/STEC														
Hbp	Hemoglobin-binding protease	ExPEC														
Hemolysin-III	Hemolysin	ExPEC														
hlyA	Pore-forming bacterial endotoxin	ExPEC														
hlyB	Pore-forming bacterial endotoxin	ExPEC														
hlyC	Pore-forming bacterial endotoxin	ExPEC														

hlyD	The T1SS is composed of an inner membrane ABC transporter HlyB, an outer-membrane channel protein TolC, and a membrane fusion protein	ExPEC													
HlyE	Pore-forming toxin	ExPEC													
HlyF	Regulates outer membrane vesicle biogenesis	ExPEC													
hutX	Heme utilization carrier protein	IPEC/ExPEC													
ibeA	Invasion protein gene	ExPEC													
ibeB	Invasin	ExPEC													
ibeC	Invasin	ExPEC													
ibrA	Immunoglobulin-binding regulator A	IPEC/ExPEC													
iha	Adhesin	IPEC and ExPEC													
ipaH	Invasion plasmid antigen	EIEC													
ireA	Iron-regulated gene that participates in adhesion and stress-resistance	ExPEC													
iroB	Encode the siderophore salmochelin operon	ExPEC													
iroC	Salmochelin siderophore system gene	ExPEC													
iroD	Encode the siderophore salmochelin operon	ExPEC													
iroE	Encode the siderophore salmochelin operon	ExPEC													
iroN	Enterobactin siderophore receptor protein	ExPEC													
irp1	Gene from the high-pathogenicity island (HPI) of Yersinia	ExPEC													
irp2	Gene from the high-pathogenicity island (HPI) of Yersinia	ExPEC													
iss	Increased serum survival gene	ExPEC													
iucA	Aerobactin uptake system	ExPEC													
iucB	Aerobactin uptake system	ExPEC													
iucC	Aerobactin uptake system	ExPEC													
iucD	Aerobactin biosynthesis gene	ExPEC													
iutA	ferric aerobactin receptor gene	ExPEC													
kpsM	Capsule polysaccharide export protein gene	ExPEC													
kpsS	Capsule polysaccharide export protein gene	ExPEC													
ler	Global regulator of virulence gene expression (LEE)	EPEC/STEC													
LktB	leukotoxin	ExPEC													
LktB (pO157)	leukotoxin	ExPEC													
lpfA	Long polar fimbria protein A	IPEC													
lpfAO113	Plays a role in the adherence of STEC O113:H21 to epithelial cells	IPEC													
ltcA	Heat-labile enterotoxin A subunit	IPEC													

malX	*Escherichia coli* pathogenicity island-marker	ExPEC														
MAP	Type III secretion system (LEE)	EPEC/STEC														
Mpc	Translation is critical for the efficient production of the LEE operon's downstream gene products	EPEC/STEC														
neuA	K1 capsule antigen	ExPEC														
neuB	K1 capsule antigen	ExPEC														
neuC	K1 capsule antigen	ExPEC														
nleA	Non-LEE-encoded type III translocated virulence factor	EPEC/STEC														
nleB	A bacterial effector with glycosyltransferase activity, targets GAPDH function to inhibit NF-κB activation.	EPEC/STEC														
nleC	T3SS effectoir	EPEC/STEC														
nleC-like gene	T3SS effector protease	EPEC/STEC														
T3SS	T3SS effector	EPEC/STEC														
nleG	T3SS effector	EPEC/STEC														
nleH	T3SS effector	EPEC/STEC														
ompA	Enable intracellular survival, evasion from the body's defense	IPEC/ExPEC														
ompW	Required for resistance to phagocytosis	IPEC/ExPEC														
ospG	T3SS effector	IPEC														
paa	Outer membrane adhesin	EPEC, STEC, ETEC														
papA	structural subunit of P-fimbriae	APEC														
papB	Major pilin subunit	ExPEC														
papC	P-fimbriae outer membrane usher protein gene	ExPEC														
papD	Periplasmic transport protein	ExPEC														
papE	Fimbrial protein precursor	ExPEC														
papF	Fimbrial adapter	ExPEC														
papG	Fimbrial adhesin	ExPEC														
papH	Minor pilin subunit	ExPEC														
papI	Pap operon regulatory protein	ExPEC														
papJ	Participant in pilus assembly	ExPEC														
papK	Minor pilin subunit	ExPEC														
pduC	Correlated with increased cellular invasion and persistence in vitro	IPEC														
perA	Adherence factor	EPEC														
pet	Autotransporter enterotoxin	EAEC														
pic	Autotransporter	Strongly associated with EAEC														
pkgA	ETT2	IPEC/ExPEC														
pFHI-NMBU-03_1 Autotransporter	Novel putative autotransporter toxin	ExPEC														
sbmC	DNA gyrase inhibitor	IPEC/ExPEC														
sat	Secreted autotransported toxin	ExPEC														

sepA	Shigella extracellular protein A	Several													
sepD	T3SS secretion switch protein SepD (LEE)	EPEC/STEC													
sepL	a protein required for enteropathogenic *Escherichia coli* type III translocation, interacts with secretion component SepD. (LEE)	EPEC/STEC													
sepQ	T3SS structure protein (LEE)	EPEC/STEC													
sepZ	T3SS	EPEC/STEC													
shiA	Involved in the pathogenesis of UPEC by down-regulating the innate inflammatory response during infection	UPEC													
shuA	Outer membrane hem/hemoglobin receptor	ExPEC													
sinH	Intimin-like inverse autotransporter	IPEC/ExPEC													
sitA	Iron/manganese transport system periplasmic binding protein gene	ExPEC													
sitB	ATP-binding component	ExPEC													
sitC	Inner membrane component	ExPEC													
sitD	Inner membrane permease	ExPEC													
stcE (O103 variant)	Contributes to intimate adherence	EHEC													
stcD	Putative fimbrial-like adhesin protein gene	IPEC													
stfD	Fimbrial protein gene	Unknown													
Stx1a	Shiga toxin 1a is the most common subtype of stx1 among cattle and human strains	STEC													
Stx1b	Shiga toxin 1b	STEC													
Stx1c	Stx1c Is the most common Shiga toxin 1 subtype among Shiga toxin-producing *Escherichia coli* isolates from sheep but not among isolates from cattle	STEC													
Stx2a	Shiga toxin 2a	STEC													
Stx2b	Shiga-toxin 2 subtype b. Stx2b and Stx2e are less frequently associated with human disease	STEC													
subA	Subtilase cytotoxin (SubAB) is the prototype of a new AB5 toxin family. SubAB has been detected only in strains of *E. coli* that also produce Stx, raising fascinating questions regarding the relative contributions to pathogenesis of these two highly potent AB5 cytotoxins.	STEC													
subB	Subtilase cytotoxin subunit B	STEC													
T1SS		ExPEC													

tia	direct noninvasive *E. coli* to adherence and invasion	IPEC														
Tir	translocated intimin receptor	EPEC/STEC														
tonB	Interacts with outer membrane receptor proteins that carry out high-affinity binding and energy dependent uptake into the periplasmic space of specific substrates such as cobalamin, and various iron compounds (such as iron dicitrate, enterochelin, aerobactin, etc.).	ExPEC														
toxB	Toxin B	EHEC														
traJ	Contributes to the ability of *E. coli* K1 to penetrate the blood-brain barrier in the neonatal rat	ExPEC														
traT	Gene encoding complement resistance protein	ExPEC														
upaB	Autotransporter known to contribute to uropathogenic **E. coli* (UPEC) colonisation of the urinary tract. Disrupted or absent in all IPEC strains.	UPEC														
upaG	Mediates adherence to human bladder epithelial cells (in vitro), promotes cell aggregation and biofilm formation	UPEC														
ureD	Urease accessory protein	UPEC														
usp	Uropathogenic specific protein gene	UPEC														
Vgr-family Protein from FHI-NMBU-03	Associated with the Type VI secretion system	IPEC/ExPEC														
vat	Vacuolating autotransporter toxin gene	ExPEC														
XhlA	Hemolysin	Unknown														
ybtA	Yersinia HPI gene associated with regulatory functions. AraC family DNA-binding transcriptional regulator	ExPEC														
ybtE	Yersiniabactin siderophore biosynthetic protein	ExPEC														
ybtQ	Enhance *E. coli* fitness during high titer cystitis	ExPEC														
ybtS	Putative salicylate synthetase	ExPEC														
ybtT	Yersiniabactin biosynthetic protein	ExPEC														
ybtU	Thiazolinyl-S-HMWP1 reductase	ExPEC														
ycbS	Novel bacterial virulence factor.	ExPEC														
ycgV	Presumed to be involved in biofilm formation	IPEC/ExPEC														
yeeJ	inverse autotransporter that binds to peptidoglycan and	IPEC/ExPEC														

	promotes biofilm formation														
yfcV	*Escherichia coli* Isolates That Carry vat, fyuA, chuA, and yfcV Efficiently Colonize the Urinary Tract	ExPEC													
ygeF	ETT2	ExPEC													
ygeG	ETT2	ExPEC													
ygeH	ETT2	ExPEC													
ygeI	ETT2	ExPEC													
ygeJ	ETT2	ExPEC													
ygeK	ETT2	ExPEC													
yghJ	Uncharacterized protein	IPEC													
yjaA	Involved in the cellular response to hydrogen peroxide and acid stress	IPEC/ExPEC													
yqeH	ETT2	ExPEC													
yqeI	ETT2	ExPEC													
yqeJ	ETT2	ExPEC													
yqeK	ETT2	ExPEC													
yqfA	hemolysin-III family protein	ExPEC													
Z2098	Z2098 is a sequence derived from OI-57, a genomic island that may be associated with increased virulence of STEC strains in humans	STEC													
Z2200 F9-fimbria gene	F9 genes appear to be common in UPEC and other types of pathogenic *E. coli*. However, their precise contribution to disease remains to be determined.	Several													

Supplementary table 2. A list of environmental isolates from *Finton et al. 2020 in prep* and the corresponding GenBank accession numbers.

NMBU_W05E18	CP042878-CP042881
NMBU_W06E18	CP047609-CP047613
NMBU-W07E18	CP042882-CP042884
NMBU_W010C18	CP044402-CP044406
NMBU-W13E19	CP043406-CP043413
NMBU-W12E19	CP042885-CP042892
NMBU_R2	JAAAYF01000000
NMBU_R16	JAAAYE010000000